Name authority data and its model for non-Latin representations
with special emphasis on Chinese characters

A dissertation
submitted to the School of Library and Information Science of Keio University

Maiko Kimura
July 2015
Abstract

Recently, the importance of authority control and sharing authority data has been increasingly appreciated. However, attempts at sharing authority data internationally have been conducted mainly within Western countries. Sharing name authority data in all languages, including non-Latin languages, is an ideal but yet insurmountable goal for library communities. Moreover, the authority data recorded by organizations in non-Latin alphabet countries are diverse, and their differences have not been investigated or clarified in full detail. Taking such differences into account for sharing authority data will help us to achieve more accurate matching results.

The purposes of this study are to 1) investigate representations and data elements recorded in name authority data constructed by organizations located in the Chinese character cultural sphere and by the Library of Congress for a comparison; 2) based on the above analysis, develop an authority data model that can address complicated representations of non-Latin languages; and 3) propose authority data formats that use the developed model in actual authority data and authority works.

In Chapter 1, trends of global authority control and issues of non-Latin representations in such global authority control are explained, and the purpose of the study is provided. Related works and existing authority data models including FRAD, MARC 21 Authority Format, RDA, and DCMI Abstract Model are reviewed in Chapter 2. The review reveals that these models are equally insufficient to handle complex representations of non-Latin languages. In Chapter 3, a new framework of name authority data that includes representations, data elements, and data structures is proposed for subsequent analysis. Characteristics of personal names in the Chinese character cultural sphere are overviewed in the first half of Chapter 3 as a basis of the framework.

In Chapter 4, research methods and research objects are explained first, and then current practices and policies of authority control in China, Japan, South Korea, and Vietnam are described mainly based on interviews. The research methods involved data collection that included face-to-face interviews and collection of cataloging rules, formats, and manuals about name authority data from each organization. Search results of authority databases or OPACs of each organization were also consulted if available. After these data were collected, checkpoints that are unique to Japanese, Chinese, Korean, and Vietnamese name authority data were set. Based on the gathered information, the checkpoints were investigated, taking into account the comparison of the current practices of each organization, and issues affecting data sharing were identified. For Vietnamese names, interviews were not conducted and limited institutions were investigated. Therefore, the research method for Vietnamese names is explained separately in Chapter 8. As the checkpoints and search terms used to search the authority databases or OPACs
of each organization differed by language, they are explained in Chapters 5–8, respectively.

The results about the representations of Chinese, Japanese, Korean, and Vietnamese names in the Chinese character cultural sphere are shown in Chapters 5–8, respectively. It was revealed that Chinese character forms are recorded in letter types that are used by each region where each organization is located. This means Chinese character forms are not always “accurate” forms that the person or corporate body uses in its native country. Romanized forms of Chinese names are recorded using Hanyu Pinyin in all organizations investigated except the ones in South Korea. However, the handling of umlauts differs by organization, and this may be an obstacle to string matching based on Romanized forms of Chinese names. Romanized forms of Japanese names, on the other hand, might vary by organization because the HepburnRomanization system adopted by each organization is slightly different. Furthermore, as Romanization systems adopted by organizations in South Korea and other countries are different, Romanized forms of Korean names may differ among organizations as well. These results show that identifying CJK names merely using the Romanized forms used by organizations is difficult. In addition, despite the importance of yomi for Japanese names, it is not recorded by organizations outside Japan, and thus, yomi cannot be used for identifying Japanese names when authority data are shared among several organizations. Similarly, organizations outside South Korea do not record Hangul forms of Korean names as a mandatory element. This may preclude the possibility of identifying Korean names using Hangul forms across organizations. In Vietnam, name authority control for author names was even not conducted. In summary, any single type of representation is insufficient as a master key for name identification when name authority data are shared. Rather, the combination of several representations seems to be helpful for name identification.

In Chapter 9, the data elements recorded by each organization were examined and compared to authority data elements defined in RDA. It was ascertained that core elements defined in RDA were recorded by most organizations. Among non-core elements, field of study, lineage (especially in Japan), gender, place of ancestry (especially in China), nature or character, and history were recorded by many organizations. RDA defines that some data elements should be recorded separately from access points. These elements are, however, recorded as additions to access points in Japan and China.

Based on the above results, a modification of the FRAD model is proposed in Chapter 10. In the presentations, three kinds of representations, namely, non-Latin transliteration, non-Latin transcription, and Romanization, were defined. Introducing the parent-child relationship into Control Access Points made it possible to determine which two representations should be shown as a pair in authority data.

Chapter 11 describes the development of two authority data formats, namely, modified MARC
Format for Authority Data and RDF/XML format, which can adopt the modified FRAD model proposed in Chapter 10 to authority data. Chapter 12 summarizes the overall results of the study.
Acknowledgements

I would like to express my special appreciation and gratitude to Shunsaku Tamura, Emeritus Professor at the School of Library and Information Science, Keio University, who has been an important mentor of mine since I was an undergraduate student. After Prof. Tamura retired in March 2015, Professor Dr. Shoichi Taniguchi became my supervisor. I deeply appreciate his kind acceptance of this role and valuable comments on my dissertation. I would also like to thank Dr. Barbara B. Tillett, who provided me with excellent comments and suggestions on my first English paper, which is a basis of Chapter six of this dissertation, and on several other chapters.

I would also like to acknowledge the library staff I interviewed in the six regions I visited. All of them kindly agreed to participate in my research interviews and provided me with sample data for this research. I would also like to thank the professors in Keio and other institutions who introduced me to each library’s relevant contact person for conducting interviews.

Parts of this research were financially supported by the Mita Society for Library and Information Science (FY 2012), the Kobayashi Fellowship Program from The Fuji Xerox Setsutaro Kobayashi Memorial Fund (FY 2013), the Sasagawa Scientific Research Grant from The Japan Science Society (FY 2014), and Japan Society of Library and Information Science (FY 2015).

A special thanks to my family. Words cannot express how grateful I am to my mother and father for all of the sacrifices that they’ve made on my behalf. Finally, I would like express my appreciation to my beloved husband Takushi, who always supported me in difficult moments, and to our daughter Kureha, who happily began to go nursery when she was three months old.
Table of contents

Chapter 1  Introduction............................................................................................................  1
1.1  Authority control in a global environment ................................................................. 1
1.2  Issues in non-Latin representations in global authority control .............................. 4
1.3  Purpose of the study.................................................................................................... 5

Chapter 2  Literature review............................................................................................... 9
2.1  History of rules and standards of authority control .................................................. 9
2.2  Authority control of names in non-Latin languages .................................................. 13
2.3  Writing systems in the Chinese character cultural sphere and its handlings in Western authority data .................................................................................................................. 17
2.4  Current conceptual models of authority data in non-Latin script languages and their inadequacies ................................................................................................................. 20
  2.4.1  FRAD model ........................................................................................................ 20
  2.4.2  Model A and Model B of MARC 21/A ................................................................ 23
  2.4.3  RDA model........................................................................................................ 26
  2.4.4  DCMI Abstract Model ....................................................................................... 27

Chapter 3  Framework of name authority data ................................................................ 40
3.1  Characteristics of personal names in the Chinese character cultural sphere .......... 40
  3.1.1  Chinese names ...................................................................................................... 40
    3.1.1.1  Current Chinese names .................................................................................. 40
    3.1.1.2  Romanized representations of Chinese names ................................................ 41
  3.1.2  Japanese names .................................................................................................... 42
    3.1.2.1  Current Japanese names ................................................................................ 42
    3.1.2.2  Romanized representations of Japanese names ................................................. 45
  3.1.3  Korean names ....................................................................................................... 45
    3.1.3.1  Current Korean names ................................................................................... 45
    3.1.3.2  Romanized representations of Korean names ................................................... 47
  3.1.4  Vietnamese names ............................................................................................... 48
  3.2  Representations, data elements, and structures ........................................................ 49

Chapter 4  Method and research objects .......................................................................... 57
4.1  Method....................................................................................................................... 57
  4.1.1  Organizations studied .......................................................................................... 57
  4.1.2  Step 1 – Data collection ....................................................................................... 58
  4.1.3  Step 2 - Setting checkpoints that are unique to authority data of each area .......... 62
  4.1.4  Step 3 - Identification of issues affecting data sharing ........................................ 62
<table>
<thead>
<tr>
<th>Chapter 8</th>
<th>Representations of Vietnamese name authority data in Chinese character cultures</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Research method</td>
<td>128</td>
</tr>
<tr>
<td>8.2</td>
<td>Search results</td>
<td>130</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Types of representations</td>
<td>132</td>
</tr>
<tr>
<td>8.2.2</td>
<td>The relating of Chinese characters to their corresponding chữ quốc ngữ</td>
<td>133</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Diacritics of chữ quốc ngữ</td>
<td>134</td>
</tr>
<tr>
<td>8.2.4</td>
<td>Representations in local languages</td>
<td>134</td>
</tr>
<tr>
<td>8.3</td>
<td>Discussion</td>
<td>134</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Comparison of data elements of authority data in the Chinese character cultural sphere and the RDA framework</td>
<td>138</td>
</tr>
<tr>
<td>9.1</td>
<td>Authority data elements in Japan</td>
<td>138</td>
</tr>
<tr>
<td>9.2</td>
<td>Authority data elements in South Korea</td>
<td>142</td>
</tr>
<tr>
<td>9.3</td>
<td>Authority data elements in Mainland China, Taiwan, and Hong Kong</td>
<td>142</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Mainland China</td>
<td>142</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Taiwan</td>
<td>144</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Hong Kong</td>
<td>145</td>
</tr>
<tr>
<td>9.4</td>
<td>A comparison to authority data elements in RDA</td>
<td>146</td>
</tr>
<tr>
<td>9.5</td>
<td>Findings and recommendations</td>
<td>151</td>
</tr>
<tr>
<td>Chapter 10</td>
<td>A proposal of a modification of FRAD model</td>
<td>154</td>
</tr>
<tr>
<td>10.1</td>
<td>Differences among transliteration, transcription, and Romanization</td>
<td>154</td>
</tr>
<tr>
<td>10.2</td>
<td>The parent-child relationship between an original form and its Romanization and/or transcription</td>
<td>155</td>
</tr>
<tr>
<td>10.3</td>
<td>A modification of the FRAD model for recording names in non-Latin languages</td>
<td>157</td>
</tr>
<tr>
<td>10.4</td>
<td>Discussion</td>
<td>164</td>
</tr>
<tr>
<td>10.5</td>
<td>Chapter conclusion</td>
<td>166</td>
</tr>
<tr>
<td>Chapter 11</td>
<td>Trial on creation of new authority data formats</td>
<td>168</td>
</tr>
<tr>
<td>11.1</td>
<td>Modified MARC 21/A</td>
<td>168</td>
</tr>
<tr>
<td>11.1.1</td>
<td>Modification of MARC 21/A</td>
<td>168</td>
</tr>
<tr>
<td>11.1.2</td>
<td>Examples of modified MARC 21/A records</td>
<td>171</td>
</tr>
<tr>
<td>11.2</td>
<td>The Resource Description Framework (RDF)</td>
<td>180</td>
</tr>
<tr>
<td>11.2.1</td>
<td>Class and property newly defined</td>
<td>181</td>
</tr>
<tr>
<td>11.2.2</td>
<td>Example of RDF/XML records</td>
<td>187</td>
</tr>
<tr>
<td>Chapter 12</td>
<td>Conclusion</td>
<td>199</td>
</tr>
</tbody>
</table>
List of Tables

Table 2-1 *Subtags* related to CJK languages registered in the Language Subtag Registry ........................................ 29
Table 4-1 Number of sample records consulted ................................................................. 58
Table 4-2 Cataloging rules, formats, and manuals consulted .................................................. 59
Table 4-3 Databases and OPACs used for search .................................................................. 63
Table 5-1 Search terms ........................................................................................................... 74
Table 5-2 Adoption and character forms of Chinese characters .............................................. 75
Table 5-3 Treatments and types of Romanization ................................................................. 78
Table 5-4 Adoption of tone marks, apostrophes, and umlaut marks ....................................... 80
Table 5-5 Examples of *pinyin* forms for names including “ü” pronunciation ....................... 81
Table 5-6 Separation between a surname and a given name etc. ............................................ 84
Table 6-1 Search terms ........................................................................................................... 92
Table 6-2 Adoption and character forms of Chinese characters .............................................. 93
Table 6-3 *Yomi* and the corresponding *Kanji* .................................................................... 96
Table 6-4 Treatments and types of Romanization ................................................................. 98
Table 6-5 Differences in Romanization among four organizations ....................................... 99
Table 6-6 Differences in Hepburn Romanization ................................................................. 100
Table 6-7 Defects of VIAF (caused by the Romanization of the moraic obstruent in NDL) .................................................. 102
Table 6-8 Separation between surnames and given names .................................................... 103
Table 6-9 Representations in local languages outside Japan .................................................. 104
Table 6-10 Names in *hiragana* .......................................................................................... 106
Table 7-1 Search terms ......................................................................................................... 112
Table 7-2 Adoptions of *Hangul* forms ................................................................................ 114
Table 7-3 Adoptions of *Hanja* forms .................................................................................... 116
Table 7-4 Adoptions and types of Romanization ................................................................. 118
Table 7-5 Separation of surname and given name with a comma ........................................ 120
Table 7-6 Representations in local languages outside Korea .................................................. 123
Table 8-1 Vietnamese names searched .................................................................................. 129
Table 8-2 Records retrieved .................................................................................................. 130
Table 8-3 Samples of authority data ...................................................................................... 131
Table 8-4 Recording of representations in Chinese characters and *chữ quốc ngữ* ............ 132
Table 9-1 Authority data elements for Japanese authors recorded by organizations in Japan .................................................................................................................. 139
Table 9-2 Authority data elements for Chinese authors recorded by organizations in Mainland China, Taiwan, and Hong Kong ........................................................................ 143
Table 9-3 Conference names as search terms for the HKCAN database ................................ 146
Table 9-4 A comparison of authority data elements defined in Section 9 of RDA to elements recorded by each organization ................................................................. 147
Table 9-5 A comparison of authority data elements defined in Section 11 of RDA to elements recorded by each organization ................................................................................. 148
Table 10-1 Existence of the parent-child relationship in Romanization, transliteration, and transcription ........................................................................................................... 156
Table 10-2 Attributes and relationships regarding representations of name and Controlled Access Point .................................................................................................................. 160
Table 11-1 Relationship identification codes ........................................................................................................... 170
Table 11-2 Romanization scheme codes for CJK languages (provisional) ................................................................. 170
Table 11-3 Properties of “ex:Representation” class newly defined ................................................................. 184
List of Figures

Figure 1-1 An example of incorrect links in Japanese names in VIAF ........................................... 5
Figure 2-1 Alternate linguistic form relationship between Names defined in FRAD ............... 21
Figure 2-2 A parallel language relationship between CAPs defined in FRAD ................. 21
Figure 2-3 Alternate script relationship between CAPs and other variant name relationship between Names defined in FRAD .................................................................................. 23
Figure 2-4 Model A. a sample authority record from NDL manual.................................. 25
Figure 2-5 Model B. Sample authority record from the HKCAN Database OPAC (retrieved 2014-01-08) ...................................................................................................................... 26
Figure 3-1 Examples of the same original form with different yomi in Web NDL Authorities. 44
Figure 3-2 Examples of the same yomi with different original forms in Web NDL Authorities 44
Figure 3-3 Examples of the same Korean personal name represented in Hanja versus Hangul 47
Figure 3-4 Framework of name authority data ........................................................................ 52
Figure 10-1 The modified FRAD model for personal names in non-Latin languages ........ 159
Figure 10-2 Several Romanization relationships for one Controlled Access Point ............ 162
Figure 10-3 A LCNAF record for Haruki Murakami ................................................................. 163
Figure 10-4 Applying the modified FRAD model to Haruki Murakami ............................... 164
Figure 10-5 The modified FRAD model for the Controlled Access Point “Гоголь, Николай Васильевич.”............................................................ 165
Figure 11-1 An example of authority data for Haruki Murakami in Model A of MARC 21/A 169
Figure 11-2 Modified authority data of Figure 11-1 ............................................................... 170
Figure 11-3 An example of authority data for Haruki Murakami with CJK representations .. 172
Figure 11-4 An errant authority data for Haruki Murakami .................................................... 173
Figure 11-5 An example of data which has two yomi forms for one Kanji form.............. 174
Figure 11-6 An example of authority data with a Romanization form in the 100 field ........ 174
Figure 11-7 An example of authority data with an English name in the 100 field .......... 175
Figure 11-8 An example of authority data for Song Qingling with CJK representations ...... 175
Figure 11-9 An example of authority data for Yi Kwang-su with CJK representations ....... 176
Figure 11-10 An errant authority data for Yi Kwang-su (1).................................................... 177
Figure 11-11 An errant authority data for Yi Kwang-su (2)................................................... 177
Figure 11-12 An example of authority data for Kim Ha-nûl with CJK representations ........ 178
Figure 11-13 An example of authority data for Nguyễn Du with CJKV representations ...... 179
Figure 11-14 An example of authority data for Dương Thu Hương with CJKV representations ................................................................................................................................. 180
Figure 11-15 An authority data of Haruki Murakami derived from Web NDL Authorities.... 182
Figure 11-16 The RDF graph of NDL’s representations of Haruki Murakami .................. 183
Figure 11-17 The modified RDF graph of representations for Haruki Murakami .......... 185
Figure 11-18 Modified authority data for Haruki Murakami in RDF/XML .................. 186
Figure 11-19 An example of RDF/XML authority data for Haruki Murakami with CJK representations ............................................................................................................. 188
Figure 11-20 The RDF graph of authority data in Figure 11-19................................. 189
Figure 11-21 The RDF graph of the authority data in Figure 11-7 ............................ 190
Figure 11-22 RDF/XML authority data of Figure 11-21 ........................................ 191
Figure 11-23 RDF/XML authority data of Song Qingling in Figure 11-8 .................... 192
Figure 11-24 RDF/XML authority data of Yi Kwang-su in Figure 11-9 ...................... 193
Figure 11-25 RDF/XML authority data of Kim Ha-nul in Figure 11-12 ...................... 194
Figure 11-26 RDF/XML authority data of Nguyễn Du in Figure 11-13 ....................... 195
Figure 11-27 RDF/XML authority data of Dương Thu Hương in Figure 11-14 ........... 196
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACR</td>
<td>Anglo-American Cataloging Rules</td>
</tr>
<tr>
<td>AACR2</td>
<td>Anglo-American Cataloging Rules, Second Edition</td>
</tr>
<tr>
<td>ALA</td>
<td>The American Library Association</td>
</tr>
<tr>
<td>ALCTS</td>
<td>The Association for Library Collections &amp; Technical Services</td>
</tr>
<tr>
<td>BAP</td>
<td>Base Access Point</td>
</tr>
<tr>
<td>BCP</td>
<td>Best Current Practice</td>
</tr>
<tr>
<td>BL</td>
<td>The British Library</td>
</tr>
<tr>
<td>CALIS</td>
<td>The China Academic Library &amp; Information System</td>
</tr>
<tr>
<td>CAP</td>
<td>Controlled Access Point</td>
</tr>
<tr>
<td>CCCNA</td>
<td>The Cooperative Committee for Chinese Name Authority</td>
</tr>
<tr>
<td>CEAL</td>
<td>The Council on East Asian Libraries</td>
</tr>
<tr>
<td>CJK</td>
<td>Chinese-Japanese-Korean</td>
</tr>
<tr>
<td>CMARC/A</td>
<td>Chinese MARC Format for Authority Records (Taiwan)</td>
</tr>
<tr>
<td>CNMARC/A</td>
<td>China MARC Format/Authorities (Mainland China)</td>
</tr>
<tr>
<td>DCAM</td>
<td>The DCMI Abstract Model</td>
</tr>
<tr>
<td>DCMI</td>
<td>The Dublin Core Metadata Initiative</td>
</tr>
<tr>
<td>EAC-CPF</td>
<td>Encoded Archival Context for Corporate Bodies, Persons, and Families</td>
</tr>
<tr>
<td>EU</td>
<td>The European Union</td>
</tr>
<tr>
<td>FRAD</td>
<td>Functional Requirements for Authority Data</td>
</tr>
<tr>
<td>FRBR</td>
<td>Functional Requirements for Bibliographic Record</td>
</tr>
<tr>
<td>FRSAD</td>
<td>Functional Requirements for Subject Authority Data</td>
</tr>
<tr>
<td>GARE</td>
<td>Guidelines for Authority and Reference Entries</td>
</tr>
<tr>
<td>GARR</td>
<td>Guidelines for Authority Records and References</td>
</tr>
<tr>
<td>HKCAN</td>
<td>The Hong Kong Chinese Authority Name Workgroup</td>
</tr>
<tr>
<td>IANA</td>
<td>The Internet Assigned Numbers Authority</td>
</tr>
<tr>
<td>ICP</td>
<td>Statement of International Cataloguing Principles</td>
</tr>
<tr>
<td>IETF</td>
<td>The Internet Engineering Task Force</td>
</tr>
<tr>
<td>IFLA</td>
<td>The International Federation of Library Associations</td>
</tr>
<tr>
<td>ISAAR (CPF)</td>
<td>International Standard Archival Authority Record for Corporate Bodies, Persons and Families</td>
</tr>
<tr>
<td>ISADN</td>
<td>International Standard Authority Data Number</td>
</tr>
<tr>
<td>ISBD</td>
<td>International Standard Bibliographic Description</td>
</tr>
<tr>
<td>ISNI</td>
<td>International Standard Name Identifier</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>JLA</td>
<td>Japan Library Association</td>
</tr>
<tr>
<td>JULAC</td>
<td>The Joint University Librarians Advisory Committee</td>
</tr>
<tr>
<td>KCR</td>
<td>Korea Cataloging Rules</td>
</tr>
<tr>
<td>Keio</td>
<td>Keio University Libraries</td>
</tr>
<tr>
<td>KERIS</td>
<td>Korea Education and Research Information Service</td>
</tr>
<tr>
<td>KORMARC/A</td>
<td>KORMARC Format for Authority Data</td>
</tr>
<tr>
<td>LC</td>
<td>The Library of Congress</td>
</tr>
<tr>
<td>LCNAF</td>
<td>The LC/NACO Authority File</td>
</tr>
<tr>
<td>LEAF</td>
<td>The Linking and Exploring Authority Files (project)</td>
</tr>
<tr>
<td>LITA</td>
<td>Library and Information Technology Association</td>
</tr>
<tr>
<td>MARBI</td>
<td>The Machine-Readable Bibliographic Information Committee</td>
</tr>
<tr>
<td>MARC 21/A</td>
<td>MARC 21 Format for Authority Data</td>
</tr>
<tr>
<td>MCT</td>
<td>The South Korean Ministry of Culture and Tourism</td>
</tr>
<tr>
<td>MOFA</td>
<td>The Ministry of Foreign Affairs (Japan)</td>
</tr>
<tr>
<td>MR</td>
<td>The McCune-Reischauer (system)</td>
</tr>
<tr>
<td>NACESTI</td>
<td>The National Centre for Scientific and Technological Information (Vietnam)</td>
</tr>
<tr>
<td>NACESTID</td>
<td>The National Centre for Scientific and Technological Information and Documentation (Vietnam)</td>
</tr>
<tr>
<td>NACO</td>
<td>The Name Authority Cooperative (Program)</td>
</tr>
<tr>
<td>NACSID</td>
<td>The National Center for Science Information Systems (Japan)</td>
</tr>
<tr>
<td>NASATI</td>
<td>The National Agency for Science and Technology Information</td>
</tr>
<tr>
<td>NCL</td>
<td>The National Central Library</td>
</tr>
<tr>
<td>NCR</td>
<td>Nippon Cataloging Rules</td>
</tr>
<tr>
<td>NDL</td>
<td>The National Diet Library (Japan)</td>
</tr>
<tr>
<td>NII</td>
<td>The National Institute for Informatics (Japan)</td>
</tr>
<tr>
<td>NLC</td>
<td>The National Library of China</td>
</tr>
<tr>
<td>NLK</td>
<td>The National Library of Korea</td>
</tr>
<tr>
<td>NLST</td>
<td>The National Library for Science and Technology (Vietnam)</td>
</tr>
<tr>
<td>NLV</td>
<td>The National Library of Vietnam</td>
</tr>
<tr>
<td>NTS</td>
<td>Nippan Library Service, Co., Ltd.</td>
</tr>
<tr>
<td>NTUL</td>
<td>National Taiwan University Library</td>
</tr>
<tr>
<td>OCLC</td>
<td>OCLC Online Computer Library Center, Inc.</td>
</tr>
<tr>
<td>OPAC</td>
<td>Online Public Access Catalog</td>
</tr>
<tr>
<td>ORCID</td>
<td>Open Researcher and Contributor ID</td>
</tr>
<tr>
<td>PCC</td>
<td>The Program for Cooperative Cataloging</td>
</tr>
<tr>
<td>RDA</td>
<td>Resource Description and Access</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RFC</td>
<td>Request for Comments</td>
</tr>
<tr>
<td>RLG</td>
<td>The Research Libraries Group</td>
</tr>
<tr>
<td>RLIN</td>
<td>The Research Libraries Information Network</td>
</tr>
<tr>
<td>SMRT</td>
<td>The Synergy of Metadata Resources in Taiwan (system)</td>
</tr>
<tr>
<td>SNUL</td>
<td>The Seoul National University Library</td>
</tr>
<tr>
<td>TRC</td>
<td>Toshokan Ryutsu Center Co., Ltd.</td>
</tr>
<tr>
<td>UBC</td>
<td>The Universal Bibliographic Control</td>
</tr>
<tr>
<td>UCS</td>
<td>Universal Coded Character Set</td>
</tr>
<tr>
<td>UNIMARC/A</td>
<td>UNIMARC Manual Authorities Format</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
</tr>
<tr>
<td>VIAF</td>
<td>Virtual International Authority File</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
<tr>
<td>YUL</td>
<td>Yonsei University Library</td>
</tr>
</tbody>
</table>
Bibliography

<English>


Groom, Linda. Converting Wade-Giles cataloging to pinyin: the development and


Kimura, Maiko. A comparison of recorded authority data elements and the RDA framework


“MARC 21 specifications for record structure, character sets, and exchange media: Character


Miyazawa, Akira. “Parallel writing in East Asian languages and its representation in metadata
“History of National Agency on Science and Technology Information (NASATI)”. National Agency on Science and Technology Information.
OCLC. “VIAF (Virtual International Authority File)”. http://viaf.org/, (accessed 2013-04-10).
“OCLC Connexion Client guides: cataloging, export or import bibliographic records”. OCLC. 2014-03.
“OCLC, Die Deutsche Bibliothek and the Library of Congress to develop Virtual International Authority File.” OCLC Abstracts. 2003, vol. 6, no. 35,
Park, Hong-Seok. “Comparative analysis of author name authority records of Seoul National University and Yonsei University (part 2)” . 日本語，中国語，韓国語の名前典拠ワークシ

Park, Hong-Seok. “The revision of the cataloging rule (KCR4) and authority control under the new rule”. 日本語，中国語，韓国語の名前典拠ワークショップ記録. 東京, 2002-03-14/18, 国立情報学研究所, 2002.


Schiff, Adam. “Non-Latin script references in name authority records”. NACO: Name Authority Cooperative Program. 2009-06-01.


x


Yu, Abraham J. Evaluation and Analysis of the Chinese Name Authority Files under Development in Beijing, Hong Kong, and Taipei. OCLC, 2002, 52p.


<Chinese>


101-104.

GB/T16159-2012. 汉语拼音正词法基本原则.
“團體權威整理作業手冊”. 國家圖書館編目園地全球資訊網.

“香港中文名稱規範數據庫公衆检索目錄: HKCAN Database OPAC”.
罗翀. 国家图书馆外文书目规范控制的实践探索. 图书馆学研究. 2011, vol. 16, p. 30-34.
万爱雯. 从国家图书馆中文名称规范库看标目组织对编目工作的影响. 图书情报工作. 2010, suppl. 1, p. 119-121, 171.
王德毅编著. 清人别名字号索引. 新文豐, 1985, 867p.
王小盾, 劉春銀, 陳義, eds. 越南漢喃文獻目錄提要. 中央研究院中國文哲研究所, 2002, 3 vols.
国家图书馆. 中国机读规范格式使用手册; 中文图书名称规范数据款目著录规则; 中文图书主题规范数据款目著录规则. 1999, 182p.
“全國 1,510 姓氏 陳林滿天下 內政部《全國姓名統計分析》全新出版”. 中華民國內政部. 2014-10-29.
“護照外文姓名拼音對照表（包括：國音第一式、漢語拼音、通用拼音、國音第二式、及 WG 拼音法）”. 外交部領事事務局:護照.
“中文譯音使用原則”. 臺灣地區地名查詢系統: 地名相關法規. 2008-12-08.

xiii

“香港地區大學圖書館編目作業近況” 中文名稱規範聯合協調委員會第八次會議 臺北, 2010-11-02. 


<Japanese>
石川史士. 著者名典拠ファイルの公開に向けて. 全国書誌通信. 1994, no. 90, p. 2.
武部良明. 日本語の表記. 角川書店, 1979, 574p.
典拠ファイルについて. 印刷カード通信. 1972, no. 14, p. 3-6.
「TRC MARC 人名典拠録」編集部編. 17人のステインソン. 図書館流通センター, 1985, 144p.
丹羽基二編. 日本苗字大辞典. 芳文館, 1996, 3 冊.
“常用漢字表”. 文化庁.

xviii
“ヘボン式ローマ字綴方表”．東京都生活文化局．
http://www.seikatubunka.metro.tokyo.jp/hebon/, (accessed 2015-02-03)．

法務省令第二号．官報．2015-01-07, no. 6446, p. 1,
http://kanpou.npb.go.jp/20150107/20150107h06446/20150107h064460001f.html (accessed 2015-02-03)．


満都拉, 藤井敦, 石川徹也．伝統的モンゴル語と現代モンゴル語を対象とした多方向的な翻字手法．情報処理学会論文誌．2006, vol. 47, no. 8, p. 2733-2745．

三浦敬子，松井幸子．欧米における著者名典拠ファイルの共同作成の動向．日本図書館情報学会誌．2001, vol. 47, no. 1, p. 29-41．

水野直樹．創氏改名：日本の朝鮮支配の中で．岩波書店．2008, 246p．

“メタデータ情報共有のためのガイドライン”．メタデータ情報基盤構築事業．

望月源, 大和加寿子, 前嶋淳子, 林俊成．書誌情報データベースにおける非ローマンアルファベット系言語の原綴り・翻字相互変換システム．情報処理学会論文誌:データベース．2005, vol. 46, no. SIG18, p. 50-58．

“常用漢字表”．文部科学省．1981-10-01.

“ローマ字のつくり方”．文部科学省．1954-12-09.

山崎賢二．目录における朝鮮人名の読み方．図書館雑誌．1980, vol. 74, no. 4, p. 162-163．

柚木利博．いまさら聞けない漢字の常識：常用漢字．月刊しにか．2002, vol. 13, no. 9, p. 14-16．

吉川雅之．香港粵語:語彙とリスニング I．白帝社．2002, 138p．

“旅券法施行規則”．電子政府の総合窓口．2014-12-01.
http://law.e-gov.go.jp/htmldata/H01/H01F03301000011.html, (accessed 2015-02-03)．

<Korean>


“외래어 표기법”．국립국어원．

xix
이길성. 漢字이름 5151字 안에서 저으세요 : 대법, 人名用 113 자 추가 이름 다섯자 넘으면 안돼. 朝鮮日報. 2007-03-05, 사회 A15 면.


“쿠사나기 츠요시”. NAVER 인물검색.
http://people.search.naver.com/search.naver?where=nex&sm=tab_ppn&query=%EC%BF%A0%EC%82%AC%EB%82%98%EA%B8%B0%EC%B8%A0%EC%9A%94%EC%8B%9C, (accessed 2015-03-31). NAVER is one of the most popular search engines in Korea.


<Vietnamese>


xxi
Chapter 1

Introduction

1.1 Authority control in a global environment

There is increasing appreciation in recent times of the importance of authority control and the sharing of authority data. *Functional Requirements for Bibliographic Records* (FRBR), proposed in 1997, aimed at recommending the functionality of bibliographic records as distinct from authority data,¹ while *Functional Requirements for Authority Data* (FRAD) was proposed in 2009 as a conceptual model for authority data. FRAD is approved by the International Federation of Library Associations (IFLA). The FRAD model provides a framework for the analysis of functional requirements for the type of authority data that are required to support authority control and for the international sharing of authority data.² Another notable development was the publication in 2011 of the *Functional Requirements for Subject Authority Data* (FRSAD), which addresses subject authority data.

The *Statement of International Cataloguing Principles* (ICP) was published in February 2009 as a substitute for the so-called Paris Principles approved in 1961. ICP was developed based on the conceptual model of FRBR and states that “a cataloguing code should take into account the entities, attributes, and relationships as defined in conceptual models of the bibliographic universe.”³ It notes that these “conceptual models” are FRBR, FRAD, and FRSAD. Thus, it can be said that FRAD and FRBR form the foundation for ICP. In addition, ICP clearly requires the construction of authority data according to the following rule: “Authority records should be constructed to control the authorized forms of names, variant forms of name, and identifiers used as access points (6.1.1.1).” In contrast to Paris Principles, which do not stipulate a rule for the construction of authority data but state that “the main entry for works entered under author’s names should normally be made under a uniform heading (6.1),” the necessity of authority control is emphasized more in ICP.

*Resource Description and Access* (RDA) was released in 2010 as a replacement for *Anglo-American Cataloguing Rules, 2nd edition* (AACR2) and was developed to align with the conceptual models for bibliographic and authority data developed by IFLA, such as FRBR, FRAD, and FRSAD (0.2.1).⁴ Although the relationships between Controlled Access Points defined in FRAD are currently out of scope for RDA (0.2.3), rules for authority data in RDA are compliant with FRAD. To identify a person, family, or corporate body, RDA stipulates that attributes such as date of birth, profession or occupation, location of conference, period of activity of the corporate body, etc., could be recorded in authority data. These attributes can be recorded as separate elements, as parts of the authorized access points representing the person,
family, or corporate body, or as both (0.6.7). AACR2 also stipulates that elements could be
added to access points, but it does not refer to recording elements apart from access points. In
other words, while AACR2 is a set of cataloging rules for bibliographic data, RDA is a set of
cataloging rules for authority data and bibliographic data. In summary, the importance of
authority control has been clearly demonstrated to the library community since FRAD clarified
the functional requirements for authority data and it was adopted by ICP and RDA.

In contrast to AACR2, the cataloging rules that were originally designed for Anglo-American
countries, RDA is designed for use in other language communities. Many libraries in
non-Latin alphabet countries have started to adopt RDA. For example, Israeli academic libraries
started preparing to adopt RDA in 2012. Additionally, the National Diet Library (NDL) of Japan
and Japan Library Association are planning to set up a bibliographic standard corresponding
to RDA, although NDL has already implemented RDA for foreign materials since April 2013.

Sharing authority data is not a particularly new trend, as it has been conducted among Western
countries. One such attempt is the Name Authority Cooperative (NACO) Program, a joint
project started in 1977 by the Library of Congress (LC) and the Government Printing Office to
construct a common name authority file. NACO is part of the Program for Cooperative
Cataloging (PCC), an international cooperative effort aimed at expanding access to library
collections by providing useful, timely, and cost-effective cataloging that meets mutually
accepted standards of libraries around the world. In the 2014 fiscal year, NACO had 710
institutional members. Since the NACO program started, NACO members have created
authority records in accordance with LC’s authority format and have provided them to LC
Name Authority File (later called the LC/NACO Authority File: LCNAF). In turn, the LC has
provided its authority records created by LC and NACO members to libraries around the
world. They are a major part of Virtual International Authority File (VIAF).

After the 1990s, the development of online library catalogs accelerated such sharing of
authority data. In 1993, the British Library (BL) started the Anglo-American Authority File
(AAAF) project, which aimed to share authority data created by the BL and LC following
AACR2. Project AUTHOR, conducted between 1995 and 1997, was an attempt to create and
share authority files among five national libraries in Europe.

The <indecs> project, which analyzes the requirements for metadata for e-commerce in
intellectual property in the network environment, was conducted between 1998 and 2000. The
InterParty project, conducted during 2002–2003, aimed to develop a mechanism that enables the
interoperation of identifiers for parties or persons across multiple domains. Both projects
necessitated cooperative work among libraries, museums, archives, and rights management
communities to share authority information.
The AUTHOR, the InterParty projects were funded by the European Commission, which is an executive body of the European Union (EU). The EU also funded the Linking and Exploring Authority Files (LEAF) project, conducted for three years from 2001. LEAF was aimed at sharing authority data among libraries, archives, and museums in Europe. VIAF is a system that links together authorized forms of names and titles among authority files of national bibliographic agencies and other regional agencies, and permits users to search and display names of a given entity in various languages and scripts. The VIAF project was initially started in 1998 by Die Deutsche Bibliothek and the LC in collaboration with OCLC, and as of July 2014, involved 34 participating agencies in 29 countries. As seen above, sharing authority data has been conducted for many years, not only among libraries, but also among museums, archives, and rights management communities. However, such institutions are mainly located in Western countries. Although VIAF aims to link authority data globally, only a few organizations in non-Latin alphabet countries are currently involved in VIAF. With intellectual activities becoming borderless, the importance of sharing and utilizing authority data in non-Latin languages is increasing. Sharing authority data in all languages is an ideal, yet insurmountable goal for library communities.

Individuals might be identified by identifiers, not by name strings. In fact, there is an increasing trend in approaches to identifying individuals by unique numbers or Uniform Resource Identifiers (URIs). For example, each authority record in VIAF has a unique identifier, which is called VIAF ID. The International Standard Name Identifier (ISNI) is an International Organization for Standardization (ISO) certified global standard for identifying contributors to creative works and those active in their distribution, including writers, artists, creators, performers, and researchers. The Open Researcher and Contributor ID (ORCID) and ResearcherID are identifiers for researchers, providing a registry of unique identifiers and generating research activities of individuals. The advantages of identifiers, compared to name strings, are recognized by library communities, because identifiers can uniquely identify a person or a corporate body that may share the same name with others or may have variant forms of the same name. However, many authors are filed in multiple databases and authority control on these identifiers is therefore still needed.

Because identifier management systems such as VIAF and ISNI collect authority data from various source databases first and then identify entities algorithmically before assigning identifiers to each entity, name identification as a precedent using name strings is unavoidable. When sharing authority data among databases, authors may be automatically disambiguated by computer algorithms. However, it is difficult to rely on name string matching to determine whether two authors represent the same person because the name form recorded in authority
databases may differ from one community and language to another, such as “Confucius” for Anglo-American communities and “孔子” for Chinese, Japanese, and Korean communities. Furthermore, Chinese-Japanese-Korean (CJK) names are particularly less amenable to disambiguation because of the high frequency of homonyms they contain. The speed and performance of disambiguation algorithms may be improved by the addition of data elements to author names. In addition, representations of names and their relationships could help with the process of author disambiguation, especially for non-Latin names.

1.2 Issues in non-Latin representations in global authority control

Sharing name authority data in non-Latin languages is more difficult than for data in Western languages, mainly because of the former’s diversity of scripts. For example, the name in Chinese characters for Mao Zedong, the founder of the People’s Republic of China, is represented as “毛泽东” in Mainland China, “毛澤東” in Hong Kong and Taiwan, and “毛沢東” in Japan. Each organization records personal and corporate names mainly in its own language and script; thus, it is difficult to identify one entity that is recorded in several databases. Even in the Chinese character cultural sphere (covering China, Korea, Japan, and Vietnam), scripts used vary depending on the area. In addition, although many non-Latin alphabet countries started to implement RDA, each organization in such countries has long been using its own cataloguing rules, which define how to establish access points in their original forms, depending on their customs governing personal and corporate names. Thus, authority data recorded by each organization in non-Latin alphabet countries are diverse. To share such data, it is important to understand how they differ from each other and to take such differences into account in the construction of an integrated system or database. However, the investigation and detailed clarification of the contents of authority data recorded by each organization, which is required in such an exercise, have not yet been undertaken.

The Romanized form of a name is often recorded in authority data, in addition to its original form. For names in non-Latin languages, the original form is more important than the Romanized form. However, in Western library communities, recording names in Romanized form remain the priority, even for non-Latin names. For example, wrong links to Japanese names sometimes surface in the VIAF, as Figure 1-1 shows. In heading No. 1, “Hayashi, Yoshitsugu” is linked to two Japanese names “林, 宜嗣” and “林, 良嗣”. However, although their birth years are the same, they are, in fact, very different people. This shows that it can be difficult to maintain accuracy in linking names that use Chinese characters, and users in non-Latin alphabet countries cannot currently place full trust in the VIAF. Such a wrong link may be created, because the VIAF system was designed without full understanding of the writing systems of non-Latin names, which should accord priority to the original form of a
name during the name disambiguation process. In this example, the VIAF algorithm assumed that the Romanized form takes priority over the Kanji form, or, at least, that the two forms have equivalent importance. However, in Japanese names, the Kanji form of the name should be given priority, because Romanized forms of names merely convey the reading of the name and several people who have different Kanji names may share the same Romanized form. This kind of misunderstanding could arise not only in VIAF, but also in all applications that address global personal and corporate names administrated in Western countries. This poses an obstacle to the successful sharing of authority data recorded by worldwide organizations.

<table>
<thead>
<tr>
<th>Heading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Hayashi, Yoshitsugu, 1951-.... 林, 宜嗣, 1951- 林, 景宣, 1951-</td>
</tr>
<tr>
<td>2 林, 景宣, 1930-</td>
</tr>
<tr>
<td>3 Hayashi, Yoshitsugu</td>
</tr>
</tbody>
</table>

Note: ©2014 OCLC Online Computer Library Center, Inc. (“OCLC”). Used by permission. VIAF® is a service mark of OCLC.

Figure 1-1 An example of incorrect links in Japanese names in VIAF.

1.3 Purpose of the study

Because sharing authority data internationally is required and is already underway, there is an urgent need to present a real picture of authority data recorded by each organization in non-Latin alphabet countries, to compare them and clarify how they differ from, or share commonalities, with each other. Such differences or commonalities will serve as tips or traps for sharing authority data internationally.

Since FRAD is the basis of ICP and RDA, its authority data model is respected by library communities worldwide, and it will form the basis of many systems’ attempts to share authority data. However, because FRAD was developed in the Western cultural sphere, the model might be biased toward Western customs. Whether FRAD could directly process authority data in non-Latin languages should thus be investigated, and if it is unable to handle non-Latin languages, the model should be modified for use in future attempts at authority data sharing.
This study narrows the research focus to the Chinese character cultural sphere, because while many languages and scripts are used within this sphere, Chinese characters are commonly used in several countries. The author considers areas and countries belonging to the Chinese character cultural sphere to be representative examples of non-Latin alphabet countries.

The purposes of study are: (1) to investigate representations and data elements recorded in name authority data constructed by organizations located in the Chinese character cultural sphere, in comparison to the LC as a representative in the Western cultural sphere; (2) based on the above analysis, to develop an authority data model that can address complicated representations of non-Latin languages; and (3) to propose authority data formats to utilize the developed model in actual authority data and authority works.

For these purposes, related works and existing authority data models including FRAD are reviewed in Chapter 2. In Chapter 3, a new framework of name authority data that includes representations, data elements, and data structures is proposed for subsequent analysis. In Chapter 4, research methods and research objects are explained first and current practices and policies of authority control in China, Japan, South Korea, and Vietnam are explained mainly based on the interviews. Results of the representations of Chinese, Japanese, Korean, and Vietnamese name authority data recorded in the Chinese character cultural sphere are presented in Chapters 5 to 8, respectively. In Chapter 9, data elements recorded by each organization are examined and compared to authority data elements defined in RDA. Based on these results, a modification of the FRAD model is proposed in Chapter 10. Chapter 11 aims to develop two authority data formats, which can apply the proposed model to authority data. Chapter 12 summarizes all the results of the present study.

Authority data addressed in this study is limited to authority data for names of persons or corporate bodies as authors of works. Authority data for subject names and titles fall outside the scope of this study. The reason for excluding family names is that, although family names may be recorded in archives and special libraries that are focused on genealogy and history, they are rarely recorded as authors in bibliographical records created by libraries. Rather, they appear as subjects, which are more properly handled by FRSAD.

Notes


The screenshot was made at 2013-04-10 from OCLC. “VIAF® (Virtual International Authority File)”. http://viaf.org/. This error was reported and fixed already.
Chapter 2

Literature review

2.1 History of rules and standards of authority control

Many past cataloging rules failed to mention authority control explicitly. However, Panizzi’s *Catalogue of Printed Books in the British Museum* published in 1841 already stated the need of cross-references from name to name, in rule LV.1 Charles A. Cutter also recognized the need for cross-references of author’s name in his *Rules for a Printed Dictionary Catalogue* published in 1876. He stated that “when an author’s name is variously spelled, select the best authorized form as heading, add the variants in parentheses, and make references from them to the form adopted.” in rule 21.2 LC devised the rule in 1899 that “an authority card is to be prepared for every person, corporation, or title of anonymous work that appears in the catalog for the first time, whether as author (main) or added heading,”3 meaning LC had authority control for at least that year.

In rule 355 of *Rules for a Dictionary Catalog, 4th ed.*, published in 1904, Cutter stated the need for authority control, noting that “the cataloger’s author list, kept alphabetically, prevents duplication of work.”4 According to Auld, although *Catalog Rules: Author and Title Entries*, published in 1908, did not mention the need to keep authority records, it prescribed the use of cross-references.5 The 1941 *A.L.A Catalog Rules* offered a guideline on the use and construction of authority cards for headings representing personal and corporate names and uniform titles.5 Similar to the 1908 rules, *A.L.A. Cataloging Rules for Author and Title Entries* (1949) and * Anglo-American Cataloguing Rules* (AACR), published in 1967, prescribed cross-references, but without providing any suggestion for keeping a record of references used.5 * Anglo-American Cataloguing Rules, 2nd ed.* (AACR2), published in 1978, on the other hand, “provided detailed instructions including an entire chapter on the making of see and see also references,” but “without suggestions as to how a library was to keep track of those references that had been made.”5 According to Burger, “all practical aspects (form of catalog, filing, authority procedures, etc.) were consciously left out of AACR2.”6

In the context of the Universal Bibliographic Control (UBC), many tools for authority control have been developed since the late 1970s. UBC became one of the core programs of IFLA in 1974.7 UBC was based on the idea that each document would be cataloged in its country of origin and that the results of that cataloging would be shared and made available throughout the world.8 It was recommended that each national bibliographic agency accepts the responsibility of establishing authoritative lists of its country’s authors’ names.9
In 1963, IFLA had already developed *Names of Persons*, which lists national practices for forms of personal names, reflecting the Paris Principles. IFLA International Office for UBC published *Names of Persons* in 1977 with a full revision and extension. In 1980, *Form and Structure of Corporate Headings* was published, which aimed to promote uniformity in the headings that appear in bibliographic records produced for international exchange within the framework of UBC.

In 1978, the Working Group on an International Authority System was established and the *Guidelines for Authority and Reference Entries* (GARE) was published in 1984. The GARE was the first international principle for the creation of authority data that defined the contents and architecture of authority records. Because GARE was only a logical guideline, its scope was confined to the overall structure and major functional components of entries.

The Library of Congress developed the preliminary edition of *Authorities: A MARC Format* in 1976 and published the first edition in 1981 (later revised in 1987 and 1993 as *US MARC Format for Authority Data*). In 2000, it was integrated with *MARC Communication Format: Authorities of Canada* and renamed *MARC 21 Format for Authority Data* (MARC 21/A). The *UNIMARC/Authorities*, which is based on GARE and published in 1991, is an authority format that aims to make the exchange of authority data internationally compatible.

The IFLA Working Group on Minimal Level Authority Records and ISADN (International Standard Authority Data Number) formed in 1996. The group published a report titled *Mandatory Data Elements for Internationally Shared Resource Authority Records* in 1998, which provides both mandatory and optional authority data elements for the purpose of internationally sharing authority records. The working group also reported whether these elements were included in 10 different kinds of authority formats. On the working group’s recommendation, the *UNIMARC/Authorities* and GARE were revised in 2001 and renamed *UNIMARC Manual Authorities Format* (UNIMARC/A) and *Guidelines for Authority Records and References* (GARR), respectively.

The IFLA Working Group on Minimal Level Authority Records and ISADN concluded that “the IFLA goal of Universal Bibliographic Control by way of requiring everyone to use the same form for headings globally is not practical.” Instead, linking “the authority records created in one country according to one set of cataloguing rules with those in another country to facilitate sharing of authority records and potentially to enable computer-assisted switching to display authorized forms” became the new method of achieving UBC.

*Resource Description and Access* (RDA) is a replacement for AACR2 and is based on *Functional Requirements for Authority Data* (FRAD), *Functional Requirements for Bibliographic Records* (FRBR), and *Functional Requirements for Subject Authority Data* (FRSAD). These are conceptual models developed by IFLA working groups. They provide a
framework to analyze the functional requirements of bibliographic and authority data. FRAD defines entities, attributes, and relationships that consist of authority data.

In archival communities, the second edition of ISAAR (CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons and Families), published in 2003, defines authority data elements for archival authority records.\textsuperscript{17} Based on ISSAR (CPF), EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families)—an Extensible Markup Language (XML)-schema for encoding names of creators of archival materials—was developed and fully adapted by the Society for American Archives in 2011.\textsuperscript{18}

Compared to Western countries, the start of authority control in CJK countries was slow. In Japan, the National Diet Library (NDL) had produce authority cards for CJK names since 1948.\textsuperscript{19} Unfortunately, Japanese libraries have not conducted authority control consistently. According to the result of a questionnaire on the technical services of Japanese libraries conducted by the Japan Library Association (JLA) in 1989, only 15.1% of 1,693 public libraries and 25.18% of 834 academic libraries had constructed name authority files for author names.\textsuperscript{20} The condition seemed to have improved by 2010, when the sixth piece of research by JLA was conducted. Of 825 public libraries providing cataloging services, 56% did authority control for names of authors, and of 840 academic libraries providing cataloging services, 63.7% of them did authority control for author names.\textsuperscript{21}

NACSIS-CAT, an online shared cataloging system for academic libraries in Japan, which is operated by the National Institute of Informatics (NII; formerly, the National Center for Science Information Systems or NACSIS), was started in 1985 and the system included a union authority database from the beginning.\textsuperscript{22}

Although the JLA's research was began in 1964 and was also conducted in 1972 and 1981, a question about authority control only first appeared in 1989, and, thus, we cannot know what the situation was before 1989. In rule 21.2.0 of 日本目録規則1987年版 (Nippon Cataloging Rules 1987 ed.; NCR1987), it states that “heading (except title headings and class number headings) should be used as an authorized form, unified in an authority file for author names, a list of subject headings, or a subject authority file.”\textsuperscript{23} It also noted that “to maintain uniform headings, an authority file that records forms and expressions of uniform headings, sources consulted in establishing headings, and references from headings that were not chosen as authorized headings, etc. is needed.”\textsuperscript{23} However, NCR1987 and its revisions did not provide procedures or any further guidance about authority control. Although cross-references from headings to headings were requested in rule 124 of the first Japanese cataloging code, 日本目録規則 (Nippon Catalog Rules) published in 1943\textsuperscript{24}, any version of NCR before NCR1987 did not refer to authority control or an authority file.
Currently, several organizations generate their own authority data in Japan: NDL, NACSIS-CAT, Keio University Libraries, Toshokan Ryutsu Center Co., Ltd. (TRC), and Nippan Library Service, Co., Ltd. (NTS). TRC and NTS are creating bibliographic and authority data for commercial use. Both companies have adopted their own original format.25

The National Library of China (NLC), on Mainland China, started to carry out a study on authority data in 1989, and it developed the draft of 中文图书名称规范数据款目著录规则 (Description Rule for Authority Data Entries), based on GARE, and the draft of 中国机读规范格式 (China MARC Format/Authorities; CNMARC/A), based on UNIMARC/Authorities, in 1990,26 both of which were revised in 1998.26 Production of Chinese name authority records in NLC were started in 1995.26 In 1997, NLC issued 中国机读规范格式使用手册 (The Handbook for CNMARC/A).27 In 2002, CNMARC/A was approved by the Ministry of Culture of the People’s Republic of China and became an industry standard (WH/T 15-2002).28 In 2003, NLC introduced the ALEPH500 integrated library system and authority data were successfully linked with bibliographic data.29

Li reports that in Tsinghua University Library, on Mainland China, authority control using authority cards began in 1994.30 Then, in 2002, Tsinghua University Library constructed a name authority file according to the INNOPAC Library System.31 However, according to Liu, by 2003, the majority of libraries in Beijing had not yet started to conduct authority control.32

The China Academic Library & Information System (CALIS), which started in 1998, is a nation-wide resource-sharing system among Chinese academic libraries.33 Its online union cataloging system was officially started in 2003, and it also includes an authority database.34

The first national cataloging code used on Mainland China was 中国文献编目规则 (Chinese Cataloging Rules) which was published in 1996. Part II of 中国文献编目规则 was for heading selections and forms. Rule 21.1.2 provided simple guidelines for constructing an authority file. These guidelines were deleted from the second edition of 中国文献编目规则, which was published in 2005, although rule 21.5 it stated that the “authority file” is one of the sources that determines the form of headings.35

In Taiwan, 中國編目規則 (Chinese Cataloging Rules) was first published in 1983.36 Rule 26.0.2 of it and its second edition, published in 1995, and its third edition, published in 2005, stated that “the form of reference entries used in bibliographic records should also be recorded in authority file so that they can be referred to when revising or deleting entries.”37, 38 中國機讀權威記錄格式 (Chinese MARC Format for Authority Records; CMARC/A) based on UNIMARC/Authorities was published in 1994 by the National Central Library (NCL).39

In South Korea, the National Library of Korea (NLK), Seoul National University Library (SNUL), Yonsei University Library (YUL), and Ewha Womans University Library are known for creating name authority data in different forms and at different levels.40, 41, 42, 43, 44 At NLK,
construction of authority data was started in 2000. According to Shim, public libraries in Korea have not instituted any form of authority control. Although SNUL, YUL, and the Ewha Womans University Library are participating in the integrated bibliographic database managed by Korea Education and Research Information Service (KERIS), KERIS does not have a national-level authority database; each of the libraries, therefore, has different forms of authority data, and there is difficulty in integration. Park and Lee indicate some reasons why authority data have not been standardized among Korean libraries. First, when library catalogs were computerized at the end of the 1980s, there were neither standardized catalog data to refer to nor a leading organization for computerization; hence, most of the libraries built a MARC database in their own way. Second, the third edition of Korea Cataloging Rules (KCR3) only had descriptive rules for monographs, and did not have rules for heading selections and forms. The fourth edition of Korea Cataloging Rules (KCR4), published in 2003, also had no rules for heading selections and forms. In addition, a particular form for an access point is not considered as a uniform heading in KCR4, because “different forms of an access point are connected to each other, and these terms are used for information retrieval.” Therefore, “a standard form for a heading need not be decided.” Although the KCR4 compilation report stated that the selecting of and the forms for access points should be defined by the authorities, further information including cross-references was not provided. The second edition of Korea Cataloging Rules (KCR2), published in 1966, on the other hand, had an instruction for cross-references although it did not mention constructing authority file. According to Kim, the first edition of Korea Cataloging Rules (KCR), published in 1964, imparted the same rules as KCR2 for heading selections and forms.

NLK asked the Society for Information Management to develop a draft of the KORMARC Format for Authority Data (KORMARC/A) in 1993, and the draft was approved as a Korean Standard (KS X6006-4) in December 1999. KORMARC/A is based on the USMARC Format for Authority Data and is widely used by Korean libraries.

2.2 Authority control of names in non-Latin languages

Both Names of Persons and Form and Structure of Corporate Headings aimed at ensuring that the authors’ names would be recorded in catalogs in a consistent way. Naturally, both gave little consideration to aspects of notational variation of names in non-Latin languages, which sometimes need to show several forms of names simultaneously. Although GARE and GARR stipulates that other language forms, variant spellings, variant transliterations, etc., can be recorded in the “see reference tracing” area (1.3.1.1), it does not provide any further guidelines about how these transliterations should be derived and recorded for names in non-Latin languages.
In 1979, American Library Association’s (ALA) Library and Information Technology Association (LITA) held two conferences entitled “Authority Control: The Key to Tomorrow’s Catalog.” Although Burns spoke about authority control in English-French bilingualism, problems regarding non-Latin languages were not dealt with in this conference.53

Both MARC 21/A and UNIMARC/A are currently the two major authority formats internationally accepted by libraries. Both indicate what kinds of authority data elements should be recorded in authority records and how to record notational variations of names in non-Latin languages. Recently, in accordance with RDA, both formats were revised and data elements increased. UNIMARC/A prepares the subfield $7 (Script of Cataloguing and Script of the Base Access Point), and $8 (Language of Cataloging and Language of the Base Access Point) that can identify the script, direction of the script, transliteration scheme, and the language for each access point. These correspond to the FRAD, which prepares the attributes “language,” “script,” and “transliteration scheme” for names and access points.

Unlike a cataloging code or implementation manual, FRAD is merely a conceptual model that relates each data element to its functions. It does not indicate which attributes and relationships are mandatory, nor does it provide an exhaustive list of all required authority data elements.54 RDA, on the other hand, is a cataloging code based on FRBR and FRAD and lists data elements that might be recorded in authority data. It also shows if each element is mandatory (i.e., a core element) or optional. As the attributes “language,” “scripts,” and “transliteration scheme” for access points are available in FRAD, RDA55 also gives some attention to variations of access points and shows some examples of access points in non-Latin languages.

Although ISAAR (CPF) states that parallel forms of the authorized form can be recoded (5.1.3),17 it does not provide any specific rules about recording notational variations of names.

Chinese-Japanese-Korean (CJK) scripts were implemented on the Research Libraries Information Network (RLIN) cataloging system originally operated by the Research Libraries Group (RLG) in 1983, and on the OCLC Online Computer Library Center’s (OCLC) system in 1986.56 RLG subsequently implemented Cyrillic script in 1986, Hebrew script in 1988, and Arabic script in 1991.56 OCLC added Arabic-script cataloging to its cataloging software in 2000, and by 2005, Cyrillic, Greek and Hebrew scripts were also introduced. OCLC added the Thai and Tamil scripts as of spring 2006.56 However, records of non-Latin data are generally catalogued with both original scripts and the Romanized equivalent, in case a system lacks non-Latin script capability.

Rules for using the MARC-8 character set were incorporated into the input functions of library utilities, such as OCLC and RLIN, and library vendor systems.57 MARC-8 was introduced in 1968 and was initially limited to essentially Latin script, although gradually it was expanded to include Arabic, Chinese, Cyrillic, Greek, Japanese, and Korean. 58 In 1998, the
Machine-Readable Bibliographic Information Committee (MARBI) of the Association for Library Collections & Technical Services (ALCTS), a division of ALA, agreed that it was acceptable for MARC 21 libraries to begin using the Unicode encoding scheme.Unicode, first released in 1993, has been jointly developed by the International Organization for Standardization (ISO) and the Unicode Consortium. Unicode is “a multilingual character set designed to combine the majority of the world’s writing systems and character set standards into a significantly larger repertoire of characters.” However, to facilitate the movement of records between MARC-8 and Unicode environments, it was recommended for an initial period that the character set repertoire be limited to those characters that could be expressed in MARC-8. According to Coyle, this “stalled the expansion of the MARC 21 standard to a wider use of vernacular expression of non-Latin languages.” Jacobs et al. claimed in 2004 that “integrated online library system vendors all claim to be on the road to Unicode, but display and input capabilities, in many cases, remain vaporware” and that “few libraries are able to mount multi-script catalogs that cover even the entire MARC-8 range.”

In 2001, OCLC started a major project to move its online union catalog called WorldCat to a new platform that supports Unicode. In January 2003, work began to move the WorldCat database. In 2002, LC also began planning for its transition to the Unicode standard for its MARC 21 bibliographic, holdings, and authority records. LC started the conversion of its records to Unicode in January 2003. In 2004, RLG began the transition from the Windows-based RLIN system to a new web-based system called RLIN21, in which data are stored in Unicode. In 2006, RLG merged with OCLC. In 2007, the restriction on the use of Unicode was no longer appropriate and the full Universal Coded Character Set (UCS) repertoire was valid for encoding MARC 21 records. Currently, both the MARC-8 character set and the UTF-8 Unicode character set can be selected when bibliographic and authority records are exported or imported by libraries via the OCLC Connexion interface. UTF-8 (UCS Transformation Format 8) is only one authorized Unicode encoding form for MARC 21 records.

The Romanization of names in non-Latin script languages has been an important issue of the Western library community. Catalog rules: author and title entries published in 1908 already included Romanization tables for Semitic, Sanskrit, Slavic Cyrillic languages, Russian, and modern Greek. Although the limitation of Romanization has been pointed out, the Romanization of names in non-Latin languages for access points was required by AACR and AACR2. This is understandable, because computerized systems in the 1970s used in Western countries could only handle data expressed in the Latin alphabet. LC provides ALA/LC Romanization Tables for languages in non-Latin scripts. Romanization schemes have undergone several modifications; for example, the replacement of
ALA/LC Romanization Tables are widely used by libraries in Western countries. However, as Vassie pointed out, the majority of Arabic-speaking users are unfamiliar with the ALA/LC Romanization Table, and so ALA/LC’s schemes are not always used in countries where the languages are spoken. For example, although the McCune-Reischauer system is used in the ALA/LC Romanization Table for Korean, a system proposed by the South Korean Ministry of Culture and Tourism (MCT) in 2000 has been adopted as the official Romanization system in South Korea. In another example, some libraries follow the Romanization rule of the Academy of the Hebrew language, which differs from the ALA/LC Romanization Table for Hebrew. These differences may be obstacles to conducting name identification using Romanized forms of names.

Another important issue on non-Latin languages is how to record original forms in authority data. Although LC’s authority cards had handwritten non-Latin script forms of names added to Romanized access points at least as early as 1920, it did not have the system capability to include non-Latin scripts for MARC authority data. Non-Latin original forms were permitted in variant access point fields of the LC/NACO Authority File (LCNAF) in 2008; however, the addition of non-Latin data was optional for NACO participants. Although many have argued for the necessity of original scripts, Romanized forms remain as the form for the authorized access points in authority data for non-Latin names in North America and in countries participating in NACO (e.g., United Kingdom, South Africa, Australia, New Zealand). This is natural as the language of cataloging in those countries is in Latin script.

Apart from LCNAF, libraries located in non-Latin alphabet countries have developed their own authority databases that can handle non-Latin scripts. For example, the Hong Kong Chinese Authority Name Workgroup (HKCAN) was set up in 1999 to establish a union authority database, because LCNAF does not show original scripts and is insufficient for their authority control. The Bibliotheca Alexandrina in Egypt started to build its local Arabic authority file in 2004 and aimed to construct an Arabic Script Union Catalogue and Authority File, which was to be called ANACO, Arabic Name Authority Cooperation.

In order to develop a basis for Chinese name authority data sharing, the Cooperative Committee for Chinese Name Authority (CCCNA) was established by the CALIS Administration Center, NLC, and the Joint University Librarians Advisory Committee (JULAC) in 2003. CCCNA has launched the Chinese Name Authority Joint Database Search System, which can search all authority records created by NLC, HKCAN, the Chinese Name Authority Database (CNAD) in Taiwan, and CALIS at the same time.

These authority databases constructed in countries using non-Latin scripts should ideally be utilized in Western countries because variant forms recorded in their original scripts are valuable.
when authority data sharing is conducted. In 2005, an agreement was signed between OCLC and JULAC, a consortium of academic libraries in Hong Kong, to make the HKCAN authority file available to OCLC Connexion clients. Unfortunately, the agreement was not renewed, and the HKCAN authority file has not been available in OCLC Connexion since April 7, 2013. One of the reasons for the non-renewal of the agreement was the inception and expansion of VIAF.

The VIAF system links authorized and variant forms of names and titles among the authority files of national bibliographic and other regional agencies. Users can search for and display names of a specified entity in various languages and scripts. Providing links between records in numerous languages and scripts allows users to search by any version of names including original form of names and Romanized names. Currently, national libraries in countries using non-Latin scripts, such as the Bibliotheca Alexandrina, NDL of Japan, and the National Library of Israel are participating in VIAF. However, the number of libraries in countries using non-Latin scripts that are involved in VIAF is relatively few. Although many of the VIAF contributors include non-Latin data as part of their authority files, the authority data of names in non-Latin languages are not yet adequately shared among the international community.

2.3 Writing systems in the Chinese character cultural sphere and its handlings in Western authority data

Since ancient times, there has been active intercommunication between people and books of countries in the so-called Chinese character cultural sphere. Within this sphere, an area including China, Japan, Korea, and Vietnam, people use or used Chinese characters as their official writing system. However, because the languages spoken in Japan, Korea, and Vietnam are different from that spoken in China, these countries did not simply adopt Chinese characters, but also modified their usage and character forms or invented new characters to write their own languages.

In Japan, hiragana and katakana syllabaries (known collectively as kana) were developed from Chinese characters during the early Heian period (794–1185). Since then, Japanese has been written in a mixture of Kanji (Chinese characters), hiragana, and katakana.

Korea has had its own script, called Hangul, since 1443, and has used both Hangul and Hanja (Chinese characters). Although Hanja is no longer commonly used in everyday communication, Korean children still learn 1800 Hanja characters at school.

The Chinese writing system adopts Hanzi characters. Since 1956, following the Chinese writing reform program, Chinese characters have been simplified in the People’s Republic of China, although Hong Kong, Taiwan, and Macau use traditional Chinese characters. Although traditional Chinese scripts can be transliterated to simplified Chinese scripts, the reverse is not necessarily true, because a simplified Chinese character subsumes several traditional characters.
For example, the simplified Chinese character “发” subsumes at least two traditional Chinese characters, “發” and “髮.”

The simplified script is less complex than the traditional script in that it contains fewer strokes. For example, the characters for “battle” are written as “战斗” in the simplified script and “戰鬪” in the traditional script. Both scripts are slightly different from Kanji or Hanja scripts. For instance, “battle” is written as “戦鬪” in Japanese Kanji and “戰鬪” in Korean Hanja.

The NDL’s authority cards include both an original (mainly in Kanji) form and its Romanized form of names. In 1979, 国立国会図書館著者名典拠録 (National Diet Library Authority File for Japanese Authors) was published in book form. In 国立国会図書館著者名典拠録, the Romanized form is recorded to show the precise yomi (pronunciation) of the name because the same Kanji may have different yomi, especially in proper names, and standardization of the yomi for the same person (sometime different yomi appear in different works by the same person) is important. In the second edition of 国立国会図書館著者名典拠録 published in 1991, yomi in katakana was used instead of Romanized forms. As yomi can be more precisely shown in katakana than in Latin alphabets, the current authority database of NDL records yomi both in katakana and in Latin scripts. In this study, the word yomi refers to yomi in katakana form.

The original forms of names, which have different forms depending on regions, even for the same person, are recorded as variant access points in authority data created by Western libraries. Authority records created by members of the CJK NACO project include such variant access points. The CJK NACO project is one of NACO funnel projects, which are groups of libraries that catalog specific subjects (e.g., art, law), languages (e.g., Arabic, Hebraica), or catalog for specific regions or locations (e.g., Alaska, Caribbean, East-Central-West Africa, Nevada) that contribute to the LCNAF together. Currently, the 27 participant institutions of the CJK NACO project contribute their CJK (Chinese-Japanese-Korean) authority records to the LCNAF. However, these contributions comprise less than 1 percent of all NACO records.

In Western countries, where the users are expected to read languages in the Latin scripts, names that would usually be in non-Latin scripts are Romanized. The Romanized form of name is recorded as an authorized access point in authority data with non-Latin scripts sometimes provided as variant forms of the name. It is in contrast to libraries in CJK countries, which regard Romanized letters as redundant, because their users expect the original script for CJK data, so access points in original forms (including yomi) provide adequate information to users in CJK countries. It should be noted that Romanized letters are not mandatory in the access points of several Online Public Access Catalogs (OPACs) in CJK countries, such as the NACSIS-CAT system in Japan, SNUL, and YUL in South Korea.
As Harrison points out, Romanization of the same Chinese character (an example would be a character meaning “forest”) can be different in Mandarin, Cantonese, Taiwanese, Japanese, Korean, and Vietnamese because of the differences in pronunciation. Even though some names in the Chinese character cultural sphere look the same because they have the same Chinese characters, their pronunciations vary from region to region. Because their Romanization is governed by the language being Romanized rather than the script being used, the resulting Romanized form is also different.

Although LC started to produce bibliographic records with Chinese and Japanese scripts in 1949 and Korean scripts in 1951, in the early 1950s, no standardized rule existed for cataloging CJK language materials in North American Libraries. Committees representing the American Library Association (ALA) and the LC amended the then-standard American national cataloging standards, the *ALA Cataloging Rules for Author and Title Entries* and the *Rules for Descriptive Cataloging in the Library of Congress* of 1949, to accommodate works written in East Asian languages. These efforts culminated in the *Preliminary Rules and Manual for Cataloging Chinese, Japanese, and Korean Materials* published in 1957, including the *Manual of Romanization, Capitalization, Punctuation, and Word Division for Chinese, Japanese, and Korean*. These rules were incorporated into the AACR of 1967 with the exception of the sections on Romanization, word division, and related items. Romanization rules for CJK languages have since been issued in the ALA/LC Romanization Tables. Following the 1957 rules, ALA/LC Romanization Tables imposed the following Romanization schemes: the Wade-Giles system for Chinese, the modified Hepburn system for Japanese, and the McCune-Reischauer system for Korean. In 1997, The Wade-Giles system was replaced by the *Hanyu pinyin* system.

It has long been debated whether North American libraries should adopt the Wade-Giles system or *Hanyu pinyin*, the official Romanization system in the People’s Republic of China since 1979, as the Romanization standard of Chinese Mandarin. The controversy finally ended in 1997 when LC negotiated the replacement of Wade-Giles by *Hanyu pinyin* and began a massive conversion project with the PCC libraries and OCLC.

In 1958, the Association for Asian Studies established the Committee on American Library Resources on the Far East, which preceded the Council on East Asian Libraries (CEAL). The CEAL Subcommittee on Technical Processing has been working closely with the LC to resolve problems of cataloging East Asian materials. For example, CEAL and LC collaborated to revise CJK examples of AACR2 and *Library Congress Rule Interpretations*. CEAL also provides CJK examples of RDA on its wiki. These works naturally include authority control issues.
2.4 Current conceptual models of authority data in non-Latin script languages and their inadequacies

To date, several authority data models have been developed for handling non-Latin script languages. This section reviews FRAD, Model A and Model B of MARC21/A, RDA, and the Dublin Core Metadata Initiative (DCMI) Abstract Model. These models are assessed for their ability to represent names in non-Latin script languages, and their deficiencies are identified.

2.4.1 FRAD model

FRAD was proposed in 2009 as a conceptual model for authority data. FRAD, which was approved by the IFLA, provides a framework for the analysis of functional requirements for authority data needed to support authority control and for the international sharing of authority data.$^{54}$ FRAD adopts the entity analysis technique of FRBR. The model defines 16 entities, their attributes, and relationships among the entities. Among the entities, Name and Controlled Access Point (CAP) are related by “is based on/is basis for.” “Has appellation/is appellation of” relates Name to Person or Corporate Body (i.e., a Person or a Corporate Body has a Name/Nomes). Based on the Name, a CAP is made. The model also defines four user tasks (i.e., Find, Identify, Contextualize, and Justify), which must be fulfilled by the authority data.$^{120}$

The FRAD concept was embraced by the ICP$^{121}$ and RDA$^{122}$ Although FRAD does not specifically focus on languages, it aspires to achieve global acceptance by designing elements such as “language,” “script,” and “transliteration scheme” for multi-language or multi-script records.

In FRAD, names expressible in several writing systems are processed by three elements: language, script, and transliteration scheme. Writing systems involve three name attributes (“language of name,” “script of name,” and “transliteration scheme of name”) and three CAP attributes (“language of base access point,” “script of base access point,” and “transliteration scheme of base access point”). The phrase transliteration scheme used here is synonymous with the Romanization scheme, because all examples for transliteration shown in FRAD are, actually, examples of Romanization. Further explanation of difference between transliteration and Romanization will be given in Chapter 10.

FRAD also applies four relationships in writing systems: “alternative linguistic form relationship” and “other variant name relationship” between names, and “parallel language relationship” and “alternate script relationship” between CAPs. The “alternative linguistic form relationship” includes other-language translations of names. For example, FRAD identifies an “alternative linguistic form relationship” between the names “Horace” in English and “Quintus Horatius Flaccus” in Latin (see Figure 2-1).$^{54}$p. 42$^2$ The “parallel language relationship” relates two or more CAPs for a given entity established in parallel languages. For example, as shown in
Figure 2-1 Alternate linguistic form relationship between Names defined in FRAD

Note. CAP—Controlled Access Point.

Figure 2-2 A parallel language relationship between CAPs defined in FRAD

Note. CAP—Controlled Access Point.
Figure 2-2, the CAP “Library and Archives Canada,” established in English, and the CAP “Bibliothèque et archives Canada,” established in French, share a “parallel language relationship.”\cite{54} The “alternative linguistic form relationship” appears for any translation of a name, while the “parallel language relationship” is used for a name in another official language.

The “alternate script relationship” relates two or more CAPs that are established as alternate linguistic scripts of the authorized forms of a name for a given entity. For example, the CAP “Gogol, Nikolai Vasilievitch,” expressed in the Latin alphabet, and the CAP “Гоголь, Николай Васильевич,” expressed in the Cyrillic alphabet, share an “alternate script relationship.”\cite{54} The “alternate script relationship” apparently includes transliteration, such as the relationship between “Гоголь, Николай Васильевич” and “Gogol’, Nikolai Vasil’evich” (Figure 2-3). Jin demonstrated this relationship using the example “Mencius,” which is related to “孟子.” However, in the present author’s understanding, this explanation is not correct, because “Mencius” and “孟子” are neither transcriptions nor transliterations of each other. Since “Mencius” is the English name of “孟子,” it seems more appropriate to suggest an “alternative linguistic form relationship” between these two names.

As specified in FRAD, the “other variant name relationship” between names includes “transliterations.” This relationship overlaps with an “alternate script relationship” (Figure 2-3). FRAD categorizes “transliterations” among “orthographic relationships,” alongside spelling, punctuation, and capitalization variations. However, transliterations are very different from spelling, punctuation, and capitalization variations because they require switching between writing systems, whereas spelling, punctuation, and capitalization variations occur within a single writing system.

FRAD is unequipped to differentiate between katakana names and their yomi in Japanese. Since both katakana and yomi are written in the same Japanese language and in the same katakana script, they are not readily differentiated by FRAD attributes. In other words, FRAD cannot represent the transcription relationship, and thus cannot represent the fact that yomi is a phonetic transcription of Kanji. Moreover, FRAD users perceive that all CAPs are equivalent. In fact, both yomi and Hepburn Romanization are mere derivations of the original script. To facilitate entity identification by users and machines, the parent-child relationship between a original name and its derivation should be separately shown.

Based on the above analysis, the FRAD model has limited capability to record names in non-Latin languages. Specifically, (1) transcriptions are not adequately represented by the attributes and relationships of names and CAPs in FRAD, (2) “other variant name relationship” and “alternate script relationship” overlap because both include the transliteration relationship,
and (3) transliterating are ambiguously treated as either orthographic relationships or Romanization.

Figure 2-3 Alternate script relationship between CAPs and other variant name relationship between Names defined in FRAD

2.4.2 Model A and Model B of MARC 21/A

MARC21/A utilizes two models that record authority data in multiple scripts: the vernacular and transliteration model (Model A), which uses fields 880, and the simple multiscript records model (Model B), which does not use fields 880. While Model A can show relationships among multiple scripts of the same name using 880, Model B merely shows one equivalent script form of an authorized access point using a 7XX heading linking entry field. Model B also allows use of 4XX only (i.e., without 7XX) for multiple script forms. According to Appendix C of MARC 21/A, “Model A is preferred if the same data is recorded in both the original vernacular script and transliteration”.123

Chan et al. proposed two authority models based on MARC 21/A for the HKCAN. In one model, field 1XX is repeated.86 The alternative model uses two different fields for Chinese
characters and LCNAF headings. Following a debate, the HKCAN decided to use field 1XX for
LCNAF headings and 7XX for Chinese scripts. Lam also reviewed the two MARC 21/A models and reported that, while Model A allows
better linkage between original scripts and their transliterations than Model B, it is not eagerly
supported by library vendors and bibliographic utilities. Thus, neither model is suitable for
global distribution. Lam suggested adopting and enhancing the MARC-XML format with the
“script” attribute added to the Field Link Control Subfield to differentiate multi-lingual
attributes of the corresponding fields. These “script” attributes take the form of
“script.language.romanization.” For example, the script attributes of the names “查良鏞” and
“Zha, Liangyong” are “cjk.chinese” and “latin.chinese.pinyin,” respectively. Lam’s idea of
adding Romanization code to authority data had been earlier proposed by Smith-Yoshimura,
who suggested that adding language/Romanization codes to the $w subfield of headings in
authority records would inform users of the most likely headings in overseas bibliographic
records.

Aliprand pointed out that in Model A of MARC 21/A, each 880 field is paired with the field
that contains the Romanization of the data in the 880 field and unlinked 880 fields containing
non-Latin scripts cannot exist. Appendix C of MARC 21/A states that “there may be unlinked
880 fields.” However, the existence of unlinked 880 fields in MARC 21/A instead of moving
these access points to the field 4XX or 5XX is unusual. She also argued that unlike
bibliographic data that allows Romanized data to be substituted for the original scripts, a paired
methodology in authority data is not needed because Romanization cannot be substituted for the
original script, and a cataloger should see the original script.

Models A and B of MARC 21/A are adopted by NDL in Japan and HKCAN, respectively.
Figures 2-4 and 2-5 illustrate authority records from the NDL manual and the HKCAN
Database OPAC, respectively. To simplify the figures, the 0XX and 6XX fields (displayed in
the actual record) are deleted.

In Figure 2-4 (Model A), the Kanji names reside in regular fields (100/400/500), and their
corresponding phonetic representations and scripts ($yomi$ and Romanizations) reside in fields
880. The subfield $s6$ includes $s6$[linking tag]-[occurrence number]/[script identification code].
The linking tags contain the tag number of the associated field, and the occurrence numbers
show sets of associated fields. For example, the fields containing “中島, 梓,” “ナカジマ, アズ
サ,” and “Nakajima, Azusa,” form a set because they share the occurrence number “01.”
Alternative scripts found in a field are marked with script identification codes (specified in
Appendix A of MARC 21/A).
In Model A, only six types of script identification codes are available: Arabic, Latin (encoded “(B”, as in Figure 2-4), Chinese/Japanese/Korean (encoded “$1”, as in Figure 2-4), Cyrillic, Greek, and Hebrew. Note that CJK languages are assigned a single code “$1”. As noted earlier, at least two types of scripts are adopted in all three languages, but these cannot be distinguished by MARC 21/A’s identification codes.

Besides being unable to differentiate different scripts, the code cannot differentiate between languages. Therefore, it cannot extract Chinese or Japanese alone from authority databases for any purposes.

In Figure 2-5 (Model B), “Jin, Yong” in field 100 is the pinyin form of “金庸” in field 700. Fields 100 and 700 form a set, and corresponding relationships exist between them. Many other forms of names are retrieved in the 400 fields. As reported by Lam, Model B cannot identify relationships among multiple scripts of the same name, i.e., only one equivalent script can reside in 700. If multiple alternate forms exist, such as yomi and Kanji, only one form is selected for field 700 in order to make the relationship with the 100 field more explicit. Moreover, the corresponding relationships among the 400 fields are excluded in this model. For example, “Zha, Liangyong” is the pinyin form of “查良鏞,” but this relationship does not appear in the record. All forms of his name, including the English name “Cha, Louis,” the real name “査良鏞,” and several Romanizations of Chinese characters are treated equivalently.

In summary, neither of the MARC 21/A models can properly record names in non-Latin languages. The following problems were identified: (1) the script identification codes only distinguish limited kinds of scripts; in particular, they cannot distinguish among scripts in CJK languages; (2) Model B displays only one corresponding script of the authorized heading; and (3) the corresponding relationships among the references are invisible in Model B.
2.4.3 RDA model

RDA stipulates rules for recording data based on FRBR and FRAD. Obviously, RDA itself is not a data model. However, Taniguchi argued that RDA includes several elements for accommodating additional information to suit modern cataloging practices and can be recognized as having a model of its own. RDA-based models (a term coined by Taniguchi) are slightly different from FRBR/FRAD models.

While FRAD refers to a single CAP, RDA distinguishes two types of controlled access points: an Authorized Access Point and a Variant Access Point. According to RDA 9.19 and 11.13 on constructing access points for persons and corporate bodies, respectively, an Authorized/Variant Access Point for Persons/Corporate Bodies is based on the Preferred/Variant Name of the Person/Corporate Body. Only one Preferred Name is chosen for a person or a corporate body, and thus, only one Authorized Access Point exists for a person/corporate body. On the other hand, as many Variant Access Points as are needed can be constructed.

Names with several writing systems are treated in two ways: Names Found in a Non-preferred Script (RDA rules 9.2.2.5.3/11.2.2.12) and Alternative Linguistic Form of Name (RDA rules 100 1 $aJin, Yong, $d1924-
400 1 $aChin, Yung, $d1924-
400 1 $aZha, Liangyong, $d1924-
400 1 $a查良鏞, $d1924-
400 1 $aCha, Louis, $d1924-
400 1 $aCha, Liang-yung, $d1924-
400 0 $aKim-Dung, $d1924-
400 1 $aKim, Dung, $d1924-
400 0 $aJinyong, $d1924-
400 1 $aYong, Jin, $d1924-
400 1 $aKin, Yō, $d1924-
400 1 $aLin, Huan, $d1924-
400 0 $aKimyong, $d1924-
400 0 $aKim Yong, $d1924-
400 1 $a林歡, $d1924-
400 1 $aYao, Fulan, $d1924-
400 1 $a姚馥蘭, $d1924-
400 1 $aYao, Jiayi, $d1924-
400 1 $a姚嘉衣, $d1924-
700 1 $a金庸, $d1924-

Figure 2-5 Model B. Sample authority record from the HKCAN Database OPAC (retrieved 2014-01-08)
9.2.3.9/11.2.3.6). According to RDA 9.2.2.5.3, “if the name of a person is found in a script that differs from a preferred script of the agency creating the data, transliterate the name according to the scheme chosen by the agency” as the Preferred Name. An example is “Yi Sŭng-man” as the Preferred Name for “李承晩,” and “if the name recorded as the preferred name for a person has one or more alternative linguistic forms, record them as variant names” (RDA 9.2.3.9). Rule 9.2.2.5.3 results in data that could be used in FRAD’s “other variant name relationship”, while names recorded according to RDA rule 9.2.3.9 could be used in FRAD’s “alternative linguistic form relationship” between Names.

In contrast to FRAD, which identifies relationships among access points, writing system relationships in RDA are identified only among Names, and access points are constructed for each Name. Thus, Names includes all variations in the writing system, in addition to pseudonyms, nicknames, and other name forms. However, this tenet is inconsistent with practices in CJK countries, where Romanized letters and phonetic versions of a name (yomi) are considered only as additions to access points.

Similar to FRAD, RDA does not properly implement transcriptions. The definition of the word “transcription” will be explained more in Chapter 10. In RDA, the word “transcription” is limited meaning “copying.” For example, RDA 0.11.2 stipulates “when the instructions for an element specify transcription, data are transcribed in the language and script in which they appear on the source of information from which the data are taken”. This is “transcribed” from the resource or other source of information, rather than re-writing the words in another script or form.

In summary, RDA encounters the following problems when recording non-Latin script names: (1) variations in writing system and name are treated similarly, (2) RDA excludes transcriptions such as Japanese yomi, and (3) the word “transcription” is defined in the sense of “copying.”

2.4.4 DCMI Abstract Model

The DCMI Abstract Model (DCAM) specifies the components and constructs used in Dublin Core Metadata. DCMI is an open organization that is managed as a project of the Association for Information Science and Technology. DCMI maintains a large set of metadata vocabularies and technical specifications. In DCAM, a record contains description sets containing one or more descriptions composed of statements. Each statement denotes a property-value pair comprising a property URI and a value surrogate. The value surrogate is either a literal value surrogate composed of a single value string or non-literal value surrogate containing either zero or a value URI. According to Zeng and Zumer, this model allows processing, exchanging, referencing, and linking of data at the statement level. When a record contains resource descriptions, the individual descriptions can also be linked to the
authority data that manage the values associated with those properties. Since a resource in DCAM can be any identifiable entity, such as bound books, concepts, or human beings, the model can describe not only components and constructs of bibliographic data, but also authority data itself.

Miyazawa reported that multiple value strings lack appropriate language tags for representing parallel writing. Although Miyazawa has clarified this problem well, DCAM is re-examined here to establish whether the problem persists.

DCAM specifies that a “value string may have an associated value string language” that is an ISO language tag (for example en-GB). The Best Current Practice (BCP) 47, published by the Internet Engineering Task Force, defines Tags that use ISO language code plus subtags, such as “en-GB”. The Request for Comments (RFC) 5646 document for Dublin Core, which replaced RFC 4646 in 2009, combines BCP 47 and RFC 4647. Valid subtags (according to RFC 5646) are registered in the Language Subtag Registry, maintained by the Internet Assigned Numbers Authority (IANA).

In RFC 5646, a language tag forms a sequence of one or more subtags, which are of various types, such as language, script, and region. Table 2-1 lists the subtags related to CJK languages in the current Language Subtag Registry. Language tags such as “zh-Latn-CN” (denoting Latin scripts of Chinese language in China) or “zh-yue-Hant” (denoting Traditional Chinese scripts of Cantonese) are compiled from the subtags in Table 2-1.

As Miyazawa has identified, although the RFC 5646 clearly distinguishes between different scripts in CJK languages (as shown in Table 2-1), it does not distinguish yomi from katakana or between two different Romanization schemas in the same language.

Another problem exists in DCAM. Since a value surrogate is compiled from several value strings, and is used to construct a statement, value strings under a statement are conjugated or related. This means that a name expressed in several writing systems may appear as a pair or set. However, because each value string with different language tags is treated equally, the model cannot imply that yomi or Romanized letters are merely additional to the access point.

In summary, the DCAM (1) can describe script variations, but cannot differentiate between yomi and names in katakana, or between two Romanization schemes, and (2) assigns equal status to names in different scripts and cannot differentiate primary (or authorized) names from their derivations.
Table 2-1 *Subtags* related to CJK languages registered in the Language Subtag Registry

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>zhx</td>
<td>Chinese (family)</td>
</tr>
<tr>
<td>language</td>
<td>zh</td>
<td>Chinese</td>
</tr>
<tr>
<td>language</td>
<td>och</td>
<td>Old Chinese</td>
</tr>
<tr>
<td>language</td>
<td>cmn</td>
<td>Mandarin Chinese</td>
</tr>
<tr>
<td>language</td>
<td>cdo</td>
<td>Min Dong Chinese</td>
</tr>
<tr>
<td>language</td>
<td>cly</td>
<td>Jinyu Chinese</td>
</tr>
<tr>
<td>language</td>
<td>cpx</td>
<td>Pu-Xian Chinese</td>
</tr>
<tr>
<td>language</td>
<td>czh</td>
<td>Hui Chinese</td>
</tr>
<tr>
<td>language</td>
<td>czo</td>
<td>Min Zhong Chinese</td>
</tr>
<tr>
<td>language</td>
<td>gan</td>
<td>Gan Chinese</td>
</tr>
<tr>
<td>language</td>
<td>hak</td>
<td>Hakka Chinese</td>
</tr>
<tr>
<td>language</td>
<td>han</td>
<td>Xiang Chinese</td>
</tr>
<tr>
<td>language</td>
<td>ltc</td>
<td>Late Middle Chinese</td>
</tr>
<tr>
<td>language</td>
<td>lzh</td>
<td>Literary Chinese</td>
</tr>
<tr>
<td>language</td>
<td>mnp</td>
<td>Min Bei Chinese</td>
</tr>
<tr>
<td>language</td>
<td>nan</td>
<td>Min Nan Chinese</td>
</tr>
<tr>
<td>language</td>
<td>wuu</td>
<td>Wu Chinese</td>
</tr>
<tr>
<td>language</td>
<td>yue</td>
<td>Yue Chinese</td>
</tr>
<tr>
<td>language</td>
<td>jpx</td>
<td>Japanese (family)</td>
</tr>
<tr>
<td>language</td>
<td>ja</td>
<td>Japanese</td>
</tr>
<tr>
<td>language</td>
<td>ojp</td>
<td>Old Japanese</td>
</tr>
<tr>
<td>language</td>
<td>ko</td>
<td>Korean</td>
</tr>
<tr>
<td>language</td>
<td>oko</td>
<td>Old Korean (3rd-9th cent.)</td>
</tr>
<tr>
<td>language</td>
<td>okm</td>
<td>Middle Korean (10th-16th cent.)</td>
</tr>
<tr>
<td>script</td>
<td>Hani</td>
<td>Han; Hanzi; Kanji; Hanja</td>
</tr>
<tr>
<td>script</td>
<td>Hant</td>
<td>Han (Traditional variant)</td>
</tr>
<tr>
<td>script</td>
<td>Hans</td>
<td>Han (Simplified variant)</td>
</tr>
<tr>
<td>script</td>
<td>Hira</td>
<td>Hiragana</td>
</tr>
<tr>
<td>script</td>
<td>Kana</td>
<td>Katakana</td>
</tr>
<tr>
<td>script</td>
<td>Hrkt</td>
<td>Japanese syllabaries (alias for Hiragana + Katakana)</td>
</tr>
<tr>
<td>script</td>
<td>Jpan</td>
<td>Japanese (alias for Han + Hiragana + Katakana)</td>
</tr>
<tr>
<td>script</td>
<td>Hang</td>
<td>Hangul; Hangüel; Hangeul</td>
</tr>
<tr>
<td>script</td>
<td>Kore</td>
<td>Korean (alias for Hangul + Han)</td>
</tr>
<tr>
<td>script</td>
<td>Latn</td>
<td>Latin</td>
</tr>
<tr>
<td>region</td>
<td>CN</td>
<td>China</td>
</tr>
<tr>
<td>region</td>
<td>HK</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>region</td>
<td>MO</td>
<td>Macao</td>
</tr>
<tr>
<td>region</td>
<td>TW</td>
<td>Taiwan, Province of China</td>
</tr>
<tr>
<td>region</td>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>region</td>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>region</td>
<td>KP</td>
<td>Democratic People’s Republic of Korea</td>
</tr>
</tbody>
</table>

Notes

3 The “authority cards” rule on cards for the Library of Congress Cataloging Rules (Suppl.), dated Dec. 16, 1907 (amended Feb. 7, 1919) was obtained from Dr. Barbara B. Tillett via a personal correspondence on June 13, 2015. There is a note referring to “replacing rules of 1899 and May 5, Aug. 8, Nov. 19, 1902)” on the cards, which means this rule dates from at least as far back as 1899.
7 図書館情報学ハンドブック編集委員会編. 図書館情報学ハンドブック. 第 2 版, 丸善, 1999, 1145p.
15 IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN.
“Mandatory Data Elements for Internationally Shared Resource Authority Records”.
16 Guidelines for Authority Records and References. 2nd ed., K.G. Saur, 2001, 46p, (UBCIM
18 Society of American Archives. “Encoded Archival Context: Corporate Bodies, Persons, and
21 日本国書館協会目録委員会編. 目録の作成と提供に関する調査報告書(2010 年調査).
22 Miyazawa, Akira. “Basic concept of authority control in NACSIS-CAT”. 日本語, 中国語,
24 青年図書館員聯盟目録法制定委員会編. 日本目録規則: N.C.R. 大阪, 関宮商店, 1943,
123p.
25 国立国会図書館書誌部編. 第 4 回書誌調整連絡会議記録集: 名称典拠のコントロール.
26 Zhou, Shengheng. “Chinese name authority in China”. 日本語, 中国語, 韓国語の名前典拠
101-104.
18-22.
30 李凤侠. 香港城市大学图书馆与清华大学图书馆书目规范控制的比较. 图书馆建设.
66-68.
46 심정. 국내 공공도서관의 인명 전기제어의 현황 및 발전 방향. 한국문헌정보학회지. 2006, vol. 40, no. 4, p. 221-244.
47 김미향. 대학도서관의 협력형 전기제어 방안 연구. 국립대학도서관보. 2009, no. 27, p. 41-70.
48 Park, Hong-Seok. “The revision of the cataloging rule (KCR4) and authority control under the new rule”. 日本語, 中国語, 韓国語の名前典拠ワークショップ記録. 東京,
2002-03-14/18, 国立情報学研究所, 2002.

49 韓國圖書館協會目錄委員會編. 韓國目錄規則. 第 4 版, 韓國圖書館協會, 2003,
576p.


51 日韓業務交流報告「国立中央図書館の人名典拠コントロールの現況及び課題」. NDL
書誌情報ニュースレター. 2012, no. 4,
2015-06-25).

52 Ki, Min-Do, Lee, Jae-Sun. “Name authority data by KORMARC”. 日本語, 中国語, 韓国

53 Ghikas, Mary W. Authority Control: the Key to Tomorrow’s Catalog: Proceedings of the
1979 Library and Information Technology Association Institutes. Phoenix, Oryx Press, 1982,
194p.

54 IFLA Working Group on Functional Requirements and Numbering of Authority Records
(FRANAR). Functional Requirements for Authority Data. as amended and corrected through
2014-08-27).

55 American Library Association; Canadian Library Association; Chartered Institute of Library
2015-05-07).

http://www.ala.org/alcts/sites/ala.org.alcts/files/content/ianda/nonenglish/07marchrpt.pdf,

57 Coyle, Karen. Managing technology: Unicode, the universal character set part 2, Unicode in

58 Cain, Jack. “Assessment of options for handling full Unicode character encodings in MARC

59 “Proposal no: 98-18, Unicode identification and encoding in USMARC records”. MARC
2015-06-22).


82 LC’s authority card for an authorized access point “China. Chün p’u tsung chü,” created in 1920, had a handwritten original script, “濬浦總局.” The card was obtained from Dr. Barbara B. Tillett via a personal correspondence on June 13, 2015.
103 Sakai, Kiyohiko; Kyoto, Toru. "NACSIS-CAT author name authority record: its function and processing". 日本語, 中国語, 韓国語の名前典拠ワークショップ記録. 東京, 2001-01-10/11, 国


Chapter 3

Framework of name authority data

In this chapter, the author proposes a new framework of name authority data that includes representations, data elements, and data structures. The reason behind the author’s adoption of the concept of “representations” for this framework is explained in this chapter. In the explanation, the characteristics of personal names in the Chinese character cultural sphere are given an overview in the first half of this chapter. Due to the variance of scripts used in this cultural sphere, the personal names in this sphere are good examples with which to explain the complexity of handling authority data in non-Latin languages all over the world. Characteristics of the names of corporate bodies are not discussed here. However, the review of personal names is sufficient for demonstrating how “representations” in this cultural sphere are important for authority data.

3.1 Characteristics of personal names in the Chinese character cultural sphere

3.1.1 Chinese names

3.1.1.1 Current Chinese names

In the wake of the formations of the Qin and Han dynasties, the common folk in China started to have surnames. Nowadays, the number of Chinese surnames is determined to be approximately 23,813 in Mainland China. Of these, 6,931 surnames consist of one Chinese character, while the others consist of 2 to 10 characters (including the surnames of ethnic minorities). A high percentage of the population is concentrated in particular surnames. As of 2013, 56.61% of the total population (1.3 billion persons) share only 23 kinds of surnames. Of these, three surnames, namely “王,” “李,” and “张,” account for approximately 21% of the population.

In People's Republic of China, the Marriage Law promulgated in 1950 stipulated that married couples can have surnames that are independent from one another’s. Since then, married couples can keep separate surnames or retain one surname based on their preference in Mainland China. Adding a husband’s surname to the beginning of a wife’s surname, or its reverse, is also allowed. As for children’s surnames, they can inherit their fathers’ surnames or their mothers’ surnames (Article 22 of the Marriage Law). However, in reality, most married couples have separate surnames and most of their children inherit their fathers’ surnames.

As of 2014 in Taiwan, there are 1,510 surnames. Of these, 1,396 surnames consist of one Chinese character and 114 surnames consist of two characters. 52.77% of the total population (about 23 million persons) share only 10 kinds of surnames.
The civil law of the Republic of China established in 1930 stipulated that married women should add their husbands’ surnames to the beginning of their surnames, although separate surnames are also allowed, and that children should inherit their fathers’ surnames as a general rule. The rule was revised in 1998, and since then, married couples have generally had independent surnames in Taiwan. Adding a husband’s surname to the start of a wife’s surname, or its reverse, is also allowed. Children can inherit their fathers’ or mothers’ surnames through their parents’ cooperative consultation. In fact, only 5.41% of the total population have added surnames, and most children inherit their fathers’ surnames.

Because Hong Kong is a former British colony, most people in Hong Kong have English names as well as Chinese names. Generally, these English names are not a Romanization of the Chinese names or even official names, but given by the family or the individuals themselves.

In Mainland China, the simplification of Chinese characters was implemented in the late 20th century. As a result, simplified Chinese characters are currently used in Mainland China. On the other hand, Hong Kong, Taiwan, and Macau, which were not under the influence of the Chinese Communist Party during that time, currently still use traditional Chinese characters. Moreover, there are some minor differences between the traditional Chinese characters of Hong Kong and Taiwan.

3.1.1.2 Romanized representations of Chinese names

Basically, each Chinese character is pronounced in only one way. However, some characters have two or more pronunciations. These characters are often pronounced differently when they appear as surnames due to the effect of old-time sounds or dialects.

Pronunciations of Mandarin are represented by Hanyu pinyin, which is an official Romanization scheme established as 汉语拼音方案 (Scheme of the Chinese Phonetic Alphabet) by People's Republic of China in 1958. Since 1979, this has been the most pervasive Romanization scheme of Mandarin used for Chinese geographical and personal names by both the government of People's Republic of China and the United Nations.

Before the prevalence of Hanyu pinyin, the Wade-Giles Romanization system developed by Thomas F. Wade and amended by Herbert Giles in the late 19th century was used internationally.

In Taiwan, many systems, such as the Wade-Giles, Chinese postal map, and Yale Romanization systems, exist along with Hanyu pinyin. In 2008, 中文譯音使用原則 (The Principle of Chinese Transcription) stated that Hanyu pinyin is the official Romanization system of Taiwan. However, for personal names, the aim is to “primarily respect personal will;” for example, Hanyu pinyin is not imposed (but recommended) by the government for the names displayed on passports.
Because dialects of Chinese languages are pronounced differently, their Romanized forms differ from that of Mandarin. Therefore, the Romanized spelling of a name written in Chinese characters depends on the spoken dialect.\textsuperscript{19,20} As of 2011 in Hong Kong, 89.5\% of the total population at age 5 or over (about 6.8 million persons) speak Cantonese as a native language, and 46.5\% of the total population speak Mandarin as a non-native language.\textsuperscript{21} Yu reported that approximately 90\% of the university students in Hong Kong do not understand Mandarin or Hanyu pinyin.\textsuperscript{22} The Romanization scheme for Cantonese is varied and not unified. For example, among other systems, the Yale Romanization system developed in 1956 has been the most pervasive method of teaching Cantonese, while the Cantonese pinyin developed by the Education Department of Hong Kong in 1988 has been used for teacher training in Hong Kong.\textsuperscript{23}

3.1.2 Japanese names  
3.1.2.1 Current Japanese names  
Since 1875, all Japanese people have had both surnames and given names. Nowadays, the number of Japanese surnames is determined to be approximately 290,000,\textsuperscript{24} which is significantly more than the 23,813 surnames in Mainland China and the 286 surnames in Korea.\textsuperscript{25}  

*Kanji* is used to write a large majority of Japanese surnames; a smaller number is written in *kana* (hiragana or katakana).\textsuperscript{24} Scripts for given names are officially recognized by *戸籍法施行規則 (The Ordinance for Enforcement of the Family Registration Law)*\textsuperscript{26} for hiragana, katakana, and the 2,998 characters\textsuperscript{27} of *Kanji*. In these characters of *Kanji*, 2,136 characters are specified in *常用漢字表 (The National List of Chinese Characters in Common Use)*,\textsuperscript{28} and the others are specified in an appended table of *戸籍法施行規則.*  

*常用漢字表* is a national list of Chinese characters in common use, which came into effect through Cabinet Notification No. 1 of 1981.\textsuperscript{29} This list is based on *当用漢字表 (The List of Chinese Characters in Daily-Use)*, a system consisting of 1,850 characters that was established in 1946 for the purpose of simplifying the tasks of reading and writing Chinese characters.\textsuperscript{30} Approximately 500 characters in *当用漢字表* were simplified compared to their traditional forms. Characters included in *当用漢字表* at this time are called *shinjitai (新字体)*, which literally means “new letter shapes,” and those used before the establishment of *当用漢字表* are called *kyūjitai (旧字体)*, which literally means “old letter shapes.”\textsuperscript{31}  

As *当用漢字表* was a list designed to restrict the number and forms of Chinese characters in general usage, it was inconvenient and inadequate for conveying Japanese people’s names. In order to lessen the public discontent with this system, an appended table designated specifically for personal names was established for *Kanji*, to which Chinese characters were added and
continually updated. In 1981, 当用漢字表 was abolished, and 常用漢字表 was established instead. 常用漢字表 was merely a guideline set by the Japanese government and, furthermore, did not restrict the use of characters. However, with regards to naming, only the characters listed in 常用漢字表 and the appended table of 戸籍法施行規則 remained officially permitted for use in given names. In 2010, 常用漢字表 was revised and characters were added. The appended table of the latest 戸籍法施行規則 shows characters specifically designated for personal names, along with itaiji (異体字, meaning variant characters). Itaiji have different character forms from those of 常用漢字 (Kanji in 常用漢字表), although their pronunciations and meanings are the same as in 常用漢字. 常用漢字表 includes eight characters of Kokuji (国字) which were original Chinese characters invented in Japan. As regards Japanese surnames, the use of characters is not restricted.

The characters of Japanese Kanji and kana may also be pronounced in multiple ways. For example, the Kanji “紅” has at least three pronunciations: “ko,” “beni,” and “kurenai.” Moreover, although each kana character is usually pronounced in a single way, exceptions exist. For example, “ほ” is pronounced as both “ho” and “o.” This dichotomy arose from the Japanese orthographic reforms following World War II. Although “ほ” is pronounced as “ho” in modern kana orthography, some personal or corporate names adopt the historical kana orthography, in which a character sequence (two or more characters) corresponds to a single sound. For example, the name “平塚, らいてう” reads as “Hiratsuka, Raichō.” Although the kana pronunciation of “て” is “te,” and “う” is generally pronounced as “u,” the character sequence of “てう” reads as “chō.” The pronunciation of Kanji character sequences is more complicated. For example, the Kanji “紅” is pronounced as neither “mo” nor “momi,” but the character sequence of “紅葉” reads as “もみじ.” In kana (hiragana or katakana), Kanji readings are often placed alongside each character to indicate the character pronunciation. These guides are called furigana. People must provide furigana for their names when they submit birth registration forms or business applications to show how these names should be pronounced in Japanese.

Since multiple pronunciations exist in kana, amended furigana called yomi are added to the access points of bibliographic and authority records in Japan. The yomi (recorded in katakana) serve to standardize and collocate access points. Figure 3-1 is an example of the same original form of names with different yomi. On the other hand, Figure 3-2 is an example of the same yomi form with different original forms.

The original and yomi forms of names in katakana may be identical in instances such as “ケンドーコバヤシ [Kendo Kobayashi]” or “サトウ, ヒロ [Sato Hiro].” However, in some cases such as “ウメマツ, カオル [Umematsu Kaoru],” since “ヲ” is a katakana pronounced as “オ [o],” the original and yomi forms (“ウメマツ, カオル”) are different. As aforementioned, for
Figure 3-1 Examples of the same original form with different yomi in Web NDL Authorities

Figure 3-2 Examples of the same yomi with different original forms in Web NDL Authorities
Japanese personal and corporate names represented in Kanji, yomi information is considered to be very important. In general, even when the same Kanji are used, names can identify different individuals when their yomi are different.

3.1.2.2 Romanized representations of Japanese names

There are several types of Japanese Romanization systems, and conflict between the advocates of the rival systems has continued to this day. Two main Romanization systems are used today in Japan: one is the Hepburn system, which is also called the modified Hepburn system in Western countries, and the other is the kunrei-shiki (訓令式) system. Kunrei-shiki is an official system that was originally designated in Cabinet Notification No. 1 of 1954 as the so-called ローマ字のつづり方 (A method of Writing Japanese in Roman Characters). However, ローマ字のつづり方 states that spelling may also be determined by the Hepburn system when “international relations and situations with prior precedent in which a sudden spelling reform would be difficult” are involved.

On the other hand, Japanese people are obliged to Romanize their names by using the Hepburn system on their passports, which is in line with 旅券法施行規則 (The Ordinance for Enforcement of the Passport Act). The Hepburn system set forth by the Ministry of Foreign Affairs of Japan for Japanese passport applications is slightly different from the usual Hepburn system; for example, it does not use macrons for long vowels. In fact, as “the Hepburn system can be reasonably considered more as a set of principles, something that serves as the bedrock of a specific Romanization method, rather than as a fixed set of rules,” it is not possible to determine the usual or definitive Hepburn system. It can be said that no unified rules govern the Romanization systems for Japanese nomenclature.

3.1.3 Korean names

3.1.3.1 Current Korean names

In the Korean Peninsula, the use of Chinese-like surnames began at the earliest in the mid-6th century. Before the Joseon dynasty (1392-1910), a limited number of people, such as nobles, had surnames, and the common folk had first names only. In 18th century, however, 70-80% of the total population had surnames and were registered in the official family registry.

In the Korean Peninsula, 훈민정음 (訓民正音, meaning “the present Hangul”) was established in the 15th century, after which it permeated gradually into the general public's daily language. Hangul took the lead in written word instead of Chinese characters at the end of the 19th century. In South Korea, the law for the exclusive use of Hangul was proclaimed for the purpose of diminishing the percentage of illiterates in 1948 and it accelerated the sole use of Hangul by the people. Nowadays, Chinese characters are hardly used in the social lives of
Korean people.

However, the Chinese character culture still exists in their personal names. According to Article 63 of 가족관계의 등록 등에 관한 규칙 (The Rule of Registration About Family Relations), the Hanja and Hangul of one’s name should be shown together in a family register called 가족관계등록부 (family-related directory). After some debates, it was decided that the Hanja name would be shown as well as the Hangul name on the resident registration cards issued to all nations by the Resident Registration Act because too many different people have the same Hangul name.

Although the 호적법시행규칙 (Ordinance for Enforcement of the Family Register Act) prohibited the use of simplified forms of Hanja and symbols for personal names, the usable range of Hanja was not defined until 1990. Thus, some people used Hanja with too many strokes or even created new Hanja that were not in the dictionary, which brought difficulty to administrative processing. Therefore, the revision of 호적법 (The Family Register Act, law No. 4298) dated December 31, 1990 proclaimed that children should be named with Hangul or Hanja that are routinely used, and that the usable range of Hanja is defined in 대법원규칙 (The Chancery Rule). The complete revision of the Ordinance for Enforcement of the Family Register Act that was issued as the Chancery Rule No. 1137 established the 2,731 Hanja characters that could be chosen for personal names. These include the 1,800 characters of “교육용 기초한자 (basic Hanja for educational use)” defined in 1972 by the Ministry of Education, plus 931 additional characters. On March 21, 1991, the Act was revised again, and the number of Hanja characters was established as 2,856 characters. Currently, the number has increased to 5,151 characters, as some original Hanja invented in Korea, such as “曺” and “乭”, have also been included.

Hanja is an ideogram, and Hangul is a phonogram. In contrast to Japanese Kanji, most Korean Hanja are pronounced in one way only. In other words, each Hanja corresponds with only one Hangul in many cases. Although some Hanja may be pronounced in multiple ways, only the pronunciations shown in Appendix 1 of 가족관계의 등록 등에 관한 규칙 and 교육용 기초한자 are allowed to be used for personal names. Thus, we can easily transliterate Hanja to Hangul. However, the reverse does not apply, because many Hanja share the same sounds (i.e., the same Hangul) and several Hanja candidates exist for a given Hangul name. Consequently, many Korean personal names are common in Hangul but vary in Hanja (see Figure 3-3).

Although most of the South Korean given names are Chinese-derived Sino-Korean words that use Hanja, some names are coined from native Korean words that cannot be represented in Hanja. Some names are even coined from foreign words. In these cases, names are represented in Hangul only.
Figure 3-3 Examples of the same Korean personal name represented in Hangul versus Hanja

On the other hand, all surnames can be represented in both Hanja and Hangul. There are 286 surnames in South Korea. Similar to China, a high percentage of the population is concentrated in particular surnames. Five surnames, namely “김 (金), ” “이 (李), ” “박 (朴), ” “최 (崔), ” and “정 (鄭), ” account for 53.9% of the total population (45,985,289 persons). There are 13 surnames consisting of two-syllable, such as “남궁 (南宮)” and “황보 (皇甫);” the number of people with these surnames account for only 0.0941% of the total.

In South Korea, the initial sound rule (두음법칙) is applied for Sino-Korean words, including Korean personal names. The rule is that when a Sino-Korean word begins with sound [r] (“ㄹ” script) followed by sound [i] or sound [j] (either with the script of ㅣ, ㅑ, ㅕ, ㅖ, or ㅛ), “ㄹ” should be changed to “ㅇ;” otherwise, “ㄹ” should be changed to “ㄴ.” Similarly, a Sino-Korean word that begins with [ny] [nyeo], [nyo], [neu], or [ni] should be converted to “ㄹ” [yeo], ㅗ [yo], ㅗ [eu], or 오 [yi], respectively. On the other hand, in Democratic People's Republic of Korea (North Korea), this rule is not applied. Thus, for example, “李” should be written and pronounced as “이” in South Korea and “리” in North Korea, and “盧” should be written and pronounced as “로” in South Korea and “로” in North Korea. This shows that even the same Hanja name might be represented in different Hangul in South and North Korea.

3.1.3.2 Romanized representations of Korean names

Even if some Korean personal names share the Hangul or Hanja forms, their Romanization may be different. One reason for this is that the Romanization scheme developed by the Korean government has been changed many times. The scheme issued in 1948, 1959 (which is the same as the system of the Korean Language Society in 1940), and 1984 (which is the same as the McCune-Reischauer system), as well as the present official system issued in 2000, are all different, respectively. Currently, the system proposed by the South Korean Ministry of Culture and Tourism (MCT system) in 2000 is adopted as the official Romanization system in South Korea. However, as this system also allows the existing Romanization of personal and corporate names that is different from the official system, it can be said that the Romanization
scheme for personal names is not defined in South Korea.

Moreover, there have been many other Korean Romanization schemes, such as the Victorian method that was used among missionaries, the McCune-Reischauer system (MR system) that was developed by G. M. McCune and E. O. Reischauer in 1939, the system proposed by the Korean Language Society in 1940, and the Yale system jointly developed and released by American and Korean researchers in 1954.\textsuperscript{57,58,59} The MR system was adopted by the LC as a standard Romanization scheme for Korean materials in 1959,\textsuperscript{60} and now it has been adopted by ALA/LC Romanization Tables.\textsuperscript{61} It has been used by a wide range of Western libraries. However, the MR system is difficult to understand, especially for native Korean speakers, because it was developed by non-natives.\textsuperscript{58,59}

The MR system, the system of the Korean Language Society in 1940, the systems released in 1948, 1959, and 1984 by the Korean government, and the MCT system are all Romanization systems based on transcriptions that represent Korean pronunciations in Latin alphabets.\textsuperscript{55} On the other hand, the Yale system used in the linguistics field and the ISO TR11941:1996 system developed under the agreement of the South and North Korean governments are Romanization systems based on transliterations that correspond each Hangul to a specific Latin alphabet string in principle.\textsuperscript{58,62} Because a full agreement between the two governments on the ISO TR11941:1996 system has not yet been reached, which renders it a Technical Report and not yet an official standard, the system is not widely used.

Although both the MCT and MR systems are based on transcription, they have certain differences: while the MCT system always Romanizes each Korean letter in the same way regardless of its pronunciation change, the MR system Romanizes Korean letters differently according to their pronunciations (Korean letters have different pronunciations depending on the letters that precede or follow them).\textsuperscript{58} In addition, while the MR system adds a hyphen between the first and second syllables of a first name, the MCT system puts two syllables together in principle, though the addition of a hyphen is also permitted.\textsuperscript{56,61} Compared to the MR system, the MCT system is easy to input because it does not use special diacritics such as breves and apostrophes.\textsuperscript{58}

As the Korean language has many Romanization schemes, people determine their Romanized names depending on their preferences. The representation of Korean personal names in Latin alphabets is called “Anglicization” by Kim and Cho.\textsuperscript{63} Thus, it can be seen that Romanized forms of Korean names are not produced under a uniform system, but rather reflect the particular “English names” of individuals, at least in South Korea.

3.1.4 Vietnamese names

Similar to China, Vietnam is a multiethnic country that has 54 ethnic groups. Of these, the
majority (85.7%) are the Kinh people. People have used Chinese-like surnames through the ages. There are 931 kinds of surnames in Vietnam. Similar to China and South Korea, a high percentage of the population is concentrated in particular surnames. Le estimated that three surnames, namely “Nguyễn,” “Lê,” and “Trần,” account for 60% of the population of Vietnam. “Nguyễn” is especially shared by approximately 48% of North Vietnamese and 28% of South Vietnamese.

Chinese characters were once used in Vietnam. At the time when Chinese characters were used, many new Chinese characters called chữ Nôm were invented in Vietnam. Contrary to the original Chinese characters called chữ Hán, chữ Nôm was used only in Vietnam, and not in other countries in the Chinese character cultural sphere.

At the end of the 19th century, under the reign of the French, chữ quốc ngữ, the modern Vietnamese script, was adopted as the official writing system of Vietnam. Nowadays, most Vietnamese people do not understand Chinese characters, and only a limited number of people, such as the researchers of classical literature, would understand them.

Before 1945, many people had names with three syllabaries, which consist of one syllable of a surname, one syllable of a middle name, plus one syllable of a first name. However, recently, an increasing number of people have two syllabaries of their middle or first names. Middle names used to have the function of distinguishing gender, such as “Van” for men and “Thi” for women. However, this is not the case now. Sometimes a mother’s surname is used as her children’s middle name, although women do not change their surnames after marriage, and habitually their children inherit their fathers’ surnames. As such, this division of surname, middle name, and first name is quite difficult for foreigners to understand.

As people tend to name their children with Chinese-derived words, most names in Vietnam can be represented in Chinese characters as well as in chữ quốc ngữ; however, in general, people do not use or even know the Chinese characters of their own names.

3.2 Representations, data elements, and structures

In this study, name authority data are divided into three parts: representations, data elements, and data structures (Figure 3-4). Attributes of entity, such as birth date, gender, and address, are essential parts of authority data because they facilitate author identification; thus, existing standards such as FRAD and RDA define the kinds of attributes that are considered to be included in authority data. These attributes are data elements. In addition, names, including real names, pseudonyms, earlier and later names, etc., are also data elements recorded in authority data.

Standards like FRAD and RDA, however, do not distinguish between data elements and representations. Existing standards assume that alternative linguistic forms, alternative script
forms, and transliterations of names are also data elements. In contrast, such notational variants of names are assumed as “representations” in the authority data framework proposed by this study.

Why should the concept of “representations” be adopted in addition to “data elements”? There are two reasons. The first is that notational variations of names, especially Romanized names, are not the “real” names used commonly in the real world. It is rather a string of Latin characters imposed by libraries or organizations by which the authority data are constructed because library systems cannot process non-Latin characters, or because users or librarians who cannot input non-Latin characters are using them. Therefore, a Romanized name is an access point rather than a “name,” and should thus be distinguished from a “name.” Name is a data element, while its Romanization is one of the “representations” of the name. For example, “박근혜” is a “real” name for the eleventh and current President of South Korea. “Pak, Kŭn-hye,” which is an authorized access point of her authority data in the Library of Congress Authorities, is an MR Romanization of “박근혜.” “Pak, Kŭn-hye” is an imposed Romanization by the Western library community and an uncommon name that is only used by libraries. On the other hand, her official English name is “Park Geun-hye.” “Park Geun-hye” is also her “real” name: as media releases in English issued by the government of South Korea use this name, it can be assumed that this name is an independent one in common use in Western countries. “Pak, Kŭn-hye” and “Park Geun-hye” should be explicitly distinguished in the following way: while the latter is one of her “real” names, or in other words, a data element, the former is merely a “representation” of “박근혜.” A “real” name should be distinguished from a name imposed by libraries because the significance of the name is different. Generally, a “real” name is more important than an imposed name because the “real” name is used by the author him/herself and is known by many people. Under the situation that authority data is used by not only libraries but also other communities such as archives, research information management systems, and online encyclopedias, the “real” name must takes priority.

The second reason is that notational variations of a name that are derived from its original forms (i.e., the “real” name), regardless of whether it is a Romanization or a non-Latin representation of the “name,” should be handled with the “name” for data identification. For example, “Sun, Wen” is a Romanization of “孫文,” the first president of the Republic of China. “Sun, Wen” is also a Romanization of “孫雯,” “孫溫,” and “孫玟.” The representation of “Sun, Wen” alone is not enough for us to identify the specific person. It should thus be represented, or handled, together with its Chinese characters. Similarly, “キクチカオル” is the Japanese yomi (non-Latin transcription) of “菊地薫.” In addition, the yomi of “菊地かおる” and “菊地かほる” are also “キクチカオル.” Again, the representation of “キクチカオル” alone is not enough to identify the person; its Kanji or kana script must be shown together with the yomi for name
identification. On the other hand, the original form is sometimes insufficient for identifying the person, because of the difference between the *yomi*. Two patterns of *yomi* for “河野明” is a good example. Analogously, sometimes one *Hanja* has more than two corresponding *Hangul*. Therefore, handling an original form of a name and its notational variations together is useful for identifying the name more precisely.

Data structures determine how authority data (including representations and data elements) should be recorded. Data structures are equivalent to authority data formats such as *MARC 21 Format for Authority Data* (MARC 21/A) and RDF (Resource Description Framework), although current MARC 21/A and RDF cannot perfectly handle “representations” of names. Cataloging rules such as RDA also affect data structures, for example, RDA 9.2.2.9 defines that “if a name consists of a surname preceded by other parts of the name, such as given names, record the surname and follow it by a comma and the parts of the name that precede it.” This rule specifies the form of the personal name included in the data structure. Therefore, in Figure 3-4, the “rules” also correspond to the data structures, which define how to record representations and data elements and are used as a base for the representations and data elements.

As noted in Chapter 1, the scope of this study is limited to author’s names, including persons and corporate bodies. Therefore, in Figure 3-4, “names” means the names of persons and corporate bodies. However, as titles of materials and geographical names also have several representations, at least in the Chinese character cultural sphere, the author trusts that this framework could also be applied to titles and geographical names.

Access points consist of names and additions that include any information denoting birth/death date, place of origin, occupation, or other characteristics of the person/corporate body. Additions also have representations when they are recorded in non-Latin scripts. However, as this study focuses on representations of names of persons and corporate bodies, representations of additions are omitted from Figure 3-4.

Based on the authority data framework shown in Figure 3-4, the representations of name authority data produced by each of the organizations under study are investigated; the results of this investigation are presented in Chapters 5 to 8. Data elements of each organization are then compared in Chapter 9. A data model that accommodates the idea of “representations” of authority data is proposed in Chapter 10. As data structures are defined by cataloging rules and formats, recommendations for the revision of MARC 21/A are presented in Chapter 11. Because BIBFRAME Authority, which is under construction by Western library communities and will be the substitute for MARC 21/A, require authority data that should be recorded in RDF format, new vocabulary with which to express the proposed authority data model is defined, and the sample authority data in RDF/XML format are shown in Chapter 11.
Figure 3-4 Framework of name authority data

Notes

7 “全国 1,510 姓氏 陳林滿天下 內政部《全國姓名統計分析》全新出版”. 中華民國內政部. 2014-10-29.
9 沢田ゆかり. “香港：植民地が生んだ多重言語社会の名前”. 第三世界の姓名: 人の名
17 “中文譯音使用原則”. 臺灣地區地名查詢系統: 地名相關法規. 2008-12-08.
18 “護照外文姓名拼音對照表（包括：國音第一式、漢語拼音、通用拼音、國音第二式、及 WG 拼音法）”. 外交部領事事務局: 護照.
22 “香港地區大學圖書館編目作業近況”. 中文名稱規範聯合協調委員會第八次會議. 臺北, 2010-11-02.
24 丹羽基二編. 日本苗字大辞典. 芳文館, 1996, 3冊.


武部良明. 日本語の表記. 角川書店, 1979, 574p.


이길성. 漢字이름 5151 字 안에서 지으세요 : 대법, 人名用 113 자 추가 이름 다섯자 넘으면 안돼. 朝鮮日報. 2007-03-05, 사회 A15 면.

According to “인명용 한자표”. 대한민국법원 전자민원센터. http://help.scourt.go.kr/nm/img/hanja/hanja.pdf, (accessed 2015-02-03), the character “欒”, for example, is defined to pronounce “락” in 教育용 기초한자 and also is defined to pronounce “악” and “요” in Appendix 1 of 가족관계의 등록 등에 관한 규칙. Therefore, “欒” is allowed to use above three kinds of pronunciations in personal names.


Chapter 4

Method and research objects

This section first explains the research methods for Chinese, Japanese, and Korean name authority data, the results for which are shown in Chapter 5-7, respectively. The research method and results for Vietnamese name authority data are shown in Chapter 8, as they differ slightly from other name authority data. Secondly, before the search results are presented, current practices and policies of authority control in China, Japan, South Korea, and Vietnam are explained mainly based on the interviews. Such information is useful for understanding and analyzing the search results of each research object.

4.1 Method

The method involves the following steps:
Step 1 - Data collection. Face-to-face interviews were conducted. In addition, cataloging rules, formats, and manuals about name authority data for each organization were gathered. Search results of authority databases or OPACs of each organization are also consulted if available.
Step 2 - Setting checkpoints which are unique to Japanese, Chinese, Korean name authority data, respectively. For example, the author set six checkpoints that are considered to be important in creating Japanese name authority data.
Step 3 - Identification of issues affecting data sharing. Using the gathered information, checkpoints were investigated based on the comparison of the current practices of each organization. Then, issues affecting the sharing of personal and corporate name authority data are pointed out.

4.1.1 Organizations studied

The author explored authority control practices in major organizations that create Chinese, Japanese, and Korean name authority data, namely, the National Library of China (NLC), the China Academic Library & Information System (CALIS), the Hong Kong Chinese Authority Name Workgroup (HKCAN), the National Central Library (NCL) of Taiwan, National Taiwan University Library (NTUL), the National Diet Library (NDL) of Japan, NACSIS-CAT, Keio University Libraries (Keio) in Tokyo, the National Library of Korea (NLK), Yonsei University Library (YUL) in South Korea. Additionally, to verify how these name authority data are dealt with in North America, the LC was also included as a research object.

Notably, not all organizations create the three kinds (Chinese, Japanese, and Korean) of name authority data. The authority control for Japanese and Korean names in the NLC has not yet
officially started.¹ The authority database of CALIS includes Chinese and Japanese authority data, but not Korean authority data.² According to the interviews, the NDL has not produced Chinese or Korean name authority data since 2012, except when these names have appeared in Japanese materials; therefore, these organizations are excluded from the results. For example, CALIS is not included in the results presented in Chapter 7 (the section addressing Korean name authority data), because it does not produce Korean name authority data.

4.1.2 Step 1– Data collection

Interviews were conducted in July 2013 (Japan), August 2013 (Taiwan), September 2013 (South Korea), and November 2013 (Mainland China). Unless otherwise cited, the results described in Chapter 5 to 7 are based on these interviews. Interviews with NACSIS-CAT, HKCAN, and the LC were not conducted. Supplemental inquiries via e-mail were also conducted as needed, and this information is indicated as such in context.

At the time of each interview, each organization was asked to provide authority record samples for the purpose of this research. The number and content of these sample records (shown in Table 4-1) differ widely across the organizations. 1,517,926 name authority records extracted on March 19, 2010 that were made between January 1, 1986 and December 31, 2009 are provided by NACSIS-CAT. These samples are also consulted for this research.

Table 4-1 Number of sample records consulted

<table>
<thead>
<tr>
<th>Organizations</th>
<th>Quantity of sample records</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>13</td>
<td>8 Chinese persons, 3 Chinese corporate bodies, 2 conferences</td>
</tr>
<tr>
<td>CALIS</td>
<td>4</td>
<td>2 Chinese persons, 1 Chinese corporate body, 1 conference</td>
</tr>
<tr>
<td>NCL</td>
<td>2</td>
<td>Chinese persons</td>
</tr>
<tr>
<td>NTUL</td>
<td>26</td>
<td>4 Chinese persons, 4 Japanese persons, 7 Korean persons, 4 Chinese corporate bodies, 1 Japanese corporate body, 6 Korean corporate bodies</td>
</tr>
<tr>
<td>Keio</td>
<td>20</td>
<td>3 Chinese persons, 4 Japanese persons, 8 Korean persons, 4 Korean corporate bodies, 1 conference</td>
</tr>
<tr>
<td>NLK</td>
<td>1</td>
<td>Korean person</td>
</tr>
<tr>
<td>YUL</td>
<td>3</td>
<td>1 Chinese person, 1 Japanese person, 1 Korean person</td>
</tr>
</tbody>
</table>

Supplemental information such as cataloging rules, formats, and manuals about name authority data for each organization were collected, as shown in Table 4-2.

Although sample records of HKCAN, NDL, and LC were not obtained, these organizations provide public access to their authority databases.³,⁴,⁵ Therefore, the search result of these databases could be used instead of sample records. In addition, as the number of sample records
Table 4-2 Cataloging rules, formats, and manuals consulted

<table>
<thead>
<tr>
<th>Organizations</th>
<th>Cataloging Rules, Formats, and Manuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>中国文献编目规则（第二版）<em>&lt;br&gt;中国机读规范格式 (CNMARC/A)</em>&lt;br&gt;中国机读规范格式使用手册*&lt;br&gt;中文图书名称规范数据款目著录规则*&lt;br&gt;个人名称规范学科附加成分选取表及说明*</td>
</tr>
<tr>
<td>CALIS</td>
<td>中国文献编目规则（第二版）<em>&lt;br&gt;CALIS联合目录规范控制过程详细说明 (更新版)</em>&lt;br&gt;中文文献规范控制原则*</td>
</tr>
<tr>
<td>HKCAN</td>
<td>AACR2 2nd ed., 2002 rev.<em>&lt;br&gt;MARC 21/A</em></td>
</tr>
<tr>
<td>NCL</td>
<td>中國編目規則(第三版)†&lt;br&gt;MARC 21/A†&lt;br&gt;中文名稱權威紀錄彙整原則†&lt;br&gt;中文權威紀錄著錄規則†&lt;br&gt;團體權威整理作業手冊†&lt;br&gt;譯名權威紀錄處理原則†&lt;br&gt;出版社、學校及社團機讀格式記錄原則†&lt;br&gt;日本作者中譯名與原名之著錄原則†&lt;br&gt;「日文書」人名標目著錄原則†</td>
</tr>
<tr>
<td>NTUL</td>
<td>中國編目規則(第三版)†&lt;br&gt;MARC 21/A†&lt;br&gt;團體權威整理作業手冊†&lt;br&gt;出版社、學校及社團機讀格式記錄原則†</td>
</tr>
<tr>
<td>NDL</td>
<td>日本目錄規則1987年版改訂3版 (NCR1987 3rd rev.)<em>&lt;br&gt;JAPAN/MARC MARC21フォーマットマニュアル典拝編</em>&lt;br&gt;「日本目錄規則1987年版改訂3版 第II部 標目」適用細則 (2012年1月)<em>&lt;br&gt;個人名標目的選択・形式基準 (2012年1月以降)</em>&lt;br&gt;団体名標目的選択・形式基準 (2012年1月以降)<em>&lt;br&gt;『JAPAN/MARC MARC21フォーマット』における片仮名読み表記要領</em>&lt;br&gt;『JAPAN/MARC MARC21フォーマット』におけるローマ字読み表記要領*</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>日本目錄規則1987 年版改訂第2版 (NCR1987 2nd rev.)<em>&lt;br&gt;目錄情報の基準 第4版</em>&lt;br&gt;目錄系統コーディングマニュアルA&lt;br&gt;中國語資料用コーディングマニュアル (案) B&lt;br&gt;韓国·朝鮮語資料の取扱いC</td>
</tr>
</tbody>
</table>
| Keio       | AACR2 2nd ed., 2002 rev.\(^g\)  
|           | 日本目録規則1987年版改訂3版 (NCR1987 3rd rev.)\(^r\) (partially applied)  
|           | MARC 21/A\(^h\)  
|           | An original manual\(^d\)  |
| NLK       | KORMARC/A\(^E\)  |
| YUL       | AACR2 2nd ed., 2002 rev.\(^g\)  
|           | MARC 21/A\(^h\)  |
| LC        | RDA\(^f\)  
|           | MARC 21/A\(^g\)  
|           | Library of Congress-Program for Cooperative Cataloging Policy Statements\(^f\)  
|           | Descriptive Cataloging of East Asian Material: CJK Examples of AACR2 and Library of Congress Rule Interpretations\(^f\)  
|           | ALA/LC Romanization Tables\(^G\)  |

**Notes.**  
c 国家图书馆图书采编部. 中国机读规范格式使用手册; 中文图书名称规范数据款目著录规则; 中文图书主题规范数据款目著录规则. 1999, 182p.  
d A non-public original manual provided in November 2013.  
<table>
<thead>
<tr>
<th>Table 4-2 (Continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D A non-public original manual provided in February 2012.</td>
</tr>
</tbody>
</table>
from some organizations is limited, authority data retrieved from CALIS\textsuperscript{6} and NCL\textsuperscript{7} authority databases and authority data that link to bibliographic data in NLC’s OPAC\textsuperscript{8} were also consulted. As for NTUL\textsuperscript{9}, Keio\textsuperscript{10}, NLK\textsuperscript{11}, and YUL\textsuperscript{12}, the headings included in bibliographic data searched on OPACs were used as an alternative to authority records. Notably, access points in bibliographic OPACs only include authorized access points, not variant access points; thus, limited information was acquired from bibliographic OPACs. The databases and OPACs used for the search are shown in Table 4-3. When searching authority databases or OPACs, the author used several personal and corporate names to reflect the known characteristics of names in each region. These search terms are shown in each chapter. All authority databases and OPACs support Unicode.

4.1.3 Step 2 - Setting checkpoints that are unique to authority data of each area
As the second step involved setting checkpoints for the study, several aspects assumed to be treated differently by organizations representing personal and corporate names were provided for investigation. Because checkpoints differ among each area, further information on the checkpoints and the reasons for setting them are explained in Chapters 5 to 7.

4.1.4 Step 3 - Identification of issues affecting data sharing
In the third step, the checkpoints for investigation were determined by comparing the authority control practices of each organization using the collected data. The result of the third step is shown in Chapters 5 to 7.

4.2 Current practices and policies of authority control in each area
4.2.1 China
4.2.1.1 Mainland China
1) The National Library of China (NLC)
As of September 30, 2013, the NLC had about 2.60 million bibliographic records for Chinese materials, 1,060,889 authority records for Chinese persons, and 75,195 authority records for Chinese corporate bodies, including conference names. NLC applies the 2nd edition of \textit{中国文献编目规则} (Chinese Cataloging Rules) and \textit{中国机读规范格式} (China MARC Format/Authorities; CNMARC/A) as an authority format. In addition, \textit{中国机读规范格式使用手册} (The Handbook for CNMARC/A) and \textit{中文图书名称规范数据款目著录规则} (Description Rule for Authority Data Entries), both of which were published in 1999, are used as manuals. Authority data that link to bibliographic data can be searched in NLC’s OPAC. NLC carries out authority control for personal names, corporate names, conference names, titles, and subjects.\textsuperscript{13} However, authority control for Japanese, Korean, and Vietnamese names has not yet
officially started.

Table 4-3 Databases and OPACs used for search

<table>
<thead>
<tr>
<th>Organizations</th>
<th>Databases and OPACs searched</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>OPAC called &quot;聯机公共目录查询系统&quot; (authority records are linked to bibliographic records)</td>
</tr>
<tr>
<td>CALIS</td>
<td>Authority database called &quot;CALIS联合目录规范OPAC&quot;</td>
</tr>
<tr>
<td>NCL</td>
<td>&quot;臺灣書目整合查詢系統&quot; (an database of both authority and bibliographic records); OPAC called &quot;國家圖書館館藏目錄查詢系統&quot; (authority records are linked to bibliographic records)</td>
</tr>
<tr>
<td>NTUL</td>
<td>OPAC called &quot;TULIPS&quot; (bibliographic records only)</td>
</tr>
<tr>
<td>HKCAN</td>
<td>Authority database called &quot;HKCAN Database OPAC&quot;</td>
</tr>
<tr>
<td>NDL</td>
<td>Authority database called &quot;Web NDL Authorities&quot;</td>
</tr>
<tr>
<td>Keio</td>
<td>OPAC called &quot;KOSMOS&quot; (bibliographic records only)</td>
</tr>
<tr>
<td>NLK</td>
<td>OPAC called &quot;dibrary&quot; (bibliographic records only)</td>
</tr>
<tr>
<td>YUL</td>
<td>OPAC called &quot;WiSearch&quot; (bibliographic records only)</td>
</tr>
<tr>
<td>LC</td>
<td>Authority database called &quot;Library of Congress Authorities&quot;</td>
</tr>
</tbody>
</table>

2) The China Academic Library & Information System (CALIS)

CALIS is a nationwide academic library consortium funded primarily by the Chinese government. In March 2000, the Chinese union catalog for CALIS members was launched, and in September 2003, CALIS started constructing an authority database project. As of April 30, 2014, CALIS had 1,103 member organizations, most of which are academic libraries. As of June 30, 2013, CALIS had more than 5.70 million bibliographic records, and CALIS Union Catalog Authorities (the authority database of CALIS) had 472,498 authority records for Chinese persons, 49,274 records for Chinese corporate bodies, and 415 authority records for Chinese conferences, 734,318 records for Western persons and corporate bodies, and 119,729 records for Japanese persons and corporate bodies. Authority records for Koreans and Vietnamese are not constructed. While authority data can be searched via the internet, the website is open only as an experimental trial and does not yet include all up-to-date authority data held by CALIS.

CALIS applies an original authority format that is not yet published but based on two public manuals for CALIS authority: 中文文献规范控制原则 (The Principle of Authority Control for
Chinese Materials\textsuperscript{17} and CALIS 联合目录规范控制过程详细说明（更新版）(Detailed Explanation for CALIS Union Catalog Authority Control Process: Updated Version).\textsuperscript{18}

CALIS has adopted the CALIS Union Catalog Authority Format based on UNIMARC for its authority data. CALIS members can download bibliographic and authority records from NACSIS-CAT; thus, many Japanese authority records are copy records of NACSIS-CAT.

In the CALIS Online Catalog, bibliographic databases for Chinese, Western, Russian, and Japanese materials are separately constructed. Authorized headings are produced for materials in each language; thus, a Japanese, a Chinese, and a Western heading, all authorized, are produced for the same person if the person is an author of materials in all three languages.

4.2.1.2 Hong Kong
Hong Kong Chinese Authority Name Project (HKCAN)

As a project of the JULAC, which is a consortium of eight academic libraries in Hong Kong, the HKCAN Workgroup was set up in 1999 to establish a union database “that would reflect the unique characteristics of the Chinese authors and organizational names.”\textsuperscript{19} HKCAN unified about 140,000 authority records from member libraries; since then, Lingnan University and the Chinese University of Hong Kong have worked to remove duplications.\textsuperscript{20} At the moment, the online authority database, HKCAN Database OPAC, is in operation. The database adopted the MARC 21 Format for Authority Data (MARC 21/A).\textsuperscript{21}

As of September 2013, the HKCAN database had 180,994 authority records for persons, 28,844 records for corporate bodies, and 1,651 records for conferences.\textsuperscript{22} This includes Japanese, Korean, and Vietnamese persons and corporate bodies. According to an e-mail response from the Chinese University of Hong Kong Library, which was a host library of HKCAN in 2012, HKCAN does not have special cataloging or authority data manuals, and as of June 2014, three out of seven member libraries of HKCAN had started to apply RDA instead of AACR2 as a cataloging rule. Thus, authority records created by these three libraries and sent to the HKCAN database are in MARC21/A format with augmentation of tags in accordance with RDA.

Hong Kong is a bilingual society, and many authors publish in both Chinese and English; therefore, both Chinese and English access points are particularly important in Hong Kong libraries.\textsuperscript{23} However, before HKCAN was established, libraries had chosen various authority record formats and, in particular, made a decision whether to adopt Chinese character access points, Romanized access points, or both because the form of authorized access points had differed from library to library.\textsuperscript{24}

As a result of discussions, to be fully compliant with MARC 21/A and to serve the wants and needs of overseas libraries adequately, HKCAN chose to use field 1XX for LC/NACO
Authority File (LCNAF) headings and 7XX (Heading Linking Entry Fields) for Chinese scripts.\footnote{25}

In the actual authority workflow, HKCAN members usually copy records from LCNAF and enhance the records by adding names in Chinese scripts to field 7XX and field 4XX (See From Tracing Fields) as appropriate.\footnote{21} As we have seen above, in following the idea of sharing authority data with Western libraries, HKCAN has been creating authority records in close co-operation with LCNAF.

4.2.1.3 Taiwan

1) The National Central Library (NCL)

As of July 31, 2013, The NCL had about 2.55 million bibliographic records and 1.33 million authority records for persons and corporate bodies. NCL applies the 3rd edition of 《中國編目規則》 (Chinese Cataloging Rules)\footnote{26} as a cataloging rule, and MARC 21/A as an authority format. In addition, the library developed its own manuals regarding authority data, such as 《中文權威紀錄著録規則》 (Descriptive Rules for Chinese Authority Data)\footnote{27}, 《中國名稱權威紀錄彙整原則》 (The Principle for Organization of Chinese Name Authority Data)\footnote{28}, and 《團體權威整理作業手冊》 (The Handbook of Authority Work for Corporate Bodies)\footnote{29}, among others, which are available online. However, these manuals were developed before NCL changed its bibliographic and authority format in December 2011; thus, examples that appear in the manuals are in 《中國機讀權威記錄格式》 (Chinese MARC Format for Authority Records; CMARC/A) rather than MARC 21/A.

Initially, simple authority data are produced by the Collection Development and Bibliography Management Division of NCL. Authority data are then sent to the Synergy of Metadata Resources in Taiwan (SMRT) system, launched in April 2013, which includes all bibliographic and authority data created by several divisions of NCL and the National Bibliographic Information Network of Taiwan.\footnote{30} Next, the Bibliographic Information Center of NCL, which is in charge of data quality control in SMRT, augments the authority data elements as necessary.

2) National Taiwan University Library (NTUL)

As of March 2013, NTUL had about 3.70 million bibliographic records, 240,000 authority records for Chinese persons, and 16,000 authority records for Chinese corporate bodies. As at NCL, NTUL applies the 3rd edition of 《中國編目規則》 as a cataloging rule and MARC 21/A as an authority format. In 1998, NTUL and NCL jointly launched the Chinese Name Authority Database which has been incorporated into NCL’s SMRT system, so NTUL applies NCL’s manuals for corporate bodies; however, NTUL has own rules for recording persons’ authority data elements, though they are not publicly available.
4.2.2 Japan

Several organizations generate their own authority data in Japan: NDL, NACSIS-CAT, Keio, Toshokan Ryutsu Center Co., Ltd. (TRC), and Nippan Library Service, Co., Ltd. (NTS). TRC and NTS are creating bibliographic and authority data for commercial use. Both companies have adopted their own original format. ³¹

1) The National Diet Library (NDL)

As of March 2014, the NDL had 810,169 authority records for personal names and 189,991 authority records for corporate names.³² Authority data produced by NDL can be retrieved via the Web NDL Authorities, which was launched in 2012.⁴ 日本目録規則 1987 年版改訂 3 版 (Nippon Cataloging Rules 1987 ed., 2006 rev.; NCR1987 3rd rev.) is applied for CJK materials as well as foreign serials, while RDA is applied for other foreign materials outside of these countries.³³ JAPAN/MARC MARC 21 Format has been used for bibliographic and authority data from January 2012.

2) NACSIS-CAT

Most university libraries in Japan maintain their own local catalogs using shared cataloging (both bibliographic and authority) data from the NACSIS-CAT system. It should be noted, however, that linking headings in bibliographic records to authority records is optional.³⁴ This means that not every organization participating in NACSIS-CAT does authority control. The NACSIS-CAT authority file has records for personal names, corporate names, conference names, and titles.

As of March 31, 2014, the NACSIS-CAT system had 1,259 member organizations³⁵ and as of February 15, 2015, it had about 1.65 million personal, corporate, and conference name authority records.³⁶ The organization is operated by the NII, which prepares dedicated manuals (available online) for users of NACSIS-CAT. The two main manuals are 目録情報の基準 (The standard for cataloging information), 4th ed., published in 1999,³⁷ and the more recent 目録システムコーディングマニュアル (The coding manual for the cataloging system), published in April 2014, though the latter undergoes nearly constant revision.³⁸ NACSIS-CAT applies 日本目録規則 1987 年版改訂 2 版 (NCR1987 2nd rev.) for CJK authority records.³⁹ [p.25] The organization applies its original bibliographic and authority formats, called CATP format.

3) Keio University Library

Keio did not generate authority data between 1998 and 2011, when the old library system
started, and April 2011, when the new library system was launched.\textsuperscript{40} Keio currently does authority work for personal names and corporate names (including conference names). As of July 11, 2013, Keio had about 2.37 million bibliographic records and 827,863 authority records. However, many records are currently under maintenance, since authority records had not been created for the previous 10 years. Keio uses MARC 21 formats for bibliographic and authority data.

While AACR2 was adopted to create authority data, \textit{NCR1987 3rd rev.} is also consulted to create Japanese authority data. The authority data of NACSIS-CAT is consulted when a new record needs to be created, with Web NDL Authorities as the second point of consultation. Keio also has an internal manual that was provided for this research, as authority data of Keio are not publicly accessible.

4.2.3 South Korea

1) The National Library of Korea (NLK)

The NLK started fully fledged authority control in 2000, following the launch of an integrated information system in the Windows environment and the establishment of the KORMARC format as a national standard.\textsuperscript{41} Previously generated authority data were concentrated heavily on foreign authors who appeared in both general and children’s book searches in Korea, in Japanese materials, and in Chinese materials.\textsuperscript{42}

As of July 2013, NLK had about 9.10 million bibliographic records and 163,369 authority records. Of these, 49,247 records are authority data for Korean persons. At the time of interview (September 2013), NLK had not yet started to produce authority data for corporate bodies, though an e-mail from NLK in April 2014 revealed that the library started producing this data in March 2014. Authority data at NLK are not publicly available.

Because the third and fourth editions of KCR had no guidelines for selections and forms of access points, NLK established its own guidelines for personal name authority data in April 2012 and for corporate bodies in March 2014. These new guidelines include rules for selections and representations of access points as well as attributes for identifying persons and corporate bodies.

2) Yonsei University Library (YUL)

As of February 2013, YUL had 1.24 million bibliographic records and 585,050 authority records for persons, 65,699 authority records for corporate bodies, and 7,580 authority records for conferences. Before a system replacement was done in August 2009, authority data were made for all access points in bibliographic records; however, after August 2009, unless an access point had variant access points, authority data was not created. Authority data at YUL are
not publicly available. YUL applies KORMARC format.

4.2.4 Vietnam

This subsection is based on the result of interviews to the National Library of Vietnam (NLV) and to the National Library for Science and Technology (NLST) conducted in April 2014.

1) The National Library of Vietnam (NLV)

As of April 2014, NLV had more than 550,000 bibliographic records but they do not produce authority records. NLV applies its original manual titled Tài liệu hướng dẫn mô tả án phẩm: Dùng cho mục录 thư viện (The Manual of Description of Printed Materials: for Library Catalogs), which is based on the International Standard Bibliographic Description (ISBD) as a cataloging rule and a translated version of MARC 21 as a bibliographic format. For personal names, access points in bibliographic records are controlled by means of Bộ Từ khóa (The Keyword List), which includes subjects, persons, corporate bodies, and geographical names. While the list is mainly used for subject access points, the personal names section is also used for author names. When this occurs, the birth/death years are omitted for authors, so only name strings from the list will be recorded as access points. Thus, even birth/death years of authors are not recorded in the bibliographic records of NLV.

2) The National Library for Science & Technology (NLST)

The NLST is a section of the National Agency for Science and Technology Information (NASATI) of Vietnam. Originally, the Central Library on Science and Technology, which was founded in 1960, and the Central Institute for Scientific and Technical Information, founded in 1972, merged to form the National Centre for Scientific and Technological Information and Documentation (NACESTID) in 1990. NACESTID was renamed as the National Centre for Scientific and Technological Information (NACESTI) in 2003, which became NASATI in 2009. NLST is the largest science and technology library in Vietnam.

NLST has more than 300,000 books and about 7,000 titles of journals. Among them, about 250,000 books and 6,000 titles of journals have bibliographic records. However, authority records were not created.

AACR2 was adopted by NLST in 2000, instead of an original cataloging rule published in 1987. Currently, NLST uses the Vietnamese version of AACR2 published in 2009. The bibliographic format adopted by NLST in 2000 was MARC 21, based on the concise Vietnamese version of MARC 21 published in 2005 by NACESTI. Subject headings of bibliographic records are controlled by the original list developed by NACESTID in 2001. However, NLST does not control author names.
Notes

3. The author could access all authority data, including MARC tag displays, through “香港中文名称規範數據庫公共檢索目錄: HKCAN Database OPAC”. http://www.hkcan.net/hkcanopac/, (accessed 2015-01-23).
7. The Synergy of Metadata Resources in Taiwan (SMRT) system (台湾書目整合查詢系統). 国家圖書館. http://metadatanci.edu.tw/blstkmc/blstkmc#tudorkmjob, accessed 2015-02-12) was mainly used; however, authority data linking to bibliographic access points in NCL’s OPAC (“國家圖書館館藏目錄查詢系統”. 國家圖書館. http://aleweb.ncl.edu.tw/, accessed 2015-02-12) were also used because of the slight differences between the authority data in the SMRT system and OPAC.
21 Yu, Abraham J. Evaluation and Analysis of the Chinese Name Authority Files under Development in Beijing, Hong Kong, and Taipei. OCLC, 2002, 52p.
31 国立国会図書館書誌部編. 第 4 回書誌調整連絡会議記録集: 名称当職のコントロール.

32 “統計からみた書誌データ”. 国立国会図書館.

33 “書誌データ作成ツール” 国立国会図書館.


36 “総合目録データベースの現況”. 国立情報学研究所目録所在情報サービス.

37 国立情報学研究所. “目録情報の基準 第4版”. 目録所在情報サービス: ドキュメン

38 国立情報学研究所. “目録システムコーディングマニュアル”. 国立情報学研究所
2015-02-04).


40 古賀理恵子. 特集, KOSMOSIII 新図書館システムの導入: Aleph を KOSMOSIII として

41 国立중앙도서관자료기확과편. 국가전자기록의 협력적 구축방안. 국립중앙도서관,

42 国立중앙도서관. 국립중앙도서관 전거레코드 표목의 구조화 기준에 관한 연구.

43 “History of National Agency on Science and Technology Information (NASATI)”. National
Agency on Science and Technology Information.
http://www.vista.vn/Default.aspx?tabid=71&IntroId=95&temidclicked=95&language=en-US,

44 Trung tâm thông tin thư viện khoa học và công nghệ quốc gia. Từ điển từ khoa học và
công nghệ: tập I bằng tra chính. Trung tâm thông tin thư viện khoa học và công nghệ quốc gia.
Chapter 5

Representations of Chinese name authority data in Chinese character cultures

5.1 Checkpoints and search terms

5.1.1 Checkpoints

The following four aspects, which are assumed to be treated differently by organizations in representing Chinese personal and corporate names, are provided for investigation:

1) Adoption and character forms of Chinese characters
2) Treatment and types of Romanization
3) Separation of surname and given name with a comma
4) Representations in local languages outside China, Hong Kong, and Taiwan

These four topics for investigation were determined for the following reasons. The first topic, adoption and character forms of Chinese characters, was chosen because access points in Chinese character forms may differ among Mainland China, Hong Kong, Taiwan, Japan, and South Korea because the Chinese characters used in each region have different letter shapes.

The second topic, treatments and types of Romanization, was chosen because although Hanyu pinyin is an official Romanization system in Mainland China, the pinyin system is not commonly used in Hong Kong and Taiwan. The handling of tone marks, apostrophes, and umlaut marks with “u” defined by 汉语拼音方案 (Scheme for the Chinese Phonetic Alphabet) is also investigated.

According to a report by the National Diet Library (NDL), in access points of authority data constructed by the National Library of China (NLC), the surnames and given names of East Asian people are not separated. In Japanese libraries, however, the surname and given name are customarily separated by a comma. Thus, the third topic, separation of surname and given name with a comma, was chosen. Yu proposed an idea for the International Chinese Name Authority File, and regarding Romanization, he pointed out that following ALA/LC Romanization Tables and pinyin guidelines, the first letter of the surnames and given names should be capitalized and the syllables in given names should be joined together in the database. Thus, the latter two points are also investigated.

The fourth topic, representations in local languages outside China, Hong Kong, and Taiwan was chosen because Park pointed out that authorized access points of Chinese persons differ among the databases of the Seoul National University Library (SNUL), Yonsei University Library (YUL), and Ewha Woman’s University Library.
5.1.2 Search terms

The author used 12 personal names and 4 corporate names as search terms for the authority databases and OPACs noted in Chapter 4. The names were selected, as much as possible, from a list of popular authors whose works are held by many Japanese, Korean, and Chinese libraries. In addition, names that have different Chinese character forms in each region or authors with pseudonyms were preferred. The names selected for the search are listed in Table 5-1.

5.2 Adoption and character forms of Chinese characters

The adoption and character forms of Chinese characters are shown in Table 5-2. The NLC uses simplified Chinese characters for all access points. Among the search terms, however, only the record for “毛泽东” has a variant access point in traditional Chinese characters “毛澤東.” As such, it seems that the traditional Chinese character forms are also recorded for limited authors. The China Academic Library & Information System (CALIS) has several authorized access points. Therefore, authorized access points in simplified Chinese characters, traditional Chinese characters, and pinyin are given for all authority records for Chinese names. Three kinds of variant access points (in simplified and traditional Chinese characters and in pinyin) are also given for one variant name.

The Hong Kong Chinese Authority Name Workgroup (HKCAN) adopts Romanization for authorized access points, and the equivalent Chinese characters are recorded in Heading Linking Entry Fields.

As traditional Chinese characters are used in Taiwan, all access points of authority records for Chinese persons and corporate bodies are recorded in traditional Chinese characters in the National Central Library (NCL) and National Taiwan University Library (NTUL). OPACs of NLC and NTUL and authority databases of CALIS and HKCAN allow both simplified and traditional Chinese characters as search terms, regardless of which characters are input; therefore, the systems return the same results. One exception among the search terms was that “衛慧” and “卫慧” in NLC’s OPAC returned different results. In the Synergy of Metadata Resources in Taiwan (SMRT) system of NCL, the results were different depending on whether traditional Chinese characters or simplified Chinese characters were used as search terms; more records were returned when traditional Chinese characters were used.

According to the manual of NACSIS-CAT, it adopts the letter type on the materials to be cataloged for the letter type of authorized access points. Therefore, character forms of authorized access points may differ among records, and these may be simplified or traditional Chinese characters or Japanese Kanji. Keio adopts simplified Chinese characters for access points, but for Taiwanese names, traditional Chinese characters are also allowed. Both
organizations allow character forms other than those adopted for authorized access points as variant access points. As the NACSIS-CAT system adopts the “Kanji integrated index” provided by the National Institute for Informatics (NII) and the OPAC of Keio University Libraries (Keio) also has a cross-reference table of Chinese characters, which was established based on “Kanji integrated index,” both systems allow any types of Chinese

Table 5-1 Search terms

<table>
<thead>
<tr>
<th>Simplified Chinese Hanzi</th>
<th>Traditional Chinese Hanzi</th>
<th>Japanese Kanji</th>
<th>Korean Hanja</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal names</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>苏轼</td>
<td>蘇軾</td>
<td>蘇軾</td>
<td>蘇軾</td>
<td>1036-1101.</td>
</tr>
<tr>
<td>康有为</td>
<td>康有為</td>
<td>康有為</td>
<td>康有為</td>
<td>1858-1927.</td>
</tr>
<tr>
<td>孙文</td>
<td>孫文</td>
<td>孫文</td>
<td>孫文</td>
<td>Sun, Yat-sen, 1866-1925.</td>
</tr>
<tr>
<td>鲁迅</td>
<td>魯迅</td>
<td>魯迅</td>
<td>魯迅</td>
<td>Real name is 周樹人(Zhou, Shuren), 1881-1936.</td>
</tr>
<tr>
<td>宋庆龄</td>
<td>宋慶齡</td>
<td>宋慶齡</td>
<td>宋慶齡</td>
<td>Sun, Wen's wife, 1893-1981.</td>
</tr>
<tr>
<td>毛泽东</td>
<td>毛澤東</td>
<td>毛澤東</td>
<td>毛澤東</td>
<td>1893-1976.</td>
</tr>
<tr>
<td>吴浊流</td>
<td>吳濁流</td>
<td>吳濁流</td>
<td>吳濁流</td>
<td>1900-1976, a writer in Taiwan</td>
</tr>
<tr>
<td>溥仪</td>
<td>溥儀</td>
<td>溥儀</td>
<td>溥儀</td>
<td>The last emperor of China in Qing dynasty, 1906-1967.</td>
</tr>
<tr>
<td>金庸</td>
<td>金庸</td>
<td>金庸</td>
<td>金庸</td>
<td>A writer in Hong Kong, 1924-. Real name is 查良鏞(Cha, Liangyong).</td>
</tr>
<tr>
<td>卫慧</td>
<td>衛慧*</td>
<td>衛慧</td>
<td>衛慧</td>
<td>A writer, 1973-. Real name is 周衛慧(Zhou, Weihui)</td>
</tr>
<tr>
<td>王健</td>
<td>王健</td>
<td>王健</td>
<td>王健</td>
<td>Several persons share the same name.</td>
</tr>
<tr>
<td>Corporate names</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>国家图书馆</td>
<td>國家圖書館</td>
<td>國家圖書館</td>
<td>國家圖書館</td>
<td>The National Library of China in Beijing, Mainland China.</td>
</tr>
<tr>
<td>香港大学</td>
<td>香港大學</td>
<td>香港大學</td>
<td>香港大學</td>
<td>The University of Hong Kong, established 1911.</td>
</tr>
<tr>
<td>中国国民党</td>
<td>中國國民黨</td>
<td>中國國民黨</td>
<td>中國國民黨</td>
<td>Kuomintang of China, a political party in Taiwan.</td>
</tr>
<tr>
<td>全国人民代表大会</td>
<td>全國人民代表大會</td>
<td>全國人民代表大會</td>
<td>全國人民代表大會</td>
<td>The National People's Congress of the People's Republic of China (Mainland China).</td>
</tr>
</tbody>
</table>

Note. *In Hong Kong, "衛" is officially should be written as "衞". However, in HKCAN, both "衛" and "衞" are used and for "衛慧", "衞" is used.
Table 5-2 Adoption and character forms of Chinese characters

<table>
<thead>
<tr>
<th>Organization</th>
<th>Adoption</th>
<th>Character forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>AAP/VAP</td>
<td>Simplified</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 0 S$a毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 S$S$S$a01S$a毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AAP/VAP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td>CALIS</td>
<td>200 0 S$y毛泽东</td>
<td>Simplified/traditional</td>
</tr>
<tr>
<td></td>
<td>400 0 S$y毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AAP/VAP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heading Linking Entry/VAP</td>
<td></td>
</tr>
<tr>
<td>HKCAN</td>
<td>ex.)</td>
<td>Basically in traditional</td>
</tr>
<tr>
<td></td>
<td>100 1 SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>700 1 Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td>NCL</td>
<td>AAP/VAP</td>
<td>Traditional</td>
</tr>
<tr>
<td></td>
<td>100 1 Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 Sa二十八畫生</td>
<td></td>
</tr>
<tr>
<td>NTUL</td>
<td>AAP/VAP</td>
<td>Traditional</td>
</tr>
<tr>
<td></td>
<td>100 1 毛，澤東</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 毛，澤東</td>
<td></td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>AAP/VAP</td>
<td>Depends on material to be cataloged</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;AAP&gt;毛，澤東</td>
<td>Mao, Ze Dong</td>
</tr>
<tr>
<td></td>
<td>&lt;VAP&gt;毛，澤東</td>
<td>Mao, Ze Dong</td>
</tr>
<tr>
<td>Keio</td>
<td>AAP/VAP</td>
<td>Basically simplified</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td>/can choose traditional for Taiwanese</td>
</tr>
<tr>
<td></td>
<td>100 1 Sa毛，澤東 S9A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 Sa毛泽东，タクトウS9A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 SaMao, Zedong S9A</td>
<td></td>
</tr>
<tr>
<td>NLK</td>
<td>AAP/VAP</td>
<td>Depends on material to be cataloged</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $a毛，澤東</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $a毛泽东，タクトウ</td>
<td></td>
</tr>
<tr>
<td>YUL</td>
<td>AAP/VAP</td>
<td>Depends on material to be cataloged</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $a毛泽东，タクトウ</td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>AAP/VAP</td>
<td>Depends on material to be cataloged</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 Sa毛泽东，タクトウ</td>
<td></td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point.
characters as search terms. However, as the index and the table cannot define all possible pairs of different types of Chinese characters, the cross-referencing may not always be successful.

The National Library of Korea (NLK) adopts Chinese character forms in a subfield of a variant access point, while YUL adopts them in a subfield of an authorized access point. In both organizations, Chinese characters may be in simplified or traditional form, depending on the materials to be cataloged. In NLK, a Chinese character form is added when “distinguishing several persons with the same name or disambiguation is needed”. In YUL, a Chinese character form is added when it is shown on the materials as being cataloged. Regardless of the types of Chinese characters, the system returns the same results in YUL’s OPAC. The same is essentially true in NLK’s OPAC, but for some search terms, the results were different. For example, “孙文” and “孫文” returned a different number of search results. Both OPACs adopt Korean Hanja rather than Chinese traditional characters, even in records of Chinese materials. For example, these OPACs displayed “康有為” instead of “康有為”.

The LC/NACO Authority File (LCNAF) started to allow the recording of non-Latin scripts in variant access point fields in 2008. At that time, access points in non-Latin scripts, which were recorded in bibliographic records of WorldCat, were automatically copied to LCNAF records as variant access points. Therefore, the Library of Congress (LC) has adopted Chinese character forms as variant access points. However, recording Chinese character forms is optional, and the types of characters are not defined. Any types of characters on materials to be cataloged can be recorded in LCNAF records. Because not all character forms for a name are recorded, some access points could only be retrieved in traditional Chinese characters, or vice versa. For example, “中国国家图书馆. 少年儿童馆” could only be retrieved in simplified Chinese characters, and “香港大學. 中文學院” could only be retrieved in traditional Chinese characters. The two types of Chinese characters are not cross-referenced.

As seen above, all of the organizations researched adopt any type of Chinese character forms, although this is not mandatory for the LC. It is equally not mandatory for NLK and YUL, but both organizations record Chinese character forms provided that these forms are known from the materials to be cataloged. For most organizations, the types of Chinese characters are not unified and attempt to resolve this complex problem by cross-referencing several types of Chinese characters. However, in some databases and OPACs, cross-referencing has partially failed—the different search results depend on the types of characters used in the search terms. Currently, a perfect cross-reference table does not exist and it will likely never exist because covering all possible pairs of simplified and traditional Chinese characters is impossible. Therefore, retrieval of some Chinese characters may be unsuccessful.

An additional problem is that the system which adopts a cross-reference table or index of Chinese characters cannot distinguish between the types of Chinese characters used in the
search terms and search results. For example, the Chinese name for the National Library of China is “国家图书馆” and it should be shown in simplified Chinese characters; if it were shown in the traditional characters, “國家圖書館”, the meaning would change to the Chinese name for the National Central Library of Taiwan. The search results for these two organizations should be differentiated, but they are shown intermixed in these systems.

5.3 Treatments and types of Romanization

5.3.1 Treatments and types

As Table 5-3 shows, Hanyu pinyin is adopted by all organizations in China. In most organizations, Romanized forms other than pinyin may be recorded as variant access points. In NLC, pinyin forms are mandatory and automatically generated by the cataloging system.

In CALIS, the pinyin form is an authorized access point as well as simplified and traditional Chinese character forms. Wade-Giles Romanization forms, Cantonese Romanized forms, or other Romanized forms may be recorded as variant access points. Although CALIS’s manual does not designate pinyin forms as being mandatory, all records retrieved by search terms use pinyin forms as access points.

According to the interview, NCL adopts pinyin as mandatory, and it is automatically generated by the system. However, the author found that the authority records for corporate bodies in the SMRT system do not have pinyin forms or any other Romanized forms. In NTUL, pinyin forms are mandatory as variant access points.

When the Chinese Name Authority Database that was jointly developed by NCL and NTUL was started in 1998, both organizations adopted the Wade-Giles Romanization system rather than pinyin. Therefore, Wade-Giles Romanized forms may be retained as variant access points in some authority records of both organizations.

Hanyu pinyin is a Romanization system for Mandarin, but not for Cantonese, which is spoken in Hong Kong. Therefore, in Hong Kong, catalogers are not always familiar with the pinyin system. Despite this, pinyin has been adopted for many authorized access points of HKCAN because HKCAN adopted LCNAF’s forms as authorized access points.

The LC adopts Hanyu pinyin for many authorized access points. In North America in the late 1970s, using Wade-Giles Romanization for Chinese in library sectors was considered a problem despite governments, the media, and ISO adopting Hanyu pinyin. Therefore, the LC proposed changing its Romanization system for Chinese from Wade-Giles to pinyin in 1979. However, the CEAL opposed this idea not only because of conversion costs but also because it was not the right time. In 1990, the LC proposed the issue again, and the conversion was finally implemented in 2000. At this time, authorized access points of LCNAF’s records, as well as access points of bibliographic records, were converted to pinyin forms. Currently, the ALA/LC
## Table 5-3 Treatments and types of Romanization

<table>
<thead>
<tr>
<th>Organization</th>
<th>Treatment</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>AAP/VAP (mandatory and automatically generated)</td>
<td>Hanyu pinyin</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 0 $Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 0 $S7bu$$Samao ze dong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 $S5e$S6a$S13Sa毛润芝</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 $S6a01$S$Samao run zhi</td>
<td></td>
</tr>
<tr>
<td>AAP/VAP</td>
<td>ex.)</td>
<td>Hanyu pinyin;</td>
</tr>
<tr>
<td>CALIS</td>
<td>200 0 $jt0yjt0ySa毛泽东</td>
<td>Other Romanization (optional as VAP)</td>
</tr>
<tr>
<td></td>
<td>200 0 $jt0yjt0ySa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 0 $celcyec0ySaMao Zedong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 $Sa二十八画生</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 $Sa二十八畫生</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 0 $SaEr shi ba hua sheng</td>
<td></td>
</tr>
<tr>
<td>HKCAN</td>
<td>AAP/VAP</td>
<td>Mainly Hanyu pinyin;</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Tse-tung</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaEr shi ba hua sheng</td>
<td></td>
</tr>
<tr>
<td>NCL</td>
<td>VAPs (mandatory and automatically generated for personal names)</td>
<td>Hanyu pinyin (mandatory);</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td>NTUL</td>
<td>Mandatory as VAP</td>
<td>Hanyu pinyin</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>A subfield of AAP/VAP (optional)</td>
<td>Hanyu pinyin;</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;AAP&gt;毛, 沛东</td>
<td></td>
</tr>
<tr>
<td>Keio</td>
<td>Mandatory as VAP</td>
<td>Hanyu pinyin;</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $Sa毛泽东</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $Sa모우, 타クト우$9A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Zedong$9A</td>
<td></td>
</tr>
<tr>
<td>NLK</td>
<td>VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>YUL</td>
<td>VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>LC</td>
<td>AAP/VAP</td>
<td>Mainly Hanyu pinyin;</td>
</tr>
<tr>
<td></td>
<td>ex.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 1 $SaMao, Zedong</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Tsetung</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 1 $SaMao, Zetong</td>
<td></td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point
Romanization Table for Chinese adopts *Hanyu pinyin*.* 17

Many authorized access points of the LC and HKCAN are in *pinyin* forms because AACR2 22.3C2 stipulates that “if the name of a person entered under surname is written in a nonroman script, romanize the name according to the table for the language adopted by the cataloguing agency.”* 18 However, the LC also applies the alternative rule of 22.3C2, namely, “choose the Romanized form of name that has become well established in English-language reference sources for a person entered under surname whose name is in a language written in a nonroman script.”* 18,19 Currently, RDA 9.2.2.5.3 follows the same rule, which the LC adopts. Therefore, an authorized access point for “孫文” of HKCAN and the LC is “Sun, Yat-sen,” which is not a *pinyin* form of “孫逸仙” (a pseudonym of 孫文). Similarly, an authorized access point of “香港大學” is “University of Hong Kong,” not “Xianggang da xue” (a pinyin form). In addition, parts of the access points include English, such as “China. Quan guo ren min dai biao da hui” for “全国人民代表大会” in HKCAN and the LC.

Many records of HKCAN and the LC have variant access points of Wade-Giles Romanization forms. Until HKCAN was developed in 1999, Romanization systems adopted by university libraries in Hong Kong were varied, with libraries adopting Wade-Giles, *pinyin*, both systems, or none at all. 20 From 2000 to 2001, at the same time as the LC’s conversion project in which the LC changed their Romanized forms of Chinese records from Wade-Giles to *Hanyu pinyin*, member libraries of HKCAN also converted their Romanized forms to *pinyin*, and most members submitted their updated records to the HKCAN database. 14 In the LC, the Wade-Giles forms were retained in variant access point fields after the conversion, except for corporate bodies (originally in field 410) and meetings (originally in field 411). 16 Wade-Giles forms still remain in current HKCAN records as variant access points because it seems that member libraries of HKCAN also retained their Wade-Giles forms in their authority records, and because many HKCAN records were copied from the LCNAF.

In Japan, both NACSIS-CAT and Keio adopt *Hanyu pinyin* for the Romanization of Chinese names. In NACSIS-CAT, a *pinyin* form can be recorded in a subfield of access points, the syntax of authorized access points for Chinese names are “Names in Chinese characters||Japanese *yomi* in *katakana*||*Hanyu pinyin,*” and *Hanyu pinyin* is optional. 21 Recording *pinyin* forms for variant access points is also optional. In Keio, *pinyin* forms are mandatory as variant access points. Romanized forms other than *pinyin* are recorded as variant access points in both organizations.

As for the two organizations in South Korea, Romanized forms are recorded as variant access points and the types are undesignated. Romanized forms on materials to be cataloged may be recorded in both organizations. In other words, recording Romanized forms is optional and may not be recorded by both organizations.
5.3.2 Adoption of Tone Marks, Apostrophes, and Umlaut Marks with “U”

The adoption of tone marks, apostrophes, and umlaut marks was investigated because it is difficult for catalogers to input these marks, and each organization may record them differently. Organizations in South Korea were not included in this investigation because they do not designate types of Romanized forms.

As Table 5-4 shows, none of the organizations adopt Tone Marks, which are prescribed to add to Hanyu pinyin in 汉语拼音方案. An apostrophe should be used when a syllable is joined to another syllable starting with a, o, or e, according to 汉语拼音方案. For example, a Hanyu pinyin for the word “西安” should be “Xi’an”, rather than “Xian” to distinguish it from another syllable “xian” in Chinese.

In NLC, an apostrophe is not used, as NLC divides all syllables in pinyin forms; thus, joined syllables do not exist in NLC’s records. CALIS, NCL, and NTUL adopt an apostrophe, but not in all records. A name “長安” (“长安” in simplified Chinese characters), for example, should be described as “Chang’an” in pinyin. However, pinyin forms for names “王長安” and “李長安” in CALIS, NCL, and NTUL’s databases do not always use an apostrophe. CALIS’s database includes five persons who have the name “王長安” and their pinyin form is “Wang Chang’an.” On the other hand, there are three persons who share the name “李長安” and their pinyin form is “Li Changan.” In the SMRT system of NCL, four “王長安” exist and three of them have “Wang, Chang an” as the pinyin form, while one of them has “WangChangan.” Similarly, among three “李長安”, two have “Li, Changan” as the pinyin form and the last has “LiChangan.” In NTUL’s OPAC, their pinyin forms are “Wang, Chang’an” and “Li, Changan,” respectively.

Table 5-4 Adoption of tone marks, apostrophes, and umlaut marks

<table>
<thead>
<tr>
<th></th>
<th>Tone marks</th>
<th>An apostrophe</th>
<th>An Umlaut mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>×</td>
<td>×</td>
<td>v/u (disunited)</td>
</tr>
<tr>
<td>CALIS</td>
<td>×</td>
<td>〇 (disunited)</td>
<td>v/u (disunited)</td>
</tr>
<tr>
<td>HKCAN</td>
<td>×</td>
<td>〇</td>
<td>ü</td>
</tr>
<tr>
<td>NCL</td>
<td>×</td>
<td>〇 (disunited)</td>
<td>ü/u (disunited)</td>
</tr>
<tr>
<td>NTUL</td>
<td>×</td>
<td>〇 (disunited)</td>
<td>ü/u (disunited)</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>×</td>
<td>×</td>
<td>ü/u (disunited)</td>
</tr>
<tr>
<td>Keio</td>
<td>×</td>
<td>〇</td>
<td>ü</td>
</tr>
<tr>
<td>LC</td>
<td>×</td>
<td>〇</td>
<td>ü</td>
</tr>
</tbody>
</table>
In Japan, NACSIS-CAT does not use an apostrophe because it divides all syllables into *pinyin* forms, as NLC does.\(^\text{23}\) Keio uses an apostrophe in accordance with 汉语拼音方案.

As the LC-ALA Romanization Table for Chinese follows 汉语拼音方案, except for tone marks\(^\text{17}\), the LC uses an apostrophe.

One more mark used in *Hanyu pinyin* is an umlaut.\(^\text{24}\) An umlaut mark is only added to “u” in *pinyin* and the pronunciations of “u” and “ü” are different in Mandarin. For example, Chinese characters“路(lu)”and “绿(lü)” should be pronounced differently.

Table 5-5 shows the *pinyin* forms of four personal and corporate names including the “ü” pronunciation in each organization in China and the LC, as examples. Although NCL defines *pinyin* forms as mandatory, the author found that any authority records for corporate bodies have *pinyin* forms in the SMRT system of NCL, as of February 2015.

Some records use “v” or “u” instead of “ü” in NLC and CALIS. The reason for using “v” is because v is the only letter of the alphabet not used by *Hanyu pinyin*. In mandarin text input software, commonly used in Mainland China, the key “v” on the keyboard is used for inputting “ü.”\(^\text{25}\) In Taiwan, on the other hand, phonetic symbols called ZhuYin (BoPoMoFo) are used to input Chinese on the computer, rather than *pinyin*. In Hong Kong, the Cangjie input method, in which each key on the keyboard corresponds to a certain graphical part of a Chinese character, is commonly used.\(^\text{26}\) Therefore, inputting “v” instead of “ü” is specific to Mainland China. HKCAN uses “ü”, as does the LC, in accordance with 汉语拼音方案. The handling of an umlaut is not a written rule in NCL and NTUL. Some *pinyin* forms of both organizations include “u” instead of “ü.”

<table>
<thead>
<tr>
<th>姓名</th>
<th>汉语</th>
<th>法律出版社</th>
<th>全国法院干部业余法律大学</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLC</td>
<td>lv yao dou</td>
<td>fa lu chu ban she</td>
<td>quan guo fa yuan gan bu ye yu fa lv da xue</td>
</tr>
<tr>
<td>CALIS</td>
<td>lv Yaodou</td>
<td>Fa lv chu ban she</td>
<td>Quan guo fa yuan gan bu ye yu fa lv da xue</td>
</tr>
<tr>
<td>HKCAN</td>
<td>Lü, Yaodou</td>
<td>Fa lu chu ban she</td>
<td>Quan guo fa yuan gan bu ye yu fa lu da xue (China)</td>
</tr>
<tr>
<td>NCL</td>
<td>LuYao; Lu, Yaodou</td>
<td>No <em>pinyin</em> form</td>
<td>No <em>pinyin</em> form</td>
</tr>
<tr>
<td>NTUL</td>
<td>Lü, Yaodou</td>
<td>Fa lu chu ban she</td>
<td>Quan guo fa yuan gan bu ye yu fa lu da xue</td>
</tr>
<tr>
<td>LC</td>
<td>Lü, Yaodou</td>
<td>Fa lu chu ban she</td>
<td>Quan guo fa yuan gan bu ye yu fa lu da xue (China)</td>
</tr>
</tbody>
</table>
NACSIS-CAT’s manual clearly defines that you do “not necessarily have to add an umlaut.” Therefore, some pinyin forms include an umlaut and others do not. Keio always use “u” instead of “ü.”

Although the LC adopts an umlaut, when NACO contributors add a new authority record, an umlaut is automatically omitted during the comparison process that checks whether a new authorized access point is clearly differentiated from existing access points. In other words, if an authorized access point “Lu, Shi” already exists in the LCNAF, a new authorized access point “Lü, Shi” cannot be established, because “ü” is standardized as “u.” This means that in practical terms, the LC does not differentiate “ü” from “u”. Currently, for example, variant access points for “Lu, Shi” include “呂侍” (pronounced as Lü, Shi) and “魯石” (pronounced as Lu, Shi) and others. This situation appears to be in the process of being resolved now, as the LC and PCC agreed in November 2013 that all personal name authority records coded RDA should be differentiated.

As seen above, all organizations except those in South Korea adopt Hanyu pinyin as a Romanization system for Chinese, whereas tone marks of pinyin are not adopted by any organization. Not using tone marks creates more homonyms of names than using tone marks. Using an apostrophe and an umlaut is disunited in many organizations. When conducting a search using pinyin, both the apostrophe and umlaut may be omitted from the search query. However, the search efficiency will be lower if “ü” is not differentiated from “u” because they are actually two different pronunciations. Moreover, using “v” instead of “ü” may bring confusion as “v” is disunited and specific to Mainland China. Some organizations retain Wade-Giles Romanization as variant access points.

Inputting Hanyu pinyin easily causes mistypes. In NLC and NCL, pinyin forms are automatically generated by the system. For OCLC connexion users, Princeton University East Asian Library offers the OCLC Connexion Pinyin Conversion Macro. However, as some Chinese characters have more than one pinyin form, confirmation by human eyes is always needed. At least, in the systems of NACSIS-CAT and Keio, pinyin forms are inputted manually. Therefore, imperfection of pinyin forms cannot be avoided.

5.4 Separation of surname and given name with a comma

As Table 5-6 shows, NTUL, NACSIS-CAT, and Keio separate a surname and its given name in Chinese characters using a comma. NCL separates the surname and given name in Chinese characters using “|”, although the separation is not observed in the SMRT system. In authority
records linked to access points of bibliographic records in NCL’s OPAC, we can see “|” marks. NCL separates the surname and given name because they had used *Chinese MARC Format for Authority Records* (CMARC/A), which prescribes that surnames and given names for Chinese, Japanese, and Korean persons should be separated with a subfield code $b$ unaccompanied by a comma.\textsuperscript{31} As NCL started to adopt *MARC 21 Format for Authority Data* (MARC 21/A) in December 2011, “|” was inserted instead of $b$ when the data conversion from CMARC/A to MARC 21/A was undertaken. This is still used by other libraries, which use CMARC/A download authority data from NCL. Other organizations do not separate surnames and given names in Chinese characters.

As 汉语拼音方案 does not regulate Romanization of names for persons or corporate bodies, *pinyin* forms in some organizations are slightly different from others; for example, connecting more than two syllables of proper names, capitalizing the head of the proper name, and separating the surname and its given name.

As noted above, NLC and NACSIS-CAT divides all syllables in *pinyin* forms and does not use any punctuation.\textsuperscript{32,33} CALIS separates a surname and its given name with a space, not with a comma in *pinyin* forms. Other organizations use a comma between the surname and given name in *pinyin* forms.

NLC and NACSIS-CAT records all syllables in lower cases. Some variant access points of NACSIS-CAT, however, connect two syllables of first names or capitalize the first letter of proper names. It seems that rules of *pinyin* forms in NCL are not too strictly applied, therefore, access points such as “Kang, You wei” (it should be “Kang, Youwei”) and “Wu, Zhuo liu” (it should be “Wu, Zhuoliu”) are evident. In access points in sample data from NLK and YUL, the first letter of a surname and its given name are capitalized. However, as the sample data is limited and recording *pinyin* forms is not mandatory for these two organizations, both “connecting” and “capitalizing” for both organizations are shown as “undesignated” in Table 5-6.

5.5 Representations in local languages outside China, Hong Kong, and Taiwan

In both NACSIS-CAT and Keio, *yomi* forms are mandatory. In NACSIS-CAT, *yomi* forms are recorded in the subfield of access points. In Keio, *yomi* forms are recorded as variant access points. *Yomi* is Japanese pronunciations of Chinese characters. One Chinese character could have several Japanese pronunciations; thus, in NACSIS-CAT, *yomi* should be assigned according to pronunciations included in the dictionary called 大漢和辞典語彙索引 (*Daikanwa jiten goi sakuin*). If the dictionary does not include the name, the commonly used pronunciation should be given.\textsuperscript{33} However, determining “the commonly-used pronunciation” is difficult.
Table 5-6 Separation between a surname and a given name etc.

<table>
<thead>
<tr>
<th>Organization</th>
<th>Chinese character form</th>
<th>Pinyin form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Separation</td>
<td>Connecting</td>
</tr>
<tr>
<td>NLC</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>CALIS</td>
<td>×</td>
<td>〇</td>
</tr>
<tr>
<td>HKCAN</td>
<td>×</td>
<td>〇</td>
</tr>
<tr>
<td>NCL</td>
<td>&quot;フォ&quot; ex.)毛澤東</td>
<td>〇</td>
</tr>
<tr>
<td>NTUL</td>
<td>comma</td>
<td>〇</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>comma</td>
<td>× (AAP)</td>
</tr>
<tr>
<td>Keio</td>
<td>comma</td>
<td>〇</td>
</tr>
<tr>
<td>NLK</td>
<td>×</td>
<td>Undesignated</td>
</tr>
<tr>
<td>YUL</td>
<td>×</td>
<td>Undesignated</td>
</tr>
<tr>
<td>LC</td>
<td>×</td>
<td>〇</td>
</tr>
</tbody>
</table>

Notes. AAP - Authorized Access Points. Connecting means connection of more than two syllables of proper names. Capital letter means using a capital letter in the head of proper name. Separation means separation of a surname and a given name.

For example, according to NACSIS-CAT’s manual, “yomi of Chinese characters “叶”, “龙”, and “向” should be “ヨウ”, “リュウ”, and “コウ”, respectively, as long as these characters appear as personal surnames.” On the other hand, note 23.3.3.2 ア) of NCR1987 and its revisions state that “for yomi of Kanji, which are used especially for personal names, use the yomi as it is.” Then, examples of “葉昌熾 (ショウ, ショウ)” and “沈復 (シン, フク)” are shown. As yomi of “叶”, which is a simplified Chinese character of “葉”, should be “ヨウ”, the yomi is different from the rule of NCR. In the authority data of NACSIS-CAT, yomi of “葉” is, in fact, disunited; yomi of “葉昌熾” is “ショウ, ショウシ.” For another person’s surname “葉”, some data have yomi “ショウ” and some others have “ヨウ.” Yomi of a Chinese character for surnames sometimes differs depending on the dictionary, and as just described, unification of the yomi for one Chinese character is different. In Keio, yomi is adopted from the authority data of NACSIS-CAT, and if NACSIS-CAT does not have the data, Kan-on pronunciations are selected as yomi. Kan-on (literally “sounds of Han”) pronunciations are one kind of various
readings which a Kanji may have.\textsuperscript{35} For example, for a Chinese character “夏”, “カ” is a Kan-on pronunciation although it also has other pronunciations such as “ガ”, “ケ”, and “ナツ”. As authority data of NACSIS-CAT do not always adopt Kan-on, yomi for one Chinese character may differ in Keio’s database. However, Keio stated in an interview that it records yomi for descriptive purposes only and does not think unification of yomi for one Chinese character is needed.

On some Japanese materials, transcriptions of Chinese pronunciations in Japanese katakana are shown. In this case, both NACSIS-CAT and Keio adopt the transcription as a yomi form, instead of Japanese pronunciations of Chinese characters, in accordance with NCR 23.3.3.2.\textsuperscript{33} For example, when a transcription of Kanji in katakana, “ウー ヤンズ”, is shown on the material as well as an author’s name “呉彦祖” in Kanji, NACSIS-CAT and Keio adopt “ウー, ヤンズ” as a yomi form of an access point “呉, 彦祖” and do not record “ゴ, ゲンソ”, which is a commonly used Japanese pronunciation of “呉彦祖”. Moreover, even though transcriptions of Chinese pronunciations are not shown on the material, if there are well-known katakana representations based on the Chinese pronunciation that have become well established in reference sources, NACSIS-CAT adopts the representation.\textsuperscript{23} Similarly, in Keio, if a transcription form of Chinese pronunciations in katakana could be found from anywhere on the material or reference sources, Keio adopts the transcription as a yomi. Therefore, it could be said that transcriptions based on Chinese pronunciations are preferred over Japanese pronunciations of Kanji, including Kan-on.

The reason for this preference, according to Miyasaka, is that after NCR was revised in 1977, there were claims that using Japanese pronunciations of Kanji for Korean Hanja names was racial discrimination.\textsuperscript{34} In fact, “for Chinese and Korean names in Kanji, record Japanese pronunciations of these Kanji”\textsuperscript{36} was the rule in 3.4.3.2.2 (4) of NCR1977. However, in additions and revisions to NCR1977, which was published in 1983, the rule was changed to the current form in NCR1987 23.3.3.2.\textsuperscript{37} As Miyasaka pointed out, transcriptions of Chinese pronunciations in katakana are not helpful access points for use because they may be different depending on the materials.\textsuperscript{34}

NLK adopts transcriptions of Chinese pronunciations in Hangul as authorized access points for people active after 1911. For people active before 1911 and all corporate bodies, transliteration of Chinese characters into Hangul is recorded as authorized access points. Before 2000, NLK adopted transliterations of Chinese characters into Hangul as authorized access points for all persons and corporate bodies; it was changed in 2001 because NLK assumes that using pronunciation of a foreigner’s native language for the foreigner’s name is an international custom.\textsuperscript{8} The rule of transcription is according to 외래어 표기법 (Orthographic Rules for Adopted Words) established by the National Institute of the Korean Language.\textsuperscript{8} For people
active after 1911, transliteration of Chinese characters into Hangul may be recorded as variant access points.\(^8\)

YUL adopts transliterations of Chinese characters into Hangul as authorized access points, and transcriptions of Chinese pronunciations in Hangul as variant access points. Although 외래어 표기법 regulates the rule for transcription of Chinese pronunciations, representations shown on materials are not always in accordance with 외래어 표기법 and thus many variant access points may exist in one record.

In some LCNAF records, Romanization of Japanese yomi, Hangul forms, and other language forms (for example, a Hebraic form for “孫文”) are recorded as variant access points.

5.6 Discussion: differences in representations and problems to be solved

As several types of Chinese characters (namely, simplified, traditional, Japanese Kanji, and Korean Hanja) exist, matching Chinese character forms among several databases is difficult. Moreover, as one simplified Chinese character may correspond to several traditional Chinese characters (for example, a simplified Chinese character “云” corresponds to the traditional Chinese character “雲” and another traditional Chinese character “雲”), there is presently no perfect cross-reference table for these characters. For perfect string matching, recording all types of Chinese character forms as access points in each database is desirable.

Pinyin is another choice for string matching, because almost all organizations adopt pinyin forms as Romanization of Chinese character forms, although it is not mandatory in NACSIS-CAT and South Korea. However, because handling of an umlaut of “u” is disunited in several organizations, it may be an obstacle to string matching. NLC and NACSIS-CAT separate more than one syllable of proper names and do not use capital letters. 汉语拼音正词法基本规则 (Basic Rules of the Chinese Phonetic Alphabet Orthography), which is the national standard of Mainland China, regulates capitalizing the first letter of a surname and a given name and connecting two syllables of a surname or a given name.\(^38\) If NLC and NACSIS-CAT adopt this rule, the form of Hanyu pinyin in all organizations will be unified.

Notes

1 Part of this chapter has already been published as 木村麻衣子. 中国人・団体著者名典拠データの表記の相違: 中国，日本，韓国を中心に. Library and Information Science. 2013, no. 69, p. 19-46.

Yu, Abraham J. Evaluation and Analysis of the Chinese Name Authority Files under Development in Beijing, Hong Kong, and Taipei. OCLC, 2002, 52p.


19 “Library of Congress Rule Interpretations, 22.3C. Names written in a nonroman script”.


24 Actually, there is one more mark used in Hanyu pinyin. It is a circumflex added to “e”, such as “ê.” However, “ê” is only used as an exclamation in Chinese, thus it was omitted from this study.


28 This problem was already pointed out by Bolick, Hsi-chu. Problems in the establishment of nonunique Chinese personal headings with special reference to NACO guidelines and vendor-supplied authority control. Library Resources & Technical Services. 1999, vol. 43, no. 2, p. 95-105. He also pointed out that an update of the software or revision of the transliteration rule is needed.

“OCLC Connexion Pinyin conversion macro”. Princeton University East Asian Library. 


32 国家图书馆图书采编部. 中国机读规范格式使用手册; 中文图书名称规范数据款目著录规则; 中文图书主题规范数据款目著录规则. 1999, 182p.


38 GB/T16159-2012. 汉语拼音正词法基本规则. Notes that this rule also stipulates using tone marks.
Chapter 6

Representations of Japanese name authority data in Chinese character cultures

6.1 Checkpoints and search terms

6.1.1 Checkpoints

The following six aspects, which are assumed to be treated differently by organizations in representing Japanese personal and corporate names, are provided for investigation:

1) Adoption and character forms of Chinese characters
2) The relating of \textit{yomi} to their corresponding \textit{Kanji}
3) Treatments and types of Romanization
4) Separation of surname and given name with a comma
5) Representations in local languages outside Japan
6) Names in \textit{hiragana}.

These six topics for investigation were determined as follows. The first topic, adoption and character forms of \textit{Kanji}, pertains because most Japanese personal names are in \textit{Kanji}. While some \textit{Kanji} have the same forms as simplified or traditional \textit{Hanzi} or Korean \textit{Hanja}, others do not. Japanese personal and corporate names should be written in Japanese \textit{Kanji}, and it should thus be investigated whether Japanese \textit{Kanji} are accurately used for authority data in China, Korea, or the LC.

In this study, \textit{Kanji} includes \textit{shinjitai} (新字体), \textit{kyūjitai} (舊字体), other \textit{itaiji} (異体字), and \textit{Kokuji} (国字). It is beyond the scope of this study to treat problems occurring in Japanese \textit{Kanji}, because the study aims to investigate representations of Japanese names in the whole Chinese character cultural sphere.

The second topic, the relating of \textit{yomi} to their corresponding \textit{Kanji}, is studied because \textit{yomi}, which shows how the \textit{Kanji} should be pronounced, is important. One \textit{Kanji} may have several \textit{yomi}; thus, a name in \textit{Kanji} and its \textit{yomi} are treated as a pair in Japanese names.

The third topic, treatments and types of Romanization, has to do with examining the different Romanized forms used by organizations because, as noted in Chapter 3, there are two prominent methods of Japanese Romanization. The author also conducted a preliminary investigation to estimate how differences in Romanization affect VIAF matching.

For the fourth topic, separating surnames and given names with a comma should be considered.
For Japanese users, in particular, it is desirable that surnames and given names are treated separately, as they prefer searched names to be collated according to surnames in Kanji on screen.

The fifth topic, representations in local languages outside Japan, aims to show what kinds of local representations for Japanese names—other than Japanese Kanji, hiragana, or katakana—are adopted in China and Korea.

The sixth topic, names in hiragana, is to investigate how Japanese names in hiragana are managed. In China and Korea, though character forms are different, Japanese Kanji can typically be converted into their own Chinese characters. However, names in hiragana and katakana cannot be replaced by Chinese characters, as they represent only sounds. It is necessary to investigate how these names are dealt with in China and Korea. In this study, only names in hiragana are investigated because katakana has the same nature as hiragana, in the sense that both represent syllables. Usually, katakana is used for foreign or foreign-derived words.

6.1.2 Search terms

The author used 11 personal names and five corporate names to reflect the known characteristics of Japanese names as found in different Romanization schemes. These are listed in Table 6-1 as search terms for authority databases and OPACs.

As noted in Chapter 4, Japanese Romanization has two main systems. Thus, the author selected names in which Romanization may differ between these two systems. The differences include long vowel pronunciations, the moraic nasal (letter “n”), and the moraic obstruent (small “tsu” in Japanese). Reasons why the respective names were chosen are also shown in Table 6-1. Some names did not appear in a specific database. A similar name was used in such cases.

6.2 Adoption and character forms of Chinese characters

The adoption and character forms of Chinese characters when names are in Kanji are shown in Table 6-2. Three Japanese organizations use Chinese characters for authorized access points and variant access points. The same is true in the cases of the China Academic Library & Information System (CALIS), the National Central Library (NCL) of Taiwan, and National Taiwan University Library (NTUL). The Hong Kong Chinese Authority Name Workgroup (HKCAN) adopts Romanization for authorized access points, and equivalent Chinese characters are recorded in Heading Linking Entry Fields.
Table 6-1 Search terms

<table>
<thead>
<tr>
<th>Kanji</th>
<th>Yomi</th>
<th>Romanization (in Hepburn system without any marks)</th>
<th>Reason for choice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>吉川英治</td>
<td>ヨシカワエイジ</td>
<td>Yoshikawa Eiji</td>
<td>The Romanization differs between Hepburn/kunrei: ji/zi</td>
</tr>
<tr>
<td>瀬戸内寂聴</td>
<td>セトウチジャクチョウ</td>
<td>Setouchi Jakucho</td>
<td>Including a long vowel; the Romanization differs between Hepburn/kunrei: ja/zya</td>
</tr>
<tr>
<td>六角恒広</td>
<td>ロッカクツネヒロ</td>
<td>Rokkaku Tsunehiro</td>
<td>Including a moraic obstruent</td>
</tr>
<tr>
<td>頼山陽</td>
<td>ライサンヨウ</td>
<td>Rai Sanyo</td>
<td>Including a syllabic-final &quot;n&quot; preceding a vowel or &quot;y&quot;</td>
</tr>
<tr>
<td>安野光雅</td>
<td>アンノミツマサ</td>
<td>Anno Mitsumasa</td>
<td>Including a syllabic-final &quot;n&quot;</td>
</tr>
<tr>
<td>大江健三郎</td>
<td>オオエケンザブロウ</td>
<td>Oe Kenzaburo</td>
<td>Pronouncing &quot;oo&quot; written as &quot;o&quot;</td>
</tr>
<tr>
<td>妹尾河童</td>
<td>セノオカッパ</td>
<td>Senoo Kappa</td>
<td>Pronouncing &quot;oo&quot; (not a long vowel)</td>
</tr>
<tr>
<td>森鷗外</td>
<td>モリオウガイ</td>
<td>Mori Ogai</td>
<td>Pronouncing &quot;ou&quot; written as &quot;O&quot;</td>
</tr>
<tr>
<td>空海</td>
<td>クウカイ</td>
<td>Kukai</td>
<td>Pronouncing &quot;uu&quot; written as &quot;u&quot;</td>
</tr>
<tr>
<td>飯田哲也</td>
<td>イイダテツナリ</td>
<td>Iida Tetsunari</td>
<td>More than one possible yomi for the same combination of Kanji</td>
</tr>
<tr>
<td>宮部みゆき</td>
<td>ミヤベミユキ</td>
<td>Miyabe Miyuki</td>
<td>(&quot;Iida Tetsuya&quot;);</td>
</tr>
<tr>
<td><strong>Corporate Names</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>講談社インターナショナル株式会社</td>
<td>コウダンシャインタラナショナルカブシキガイシャ</td>
<td>Kodansha Intanashonaru Kabushiki Gaisha</td>
<td>Including a long vowel.</td>
</tr>
<tr>
<td>お茶の水女子大学</td>
<td>オチャノミズジョーシンダイガク</td>
<td>Ochanomizu Joshi Daigaku</td>
<td>Including hiragana; the Romanization differs between Hepburn/kunrei: cha/tya</td>
</tr>
<tr>
<td>文化出版局</td>
<td>ブンカシュッパンキュウ</td>
<td>Bunka Shuppanyoku</td>
<td>The Romanization differs between Hepburn/kunrei: shu/syu</td>
</tr>
<tr>
<td>内閣官房</td>
<td>ナイカクカンボウ</td>
<td>Naikaku Kanbo</td>
<td>Including a moraic nasal &quot;n&quot; preceding &quot;m&quot;, &quot;b&quot;, &quot;p&quot;</td>
</tr>
<tr>
<td>日中経済協会</td>
<td>ニッチュウケイサイキョウカイ</td>
<td>Nitchu Keizai Kyokai</td>
<td>Including a moraic obstruent preceding &quot;ch&quot;</td>
</tr>
<tr>
<td>Organization</td>
<td>Adoption</td>
<td>Character form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>CALIS</td>
<td>AAP/VAP</td>
<td>Japanese Kanji for Japanese access points; simplified/traditional Chinese characters for Chinese access points</td>
<td></td>
</tr>
<tr>
<td>HKCAN</td>
<td>Heading linking entry/VAP</td>
<td>Depends on material to be cataloged</td>
<td></td>
</tr>
<tr>
<td>NCL</td>
<td>AAP/VAP</td>
<td>Depends on material to be cataloged</td>
<td></td>
</tr>
<tr>
<td>NTUL</td>
<td>AAP/VAP</td>
<td>Traditional Chinese characters (AAP); Japanese Kanji (a VAP)</td>
<td></td>
</tr>
<tr>
<td>NDL</td>
<td>AAP/VAP</td>
<td>Japanese Kanji</td>
<td></td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>AAP/VAP</td>
<td>Japanese Kanji</td>
<td></td>
</tr>
<tr>
<td>Keio</td>
<td>AAP/VAP</td>
<td>Japanese Kanji</td>
<td></td>
</tr>
</tbody>
</table>
| NLK          | A subfield of VAP | Depends on material to be cataloged 
|              | ex) 400 1# $a$ 대강건삼랑=$h$大江健三郎 |
| YUL          | A subfield of AAP | Depends on material to be cataloged 
|              | ex) 100 1# $a$대강건삼랑$Sh$大江健三郎 |
| LC           | VAP      | Depends on material to be cataloged 
|              | ex) 400 1# $a$大江健三郎 |

Note. AAP - Authorized Access Point; VAP - Variant Access Point.
In the National Library of Korea (NLK), Chinese characters are added as designations associated with variant access points whenever differentiations between persons with the same names or clearer identifications are needed. In Yonsei University Library (YUL), Chinese characters are added as designations associated with authorized access points. LC adopts Romanization for the authorized access point, and Chinese characters are recorded as variant access points.

The three organizations in Japan adopt Japanese Kanji for access points. CALIS adopts Kanji for Japanese authorized access points, and Chinese simplified and traditional characters for Chinese authorized access points. According to the CALIS manual, authorized access points in different languages, but representing the same entity, should be shown together in Authorized Headings Fields (2XXs) under one authority record to ensure user accessibility. However, in the actual authority database, these access points sometimes appear in different records. There are two records for Senoo, Kappa in CALIS, for example: one for the Japanese authorized access point “妹尾河童 (セノオ, カッパ), 1930-,” and the other for the Chinese authorized access point “妹尾河童, 1930-.” By way of contrast, only one record for Oe Kenzaburo exists in CALIS, and the record has authorized access points in simplified Chinese script, traditional Chinese script, Hanyu Pinyin, and Japanese.

According to the interview, NCL constructs both Japanese authorized access points for Japanese materials and Chinese authorized access points for translated materials by taking character forms from the resources being cataloged and cross-referencing them, using Heading Linking Entry Fields. However, in the case of Rai Sanyo, search results differ between search terms “頼山陽” (in Kanji) and “賴山陽” (in traditional Chinese characters). In this case, cross-referencing seems not to have been done.

NTUL adopts traditional Chinese character forms for authorized access points and Japanese Kanji for a variant access point.

In the HKCAN database, Chinese characters corresponding to authorized access points are recorded in Heading Linking Entry Fields. According to an e-mail response from HKCAN, the types of Chinese characters generally depend on the resources being cataloged. For example, the authorized access point of “Nitchū Keizai Kyōkai”, “日中経済協会” (in Kanji), is recorded in the Heading Linking Entry Field instead of in its traditional Chinese characters, which should be “日中経済協会.” On the other hand, the authorized access point of “Kōdansha Intānashonaru Kabushiki Kaisha”, “講談社インターナショナル株式會社” (“會” is a traditional Chinese character, and in Kanji it should be “会”), is recorded for the heading linking entry. In both
NLK and YUL, Chinese characters are transcribed (i.e. without any change) from the resources being cataloged. This means that several types of Chinese characters might be recorded.

LC does not prescribe forms of Chinese characters. Thus, some records do not have variant access points in Kanji even if the authorized access points are for Japanese names. For instance, a record with the authorized access point “Setouchi, Jakuchō, 1922-” (record number: nr 94028021) has six variant access points in Chinese characters: “瀨戸内寂聴”, “瀨戶內寂聴”, “瀨戸内寂聴”, “瀨戸内寂聴” and “瀨戸内寂聴.” However, none of these forms is in the correct Kanji, “瀨戸内寂聴.”

6.3 Relating yomi to their corresponding Kanji

As noted already, yomi is highly important for Japanese names in order to distinguish one name from a similar name that uses the same characters but that is pronounced differently. As Table 6-3 shows, three Japanese organizations construct authority data relating author names to yomi and their corresponding Kanji entries. In this study, the word yomi refers yomi in katakana form.

Conforming to “Appendix C: Multiscript Records Model A: Vernacular and transliteration” of MARC 21/A, the NDL employs Kanji for regular fields, adds its yomi in katakana in 880 fields, and links both fields by using a linkage subfield entitled $6. The linkage subfield $6 contains a linking field and an occurrence number, so that users can identify which yomi entry is associated with which Kanji entry. When a name is in Kanji, 880 fields are mandatory.

NACSIS-CAT separates names in Kanji from their yomi using “||” in a authorized access points and variant access point fields. According to its manual, yomi is mandatory when the information is applicable or readily available. In the system at Keio, Kanji with yomi are indexed together in order to distinguish between the same characters with different pronunciations. Additionally, Keio prepares extended fields for Japanese data in addition to regular fields. Extended fields repeat regular fields and add a subfield, $9K, for yomi entries. Subfield $6 contains occurrence numbers to indicate which yomi entry corresponds to which Kanji entry. CALIS adds yomi in parenthesis to designate authorized access points.

According to an interview, it is recorded by way of variant access point if the cataloger judges it necessary in NLK. In YUL, yomi shown on the authority database of NACSIS-CAT which YUL usually consults for or yomi on the materials being cataloged will be recorded as variant access points. However, the correspondences between the Kanji access points and their yomi are not usually indicated in both NLK and YUL. In other organizations, yomi is generally not employed.
Table 6-3 *Yomi* and the corresponding *Kanji*

<table>
<thead>
<tr>
<th>Organization</th>
<th><em>Yomi</em> and corresponding <em>Kanji</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>CALIS</td>
<td>Using () for <em>yomi</em> as a designation associated with the AAP</td>
</tr>
<tr>
<td></td>
<td>ex.</td>
</tr>
<tr>
<td>HKCAN</td>
<td>No <em>yomi</em></td>
</tr>
<tr>
<td>NCL</td>
<td>No <em>yomi</em></td>
</tr>
<tr>
<td>NTUL</td>
<td>No <em>yomi</em></td>
</tr>
<tr>
<td>NDL</td>
<td>880 for <em>yomi</em>, using $6$ to pair <em>Kanji</em> and <em>yomi</em> (mandatory if applicable)</td>
</tr>
<tr>
<td></td>
<td>ex.</td>
</tr>
<tr>
<td></td>
<td>100 1# $880-01$a大江, 健三郎</td>
</tr>
<tr>
<td></td>
<td>880 1# $6100-01/$1$aオオエ, ケンザブロウ</td>
</tr>
<tr>
<td></td>
<td>880 1# $6100-01/(B$aOoe, Kenzaburo</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>Using</td>
</tr>
<tr>
<td></td>
<td>ex.)大江, 健三郎</td>
</tr>
<tr>
<td></td>
<td><em>Yomi</em> at subfield&quot;$A&quot; of AAP with $9A + <em>yomi</em> as one of the AAPs with $9K, both are mandatory</td>
</tr>
<tr>
<td>Keio</td>
<td>ex.</td>
</tr>
<tr>
<td></td>
<td>100 1# $880-01$a大江, 健三郎</td>
</tr>
<tr>
<td></td>
<td>880 1# $6100-01/$1$aオオエ, ケンザブロウ</td>
</tr>
<tr>
<td></td>
<td>100 1# $6100-01/(B$aOoe, Kenzaburo</td>
</tr>
<tr>
<td></td>
<td>880 1# $6100-01/(B$aOoe, Kenzaburo</td>
</tr>
<tr>
<td>NLK</td>
<td>Sometimes as VAP (with no correspondences to the AAP)</td>
</tr>
<tr>
<td>YUL</td>
<td>Sometimes as VAP (with no correspondences to the AAP)</td>
</tr>
<tr>
<td>LC</td>
<td>No <em>yomi</em></td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point
6.4 Treatments and types of Romanization

6.4.1 Treatments of Romanization

As Table 6-4 shows, NDL and Keio provide Romanization in the same manner as *yomi*. NTUL adopt Romanization as mandatory for a variant access point. NACSIS-CAT occasionally employs Romanization as an optional variant access point; likewise, Romanization is not mandatory in CALIS, NCL, NLK, or YUL. In YUL, Romanization is automatically generated by the system based on Hepburn Romanization from *yomi* if it exists. In many search terms, Romanization in NACSIS-CAT also exists in variant access points in CALIS, in that CALIS copies Japanese authority data from NACSIS-CAT. HKCAN and LC employ Romanization for authorized access points.

6.4.2 Types of Romanization

Except for those that do not adopt Romanization and NTUL, all organizations studied employ the Hepburn system for Romanization of Japanese. NTUL adopt Romanization as a variant access point, but a Romanization system is not designated. According to the interview, variant access points are came from books in hand, the authority database of NDL, the Kyoto University Library (it uses NACSIS-CAT), and the Internet. From sample records of NTUL, a record for “村瀬, 秀甫” has a variant access point “Murase, Shuwuho” which does not accord to Hepburn Romanization system, although most of records have access points in Hepburn Romanization forms.

NDL used to follow the *kunrei-shiki* (訓令式) system, but since November 2011, it has been using the Hepburn system also. Since HKCAN complies with AACR2, MARC 21, and LCNAF, it is natural that HKCAN should also use the ALA/LC Romanization Tables.

The ALA/LC Romanization Tables were revised in 2012. The new ALA/LC Romanization Tables set out the rules in more detail but do not attempt to modify the existing rules. Both old and new versions are based on *Kenkyusha’s New Japanese-English Dictionary* and the American National Standard (ANSI Z39.11-1972).

Table 6-5 shows differences among four organizations that adopt Hepburn Romanization for authorized access points or mandatory elements.

Even the same “Ō” can be used for two different *yomi*. One is “オオ” in “O”e Kenzaburo. Another is “オウ” in “O”gai Mori. For Ogai Mori, disregarding the use of a macron, four organizations use the same representation, “Mori, Ōgai”. For Oe Kenzaburo, however, NDL uses “Ooe, Kenzaburo”, despite the fact that in *JAPAN/MARC MARC21 フォーマット* におけるローマ字読み表記要領 (Romanization Rule in JAPAN/MARC MARC21 Format),
### Table 6-4 Treatments and types of Romanization

<table>
<thead>
<tr>
<th>Organization</th>
<th>Treatment</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALIS</td>
<td>Optional as VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>HKCAN</td>
<td>ex.)</td>
<td>Hepburn</td>
</tr>
<tr>
<td>NCL</td>
<td>Optional as VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>NTUL</td>
<td>ex.)</td>
<td>Undesignated</td>
</tr>
<tr>
<td>NDL</td>
<td>ex.)</td>
<td>Hepburn</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>Optional as VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>Keio</td>
<td>ex.)</td>
<td>Hepburn</td>
</tr>
<tr>
<td>NLK</td>
<td>Optional as VAP</td>
<td>Undesignated</td>
</tr>
<tr>
<td>YUL</td>
<td>VAP (if the yomi exists)</td>
<td>Hepburn</td>
</tr>
<tr>
<td>LC</td>
<td>ex.)</td>
<td>Hepburn</td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point
which NDL has applied since December 2011, the representation should “omit the representation of prolonged sounds or long vowels”; thus Oe Kenzaburo should be presented as “Oe, Kenzaburo”. According to an e-mail response from NDL, when re-generating Romanization from *yomi* at the time of data migration, in line with the introduction of the new system in 2012, all of the forms “オオ”, “コオ”, “ソオ”, “トオ” etc. were converted to “oo”, “koo”, “soo”, “too” etc., without exception. This rule is correct for names like “Senoo, Kappa” or “Tamura, Naoomi” (田村直臣), in which “o” is not a long vowel. In this case, batch conversion caused some data to be described outside the Hepburn system. Meanwhile, new data constructed by NDL after 2012 use “o” for long vowels, resulting in a variety of representations.

According to *ローマ字のつづり方* (*A method of Writing Japanese in Roman Characters*), the moraic obstruent (small “tsu” in Japanese) before “ch” should be represented by doubled consonants. On the other hand, the Hepburn system set forth by the Ministry of Foreign Affairs of Japan for Japanese passport applications (hereafter referred to as “MOFA Hepburn”) has a special provision which states that “before “ch,” the moraic obstruent should be described by inserting “t” in the same manner as ALA/LC Romanization Tables”. Of the four organizations shown in Table 6-5, only NDL reflects the policy of *ローマ字のつづり方*.

Additionally, ALA/LC Romanization Tables and *ローマ字のつづり方* have a rule that, when preceding a vowel or “y” within a single word, a syllabic-final “n” should be followed by an apostrophe: for example, “Rai, San’yō”, rather than “Rai, Sanyō”; but “Senoo, Kappa”, where the “n” is syllabic-initial. Neither Keio nor MOFA Hepburn applies the rule.

<table>
<thead>
<tr>
<th>Names examined</th>
<th>NDL</th>
<th>Keio</th>
<th>HKCAN</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>講談社インターナショナル株式会社</td>
<td>Kodansha Intanashonaru Kabushiki Gaisha</td>
<td>Kodansha intanashonaru kabushiki gaisha</td>
<td>Kōdansha Intā nashonaru Kabushiki Kaisha</td>
<td>Kōdansha Intā nashonaru Kabushiki Kaisha</td>
</tr>
<tr>
<td>大江健三郎</td>
<td>Ooe, Kenzaburo</td>
<td>Oe, Kenzaburo</td>
<td>Ōe, Kenzaburō</td>
<td>Ōe, Kenzaburō</td>
</tr>
<tr>
<td>日中経済協会</td>
<td>Nicchu Keizai Kyokai</td>
<td>Nitchu keizai kyokai</td>
<td>Nitchū Keizai Kyōkai</td>
<td>Nitchū Keizai Kyō kai</td>
</tr>
<tr>
<td>頼山陽</td>
<td>Rai, San'yo</td>
<td>Rai, Sanyo</td>
<td>Rai, San’yō</td>
<td>Rai, San'yō</td>
</tr>
</tbody>
</table>
Moreover, HKCAN and LC Romanize “株式会社” (meaning “a joint-stock corporation”) as “kabushiki kaisha”, applying the rule of ALA/LC Romanization Tables even though the Japanese pronunciation should be “kabushiki gaisha”.

Except for the search terms shown in Table 6-5, all organizations describe each Romanization in Table 6-1 in the same way, with the exclusion of long vowels. According to 日本目録規則 1965年版 (NCR 1965ed.), the moraic nasal “n” before the letters “m”, “b”, or “p” is written “m” rather than “n”. Although MOFA Hepburn applies this rule, the ALA/LC Romanization Tables and ローマ字のつづり方 do not apply it, and all organizations describe “内閣官房” (meaning “Cabinet Secretariat”) either as “Naikaku Kanbo” or “Naikaku Kanbō”, not “Naikaku Kambo”.

Table 6-6 summarizes the differences in the Hepburn system between ローマ字のつづり方, NCR 1965ed., MOFA Hepburn, ALA/LC Romanization Tables, and those organizations that adopt Romanization in their authority data. Although all of these organizations adopt the Hepburn system, some differences were found. For instance, according to ローマ字のつづり方, a circumflex accent mark should be used to describe long vowels, whereas no circumflex should be used in MOFA Hepburn. NDL used the circumflex accent mark until March 2002, but ceased in April 2002. Keio does not use any circumflexes. On the other hand, HKCAN and LC use macrons for long vowels, based on the ALA/LC Romanization Tables.

Table 6-6 Differences in Hepburn Romanization

<table>
<thead>
<tr>
<th>로마자 쓰는 방법</th>
<th>NCR 1965ed.</th>
<th>MOFA Hepburn</th>
<th>ALA/LC Romanization Tables*</th>
<th>NDL</th>
<th>Keio</th>
</tr>
</thead>
<tbody>
<tr>
<td>The moraic nasal “n” preceding &quot;m&quot;, &quot;b&quot;, or &quot;p&quot;</td>
<td>Use &quot;n&quot;</td>
<td>Use &quot;m&quot;</td>
<td>Use &quot;m&quot;</td>
<td>Use &quot;n&quot;</td>
<td>Use &quot;n&quot;</td>
</tr>
<tr>
<td>Syllabic-final &quot;n&quot; preceding a vowel or &quot;y&quot;</td>
<td>Followed by an apostrophe</td>
<td>Followed by an apostrophe</td>
<td>No apostrophe</td>
<td>Followed by an apostrophe</td>
<td>No apostrophe</td>
</tr>
<tr>
<td>The moraic obstruent before &quot;ch&quot;</td>
<td>Double consonants</td>
<td>Use &quot;t&quot;</td>
<td>Use &quot;t&quot;</td>
<td>Use &quot;t&quot;</td>
<td>Double consonants</td>
</tr>
<tr>
<td>Marks for long vowels</td>
<td>Circumflex</td>
<td>Circumflex</td>
<td>Not used</td>
<td>Macrons</td>
<td>Not used</td>
</tr>
</tbody>
</table>

NOTE. * HKCAN and LC conform to it.
6.4.3 Defects in VIAF matching

As Table 6-6 indicates, NDL Romanization is different from that used by LC. It is probable that the difference influences data matching.

To estimate how differences in Romanization affect VIAF matching, the following preliminary investigation was conducted. First, authorized access points that include the moraic obstruent (small “tsu”) before “chi” – that is to say, “ッチ” in their *yomi* – were searched December 15, 2012, in Web NDL Authorities using a SPARQL query. The search result set of 702 authority records was extracted. Second, both NDL Romanization (many including the letters “cc”) and Hepburn Romanization (converting the letters “cc” to “tc”) of these target records were searched on VIAF. The results were examined manually to determine whether the target entities were searched correctly.

Although many entities could be searched using both NDL and Hepburn Romanization – that is, the search results were the same – in 111 cases (about 15.8% of the records), the search results were different and had some defects. These 111 cases did not include the nine names used only as subject access points by NDL. Unmatched data that could not be confirmed as belonging to the same person because of lack of information in the records were omitted from this investigation. Table 6-7 shows details of the defects.

In Table 6-7, “Cannot retrieve when using ‘tc’” means either that no result was returned, or that a wrong entity appeared when Hepburn Romanization was used. “Duplicates” means that different records representing the same entity appeared when searched using both NDL Romanization and Hepburn Romanization. Separate records should be merged. “Cannot retrieve See Also records when using ‘tc’”, means that related records (See Also records) appear only when NDL Romanization is used. “Mismatched” means that NDL access points are incorrectly matched and that there are other records that should be merged.

The ratio of author names that include “ッチ” in their *yomi* to the total number of NDL authority records is small. However, 25,222 authorized access points that include “オオ” in their *yomi* in the Web NDL Authorities are supposed to have the letters “oo” in their Romanization, which differs from the Hepburn Romanization. Taking into account the authorized access points, including “コオ”, “トオ”, etc., in their *yomi* that have letters like “koo”, “too”, etc. in their Romanization, there must be more Romanizations in NDL that differ from the Hepburn system. It can be presumed that there are a certain number of defects in these records. The result from this investigation indicates that the differences that emerge from Romanization may constitute barriers to data matching.
Table 6-7 Defects of VIAF (caused by the Romanization of the moraic obstruent in NDL)

<table>
<thead>
<tr>
<th>Situations of defects</th>
<th>Records</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot retrieve when using &quot;te&quot;</td>
<td>91</td>
<td>“グッチ裕三, 1952-(VIAFID255899377)” does not appear when using &quot;te&quot; to search.</td>
</tr>
<tr>
<td>Duplicates</td>
<td>11</td>
<td>Separate records “七珍万宝, 1762-1831(VIAFID54069766)” and “Shitchin Manpō, 1762-1831(VIAFID265159018)” exist.</td>
</tr>
<tr>
<td>Cannot retrieve See Also records when using &quot;te&quot;</td>
<td>8</td>
<td>“坂元, 雪鳥, 1879-1938(VIAFID259548962)” has a See Also Record “坂元, 三郎, 1879-1938(VIAFID251461743)” which only appear when using &quot;cc&quot; to search.</td>
</tr>
<tr>
<td>Total:</td>
<td>111</td>
<td></td>
</tr>
</tbody>
</table>

6.5 Separation of surname and given name with a comma

As Table 6-8 shows, the three Japanese organizations and NTUL separate surnames and given names with a comma for all names consisting of a surname and a given name. CALIS separates surnames and given names with a comma in yomi forms, which are recorded as designations associated with authorized access points. However, neither CALIS nor HKCAN separate them in Chinese character forms. NCL separates surnames and given names using the “|” mark in its OPAC, although in the SMRT system of NCL, the “|” mark is omitted from all access points.

NLK inserts a space between surnames and given names in Hangul transcribed from Japanese pronunciation forms. YUL inserts a comma between surnames and given names in Hangul transcribed Japanese pronunciations, and yomi forms. However, neither organization separates them in Chinese character forms and Hangul transliterated from Chinese character forms. The LC does not separate them in Chinese character forms. All organizations adopting Romanized forms separate them in Romanized forms.

6.6 Representations in local languages outside Japan

As Table 6-9 shows, NLK adopts Hangul transcribed from Japanese pronunciations for personal authorized access points, and Hangul transliterated from Chinese characters for variant access points. For corporate names, NLK adopts Hangul transliterated from Chinese characters for authorized access points. However, for corporate names that include hiragana or katakana,
<table>
<thead>
<tr>
<th>System</th>
<th>Instructions</th>
<th>Examples</th>
</tr>
</thead>
</table>
| CALIS | Separate with comma in *yomi* forms (as designations associated with AAP); No separation in Chinese character forms. | ex.)  
200 $a$大江健三郎$g$(オオエ, ケンザブロウ)$f1935- |
| HKCAN | Separate with comma in Romanized forms; No separation in Chinese character forms. | ex.)  
100 1# $a$オオエ, ケンザブロウ  
700 1# $a$大江健三郎 |
| NCL | Separate with ‘|’ in Chinese character forms (not applicable in the SMRT system). | ex.)  
大江健三郎 |
| NTUL | Separate with comma in all forms | ex.)  
100 1# 大江, 健三郎 |
| NDL | Separate with comma in all forms | ex.)  
100 1# $a$大江, 健三郎 |
| NACSIS-CAT | Separate with comma in all forms. | ex.)  
AAP: 大江, 健三郎 || オオエ, ケンザブロウ  
VAP: $a$Ôe, Kenzaburô; $a$Ooe, Kenzaburo |
| Keio | Separate with comma in all forms. | ex.)  
100 1# $a$大江, 健三郎 $g$オオエ, ケンザブロウ $9A$  
100 1# $a$大江, 健三郎 $g$オオエ, ケンザブロウ $9O$  
100 1# $a$大江, 健三郎 $g$オオエ, ケンザブロウ $9W$  
100 1# $a$大江, 健三郎 $g$オオエ, ケンザブロウ $9K$  
100 1# $a$大江, 健三郎 $g$オオエ, ケンザブロウ $9R$ |
| NLK | No separation in *Hangul* transliterated from Chinese character forms and Chinese character forms; Space in *Hangul* transcripted from Japanese pronunciation forms. | ex.)  
100 1# $a$대강건삼랑  
400 1# $a$대강건삼랑 $g$大江健三郎 |
| YUL | No separation in *Hangul* transliterated from Chinese character forms and Chinese character forms; Separate with comma in *Hangul* transcripted from Japanese pronunciation forms. | ex.)  
100 1# $a$대강건삼랑  
400 1# $a$대강건삼랑, $g$大江健三郎 |
| LC | Separate with comma in Romanized forms; No separation in Chinese character forms. | ex.)  
100 1# $a$Oe, Kenzaburô  
400 1# $a$Oe, Kenzaburô |

Note. AAP - Authorized Access Point; VAP - Variant Access Point
Table 6-9 Representations in local languages outside Japan

<table>
<thead>
<tr>
<th></th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALIS</td>
<td>偶尔使用 Hanyu Pinyin 或其他变体形式作为 VAP (日本权威记录)</td>
</tr>
<tr>
<td>HKCAN</td>
<td>偶尔使用英文形式或 Hanyu Pinyin 形式作为 VAP</td>
</tr>
<tr>
<td>NCL</td>
<td>没有本地形式</td>
</tr>
<tr>
<td>NTUL</td>
<td>Hangul 从日本发音转录用于个人 AAP</td>
</tr>
<tr>
<td>NLK</td>
<td>Hangul 从汉字字符转录用于 VAP和个人 AAP</td>
</tr>
<tr>
<td>YUL</td>
<td>Hangul 从日本发音转录用于 VAP</td>
</tr>
<tr>
<td>LC</td>
<td>Many variant forms as VAP</td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point
NLK records the hiragana or katakana parts in their original forms. For example, for “講談社インターナショナル”, NLK records it as “강담사인터나ショナル.” YUL adopts Hangul transcribed from Japanese pronunciations for variant access points, and Hangul transliterated from Chinese characters for authorized access points. In YUL, Hangul forms are automatically generated by the system using the rules for foreign language notation from yomi if it exists.

In the CALIS database, Hanyu Pinyin is one of the authorized access points in Chinese authority records. In Japanese authority records, Hanyu Pinyin or other variant forms in such scripts as Hangul are recorded in variant access point fields of partial records. However, most of these variant access points have been copied from NACSIS-CAT, which can be deduced from the fact that such variant access points also exist in NACSIS-CAT records.

English names and Hanyu Pinyin are sometimes recorded in the variant access point fields of HKCAN records. When comparing them with LCNAF records, it is found that many variant access points in HKCAN are the same as variant access points in LCNAF. However, HKCAN has added some variant access points on its own accord. For instance, a record for “Rai, San’yō” has a Reference Entry “賴襄, 1780-1832” in the HKCAN database, but it does not exist in the LCNAF record. There are many variant forms in non-Latin scripts—such as Hangul, Cyrillic, and so on—in the variant access point fields of LCNAF records.

NTUL records a form in traditional Chinese characters as an authorized access point, when it is different from Japanese Kanji form.

6.7 Names in hiragana

To investigate how organizations deal with names in hiragana, five personal names (four new names added to “Miyabe, Miyuki” of Table 6-1) that include hiragana script were searched in the seven non-Japanese databases. Table 6-10 shows the search results. With regard to the two Korean organizations and NTUL, OPACs were searched instead of the authority databases because the latter are not accessible.

In the Japanese authority records of CALIS, all names are described correctly in Japanese. However, some authors also have Chinese authority records, in which the hiragana parts are converted to Chinese characters. Their representations may be a result of the resources’ being cataloged, such as Japanese books translated into Chinese, in which Japanese author names in hiragana are usually substituted by Chinese characters. The problem is that some Japanese and Chinese authority records are not linked or merged, although the CALIS manual provides for the fact that several authorized access points should be placed together under one authority.
Table 6-10 Names in *hiragana*

<table>
<thead>
<tr>
<th>Names</th>
<th>CALIS</th>
<th>HKCAN</th>
<th>NCL</th>
<th>NTUL</th>
<th>NLK</th>
<th>YUL</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>あさのあつこ</td>
<td>AAP(J): あさのあつこ (アサノ, アツコ)</td>
<td>AAP: Asano, Atsuko HLE: [Characters are garbled] VAP: 淺野敦子</td>
<td>AAP: 淺野 敦子 VAP: あさの, あつこ</td>
<td>아사노 아쓰코(16); あさのあつこ(11)</td>
<td>아사노 아쓰코(24); 아사노의 아동(24)</td>
<td>浅野敦子, あさのあつこ</td>
<td>浅野敦子</td>
</tr>
<tr>
<td>さくらももこ</td>
<td>AAP(J): さくらももこ (サクラ, モモコ)</td>
<td>AAP: Sakura, Momoko HLE: 櫻桃子 VAP: [Characters are garbled]</td>
<td>AAP: 櫻桃子 VAP: さくら, さくらももこ</td>
<td>사쿠라, 사쿠라모모코(5) VAP: 櫻桃子</td>
<td>N/A</td>
<td>봉화사쿠라모모코(5) VAP: 櫻桃子</td>
<td></td>
</tr>
<tr>
<td>藤本ひとみ</td>
<td>AAP(J): 藤本ひとみ (フジモト, ヒトミ)</td>
<td>AAP: Fujimoto, Hitomi HLE: 藤本瞳 VAP: 藤本瞳</td>
<td>AAP: 藤本, ひとみ AAP2: 藤本, 瞳</td>
<td>후지모토 히토미(5); 藤本ひとみ(2) AAP: 櫻桃子 VAP: 櫻桃子</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>宮部みゆき</td>
<td>AAP(J): 宮部みゆき (ミヤベ, ミユキ)</td>
<td>AAP: Miyabe, Miyuki HLE: 宮部美幸 VAP: 宮部美幸</td>
<td>AAP: 宮部, 美幸 AAP2: 宮部美幸</td>
<td>미야베 미유키(51); 宮部미유키(82) AAP: 櫻桃子 VAP: 櫻桃子</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>よしもとばなな (Former pseudonym is 吉本ばなな)</td>
<td>AAP(J): 吉本ばなな (ヨシモト, バナナ) VAP(J): 吉本香蕉</td>
<td>AAP: Yoshimoto, Banana HLE: 吉本真秀子 VAP: 吉本真秀子</td>
<td>AAP1: 吉本真秀子 AAP2: 吉本芭娜娜</td>
<td>요시모토 바나나(38); 요시모토바나나(35); 요시모토바나나(35); 요시모토바나나(18) AAP: 櫻桃子 VAP: 櫻桃子</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.** AAP - Authorized Access Point; VAP - Variant Access Point. Only VAPs in Chinese characters or Japanese are shown.

In the column for CALIS, "(J)" means Japanese access points; "(C)" means Chinese access points.

In the column for HKCAN, "HLE" means a heading linking entries.

In the column for NCL and NTUL, numbers added to AAP or VAP indicate that several authority records exist.

In the columns for NLK and YUL, the number of hits for each form in each OPAC is provided in parentheses.
In the HKCAN database, Japanese forms that correspond to their authorized access points are generally described correctly in Heading Linking Entry Fields; thus, both Japanese forms and Romanization can be used in a search. Only in the record for “Sakura, Momoko” are Chinese characters described in the Heading Linking Entry Field.

Despite the rule on cross-referencing Japanese and Chinese authorized access points, such cross-references could not be confirmed in the OPAC of NCL. In NTUL’s OPAC, although “藤本, ひとみ” is not cross-referenced with “藤本, 映” and an access point “さくら, ももこ” is not exist, regarding other three names, Japanese names in hiragana and its Chinese character forms are cross-referenced.

In the OPAC of NLK, the number of retrieval hits differs according to search terms. Results for Romanization and Japanese forms were the same in the OPAC of YUL. Since Japanese forms are described as variant access points in LCNAF records, both Romanization and Japanese forms could be searched in LCNAF.

6.8 Discussion: differences in representations and problems to be solved

First, no organization adopts Kanji for access points of Japanese names except Japanese organizations, CALIS, and NTUL. However, HKCAN, NCL, NLK, YUL, and LC adopt Chinese characters from the resources being cataloged; there are mixed access points in Kanji and other Chinese characters in databases, depending on the characters found in the resources. It is natural for Japanese authors to wish to be indicated in correct Kanji. However, systems or tables which convert variant characters cannot avoid some mistakes because they are not perfect, and using many characters involves many risks in retrieval. Sharing authority data may facilitate the addition or linking of names in correct Kanji, thus potentially solving this problem.

Furthermore, few organizations display correspondences between Kanji and their yomi. Yomi provides more precise pronunciations than the Romanization of Japanese. For instance, for the same “ō,” its yomi can either be “オウ” or “オオ.” Yomi is one of the most effective tools for identifying names with the same Kanji. Thus, for the purpose of sharing authority data for Japanese names, the existence of yomi in each authority data would be ideal. Otherwise, both Romanization and other data elements, such as birth year, would be necessary for matching. Moreover, correspondences between yomi and Kanji should be shown. The relationships could be shown in authorized access points like those found in records from Keio or CALIS, showing yomi as qualifiers of names. It also could be accomplished using the linkage subfield like NDL. The relationship helps with identification of names because matching algorithms should check
the combination of Kanji and yomi, and if either one does not match the candidate, a negative would result.

Although the Romanization of Japanese is not a perfect surrogate for yomi, it is still a strong guide to identifying names with the same Kanji. Nevertheless, Romanization is not mandatory in NACSIS-CAT, CALIS, NCL, NLK, and YUL. Additionally, Romanization is different among organizations, even though each organization adopts the Hepburn Romanization system. For the purpose of sharing authority data, unification or conversion of a Romanization system is needed. It is problematic that Romanization is not consistent, even in Japan. In addition, some organizations provide Romanization without presenting its correspondence to Kanji. This is also a problem in the case of yomi.

As the above investigation using VIAF shows, the Hepburn system of NDL differs from that of other VIAF participants, which could be one of the impediments to data sharing. In the research on VIAF, many records were successfully matched even when Romanized forms were different. The existence of other data elements which prompt an algorithm for identification might be the reason for this. Thus, enriching other data elements is needed when Romanization cannot be the chief factor in identification. Because the details of the VIAF matching algorithm are not clarified, the data elements weighted by the algorithm need to be discussed between the VIAF participants and OCLC. However, as FRAD and RDA advocate, recording the various attributes of authority data is very important, and doing so would be helpful in terms of any matching algorithm because more information helps to achieve more accurate matching results.

It is natural that some organizations adopt representations in their local language or script. However, as the names in hiragana show, there are some cases in which local and Japanese forms are not linked. There is a fear that some entries might escape a search. Adding access points or linkage to the Japanese forms from other databases is needed in such cases.

Notes

1 Part of this chapter has already been published as Kimura, Maiko. Differences in representations of Japanese name authority data among CJK countries and the Library of Congress. Information Processing and Management. 2014, vol. 50, issue 5, p. 733-751.
4 “Appendix C: multiscript records”. MARC 21 format for authority data.
6 国立情報学研究所, “目録システムコーディングマニュアル”. 国立情報学研究所目録
2015-02-04).
7 Kudo, Yuko. Modified Hepburn Romanization system in Japanese language cataloging: where
8 “『JAPAN/MARC MARC21 フォーマット』におけるローマ字読み表記要領”. 国立国
会図書館.
http://www.ndl.go.jp/jp/library/data/pdf/JM_MARC21revision_romajiyomiyohiyoryo.pdf,
(accessed 2015-02-04).
10 “アクセス・ポイントのローマ字表記要領（2002 年 3 月以前）”. 国立国会図書館.
11 “외래어 표기법”. 국립국어원.
2015-02-05).
Chapter 7

Representations of Korean name authority data in Chinese character cultures

7.1 Checkpoints and search terms

7.1.1 Checkpoints

The following five aspects, which are assumed to be treated differently by organizations in representing Korean personal and corporate names, are examined:

1) Adoption of Hangul forms
2) Adoption of Hanja forms
3) Adoption and types of Romanization
4) Separation of surname and given name with a comma
5) Representations in local languages outside Korea

These five topics for investigation were determined as follows.

(1) First, concerning the adoption of Hangul forms, access points are likely recorded in Hangul within South Korea to this day because the second edition of Korea Cataloging Rules (KCR2) prescribed that all access points be described in Hangul. However, the author nevertheless established this topic because there could exist national and regional variation, such as NACSIS-CAT, not requiring Hangul forms. Furthermore, the author decided to investigate this topic because the difference between whether a Hangul form is adopted as an authorized access point or as one of several variant access points should indicate how much each institution emphasizes Hangul forms.

(2) The author established the topic of the adoption of Hanja (i.e., the Chinese characters used in Korea) forms because it appears that this varies within present-day South Korea. While a Hanja form was not mandated in KCR2, based on examples raised by Park, Seoul National University Library treats Hanja forms as variant access points, while Yonsei University Library (YUL) regards them as authorized access points and as additions to variant access points.

(3) The author established the topic of the adoption and types of Romanized forms suspecting regional variation, as domestic and international standards for the Romanization of the Korean language differ in the present day. 김성원 has brought attention to three primary factors causing formatting differences to arise in the Romanization of Korean personal names: 1) the written order of the surname and given name; 2) whether a comma is used between the surname and given name; and 3) the format of the representation of the name. In 3), in cases of a name with two syllables, 김성원 included the factors of a) whether there is word-by-word separation, b) whether a
(4) The author established the topic of the presence or absence of word-by-word separation and commas between the names for comparative purposes. Whether or not a comma is present should have little actual effect on searches. However, in the case of Japanese names, it is typical to separate the family and first names to suit collocates in which the family and first names are delimited.

(5) The author established the topic of representations in local languages outside South Korea (e.g., katakana) to investigate how they are handled. This was because in regions outside South Korea, a name’s notation may change into that of a local language that applies only to that region. (In NACSIS-CAT, for example, a Korean name written in Chinese characters is assigned its corresponding Japanese yomi.)

7.1.2 Search terms

Search terms were selected from names for which the initial sound rule is applicable, including words with no corresponding Hanja, surnames consisting of two Hanja characters, and persons and corporate bodies that are popular in Japan and China as well as South Korea. Because holdings of each organization are different, not all search terms could obtain a search result in all authority databases or OPACs. When the name could not be retrieved, similar terms were used instead of the search term. Search terms are as Table 7-1 shows.

7.2 Adoptions of Hangul forms

As Table 7-2 shows, three institutions were identified that require access points in Hangul form. In South Korea, a Hangul form is mandated as an authorized access point; however, this is not necessarily the case in institutions outside of South Korea. NACSIS-CAT adopts the rule that, “As a general rule, script is recorded as it is presented in the materials utilized when an authority record is first created. However, noted authors and the like are recorded in the script that is most well-known.” When an authorized access point is in Chinese character form, the Hangul form is appended; but when an authorized access point is in katakana, Hangul is not appended. When a name of an organization in Hangul is adopted as an authorized access point, the Hangul form, separated word-by-word, is appended, even in cases in which the authorized access point is in Hangul form, in order to make searches by word units possible. Keio adopts a Chinese character form as an authorized access point, but mandates that variant access points be recorded in
<table>
<thead>
<tr>
<th>Hanja</th>
<th>Hangul</th>
<th>Notes*</th>
<th>Reason for choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>金大中</td>
<td>김대중</td>
<td>1924-2009, a politician. The 15th President of South Korea.</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>安正孝</td>
<td>안정효</td>
<td>1941-, a writer and translator.</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>徐廷柱</td>
<td>서정주</td>
<td>1914-, a poet. His anthology was translated into Chinese, English, and French.</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>朴正熙</td>
<td>박정희</td>
<td>1917-1979, the 5th to 9th President of South Korea.</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>尹東柱</td>
<td>윤동주</td>
<td>1917-1945, a poet. Came to Japan in 1942, was arrested as a thought criminal by the Japanese police, and dead in prison.</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>金達壽</td>
<td>김달수</td>
<td>1919-1997, a writer. His pseudonym was 大沢達雄. Came to Japan when he was 10 years old. After the second World War, joined to establish Zai Nichi Chosenjin Renmei (Korean League in Japan).</td>
<td>Famous in Korea and other countries.</td>
</tr>
<tr>
<td>李文烈</td>
<td>이문열</td>
<td>1948-, a writer representing the literary community in current South Korea.</td>
<td>The beginning-sound rule applied.</td>
</tr>
<tr>
<td>李光洙</td>
<td>이광수</td>
<td>1892-1953(?), a writer. So-called a founder of Korean literature. His Japanese name is 香山光郎.</td>
<td>The beginning-sound rule applied.</td>
</tr>
<tr>
<td>李御寧</td>
<td>이어령</td>
<td>1934-, a literary critic, a writer, and a symbolist. The first Korean Minister of Culture.</td>
<td>The beginning-sound rule applied.</td>
</tr>
<tr>
<td>盧武鉉</td>
<td>노무현</td>
<td>1946-2009, the 16th President of South Korea.</td>
<td>The beginning-sound rule applied.</td>
</tr>
<tr>
<td>南宮槿</td>
<td>남궁근</td>
<td>The president of the Seoul National University of Technology.**</td>
<td>A surname consisting of two Hanja characters.</td>
</tr>
<tr>
<td>鮮于煇</td>
<td>선우휘</td>
<td>1922-1986, a writer, a journalist, was a chief editor of &quot;조선일보&quot; (The Chosun Ilbo).</td>
<td>A surname consisting of two Hanja characters.</td>
</tr>
<tr>
<td>金하늘</td>
<td>김하늘</td>
<td>Several persons share the same name.</td>
<td>Including a given name with no corresponding Hanja.</td>
</tr>
</tbody>
</table>
### Table 7-1 (continued)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>韓國圖書館協會 한국도서관협회</td>
<td>Seoul National University. Established in 1946 with merging Keijō Imperial University and several colleges.****</td>
<td>Including a word with no corresponding Hanja.</td>
</tr>
<tr>
<td>嶺南大學校 영남대학교</td>
<td>Yeungnam University. Established in 1947.******</td>
<td>The beginning-sound rule applied; a university of the same name (嶺南大學校) exists in Hong Kong.</td>
</tr>
</tbody>
</table>

When a Chinese character form is unclear, a *Hangul* form is recorded as the authorized access point.

### Table 7-2 Adoptions of *Hangul* forms

<table>
<thead>
<tr>
<th>Organization</th>
<th>Adoption</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLK</td>
<td>Mandatory as AAP</td>
<td>100 1# $a이광수=$h李光洙</td>
</tr>
<tr>
<td>YUL</td>
<td>Mandatory as AAP</td>
<td>100 1# $a이광수=$h李光洙</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>AAP/a subfield of AAP/VAP</td>
<td>$&lt;HDNG&gt;이,광洙</td>
</tr>
<tr>
<td></td>
<td>AAP/VAP (when Chinese character form is uncertain)</td>
<td>100 1# $a이,광洙</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400 1# $a이,광수</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 1# $a정,제두</td>
</tr>
<tr>
<td>Keio</td>
<td>AAP (when Chinese character form is uncertain)</td>
<td>(Search terms returned no applicable examples)</td>
</tr>
<tr>
<td>NCL</td>
<td>AAP (when Chinese character form is uncertain)</td>
<td>100 1# $a이광수</td>
</tr>
<tr>
<td>NTUL</td>
<td>VAP (when found in material to be cataloged)</td>
<td>100 1# $a이,광洙</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400 1# $a이,광수</td>
</tr>
<tr>
<td>HKCAN</td>
<td>Heading linking entry/VAP</td>
<td>100 1# $aYi, Kwang-su</td>
</tr>
<tr>
<td></td>
<td></td>
<td>700 1# $a이광수</td>
</tr>
<tr>
<td>LC</td>
<td>VAP</td>
<td>100 1# $aYi, Kwang-su</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400 1# $a이광수</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400 1# $a이광수</td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point.

The National Central Library (NCL) of Taiwan adopts the rule that, in cases where only a *Hangul* form is found, the *Hangul* form should be recorded as an authorized access point. However, when the author searched the actual authority data using the search terms given in Table 7-1, the authorized access points were all in Chinese character form; moreover, the author was unable to find any examples in which an authorized access point was in *Hangul* form. In addition, there exists no rule dictating the recording of a *Hangul* form when one is found, and among the search terms, there was not even a single example of a *Hangul* form as a variant access point corresponding to the authorized access point in Chinese character form. On the other hand, in authority data for Western books where the authorized access point was in Romanized form, there were cases in which the record had been described in *Hangul* form as a variant access point. (For example, the authorized access point of the authority record for a Western book by “安正孝” was recorded as “An, Chŏng-hyo”, and “안정효” was recorded as one of its variant access points.) In the National Taiwan University Library (NTUL), variant access points are described in *Hangul* form only when the *Hangul* form is found in the
resources being catalogued; it is not mandatory.

The Hong Kong Chinese Authority Name Workgroup (HKCAN) assigns the Romanized form to an authorized access point (field 1XX), and records another notation form corresponding to the authorized access point in a heading linking entry field (field 7XX). The author searched the actual authority data using the search terms given in Table 7-1. The results showed that access points recorded in the heading linking entry field were almost all Hangul forms, but there were also cases in which a Chinese character form was recorded, e.g. “李御寧”. The Library of Congress (LC) has no rules related to the adoption of Hangul forms and Chinese character forms, aside from stating that they can be recorded as variant access points. However, for all of the search terms in Table 7-1 Hangul forms had been recorded as variant access points in the actual authority data. For names to which the initial sound rule had been applied in South Korea, there was an example where both an access point to which the rule had been applied and one to which it had not been were recorded; both “이광수” and “리광수” were recorded in records for “李光洙”. This kind of case is believed to arise when the access point is recorded for materials that were published in North Korea, where the initial sound rule is not applied.

As observed above, despite Hangul being the most basic notation system for describing Korean personal names and organization names, only the two Korean institutions and Keio require the Hangul form as a mandatory element. That being noted, NACSIS-CAT has a rule stipulating the assignment of a Hangul form corresponding to a Chinese character form, and NTUL has one requiring that a Hangul form be recorded if one is found. While clearly specified rules do not exist in HKCAN and LC, Hangul forms were recorded in their actual data under heading linking entry fields and variant access points. Accordingly, except for NCL, it is considered likely that a Hangul form is recorded somewhere in an authority record.

No institutions had precise rules related to the initial sound rule; the only institution for which were found cases in which both forms—with and without the rule applied—were recorded in the data, was LC. When authority data has been created from materials published in North Korea, the author name takes a form where the initial sound rule is not applied; if a user searches with a form where it is applied, it is possible that no hits will be returned.

7.3 Adoptions of Hanja forms

As Table 7-3 shows, The National Library of Korea (NLK) and YUL record a Hanja form as an addition to the authorized access point (field 1XX, subfield code $h). Even in cases where the materials to be cataloged lack a Chinese character representation, both institutions record a Hanja form if one is identified from reference materials or information on the Internet. In modern times, in many cases a Chinese character notation for the author name often does not appear in materials, and thus catalogers cannot identify and record a Hanja form. In NCASIS-CAT, when an authorized access point is not in Chinese characters, it is possible (but not mandatory) to record a Chinese character form as a variant access point. Keio takes a
Chinese character form as an authorized access point except in cases when a Chinese character form is not found. NCL adopts a Chinese character form as an authorized access point when both it and a Hangul form are identified; however, sometimes the Chinese character form is not recorded in the authority data if only a Hangul form is identified. NTUL takes a Chinese character form as an authorized access point. When the material to be cataloged lacks Chinese character notation, catalogers search for one using e.g. the Internet. If such a search proves unsuccessful, the Hangul is transliterated into Chinese characters and recorded as such.

Table 7-3 A doptions of Hanja forms

<table>
<thead>
<tr>
<th>Organization</th>
<th>Adoption</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLK</td>
<td>A subfield of AAP</td>
<td>100 1# $a이광수=$h李光洙</td>
</tr>
<tr>
<td>YUL</td>
<td>A subfield of AAP</td>
<td>100 1# $a이광수$h李光洙</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>AAP/VAP</td>
<td>&lt;HDNG&gt;이, 光洙 ㅣ, 光洙</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;HDNG&gt;이, 근관</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;SF&gt;李, 根寛</td>
</tr>
<tr>
<td>Keio</td>
<td>AAP</td>
<td>100 1# $a李, 光洙</td>
</tr>
<tr>
<td>NCL</td>
<td>AAP</td>
<td>100 1# $a李</td>
</tr>
<tr>
<td>NTUL</td>
<td>AAP (mandatory)</td>
<td>100 1# $a李, 光洙</td>
</tr>
<tr>
<td>HKCAN</td>
<td>VAP</td>
<td>400 1# $a李光洙</td>
</tr>
<tr>
<td>LC</td>
<td>VAP</td>
<td>400 1# $a李光洙</td>
</tr>
</tbody>
</table>

Note. AAP - Authorized Access Point; VAP - Variant Access Point.

In the authority data of HKCAN, a Chinese character form was often recorded as a variant access point except when recorded as a heading linking entry field; however, one was not recorded for “安正孝”. For “김하늘”, which lacks applicable Chinese characters, the phonetic equivalent “金荷娜” was utilized as a variant access point. “荷娜” appears to be a Chinese phonetic transcription of “하늘”. While “曹薰鉉” returned no hits, phenomena where the native script of Korea was insufficiently expressed were observed among multiple access points thought to be the same surname “曺”. Namely, in a variant access point for “Cho, Kuk (조국)”, which should be “曹國”, “曹” had been replaced by the geta mark “〓”; in variant access points for “Cho, Nam-hyŏn (조남현)” and “Cho, Hŭi-ung (조희웅)”, which should respectively be “曹南鉉” and “曹喜雄”, it had been replaced to yield “曹南鉉”, “曹喜雄”, and “[Cho]喜雄”. In the authority data of LC as well, a Chinese character form was recorded in a variant access point for all of the search terms except “김하늘”. However, examples were seen where an incorrect Chinese character was recorded together with the correct one: e.g., the incorrect “徐廷桂” in addition to “徐廷柱”, and the incorrect “李御寧” in addition to “李御寧”. In the same way as in HKCAN, the character “曺” had been replaced with the characters “曹”, “〓”, etc. The surveyed databases all support Unicode; however, there are
likely access points in data created prior to Unicode support that do not correctly notate the native scripts of Korea.

_Hanja_ are close to the traditional _Hanzi_ (Chinese characters) used in Taiwan and Hong Kong, although there are minor differences. Therefore, in Taiwan and Hong Kong, the result is that data would likely be created based on _Hanja_, irrespective of whether or not the cataloger himself recognizes the prerogative to input _Hanja_. In NACSIS-CAT, because the script is not uniform within authorized and variant access points, examples were seen where e.g. “서울대학교박물관” (i.e., the _Hanja_ form) was not recorded in a variant access point, but the form “서울대학교박물관” was, where “학” is _Kanji_. At Keio as well, the character form may change depending on the material to be cataloged. At LC, character forms thought to have been obtained from Chinese language materials—for example, “李御宁” and “曹薰铉” (“宁” and “铉” being simplified Chinese _Hanzi_ from mainland China)—were sometimes described in variant access points. _Hanja_ may not necessarily be recorded depending on the material to be cataloged.

As observed above, none of the institutions mandated the Chinese character form except for NTUL, which requires catalogers to derive Chinese characters from _Hangul_. If a Chinese character form is not listed in the material to be cataloged, catalogers have no choice but to investigate using reference materials, but there are nonetheless cases where they will not be able to find it, and so requiring a Chinese character form would be unreasonable. At Keio and NCL, however, if a Chinese character form is identified they prioritize another form for the authorized access point. At HKCAN and LC too, there were many cases of a Chinese character form being recorded in a variant access point in the actual data. Based on these observations, it seems that each institution observes a policy of making a best effort to record a Chinese character form. However, one must bear in mind that _Hanja_ are not invariably recorded at Japanese institutions and at LC.

### 7.4 Adoptions and types of Romanization

As Table 7-4 shows, NLK mandates recording a Romanized form in a variant access point. At least one of the following types is recorded: the Romanized representation in the material to be cataloged, the Romanization according to the MCT system, or the Romanization according to the MR system. The Romanized form in the material to be cataloged is adopted if one is present; if not, NLK records it using the MCT system. From the desire to bring diversity to access points, the MR Romanization is also recorded when it differs from the Romanized notation in the material to be cataloged. However, these are goals, not rules, and require extra effort. YUL records a Romanized form as a variant access point if present in the material to be cataloged; however, it is not mandatory. YUL appends an LCNAF authorized access point (hereafter, LC form) to a variant access point when e.g. translated works from Korean to English are recorded. NACSIS-CAT uses a Romanized form as an authorized access point when the representation shown in the materials used at the initial creation of the authority record is a Romanization, or when the
Romanized form is the most well-known. In other cases, the Romanized form may be recorded in a variant access point if one is identified, but this is not mandatory. The Romanization system is not prescribed because no specific Romanization method in particular has been set as the standard.9

Table 7-4 Adoptions and types of Romanization

<table>
<thead>
<tr>
<th>Organization</th>
<th>Adoption</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLK</td>
<td>VAP (mandatory)</td>
<td>Representation in material to be cataloged/MR/MCT</td>
<td>100 1# $a윤동주=$h尹東柱 400 1# $aYun, Dongju 400 1# $aYun, Tong-ju 400 1# $aYun, Tong-chu</td>
</tr>
<tr>
<td>YUL</td>
<td>VAP</td>
<td>Representation in material to be cataloged/LC form</td>
<td>400 1# $aYun, Tong-ju</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>AAP/VAP</td>
<td>No rule</td>
<td>&lt;SF&gt;Yoon, Dong-joo</td>
</tr>
</tbody>
</table>

| Keio         | VAP (mandatory) | MR;LC form recorded if different from MR | 400 1# $aYun, Tong-ju |
| NCL          | Not recorded   | N/A | N/A |
| NTUL         | VAP          | Representation in material to be cataloged | (No Romanized form for 尹東柱) |
| HKCAN        | AAP (mandatory)/VAP | LC form for AAP | 100 1# $aYun, Tong-ju 400 1# $aYun, Dong-ju |
| LC           | AAP (mandatory)/VAP | MR for AAP (exceptions exist) | 100 1# $aYun, Tong-ju 400 1# $aYun, Dong-ju |

Note. AAP - Authorized Access Point; VAP - Variant Access Point.

Keio, HKCAN, and LC adopt the MR system prescribed by the ALA/LC Romanization Tables as a general principle. At Keio, the Romanized form is treated as a variant access point, but its recording is mandatory; at HKCAN and LC, it is an authorized access point. In LC, non-MR Romanizations are sometimes recorded as an authorized access point—such as “Kim, Dae Jung (金大中)” and “Park, Chung Hee (朴正熙)”. HKCAN follows the LC form. How LC handles this is identifiable as based on the Alternative Rule in Section 22.3C2 of the AACR2: “choose the Romanized form of name that has become well-established in English-language reference sources for a person entered under surname whose name is in a language written in a nonroman script.”10 This rule survives in the RDA, as an Alternative in Section 9.2.2.5.3: “if there is a well-established form of name in reference sources in a language preferred by the agency creating the data, choose that form of name as the preferred name.”10 This shows that the Romanized form taken by an LC authorized access point will not necessarily be based on a single
Romanization system. Keio requires recording both the LC form and the MR format when they differ. In the authority data of HKCAN and LC, there were many cases where Romanized forms not adopted as the authorized access point were multiply recorded as variant access points.

NCL does not record any kind of Romanized form in authority data corresponding to Chinese, Korean, or Japanese language materials. Upon confirming the actual authority data, in the authority records for Western works, there were places where Romanized forms seeming to have been downloaded from OCLC had been recorded in both an authorized access point and a variant access point. NTUL describes a Romanized form only in cases where one appears in the material to be cataloged, but it is not mandatory.

As observed above, about half the institutions require Romanized forms, while the other half do not. The only institutions unified in their use of a Romanization type were Keio, HKCAN, and LC. Keio even records the MR format in addition to the LC form when they differ, taking great pains to ensure consistency with the MR scheme. HKCAN and LC adopt the MR format, but authorized access points that do not conform to the MR format also exist. The other institutions transcribe the Romanized form described in the material to be cataloged, and the type of Romanization scheme is not unified. Accordingly, data identification may not succeed between institutions, even if both mandate a Romanized form.

7.5 Separation of surname and given name with a comma

As Table 7-5 shows, for Hangul forms and Hanja forms, NLK and YUL write the surname and given name continuously, without word separation, to record them. Only for Romanized forms do they separate the surname and first name, by means of a comma. 정옥경 had stated that although KCR2 had prescribed inserting a comma between the surname and given name, this prescription ought to be revised because it does not fit with Korean convention.2 According to 김성원 et al., almost all Korean surnames are one-syllable surnames, and there are only 13 two-syllable surnames. Since everyone in Korea knows this, there is no need to sow confusion with this surname-first name separation.11 Taking these points together, one could say that inside Korea, inserting a comma between the surname and given name is unnecessary, and even considered unnatural.

In Japan, on the other hand, 日本目録規則 1965 年版 (Nippon Cataloging Rules 1965 ed.: NCR1965) established the standard of giving the yomi form via kana or else Roman script for all access points.12 This was the first time that commas were inserted between the surname and given name in yomi access points. Furthermore, 日本目録規則 新版予備版 (Nippon Cataloging Rules Preliminary New Edition: NCR1977) established that for personal name access points, catalogers should “register entries in the order of surname followed by given name, and separate the two with a comma (,).”13 This rule has been maintained since then and is still in vigor as of the NCR1987 3rd rev. ed., the current version. There are no rules concerning collocations in 日本目録規則 1952 年版 (Nippon Cataloging Rules 1952 ed.: NCR1952), which preceded NCR1965. However, 植村長三郎, who served as an advisory committee
Table 7-5 Separation of surname and given name with a comma

<table>
<thead>
<tr>
<th>Organization</th>
<th>Separation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLK</td>
<td>Separation with a comma in Romanized forms</td>
</tr>
<tr>
<td>YUL</td>
<td>Separation with a comma in Romanized forms</td>
</tr>
<tr>
<td>NACSIS-CAT</td>
<td>Separation with a comma in all forms</td>
</tr>
<tr>
<td>Keio</td>
<td>Separation with a comma in all forms</td>
</tr>
<tr>
<td>NCL</td>
<td>Separation with “</td>
</tr>
<tr>
<td>NTUL</td>
<td>Separation with a comma in all forms</td>
</tr>
<tr>
<td>HKCAN</td>
<td>Separation with a comma in Romanized forms</td>
</tr>
<tr>
<td>LC</td>
<td>Separation with a comma in Romanized forms; Hangul forms sometimes have the surname and given name separated by a space</td>
</tr>
</tbody>
</table>

nevertheless using it is solely for convenience within the collocation system. Specifically, it is because I want to establish the general principle to arrange data in the order of surname first, thinking of the surname as a ‘surname-only’ group, and the given name after that, if there is someone with the same surname”.\textsuperscript{15} NCR1965 likely established comma insertion for the purposes of aligning with Western conventions and for convenience in collocations—similarly to 植村’s ideas—because it took the standpoint of dealing with Western works and Sino-Japanese works with the same principles, unlike NCR1952, which targeted only Sino-Japanese works. Both NACSIS-CAT and Keio insert a comma between the surname and given name for Hangul, Hanja, and Romanized forms, presumably in order to comply with the provisions of the NCR. On the other hand, Chapter 22 of Chinese Cataloging Rules 3rd ed. Personal Name Access Points does not prescribe the separation of the surname and given name of Chinese character forms. Yet, NTUL still inserts a comma between the surname and given name for all forms. The NCL inserts a comma between the
surname and first name in authority data for Western works, and the symbol “|” for Chinese, Korean, and Japanese language materials. The NCL stated that because the surname and given name were separated depending on the subfield code in Chinese MARC Format for Authority Records (CMARC/A) format they had previously used, they decided to use “|” when they converted to MARC 21 Format for Authority Data (MARC21/A). They adopt this kind of policy in consideration of times when other libraries, which use the CMARC/A format even today, download data from the NCL. However, the separation is not seen in the SMRT system of NCL. HKCAN and LC use comma separation only for Romanized forms. When confirming the LC authority data, there were cases in Hangul form where a space was inserted between the surname and first name; in these cases, both were created in variant access points (e.g., “김 대중” and “김대중”). There seems to be little meaning in the fact that with-space and without-space entries were recorded as different variant access points.

As observed above, the two Japanese institutions and NTUL separate the surname and given name with a comma for all forms, including Hangul and Chinese character forms; the NCL separates them using the “|” sign, and other institutions use comma separation only for Romanized forms. Given the reality that almost all surnames in Korea consist of one syllable and that there are few varieties, searching only an author’s surname is unthinkable. Thus, it is presumably unnecessary to separate the surname and given name of Hanja forms and Hangul forms for the purposes of surname-only collocations. However, where foreign users are concerned, a comma or space is thought to be permissible since distinguishing two-syllable surnames, which occasionally appear, is difficult to do on the spot. All institutions used comma separation for Romanized forms.

7.6 Representations in local languages outside Korea

As Table 7-6 shows, in NACSIS-CAT, there were cases where Korean *yomi* and Japanese *yomi* represented in katakana (e.g., for 李御寧, イー・オリョン as Korean *yomi* and リ・ギョネイ as Japanese *yomi*) were as shown in the materials used at the time of the initial creation of the authority record. There were also cases where the most well-known form of a prominent author’s name was in katakana form. In these cases, the Korean *yomi* or Japanese *yomi* was adopted in authorized access points. Similarly, there were cases where the Romanized form was entered in an authorized access point. A Romanized form can be described as a variant access point if not adopted as an authorized access point, but it is not mandatory.

Keio mandates recording the *yomi* as a variant access point when an authorized access point is in Chinese character form. *Yomi* are represented by katakana: *yomi* are preferentially adopted in the order of: (1) *yomi* that are generally recognized in Japan and invariant; (2) *yomi* adopted in authorized access points in NACSIS-CAT; (3) *yomi* described in the actual work; and (4) *yomi* adopted in variant access points of authority data of NACSIS-CAT. However, *yomi* in (2) and (3) would not be adopted when a problem is
clearly expected: e.g., when a yomi does not appear to be a typical reading. When there are multiple yomi in (4), the yomi closest to the Korean pronunciation is adopted. In cases when representation varies and it is not clear which yomi to adopt, the Kan-on (漢音) yomi (i.e., Han reading) of the Kanji is chosen. A yomi is not entered when the Chinese character form is uncertain and so the authorized access point is in Hangul, or when a name lacks Hanja.

Section 23.3.3.2 イ of NCR1987 prescribes: “Notate Kanji with the native-language yomi for Chinese personal names and Korean personal names shown in a form where Kanji appear together [in the material to be cataloged] with their yomi in the native language.” However, NACSIS-CAT judged that “In cases of native words that originally lack Hanja notation, katakana yomi cannot necessarily be given to all Korean language materials. Providing katakana yomi for all names is thus realistically impossible,” and that “Adopting searches using Hangul should be appropriate as a general rule, because this treats all Korean language materials in a unified way.” Thus, NACSIS-CAT does not mandate providing yomi. At Keio, Korean yomi are highly likely to be recorded since in most cases they are displayed in works originally written by Korean authors and published in Japan; however, yomi are never displayed for a work in the Korean language, thus the Kan-on yomi (i.e., the Japanese yomi) is sometimes recorded. NCR rules changed from stipulating Japanese yomi to Korean yomi with the publication of 日本目録規則新版予備版追録および修正 (Nippon Cataloging Rules Preliminary New Edition - Addenda and Revisions) in 1983. This was prompted by 崔昌華’s 1975 lawsuit against the NHK (Japan Broadcasting Corporation) concerning the pronunciation of his true name, in which the adoption of Korean yomi in materials was argued from the standpoint of respecting fundamental human rights. However, because the Korean yomi represented in materials are actually adaptations of Korean pronunciations to Japanese katakana, it is easy to imagine cases in which notation for the same individual might differ depending on the material, or that the notation a user has in mind might differ from the actual notation. Therefore, this is not a very effective access point from a user perspective.

NCL does not specially add notation aside from the authorized access point. NTUL records a Chinese character form as an authorized access point, and so when the Hanja for an author are unclear, a cataloger transliterate from the Hangul form into a Chinese character form to record it. This ‘transliteration’ should properly be called a Chinese translation of a Korean personal name. This creates a few problems: the record may take a Chinese character form that differs from the specific Hanja used in the name of the author himself, and the specific Chinese characters may differ depending on the personnel who performed the conversion. For these reasons, in the same way as yomi, users see this method as unlikely to produce useful access points.

Upon inspection, the authority data of HKCAN and LC included variant access points with Japanese yomi shown in Roman script (“Kin, Daichū”), with Chinese pinyin (“Jin, Dazhong”), and Wade–Giles notation (widely used internationally before the spread of pinyin) (“Chin, Ta-chung”). These are believed to be representations from materials published outside Korea, or that were
### Table 7-6 Representations in local languages outside Korea

<table>
<thead>
<tr>
<th>Organization</th>
<th>Adoption</th>
<th>Example</th>
</tr>
</thead>
</table>
| NACSIS-CAT   | Japanese yomi; Korean yomi | <SF>李, 光洙||リ,コウシュ
<SF>李, 光洙||イ,グアンス |
|              | Yomi in katakana as VAP (mandatory if AAP is in Chinese character). | 100 1# Sa李, 光洙
400 1# Saイ, グアンス |
|              | When LC form is adopted for government institution names, the parent organization name takes the country name in English. | 110 1# SaKorea (South).B水産庁
410 1# SaKorea(South).B水産庁 |
| Keio         | None in particular | 100 1# Sa韓, 相權
400 1# Sa한, 상권 |
|              | Chinese character form transliterated from Hangul form and recorded when a Chinese character form is unclear. | 110 1# Sa韓國.Sb文化體育觀光部 |
|              | For government institution names, the parent organization name takes the (conventional) country name in Hanzi. | |
| NCL          | None in particular | N/A |
| NTUL         | Japanese Roman script yomi and Chinese pinyin included in VAP. | 400 1# SaLee, Kwang Soo
400 1# SaL, Gwansu |
|              | For government institution names, the parent organization name takes the country name in English. | 110 1# SaKorea (South).B水産庁
710 1# SaKorea (South).B水産庁 |
| HKCAN        | Japanese Roman script yomi and Chinese pinyin included in VAP. | 400 1# SaLee, Kwang Soo
400 1# SaL, Gwansu |
|              | For government institution names, the parent organization name takes the country name in English. | 110 1# SaKorea (South).B水産庁
710 1# SaKorea (South).B水産庁 |
| LC           | Japanese Roman script yomi and Chinese pinyin included in VAP. | 400 1# SaLee, Kwang Soo
400 1# SaL, Gwansu |
|              | For government institution names, the parent organization name takes the country name in English. | 110 1# SaKorea (South).B水産庁
710 1# SaKorea (South).B水産庁 |

Note. AAP - Authorized Access Point; VAP - Variant Access Point.

* Transliterated form from Hangul to Chinese characters, created by a cataloger.
appropriated from the OPACs of libraries in those regions.

When an organization name is a government agency, HKCAN and LC make it an access point with the name of the government as the parent body, according to the rules of AACR2, Section 24.18.10 This prescription is carried over almost entirely unchanged into RDA Section 11.2.2.1410; at LC, the authorized access points for government agencies continue to be created in this form. NTUL as well sets the country name as the parent institution for government institutions, but the country name is represented in the Hanzi conventionally used in Taiwan. This kind of handling of government agencies was not observed at NACSIS-CAT or the NCL. Keio adopts the form of the authorized access point from pre-existing authority data from NACSIS-CAT if such data exists. If not, but there is pre-existing data from LCNAF, Keio adopts the LC form (i.e., the form having the name of the government as the parent institution) as the authorized access point. However, subordinate institution names of $b and lower are recorded in Kanji form. Since adoption methods for authorized access points of organization names differ according to cataloging rules, national and regional variations seem to arise more easily than for personal names.

7.7 Discussion: differences in representations and problems to be solved

Each institution gives an ID to each authority record within their respective authority databases, but these IDs are only valid within their respective database. Therefore, identifiers shared by all countries and regions do not exist at present. Given this status quo, identification operations using strings are thought to be necessary to achieve international interoperability of authority data. It would be desirable to conduct authority data creation at each institution in the future with an eye to identification operations. Namely, if each institution prioritized recording those character types advantageous to identification from among Hangul, Chinese characters, and Roman script, they would be able to raise the degree of identification accuracy.

All of the surveyed institutions adopted the Chinese pinyin form for Chinese personal and organizational names: thus, it was considered advantageous to conduct identification operations with the Chinese pinyin as identification keys (hereafter, ‘keys’) while continuing to use the Chinese character form as reference.24 For Korean personal and organizational names, however, it is difficult to set Romanized forms as keys at present. This is because although there are more institutions that require a Romanized form than do a Hangul form or Chinese character form, the Romanization scheme is not unified. In Korea, consolidation to one Romanization scheme is likewise expected to be extremely difficult, since the concept of uniformly transliterating/transcribing surnames and given names from Hangul has not been adopted. However, it seems possible for libraries to record forms Romanized in a consistent way as variant access points: this is in addition to Romanized forms described in materials to be cataloged or of an author’s choice. In this case, it is necessary to carefully
consider which to apply as the Romanization scheme: the MCT system used domestically in Korea, or the MR system utilized internationally. The ALA/LC Romanization Tables must be modified if the MCT system is to gain currency in the Western library community, and more-detailed discussion must be held to achieve this.

In the present state, where the Romanized form will not work as the identification key, the most desirable one would be the **Hangul** form. There is more variation in Chinese characters than in **Hangul**, and so it is easier to eliminate incidents where different people are identified by the same name. However, sometimes names lack Chinese characters, or the correct ones are unclear. On the other hand, setting **Hangul** as a key is realistic since many institutions record a **Hangul** form in a variant access point. Four actions can be taken towards this end. (1) At institutions that do not prescribe recording **Hangul** form, a rule should be established to record a **Hangul** form as often as possible. Moreover, because there may sometimes be multiple possible **Hangul** forms depending on the initial sound rule, (2) support should be provided regarding whether to record both forms or to create a conversion table so that both of the search terms produce a hit. In addition, because there are more instances of an individual having the same surname and given name in **Hangul** form compared with in **Hanja** form, (3) it would be beneficial to record a **Hanja** form when one is found, and to think of the **Hangul** form and **Hanja** form as one pair for use in identification. Specifically, if one initially ties together a **Hangul** form and its **Hanja** form with an association specifier, one can determine that those access points identify the same entity with high probability when both the **Hangul** form and **Hanja** form agree between multiple databases. The present author has asserted that grouping the **Kanji** form with its **yomi** should assist the identification of Japanese personal and organization names as well. The same could be said for Korean personal and organizational names, since many individuals share the same surname and given name in **Hangul**. Actually, NLK and YUL adopted a **Hanja** form as an addition to an authorized access point, showing it grouped together with the **Hangul** form. NACSIS-CAT too recorded a Chinese character form as an authorized access point: where it recorded a **Hangul** form in the **yomi** field, the Chinese character form and **Hangul** form can be said to have been adopted as a pair. On the other hand, at other institutions there were instances where the connection between the **Hangul** form and Chinese character form was not particularly apparent, and where multiple **Hangul** forms and multiple Chinese character forms had been input into variant access points because of differences in names. In these cases, it is not evident which **Hangul** form corresponds to which Chinese character form. Improvements in identification accuracy and speed can be expected from establishing some sort of association specifier(s) in formats to try to display correspondence relationships between **Hangul** forms and **Hanja** forms.

However, what makes Korean personal or organization names different from Japanese personal and organization names is that cases where a Chinese character form is absent or uncertain are possible. Therefore, (4) dates of birth and death should be proactively recorded as an addition to names. Records with additions recorded in this way take on major significance for the identification of Korean personal and
organization names that lack an absolute identification key, compared with Chinese and Japanese personal and organizational names.

Notes

1 Part of this chapter has already been published as 木村麻衣子. 韓国人・団体著者名典拠データの表記の相違: 韓国, 日本, 台湾, 香港を中心に. Library and Information Science. 2014, no.72, p. 63-93.


7 Schiff, Adam. “Non-Latin script references in name authority records”. NACO: Name Authority Cooperative Program. 2009-06-01.

http://www5.cuhk.edu.hk/jas/jas_media/staff/cojiima/04-korean-kanji-style.pdf, (accessed 2014-08-26). According to 児島, there are differences in the shapes of Radicals 162 (⿰) and 140 (艹). However, they take the same shape in computer displays because the character codes do not currently reflect differences at this level of detail.


菅野裕臣. 朝鮮人名の朝鮮語読みについて: 崔昌華氏の主張と関連して. 朝鮮研究. 1984, no. 238, p. 2-13. According to 菅野, the problem of pronouncing Korean personal names in Japanese yomi is “simply a problem of language and script”, and 崔昌華's claim, that it was a modern-day version of the enforced renaming of Koreans from Japan’s colonial period, is erroneous.


Chapter 8

Representations of Vietnamese name authority data in Chinese character cultures

The aim of this chapter is to compare representations of Vietnamese name authority data in the Chinese character cultural sphere. According to the results of the interview in Chapter 4, libraries in Vietnam have not yet started to implement name authority control, although they have been conducting subject authority control. In addition, as Vietnamese authors are relatively few, the present author has not focused on Vietnamese name authority control while conducting the interviews with libraries. Therefore, this research on Vietnamese names is conducted through searches of actual authority data rather than through interviews. Six available authority data for Vietnamese names are used and the representations of each authority database are compared.

8.1 Research method

Firstly, Vietnamese names that have representations in both chữ quốc ngữ and Chinese characters (including chữ Hán and chữ Nôm) are selected from 越南漢喃文獻目錄提要 (Han-Nom bibliographies in Vietnam).¹ The index of this bibliography for classic books in chữ Hán and chữ Nôm lists 2,695 author names. Among these, authors who have contributed more than eight works are selected, although some non-Vietnamese names, such as “朱熹,” are excluded. As a result, 76 author names are selected (Table 8-1). In 越南漢喃文獻目錄提要, author names are recorded only in Chinese characters because it is a translated and enhanced version of the original catalog in Vietnamese and French that is entitled Di san Hán nôm Việt Nam.² Thus, representations in chữ quốc ngữ are checked by the author, mainly by using the original catalog and then, as supplements, the database of the digital collections of the Vietnamese Nôm Preservation Foundation,³ Google searches, or a Chinese-Vietnamese Dictionary.⁴

Secondly, the authority data for these names are sought and extracted from the LC/NACO Authority File (LCNAF),⁵ the CALIS Union Catalog Authorities (CALIS),⁶ the HKCAN Database OPAC (HKCAN),⁷ the online database of the National Central Library in Taiwan (NCL),⁸ the Web NDL Authorities (NDL),⁹ and the “Author Search” of CiNii Books,¹⁰ which includes the authority data of NACSIS-CAT. As libraries in South Korea, the National Library of China, the Keio University Libraries, and the National Taiwan University Library do not provide public access to their authority data, the author has not included these organizations as research subjects. When conducting the search, both Chinese characters and chữ quốc ngữ representations of the names are used as keywords. If a person’s name has several
### Table 8-1 Vietnamese names searched

<table>
<thead>
<tr>
<th>No.</th>
<th>Names in chữ quốc ngữ</th>
<th>Names in Chinese characters</th>
<th>No.</th>
<th>Names in chữ quốc ngữ</th>
<th>Names in Chinese characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bùi Dương Lị</td>
<td>裴楊瀝; 裴楊瓑; 裴楊歷</td>
<td>39</td>
<td>Nguyễn Huy Ông</td>
<td>赵敬; 赵敬.bin</td>
</tr>
<tr>
<td>2</td>
<td>Bùi Huy Bích</td>
<td>裴輝碧; 裴輝璧</td>
<td>40</td>
<td>Nguyễn Khuyên</td>
<td>赵勤</td>
</tr>
<tr>
<td>3</td>
<td>Bùi Văn Đi</td>
<td>裴文欽</td>
<td>41</td>
<td>Nguyễn Miên Thảo</td>
<td>赵绵炤; 赵绵炤.bin</td>
</tr>
<tr>
<td>4</td>
<td>Cao Bá Quốc</td>
<td>高伯适</td>
<td>42</td>
<td>Nguyễn Thuật</td>
<td>赵述</td>
</tr>
<tr>
<td>5</td>
<td>Cao Huy Diệu</td>
<td>高輝耀</td>
<td>43</td>
<td>Nguyễn Thường Hiện</td>
<td>赵尚賢</td>
</tr>
<tr>
<td>6</td>
<td>Cao Xuân Dương</td>
<td>高春陽</td>
<td>44</td>
<td>Nguyễn Trần</td>
<td>赵</td>
</tr>
<tr>
<td>7</td>
<td>Đặng Huy Trữ</td>
<td>鄧輝禧</td>
<td>45</td>
<td>Nguyễn Trọng Hợp</td>
<td>赵仲合</td>
</tr>
<tr>
<td>8</td>
<td>Đặng Xuân Khanh</td>
<td>鄧春卿</td>
<td>46</td>
<td>Nguyễn Tư Giản</td>
<td>赵文简; 赵文简.bin</td>
</tr>
<tr>
<td>9</td>
<td>Đặng Xuân Bằng</td>
<td>鄧春隆</td>
<td>47</td>
<td>Nguyễn Vấn Giao</td>
<td>赵文交</td>
</tr>
<tr>
<td>10</td>
<td>Đinh Nhật Thận</td>
<td>丁日慎</td>
<td>48</td>
<td>Nguyễn Vấn Lý</td>
<td>赵文理</td>
</tr>
<tr>
<td>11</td>
<td>Đỗ Văn Tâm</td>
<td>杜文心</td>
<td>49</td>
<td>Nguyễn Văn Siêu</td>
<td>赵文超</td>
</tr>
<tr>
<td>12</td>
<td>Đoàn Trân</td>
<td>段展</td>
<td>50</td>
<td>Nguyễn Bình</td>
<td>赵</td>
</tr>
<tr>
<td>13</td>
<td>Dương Lâm</td>
<td>楊琳; 楊林</td>
<td>51</td>
<td>Nguyễn Nghiem</td>
<td>赵</td>
</tr>
<tr>
<td>14</td>
<td>Hà Huy Chương</td>
<td>何輝璋</td>
<td>52</td>
<td>Nhữ Bá Si</td>
<td>汝伯仕</td>
</tr>
<tr>
<td>15</td>
<td>Hà Tông Quyền</td>
<td>何宗權</td>
<td>53</td>
<td>Phạm Đình Hợp</td>
<td>范廷琥</td>
</tr>
<tr>
<td>16</td>
<td>Hồ Xuân Hương</td>
<td>胡春煌</td>
<td>54</td>
<td>Phạm Nguyễn Du</td>
<td>范阮攸</td>
</tr>
<tr>
<td>17</td>
<td>Hoàng Cao Khải</td>
<td>黃高啟</td>
<td>55</td>
<td>Phạm Quốc Thích</td>
<td>范国超</td>
</tr>
<tr>
<td>18</td>
<td>Hoàng Chiến</td>
<td>黃擅</td>
<td>56</td>
<td>Phạm Văn Nghị</td>
<td>范文卿</td>
</tr>
<tr>
<td>19</td>
<td>Hoàng Hữu Xùng</td>
<td>黃有굉</td>
<td>57</td>
<td>Phạm Văn Thu</td>
<td>范文樹</td>
</tr>
<tr>
<td>20</td>
<td>Kửu_GROUPS.</td>
<td>喬璧懋</td>
<td>58</td>
<td>Phạm Xuân Lộc</td>
<td>范春祿</td>
</tr>
<tr>
<td>21</td>
<td>Lê Hữu Trác</td>
<td>黎有晫</td>
<td>59</td>
<td>Phạm, Phù Thứ</td>
<td>范富庭</td>
</tr>
<tr>
<td>22</td>
<td>Lê Quy Đơn</td>
<td>黎貴倬</td>
<td>60</td>
<td>Phan Bội Châu</td>
<td>潘佩珠</td>
</tr>
<tr>
<td>23</td>
<td>Lê Tùng</td>
<td>黎嵩</td>
<td>61</td>
<td>Phan Thanh Giản</td>
<td>潘清簡</td>
</tr>
<tr>
<td>24</td>
<td>Lê Thanh Tông</td>
<td>黎震宗</td>
<td>62</td>
<td>Phan Huy 交易中心</td>
<td>潘辉益</td>
</tr>
<tr>
<td>25</td>
<td>Lý Văn Phục</td>
<td>李文馥</td>
<td>63</td>
<td>Phúc Đình</td>
<td>福田</td>
</tr>
<tr>
<td>26</td>
<td>Mac Đình Chi</td>
<td>麦挺之</td>
<td>64</td>
<td>Phùng Khắc Khoan</td>
<td>阮克寛</td>
</tr>
<tr>
<td>27</td>
<td>Ngô Giáp Đậu</td>
<td>吴甲豆</td>
<td>65</td>
<td>Trần Hạnh</td>
<td>清亨</td>
</tr>
<tr>
<td>28</td>
<td>Ngô Sĩ Liên</td>
<td>吴士連</td>
<td>66</td>
<td>Trần Công Hiền</td>
<td>陈公憲</td>
</tr>
<tr>
<td>29</td>
<td>Ngô Thế Vinh</td>
<td>吴世荣</td>
<td>67</td>
<td>Trần Danh Ấn</td>
<td>陈名案</td>
</tr>
<tr>
<td>30</td>
<td>Ngô Thệ Nhiệm; Ngô Thới Nhiệm</td>
<td>吳時任</td>
<td>68</td>
<td>Trần Duy Vôn</td>
<td>陈维○*</td>
</tr>
<tr>
<td>31</td>
<td>Ngô Thị Sĩ</td>
<td>吳時仕</td>
<td>69</td>
<td>Trần Văn Phùng</td>
<td>陈文逢</td>
</tr>
<tr>
<td>32</td>
<td>Nguyễn Bá Nghị</td>
<td>阮伯儀</td>
<td>70</td>
<td>Trần Hưng Đạo</td>
<td>陈興道</td>
</tr>
<tr>
<td>33</td>
<td>Nguyễn Bạo</td>
<td>阮保</td>
<td>71</td>
<td>Trường Đăng Quế</td>
<td>张登桂</td>
</tr>
<tr>
<td>34</td>
<td>Nguyễn Bình Khâm</td>
<td>阮秉謙</td>
<td>72</td>
<td>Trường Quốc Dung</td>
<td>张国立</td>
</tr>
<tr>
<td>35</td>
<td>Nguyễn Công Trù</td>
<td>阮公著</td>
<td>73</td>
<td>Trường Hán Siêu</td>
<td>张漢超</td>
</tr>
<tr>
<td>36</td>
<td>Nguyễn Du</td>
<td>阮攸</td>
<td>74</td>
<td>Tự Đức</td>
<td>肆德</td>
</tr>
<tr>
<td>37</td>
<td>Nguyễn Đức Dat</td>
<td>阮德達</td>
<td>75</td>
<td>Vũ Phạm Hân</td>
<td>武范誠</td>
</tr>
<tr>
<td>38</td>
<td>Nguyễn Hiền</td>
<td>阮賢</td>
<td>76</td>
<td>Vũ Phạm Khải</td>
<td>武范殿</td>
</tr>
</tbody>
</table>

Note: * is a chữ Nôm; it looks like the character “丿,” but its lower left section is “禾,” rather than “木.”
representations in Chinese characters or chữ quốc ngữ, these variations are also searched. In addition, names in chữ quốc ngữ without diacritics are also searched because some systems may not accept diacritics.

Subsequently, the search results of each database are compared. Four checkpoints of the search results are as follows: 1) the types of representations (in other words, this determines whether Chinese characters are recorded as well as chữ quốc ngữ); 2) the relation of Chinese characters to their corresponding chữ quốc ngữ; 3) the diacritics of chữ quốc ngữ; and 4) the representations in local languages.

8.2 Search results

Among 76 author names, 18 names were not retrieved in any of the databases searched. Therefore, these names are excluded from subsequent analyses. Of the 58 author names, 53 records are retrieved from LCNAF, 3 records are retrieved from CALIS (1 duplicate record is excluded), 10 records are retrieved from HKCAN, 14 records are retrieved from NCL (2 duplicated records are excluded), 2 records are retrieved from NDL, and 16 records are retrieved from CiNii (2 duplicated records are excluded, see Table 8-2). Subsequent tables and analyses are produced from the search results in each database. Samples of the authority data of each organization are shown in Table 8-3. As NDL and CALIS do not create authority records for Vietnamese materials, it seems that their records were created in situations in which translated works are acquired by libraries.

![Table 8-2 Records retrieved](image)

<table>
<thead>
<tr>
<th>Database</th>
<th>Records that should be retrieved</th>
<th>Records retrieved in Chinese characters</th>
<th>Records retrieved in chữ quốc ngữ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCNAF</td>
<td>53</td>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td>CALIS</td>
<td>3*</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>HKCAN</td>
<td>10</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>NCL</td>
<td>14*</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>NDL</td>
<td>2</td>
<td>2</td>
<td>2**</td>
</tr>
<tr>
<td>CiNii Books</td>
<td>16*</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

Notes. * Duplicated records are excluded.
** Without diacritics.
Table 8-3 Samples of authority data

<table>
<thead>
<tr>
<th>Database</th>
<th>Data for Nguyễn Du</th>
<th>Data for Phạm Đình Hổ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCNAF</td>
<td>100 $aNguyễn, Du</td>
<td>100 $aPhạm, Đình Hổ</td>
</tr>
<tr>
<td></td>
<td>400 $wnna$aNguyễn-Du,$d1765-1820</td>
<td>400 $aPhạm Đình Hổ</td>
</tr>
<tr>
<td></td>
<td>400 $aGen Yū,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $aTô Như,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $aThánh Hiền,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $aHồng Sơn Lệ Hộ,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $a阮攸,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td>CALIS*</td>
<td>200 @7t0y0y@a阮攸,$@f1765-1820</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>200 @7t0y0y@a阮攸,$@f1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 @7ec0yec0y@a阮攸,$@f1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 @7da0yda0y@a阮攸,$@f1765-1820</td>
<td></td>
</tr>
<tr>
<td>HKCAN</td>
<td>100 $aSaNguyễn(228)ën, Du,$d1765-1820</td>
<td>100 $aPh(242)am, Đinh H?o,$d1768-1839</td>
</tr>
<tr>
<td></td>
<td>400 $aSaNguyễn(228)ën-Du,$d1765-1820$wnna</td>
<td>400 $aPham, Tinghu,$d1768-1839</td>
</tr>
<tr>
<td></td>
<td>400 $aSaGen Yū,$d1765-1820</td>
<td>400 $aPham Đinh Hộ</td>
</tr>
<tr>
<td></td>
<td>400 $aT(228)ơ Nh?u,$d1765-1820</td>
<td>400 $aPham(242)am Đinh H?o,$d1768-1839</td>
</tr>
<tr>
<td></td>
<td>400 $aThánh Hiền,$d1765-1820</td>
<td>700 $a范廷琥,$d1768-1839</td>
</tr>
<tr>
<td></td>
<td>400 $aH(225)ơng Sơn Lệ H(242)ơc H(242)ơc,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td></td>
<td>700 $aSa阮攸,$d1765-1820</td>
<td></td>
</tr>
<tr>
<td>NCL**</td>
<td>[AAP]阮攸</td>
<td>[AAP]范廷琥</td>
</tr>
<tr>
<td></td>
<td>[VAP]RuanYou; 阮素如; RuanSuru; 素如; Suru; 阮清軒; RuanQingxuan; 清軒; Qingxuan</td>
<td>[VAP]Fan, Tinghu.</td>
</tr>
<tr>
<td></td>
<td>[AAP]阮攸</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[VAP]RuanYou.</td>
<td></td>
</tr>
<tr>
<td>NDL***</td>
<td>100 $6880-01Sa阮攸, 攸,$d1765-1821</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>880 $6100-01/S1$aゲン, ユオ,$d1765-1821</td>
<td></td>
</tr>
<tr>
<td></td>
<td>880 $6100-01/I$S$aGen, Yu,$d1765-1821</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $aNguyễn, Du</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 $aゲン, ユオ</td>
<td></td>
</tr>
<tr>
<td>CiNii Books**</td>
<td>[AAP]Nguyễn, Du</td>
<td>[AAP]范, 廷琥</td>
</tr>
<tr>
<td></td>
<td>[VAP]Nguyễn-Du; Gen Yū, Tô Như; Thanh Hiền; Hồng Sơn Lệ Hộ; Du, Nguyên; 阮, 阮</td>
<td></td>
</tr>
</tbody>
</table>

Note. Indicators and other data, excluding access points, are omitted from this table.

*Data are reconstructed by the author based on the sample records provided by CALIS during the interview conducted in 2013.

** AAP - Authorized Access Point. VAP- Variant Access Point.

8.2.1 Types of representations

As Table 8-4 shows, the search result shows that all organizations record representations in Chinese characters in at least some of their authority records. However, CALIS and NCL do not record representations in chữ quốc ngữ.

Table 8-4 Recording of representations in Chinese characters and chữ quốc ngữ

<table>
<thead>
<tr>
<th>Database</th>
<th>Recording of Chinese characters</th>
<th>Recording of chữ quốc ngữ</th>
<th>Showing their relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCNAF</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>CALIS</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>HKCAN</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>NCL</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>NDL</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>CiNii Books</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

Note. “〇” does not mean that it is included in all records searched.

Most of the authority records retrieved from LCNAF include names in chữ quốc ngữ, while only two records include representations in Chinese characters.

Among the three records of CALIS, a record for “Ngô Sĩ Liên” has an authorized access point of “Ngô, Sĩ Liên, 15th cent.” with a variant access point of “呉士連, 15th cent.” However, this record does not retrieve hits in CALIS’s database with the use of chữ quốc ngữ as a search term, even without diacritics. Records for the other two names do not include representations in chữ quốc ngữ.

In HKCAN, both representations are recorded in most cases. The only exceptions are “Lê Quý Đôn” and “Phạm Nguyên Du,” the Chinese characters of which, such as “黎貴惇, 1726-1784,” are recorded as an equivalent heading (the 700 field), while its Hanyu pinyin form, such as “Li Guidun, 1726-1784,” are recorded as an authorized access point. For the other eight cases, representations in chữ quốc ngữ are recorded as authorized access points and their Chinese character forms are recorded in field 700s, with some garbling occurring in some chữ quốc ngữ forms, such as “Ngô, Giáp D{242}âu, 1853-1929” and “Ngô, Thị Nh{242}âm, 1746-1803.” Some records can be searched despite the occurrence of such garbling, but some others cannot.

In NCL, only the representations in Chinese characters could be used for searches. Other representations were not recorded.

The two records retrieved from NDL are recorded in Chinese characters. Both records also
include the name in chữ quốc ngữ without diacritics. One of the two records seems to have tried
to record the name in chữ quốc ngữ, but failed: it omitted a diacritic so that it recorded “Phan
Bội Châu” instead of the proper “Phan Bội Châu.” In CiNii Books, records including both
representations are retrieved for only five names. Other five names are retrieved only when
searched in their Chinese character forms, while other six names are retrieved when searched in
their chữ quốc ngữ forms. This is because the NACSIS-CAT manuals do not specify which
representation should be adopted for the authorized access point and which should be adopted
for the variant access points. For example, for “Ngô Sĩ Liên,” “呉, 士連||呉, シレン” is an
authorized access point and “呉, 士連||呉, シレン” is a variant access point. On the other hand,
“Ngô Giáp Đậu,” has “呉, 甲豆||呉, コウトウ” as an authorized access point with no variant
access points.

8.2.2 The relating of Chinese characters to their corresponding chữ quốc ngữ

As Table 8-4 shows, only HKCAN shows the corresponding relationships between the names
in Chinese characters and the names in chữ quốc ngữ. HKCAN basically records names in chữ
quốc ngữ as authorized access points (field 100) and names in Chinese characters as equivalent
headings (field 700). This means that HKCAN shows the corresponding relationships between
fields 100 and 700.

In LCNAF, names in chữ quốc ngữ are recorded as authorized access points (field 100), and
names in Chinese characters are recorded (if available) as variant access points (field 400).
Evidently, the relationship between them is not shown in the data. For example, Phan Bội
Châu’s name in Chinese characters, “潘佩珠,” is recorded as well as his real name, “Phan Văn
san,” and his pseudonym, “Sào-Nam,” and others. No designators signal that “潘佩珠” is a
corresponding form of “Phan Bội Châu.”

In CALIS, “呉士連” is recorded as a variant access point of “Ngô, Sĩ Liên,” which is treated as
an authority record of a Western name. A pinyin form of “呉士連” is also recorded as a variant
access point; thus, the corresponding relationship between “呉, 士連” and “呉士連” is not
shown in the record. Although “呉甲豆” is recorded as an authority record in Chinese
characters, its form in chữ quốc ngữ is not recorded. As for “阮攸,” two records (one of a
Chinese name and the other of a Japanese name) exist but are not merged. Both records do not
include the name in the form of chữ quốc ngữ. In NCL’s records, names in chữ quốc ngữ do not
appear. Therefore, the corresponding relationships are not shown.

NDL and CiNii Books also do not show the relationships between the names in Chinese
characters and those in chữ quốc ngữ. NDL records names in Chinese characters as authorized
access points and names in chữ quốc ngữ without (or with failed) diacritics as variant access
points. In CiNii Books, both representations can be recorded as either authorized access points
or variant access points.

8.2.3 Diacritics of chữ quốc ngữ

It is hard for foreigners to type diacritics in the Vietnamese language when conducting searches. In LCNAF, the inclusion or exclusion of diacritics in search terms has no influence on the search results. On the other hand, diacritics are recorded in authority data with precision.

As some diacritics are substituted by brackets with numbers, diacritics are not precisely shown in the HKCAN database. Moreover, some letters with horns, such as “ư” and “ơ,” are not precisely shown in the data. This produces the undesirable result of preventing some names from retrieving hits with the use of their chữ quốc ngữ forms as search terms, either with or without diacritics.

As aforementioned, one record in NDL fails to record the precise diacritics of “Phan Bội Châu” in chữ quốc ngữ, and both records include forms in chữ quốc ngữ without diacritics. It is noted in the record that “Nguyen, Du” is an English name of “阮攸,” and thus, it is not intended to record a Vietnamese name in this record. Search terms with diacritics do not retrieve hits in the database.

Diacritics are precisely recorded in the authority data of CiNii Books. Sometimes forms without diacritics are recorded in variant access point fields. The inclusion or exclusion of diacritics in search terms does not influence the search results.

8.2.4 Representations in local languages

In CALIS, HKCAN, and NCL, some records include a pinyin form of the name. NCL has Wade-Giles Romanization forms as variant access points in some records, for example, “Chang, Teng-kuei” for “張登桂.” As Vietnamese classical works are in Chinese characters, the authority records linked to them were probably made in the same fashion as the Chinese authority records. Similarly, because NDL created these records for Japanese materials, these records were made to conform to Japanese authority records. As Japanese authors have yomi as well as Kanji forms, Vietnamese authors are also given yomi, such as “ゲン, ユウ” for “Nguyễn Du” in NDL’s records. In CiNii Books, yomi is also provided for the names in Chinese characters as an authorized access point. In terms of the variant access points in Chinese characters, yomi is not always given. Yomi might take the form of a Japanese reading of Chinese characters, such as “ハン, ハイシュ” for “潘佩珠,” or the form of a Japanese transcription of a Vietnamese reading, such as “ファン, ボイチャウ” for “Phan Bội Châu.”

8.3 Discussion

The following reasons explain why representations in chữ quốc ngữ are very important for
Vietnamese names. The first reason is that, as Table 8-1 shows, an author’s name in Chinese characters may have more than one representation. For “Bùi Dương Lịch,” for example, “裴楊瀝,” “裴楊瀝,” and “裴楊瀝” may be used. Sometimes, in the case of classic books in Chinese characters, different (but similar) representations of Chinese characters are used in each of the works of an author. Unless all representations are recorded in authority data, some records cannot be retrieved. Another reason is that some names use chữ Nôm, which is not included in Unicode. These characters cannot be recorded and certainly cannot be used for searches. Regardless of its importance, some organizations do not record representations in chữ quốc ngữ, or record them imperfectly. This situation should be redressed.

Although some problems exist in the Chinese character representations of Vietnamese names, they are still important for the following reasons: one alphabet of chữ quốc ngữ may correspond to many Chinese characters, and many people who share the same representations in chữ quốc ngữ may have different Chinese character representations. Moreover, sometimes Vietnamese names in Chinese characters have several representations in chữ quốc ngữ such as “吳時任” (Ngô Thị Nhâm or Ngô Thời Nhiệm in chữ Quốc ngữ). Therefore, representations in Chinese characters should be recorded if available, especially for those who were active before the end of the 19th century and have works penned in Chinese characters. These will prove to be useful access points when authority data sharing is being conducted, especially in the Chinese character cultural sphere or among the databases of classic books in Chinese or Vietnamese.

The corresponding relationship between Chinese characters and chữ quốc ngữ is only shown in HKCAN. In the cases in which many variant access points are recorded, allowing the authors to identify which form is a transcription of which form in Chinese characters is ideal. However, to realize this possibility, catalogers would need to have knowledge of both Vietnamese and Chinese characters. As modern people in Vietnam do not have knowledge of Chinese characters, and it is difficult for foreign catalogers to understand which Chinese characters correspond to which alphabet of chữ quốc ngữ, the fulfillment of this request would likely be very difficult.

Another problem lies in the duplicated records of CiNii Books, CALIS, and NCL. One of the duplications in CiNii Books occurs due to a difference in the type of Chinese character. For “Cao Bá Quát,” one record has an authorized access point of “高, 伯適||Cao, Bá Quat,” while another record has an authorized access point of “高, 伯適||コウ, ハクテキ.” Since “適” is a simplified Chinese character of the traditional Chinese character of “適,” the latter representation is sometimes used for this name, especially in Taiwan. In fact, “適” itself is also another traditional Chinese character, and this name in Chinese characters in Vietnam appears as “高伯適.”

Under the situation in which no national authority database exists in Vietnam, organizations in the Chinese character cultural sphere would cover this absence. However, currently, there is no
perfect substitute for the authority database of Vietnamese names. Some lack Chinese characters and some are insufficient as representations of chữ quốc ngữ. For modern Vietnamese names that lack Chinese forms, LCNAF could step in as a substitute because representations in Chinese characters would not need to be recorded. Adding Chinese character forms to Vietnamese authority records in LCNAF might be a tentative solution although the corresponding relationship between Chinese characters and chữ quốc ngữ is not shown in LCNAF.

Notes

1 王小盾, 劉春銀, 陳義, eds. 越南漢喃文獻目錄提要. 中央研究院中國文哲研究所, 2002, 3 vols.
11 According to the e-mail response from NDL in January 2015, NDL does not create authority data for Vietnamese materials, as well as Chinese and Korean materials.
13 For example, “倪良耀” sometimes uses the representation of “倪良曜”. More examples could be identified in 王德毅編著. 清人別名字號索引. 新文豐, 1985, 867p.
Another example is on p. 7 of 川本邦衛. 傳奇漫録刊本校. 慶應義塾大学言語文化研究所, 1998, 280p.
Chapter 9

Comparison of data elements of authority data in the Chinese character cultural sphere and the RDA framework

Chapters 5 to 8 studied representations of authority data. This chapter investigates data elements of authority data in the Chinese character cultural sphere. Using the same research method explained in Chapter 4, data elements recorded by each library were examined and compared to the authority data elements defined in the standard RDA (Resource Description and Access) design. Then, authority data elements not recorded in an authority database in the Chinese character cultural sphere were designated, as were authority data elements recorded in the cultural sphere but not prepared in RDA. Recommendations were then made to libraries within this cultural sphere to improve and internationally standardize their authority data. In addition, suggestions are provided to modify RDA in an effort to increase compatibility with authority data in the Chinese character cultural sphere. Since two of the libraries in Vietnam do not conduct authority control for author names, they are excluded from this study.

The phrase “authority data elements” means data elements recorded in authority data records created by the organizations chosen as research subjects. Authority data elements include birth date, affiliation, and country, among others, but exclude name strings and variant names. Note that subject authority data and title/series authority data were excluded from the scope of this study.

It could be said that each organization is competent in recording authority data for authors in its corresponding region or country. As such, the author mainly researched, for example, what kinds of authority data elements for Chinese authors are recorded by organizations in China; likewise, which authority data elements for Japanese authors are recorded by organizations in Japan. If an organization records a special element for foreign authors, it is mentioned in context but not included in the results tables.

Some parts of authority data recorded by each organization may be for internal use only due to privacy issues. Although this chapter mentions examples of closed information when obviously recorded, they were excluded from the results tables.

9.1 Authority data elements in Japan

Table 9-1 shows authority data elements for Japanese authors that were recorded by the research subject organizations in Japan. The field (e.g., access points, notes) in which the element is recorded is also shown in the table. Sometimes variant names of entities appear in the note or other fields, but because the purpose of this study is to examine authority data elements other than name strings, these variant names were omitted from the table.
Table 9-1 Authority data elements for Japanese authors recorded by organizations in Japan

<table>
<thead>
<tr>
<th>NDL</th>
<th>NACSIS-CAT</th>
<th>Keio</th>
</tr>
</thead>
<tbody>
<tr>
<td>birth/death year</td>
<td>birth/death year</td>
<td>birth/death year (+date)</td>
</tr>
<tr>
<td>field of study</td>
<td>field of study</td>
<td>business title</td>
</tr>
<tr>
<td>lineage</td>
<td>lineage</td>
<td>field of study</td>
</tr>
<tr>
<td>occupation</td>
<td>occupation</td>
<td>lineage</td>
</tr>
<tr>
<td>period of Japanese history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>year (+month) when the first material for cataloging was published</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: location</td>
<td>c: location</td>
<td>c: location</td>
</tr>
<tr>
<td>c: year of establishment</td>
<td>c: number of conference</td>
<td>c: number of conference</td>
</tr>
<tr>
<td>c: year (+month) when the first material for cataloging was published</td>
<td>c: place of conference</td>
<td>c: place of conference</td>
</tr>
<tr>
<td>c: year of conference</td>
<td>c: year (+date) of conference</td>
<td></td>
</tr>
<tr>
<td>c: year of establishment</td>
<td>c: year of establishment</td>
<td></td>
</tr>
<tr>
<td>affiliation</td>
<td>affiliation</td>
<td></td>
</tr>
<tr>
<td>awards received</td>
<td>birth/death year</td>
<td></td>
</tr>
<tr>
<td>birth/death year</td>
<td>business title</td>
<td></td>
</tr>
<tr>
<td>business title</td>
<td>field of study</td>
<td></td>
</tr>
<tr>
<td>field of study</td>
<td>occupation</td>
<td></td>
</tr>
<tr>
<td>license</td>
<td>place of employment</td>
<td></td>
</tr>
<tr>
<td>name of the group over which the person presides</td>
<td>type of doctorate degree</td>
<td></td>
</tr>
<tr>
<td>occupation</td>
<td>universities etc. latest graduated</td>
<td></td>
</tr>
<tr>
<td>period of activity</td>
<td>c: activity</td>
<td></td>
</tr>
<tr>
<td>c: history</td>
<td>c: date of conference</td>
<td></td>
</tr>
<tr>
<td>c: nature or character</td>
<td>c: history</td>
<td></td>
</tr>
<tr>
<td>c: type of corporate status</td>
<td>c: mission</td>
<td></td>
</tr>
<tr>
<td>c: year of establishment</td>
<td>c: nature or character</td>
<td></td>
</tr>
<tr>
<td>c: year of termination</td>
<td>c: number of conference</td>
<td></td>
</tr>
<tr>
<td>c: year of establishment</td>
<td>c: place of conference</td>
<td></td>
</tr>
<tr>
<td>c: year of termination</td>
<td>c: place of conference</td>
<td></td>
</tr>
</tbody>
</table>

Other fields

birth/death year
birthplace
c: location
c: place of establishment
c: year of conference
c: year of establishment
c: year of termination

Note. "c:" at the head of an element means the element is for corporate bodies or conferences.
1) The National Diet Library (NDL)

While JAPAN/MARC MARC 21 Format has been used for bibliographic and authority data from January 2012, it has not been revised in accordance with RDA for authority data applied by NDL; thus, many author-related data elements are recorded in note fields, primarily in the biographical or historical data field (tag 678). Elements that might be recorded in this field for persons include birth/death year, period of activity, occupation, affiliation, name of the group over which the person presides, business title, license, field of study, awards received, and shared pseudonyms, among others. When these elements appear in access point fields, they are not redundantly recorded in tag 678. Elements that might be recorded in this field for corporate bodies include type of corporate status, nature or character of the corporate body, or history, including the year of establishment and termination. While information such as age, birthplace, place of residence, graduated/enrolled in university, gender, and year arrived in Japan for persons might be recorded in the nonpublic general note field (tag 667), these are out of scope of this research because they are closed information.

Elements that appear in access point fields for persons include lineage (tag 100/400/500 $b), other designation associated with the person (tag 100/400/500 $c), and birth/death year (tag 100/400/500 $d). Lineage appears as numerals, such as “Frederick II, Holy Roman Emperor” in Western countries, though in Japan, several patterns exist; for example, 歌川豊国 2世 (Toyokuni Utagawa II, an ukiyo-e painter), 三遊亭円楽 6代目 (Sanyutei Enraku VI, a teller of rakugo comic stories), 森川弥平次 3代 (Yaheiji Morikawa III, a draper). Kanji characters such as “世,” “代目,” and “代” equally mean lineage, but they cannot be substituted for each other because these representations are succeeded from different ancestries.

In NDL, other designation associated with the person (tag 100/400/500 $c) would be recorded only when a lineage and a birth/death year are insufficient for identification. Elements that might be recorded here include occupation, field of study, or the year (and month, if needed) when the person’s first material for cataloging was published, the latter of which is recorded temporarily until other designations are available. Tag 100/400/500 $d could include names of different periods of Japanese history during which the person was active; for example, “平安時代” (Heian period) might be recorded when birth/death year for the person is undetermined.

In NDL, other designation associated with the person (tag 100/400/500 $c) would be recorded only when a lineage and a birth/death year are insufficient for identification. Elements that might be recorded here include occupation, field of study, or the year (and month, if needed) when the person’s first material for cataloging was published, the latter of which is recorded temporarily until other designations are available. As a general rule, NDL does not establish conference names as access points.

In general, terms for field of study and occupation for persons are picked from the cataloging materials at hand rather than being controlled by a vocabulary list. For occupation, commonly accepted terms are often selected by the catalogers. If authority data elements recorded in these access point fields are derived from materials other than cataloging materials, a citation for the consulted source is recorded in the source data found field (tag 670).
2) NACSIS-CAT

In the CATP format for authority data adopted by NACSIS-CAT, place—birthplace for persons and location or place of establishment for corporate bodies—and date fields—birth/death year for persons, year of establishment and/or termination for corporate bodies, and year of convocation of a conference—are defined, as are fields for access points and a note field. Due to privacy concerns, authority data cannot record birth/death months and days for persons. For the same reason, the birthplace of persons is recorded up to the level of municipality to avoid being too specific.\(^5\)

While birth/death year, lineage, field of study, and occupation might be added as access point designations for persons, their field of study and occupation are recorded only when identification is needed for persons having the same name and either the same birth/death year or when birth/death year is unknown. Similarly, the year of establishment and location are added as designations for corporate bodies when another corporate body having the same name exists. In accordance with rule 23.2.2.6H of *NCR1987 3rd rev.*, the number (if recurring), year, and place of convocation of conferences might be added in access point fields.\(^5\) As at the NDL, terms for field of study and occupation in access points for persons are not controlled by a vocabulary list.

In the note field, any data considered necessary for author identification could be recorded. The *Mokuroku shisutemukodingumanyuaru* manual provides some sample data elements for this purpose, such as type of doctorate degree, field of study, occupation, place of employment, business title, affiliation, birth/death year other than that recorded in date field, and name of the university or educational institution latest graduated for persons; for corporate bodies, the nature or character, mission of the organization, activities, and previous/subsequent name(s) with their relationships; and for conferences, number (if recurring), date, and place of convocation.\(^6\) Previous/subsequent name(s) with their relationships are represented as “history” in Table 9-1. Due to privacy concerns, phone number and home address for persons cannot be recorded. A citation (i.e., title, publisher, and year of publication) for the cataloging material used when the authority record is first created, as well as citations for other materials consulted, must be recorded in note field.\(^5\)

3) Keio University Libraries

Keio applies the MARC 21 format for bibliographic and authority data; however, new tags added in accordance with RDA are not currently applied when creating new authority records. Authority data elements that usually appear in access points for persons include lineage (tag 100/400/500 $b$ or $c$) and birth/death year ($d$). For persons who have the same name and birth year, the birth date might be added for clarity, while business title or field of study might be recorded only for persons having the same name when birth year is unknown (tag 100/400/500 $c$). Terms used for such elements are not controlled by a vocabulary list.

For corporate bodies, year of establishment and location may be added as designations in access points. For conference names, place of the conference (tag 111/411/511 $c$), year (and day/month, if needed) (tag 111/411/511 $d$) and number of conferences (tag 111/411/511 $n$) might be recorded. In
the source data found field (tag 670), the title and the publication year are automatically filled by the system when the authority record is created using a bibliographic record.

9.2 Authority data elements in South Korea

1) The National Library of Korea (NLK)

In NLK, authority data elements for persons are limited to birth/death year, and it is worth noting that for persons who share the same name and birth/death year, their access points are not distinguished (i.e., separate authority records that have the same authorized access point will be made); however, in the epitome field (tag 678 of KORMARC), the person’s position, birth/death year, field of activity, occupation, birthplace, associated corporate body, and biographical information might be recorded. In KORMARC format for authority data, which NLK applies, tag 678 is “a tag basically for catalogers’ reference,” meaning that NLK does not currently release information recorded in this tag; however, it is considering ways to show these data in their OPAC. In the source data found field (tag 670), citations for consulted sources are recorded.

2) Yonsei University Library (YUL)

Although Yonsei University Library applies KORMARC format, generally it does not use fields other than access point fields and the field for source data found (tag 670), and no data elements are considered mandatory in access point fields. With the exception of a few access points that include birth/death date, almost no authority data elements are recorded in Yonsei University Library. In tag 670, a citation is automatically filled by the system using the bibliographic data to which the authority record was first linked.

9.3 Authority data elements in Mainland China, Taiwan, and Hong Kong

Table 9-2 shows authority data elements for Chinese authors recorded by organizations in Mainland China, Taiwan, and Hong Kong.

9.3.1 Mainland China

1) The National Library of China (NLC)

In CNMARC/A, which was adopted by NLC, authority data elements might be added in access point fields for persons, such as the abbreviated name of Chinese ethnic groups (e.g., “维” for Uyghurs), dynasty (e.g., “三国蜀” for Shu Han in the Three Kingdoms period), field of study, occupation, and place of ancestry in tag 200/400/500 $c. For women, the Chinese character meaning “female” (“女”) is added in $c. However, this rule is currently being considered for termination and moved instead to the information note field (tag 300). Birth/death year is recorded in tag 200/400/500 $f, with the month as needed. For foreign people’s names recorded in Chinese characters, abbreviated names of countries, and original or other name strings are also added in $c. Terms for field of study and occupation are controlled by choosing from a list of terms established by NLC. The list is not publicly
Table 9-2 Authority data elements for Chinese authors recorded by organizations in Mainland China, Taiwan, and Hong Kong

<table>
<thead>
<tr>
<th>Access point fields</th>
<th>NLC</th>
<th>CALIS</th>
<th>NCL</th>
<th>NTUL</th>
<th>HKCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>birth/death year (+month)</td>
<td>academic degree</td>
<td>birth/death date</td>
<td>birth/death year</td>
<td>academic degree</td>
<td></td>
</tr>
<tr>
<td>Chinese ethnic group (abbreviated)</td>
<td>dynasty</td>
<td>dynasty</td>
<td>dynasty</td>
<td>birth/death year (+month)</td>
<td></td>
</tr>
<tr>
<td>dynasty</td>
<td>business title</td>
<td>field of study</td>
<td>field of study</td>
<td>occupation</td>
<td></td>
</tr>
<tr>
<td>field of study</td>
<td>dynasty</td>
<td>place of ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender (for female)</td>
<td>occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>place of ancestry</td>
<td>occupation</td>
<td>period of activity</td>
<td>place of ancestry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>period of activity</td>
<td>title</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c: location</td>
<td>c: location</td>
<td>c: location</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c: date of conference</td>
<td>c: number of conference</td>
<td>c: number of conference</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c: number of conference</td>
<td>c: place of conference</td>
<td>c: place of conference</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c: place of conference</td>
<td>c: year (+month) of conference</td>
<td>c: year (+date) of conference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: place of conference</td>
<td>c: year of conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Note fields</th>
<th>NLC</th>
<th>CALIS</th>
<th>NCL</th>
<th>NTUL</th>
<th>HKCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>affiliation</td>
<td>biography</td>
<td>academic background</td>
<td>awards received</td>
<td>biography</td>
<td></td>
</tr>
<tr>
<td>birthplace</td>
<td>activity</td>
<td>biography</td>
<td>birth/death date</td>
<td>birthplace</td>
<td></td>
</tr>
<tr>
<td>field of study</td>
<td>biography</td>
<td>birth/death date</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>occupation</td>
<td>published works</td>
<td>dynasty</td>
<td></td>
<td>place of ancestry</td>
<td></td>
</tr>
<tr>
<td>place of ancestry</td>
<td>related persons</td>
<td>ethnic group</td>
<td></td>
<td>place of death</td>
<td></td>
</tr>
<tr>
<td>published works</td>
<td>URL of the webpage</td>
<td>famous work written by the person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>social activity</td>
<td>related to the person</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: date of establishment</td>
<td>c: history</td>
<td>gender</td>
<td></td>
<td>c: date of conference</td>
<td></td>
</tr>
<tr>
<td>c: date of termination</td>
<td>c: mission</td>
<td></td>
<td>graduated university etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: history</td>
<td>c: nature or character</td>
<td>occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: location</td>
<td>c: related corporate body</td>
<td>place of ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: nature or character</td>
<td>c: type of corporate body</td>
<td></td>
<td>c: date of establishment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: URL of the webpage</td>
<td>c: history</td>
<td></td>
<td>c: location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>related to the corporate body</td>
<td>c: nature or character</td>
<td></td>
<td>c: type of corporate status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c: nature or character</td>
<td>URL of the webpage related to the entity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other fields</th>
<th>NLC</th>
<th>CALIS</th>
<th>NCL</th>
<th>NTUL</th>
<th>HKCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL of the webpage related to the entity</td>
<td>associated group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>birth/death date</td>
<td>birthplace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>birthplace</td>
<td>gender</td>
<td></td>
<td>language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>country</td>
<td>occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>field of study</td>
<td>gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>graduated university etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>occupation</td>
<td>place of ancestry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>place of residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. "c:" at the head of an element means the element is for corporate bodies or conferences.
Authority data elements might be recorded in access point fields for corporate bodies and conferences, such as number (if recurring) (tag 210/410/510 $d), place ($e), and date ($f) of convocation. The countr(ies) for foreign corporate bodies are added in tag 210/410/510 $c.

In NLC, catalogers record information in the note field (tag 300) using cataloging information at hand. Observations from sample records show that authority data elements recorded in tag 300 for persons include occupation, affiliation, social activity, pseudonym, real name, alternative name, place of ancestry, birthplace, field of study, and works published; and for corporate bodies, nature or characters of the organization, date of establishment and/or termination, location, history, and previous/subsequent name(s). Citations for consulted sources are recorded in the source data found field (tag 810).

2) The China Academic Library & Information System (CALIS)

CALIS has adopted the CALIS Union Catalog Authority Format. In access point fields for persons, period of activity, dynasty, field of study, business title, occupation, title, academic degree, and place of ancestry might be recorded in tag 200/400/500 $c when birth/death year (recorded in $f) is unknown. Terms for field of study are chosen from a list developed by the Ministry of Education of the People's Republic of China. Gender is recorded in $c but not very often, because there are many people who share the same name and the same gender in China. Location of corporate bodies might be added in access point fields when differentiation is necessary. As for conferences, number (if recurring) (tag 210/410/510 $d), place ($e), and date ($f) of convocation might be recorded.

In the information note field (tag 300), biography/history of the entity might be recorded; however, sample records show that biographical/historical information often appears in the general cataloger’s note field (tag 830). Citations for consulted sources are recorded in the source data found field (tag 810), and a URL of the webpage related to the entity might be recorded in the electronic location and access field (tag 856). Although tags 810, 830, and 856 are defined as fields for internal use in the CALIS 联合目录规范控制过程详细说明 (更新版) manual, this information is publicly available via the CALIS Union Catalog Authorities service.

9.3.2 Taiwan

1) The National Central Library (NCL)

In NCL, simple authority data are produced firstly by the Collection Development and Bibliography Management Division of NCL and then sent to the Synergy of Metadata Resources in Taiwan (SMRT) system and the Bibliographic Information Center of NCL augments the authority data elements as necessary. In the first step, access point fields, birth/death year, field of study, and place of ancestry for persons might be recorded. Terms for field of study are controlled by a list developed by NCL. If the person was born before 1911 (i.e., up until the Qing dynasty), the name of the dynasty is used instead of the field of study. 8
According to an e-mail from the Bibliographic Information Center in August 2014, authority data elements for persons recorded by the Center include field of study and birth/death date in access point fields; biography (tag 678) including academic background, URL of the webpage related to the person, activity, published works, and related persons in note fields; birth/death date (tag 046); birthplace, country, place of residence, place of ancestry, field of study, associated group including graduated university etc., occupation, and gender, which are recorded in tag 3XXs. Data elements of 3XX are recorded only when this information is available in cataloging materials. Birth/death date, biography and webpage URL are recorded as much as possible from cataloging materials or the internet. The source data found field (tag 670) is a mandatory element.  

As for corporate bodies, authority data elements recorded by the Center were not yet defined as of August 2014. According to the manual 团体权威整理作业手册, NCL might record history, such as prior/subsequent names and names of related corporate bodies with their relationships, type of corporate body, nature or characters, and mission of the corporate body in note fields. In access point fields of a conference, location, number (if recurring), place, and year of convocation (if the same conference occurs more than once in the same year, the month should be added) might be recorded. Source data found and URL of the webpage related to the corporate body also might be recorded.  

2) National Taiwan University Library (NTUL)  
In NTUL, data elements might be recorded in the biographical or historical data field (tag 678); for example, dynasty, birth/death date, place of ancestry, courtesy name (“字” in Chinese), pseudonym, related name, graduated university or graduate school, biography, occupation, ethnic group, gender, field of study, awards received, famous works written by the person might be recorded. In access point fields, if persons share the same name, birth/death year should be added as a distinguishing factor. If the person was born before 1911 (i.e., until the Qing dynasty), the dynasty name is used. If dynasty or birth/death year are the same or unknown, the field of study could be added. Terms for field of study are controlled by a list developed by NCL.  
If corporate bodies share the same name, location (e.g., names of local autonomous bodies) might be added in access point fields. For conferences, number (if recurring), place, and year (if the same conference occurs more than once in the same year, date should be added) of convocation are recorded in access point fields.  
Date of establishment, location, nature or character, type of corporate status, and history, including prior or subsequent names, might be recorded in tag 678. In the source data found field (tag 670), titles and years of published source data or a cited URL might be recorded.  

9.3.3 Hong Kong  
Hong Kong Chinese Authority Name Project (HKCAN)  
Authority data elements of HKCAN were extracted on June 18, 2014, by means of retrieving search terms in Table 5-1 in traditional Chinese Hanzi. Four conference names, shown in Table 9-3, were added as search terms. In sum, the search terms consist of 12 personal names, 4 corporate names, and
4 conference names. If a data element was included in any of authority records retrieved, it was regarded as an element recorded in HKCAN. Because the name “Wang, Jian (王健)” is shared by 19 persons in the database, 38 authority records in all were retrieved.

Table 9-3 Conference names as search terms for the HKCAN database

<table>
<thead>
<tr>
<th>Names</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>粤劇學術研討會 Yue ju xue shu yan tao hui</td>
<td>A conference held in Hong Kong, in 1992.</td>
</tr>
<tr>
<td>先進製造技術合作與交流大會 Xian jin zhi zao ji shu he zuo yu jiao liu da hui</td>
<td>A conference held in Hangzhou, Mainland China, in 2005.</td>
</tr>
<tr>
<td>粵港澳重要天氣研討會 Yue Gang Ao zhong yao tian qi yan tao hui</td>
<td>A conference held several times.</td>
</tr>
</tbody>
</table>

HKCAN already applies augmented MARC 21/A tags in accordance with RDA, so some records include the special coded dates field (tag 046), associated place field (tag 370), occupation field (tag 374), gender field (tag 375), and the associated language field (tag 377). Only three of the 38 records retrieved, however, included tag 046 and only 1 record included 3XX tags. In access point fields, birth/death year of persons are often added, and if several persons share the same birth/death year, birth/death month (e.g., “Wang, Jian, 1965 May”), occupation (e.g., “Wang, Jian, writer on Chinese philosophy”), or academic degree (e.g., “Wang, Jian, PhD”) might be added as distinguishing factors. For corporate bodies, location might be added in access point fields, and for conference names, the number (if recurring), place, and year of convocation might be added. In the source data found field (tag 670), citations for consulted sources are recorded. In addition, information for persons such as birth/death date, birthplace, place of death, place of ancestry, courtesy name, pseudonym, biography might be added, and date of conference and history, including prior/subsequent names of corporate bodies, could be recorded in the information found subfield ($b$) of tag 670.

9.4 A comparison to authority data elements in RDA

In Table 9-4 and Table 9-5, authority data elements for persons and corporate bodies recorded by each organization are compared to elements defined in Sections 9 and 11 of RDA, respectively. Since the NLK and the Yonsei University Library in South Korea record birth/death year only these two organizations have been omitted from the tables. Name strings such as Variant Name for the Person (RDA 9.2.3) and Fuller Form of Name (RDA 9.5) have also been omitted from the tables. Identifier for the Person (RDA 9.18) and Identifier for the Corporate Body (RDA 11.12), which are core elements defined in RDA, were also excluded from the tables because all of the organizations
Table 9-4 A comparison of authority data elements defined in Section 9 of RDA to elements recorded by each organization

<table>
<thead>
<tr>
<th>Chapter Element</th>
<th>RDA</th>
<th>NDL</th>
<th>NACSIS-CAT</th>
<th>Keio</th>
<th>NLC</th>
<th>CALIS</th>
<th>NCL</th>
<th>NTUL</th>
<th>HKCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2.2.18 A roman numeral associated with a given name</td>
<td></td>
<td>lineage</td>
<td>lineage</td>
<td>lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.3 Date Associated with the Person</td>
<td></td>
<td>year (+month)</td>
<td>birth year</td>
<td>birth year (+date)</td>
<td>birth year (+month)</td>
<td>birth year</td>
<td>birth date</td>
<td>birth date</td>
<td>birth date</td>
</tr>
<tr>
<td>9.3.2 Date of Birth#</td>
<td></td>
<td>birth year</td>
<td>birth year</td>
<td>birth year (+date)</td>
<td>birth year (+month)</td>
<td>birth year</td>
<td>birth date</td>
<td>birth date</td>
<td>birth date</td>
</tr>
<tr>
<td>9.3.3 Date of Death#</td>
<td></td>
<td>death year</td>
<td>death year</td>
<td>death year (+date)</td>
<td>death year (+month)</td>
<td>death year</td>
<td>death date</td>
<td>death date</td>
<td>death date</td>
</tr>
<tr>
<td>9.3.4 Period of Activity for the Person##</td>
<td></td>
<td>period of activity</td>
<td>period of activity</td>
<td>period of activity</td>
<td>period of activity</td>
<td>period of activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.3.5 Dynasty</td>
<td></td>
<td>dynasty</td>
<td>dynasty</td>
<td>dynasty</td>
<td>dynasty</td>
<td>dynasty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.5 Title of the Person#</td>
<td></td>
<td>title</td>
<td>title</td>
<td>title</td>
<td>title</td>
<td>title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.7 Gender</td>
<td></td>
<td>gender (for female)</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.9 Place of Death</td>
<td></td>
<td>birthplace</td>
<td>birthplace</td>
<td>birthplace</td>
<td>birthplace</td>
<td>birthplace</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.10 Country Associated with the Person</td>
<td></td>
<td>country</td>
<td>country</td>
<td>country</td>
<td>country</td>
<td>country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.11 Place of Residence, etc.</td>
<td></td>
<td>place of employment</td>
<td>place of residence</td>
<td>place of employment</td>
<td>place of ancestry</td>
<td>place of ancestry</td>
<td>place of ancestry</td>
<td>place of ancestry</td>
<td></td>
</tr>
<tr>
<td>9.12 Address of the Person</td>
<td></td>
<td>affiliation</td>
<td>affiliation</td>
<td>affiliation</td>
<td>affiliation</td>
<td>affiliation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.13 Affiliation</td>
<td></td>
<td>name of the group over which the person presides</td>
<td>associated group</td>
<td>associated group</td>
<td>associated group</td>
<td>associated group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.14 Language of the Person</td>
<td></td>
<td>language</td>
<td>language</td>
<td>language</td>
<td>language</td>
<td>language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.15 Field of Activity of the Person</td>
<td></td>
<td>field of study</td>
<td>field of study</td>
<td>field of study</td>
<td>field of study</td>
<td>field of study</td>
<td>field of study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.16 Profession or Occupation##</td>
<td></td>
<td>occupation</td>
<td>occupation</td>
<td>occupation</td>
<td>occupation</td>
<td>occupation</td>
<td>occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.17 Biographical Information</td>
<td></td>
<td>biography</td>
<td>biography</td>
<td>biography</td>
<td>biography</td>
<td>biography</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.18 Published works</td>
<td></td>
<td>published works</td>
<td>published works</td>
<td>published works</td>
<td>published works</td>
<td>published works</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.19 University etc. latest graduated</td>
<td></td>
<td>academic background</td>
<td>graduated university etc.</td>
<td>academic background</td>
<td>graduated university etc.</td>
<td>academic background</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.20 Licenses and Awards received</td>
<td></td>
<td>academic degree</td>
<td>academic degree</td>
<td>academic degree</td>
<td>academic degree</td>
<td>academic degree</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. An element with # is a core element. An element with ## is a core element with reservations.
Table 9-5 A comparison of authority data elements defined in Section 11 of RDA to elements recorded by each organization

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Element</th>
<th>NDL</th>
<th>NACSIS-CAT</th>
<th>Keio</th>
<th>NLC</th>
<th>CALIS</th>
<th>NCL</th>
<th>NTUL</th>
<th>HKCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.3</td>
<td>Place Associated with the Corporate Body</td>
<td>place of establishment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.3.2</td>
<td>Location of Conference, etc.##</td>
<td>place of conference</td>
<td>place of conference</td>
<td>place of conference</td>
<td>place of conference</td>
<td>place of conference</td>
<td>place of conference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.3.3</td>
<td>Location of Headquarters##</td>
<td>location</td>
<td>location</td>
<td>location</td>
<td>location</td>
<td>location</td>
<td>location</td>
<td>location</td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>Date Associated with the Corporate Body</td>
<td>year (+month) when the first material for cataloging was published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4.2</td>
<td>Date of Conference, etc.#</td>
<td>date of conference</td>
<td>year (+date) of conference</td>
<td>date of conference</td>
<td>year (+month) of conference</td>
<td>year (+date) of conference</td>
<td>date of conference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4.3</td>
<td>Date of Establishment##</td>
<td>year of establishment</td>
<td>year of establishment</td>
<td>year of establishment</td>
<td>date of establishment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4.4</td>
<td>Date of Termination##</td>
<td>year of termination</td>
<td>year of termination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td>Period of Activity of the Corporate Body##</td>
<td></td>
<td>related corporate body</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.6</td>
<td>Number of a Conference##</td>
<td>number of conference</td>
<td>number of conference</td>
<td>number of conference</td>
<td>number of conference</td>
<td>number of conference</td>
<td>number of conference</td>
<td>number of conference</td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>Other Designation Associated with the Corporate Body##</td>
<td>type of corporate status</td>
<td>type of corporate status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7.1.4</td>
<td>Type of Corporate Body</td>
<td>type of corporate status</td>
<td>type of corporate status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7.1.5</td>
<td>Type of Jurisdiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7.6</td>
<td>Other Designation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td>Language of the Corporate Body</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.9</td>
<td>Address of the Corporate Body</td>
<td>URL of the webpage related to the corporate body</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.10</td>
<td>Field of Activity of the Corporate Body</td>
<td>nature or character</td>
<td>nature or character</td>
<td>nature or character</td>
<td>nature or character</td>
<td>nature or character</td>
<td>nature or character</td>
<td>nature or character</td>
<td>mission</td>
</tr>
<tr>
<td>11.11</td>
<td>Corporate History</td>
<td>history</td>
<td>history</td>
<td>history</td>
<td>history</td>
<td>history</td>
<td>history</td>
<td>history</td>
<td></td>
</tr>
</tbody>
</table>

Note. An element with # is a core element. An element with ## is a core element with reservations.
researched in this study have a record identifier or control number for each authority record.

Some data elements could represent relationships, defined in Sections 30 and 32 in RDA, rather than attributes, in Sections 9 and 11. For example, “associated group” in NCL, shown in Table 9-4, could also be seen as a relationship defined in RDA 32.1, “related corporate body.” However, in this study, all data elements were sorted at the attribute level, because the occurrence of such data elements was relatively small, and the handling of each data element as an attribute or a relationship might differ across organizations. In addition, relationships between persons, families, and corporate bodies are not core elements of RDA.

Core elements for persons defined in Section 9 of RDA (except name strings and identifier) include Date of Birth, Date of Death, Title of the Person, and Other Designation Associated with the Person. Period of Activity for the Person and Profession or Occupation are core elements “when needed to distinguish a person from another person with the same name” (RDA 9.3.4; RDA 9.16).

Lineage is recorded in Japanese records, and while it is treated the same as “a roman numeral associated with a given name” (RDA 9.2.2.18), lineage also occurs with some Chinese characters, such as “世,” “代目,” or “代.” Date of Birth and Date of Death were recorded by all organizations, though three of them only record the year.

RDA’s glossary states that Title of the Person is “a word or phrase indicative of royalty, nobility, or ecclesiastical rank of office, a term of address for a person of religious vocation, or another term indicative of rank, honor, or office.” Business titles such as “chief executive” or “professor” are included in Title of the Person in Table 9-4, although whether business titles can be included is not completely clear in RDA's explanations. “Business title” was recorded by CALIS and three organizations in Japan (NDL, NACSIS-CAT, and Keio).

In RDA 9.6, Other Designation Associated with the Person includes the term “Saint” for Christian saints; “Spirit” for a spirit; specific designations for persons named in sacred scriptures, apocryphal books, and fictitious and legendary persons; and a designation for type, species, or breed for non-human entities; and “other designation.” Above-mentioned designations other than “other designation” are not recorded in the Chinese character cultural sphere, so while various “other designations” are recorded in many access points fields, as shown in Table 9-1 and Table 9-2, these designations are sorted in each data element in Table 9-4.

“Period of Japanese history” or “dynasty” is not a particular indication of Period of Activity for the Person, but these would help users know a general frame during which the person was alive. Five of the eight organizations recorded this element when necessary, while occupation was recorded by seven of the eight organizations.

Of these non-core elements, “field of study” was recorded by seven of the eight organizations. “Place of ancestry” and “gender” were recorded by all organizations in China. As such, these elements can be considered as strong identifiers for persons in the Chinese character cultural sphere.

In Table 9-4, “place of ancestry” is included in Place of Residence for five of the organizations, although place of ancestry is not necessarily the place where the person lived or lives. In China and
South Korea, however, the ancestral place can be considered “another significant place associated with the person” (as defined in RDA 9.11) due to the depth of this relationship. RDA defines Date Associated with the Person as including Date of Birth, Date of Death, and Period of Activity for the Person. Table 9-4 shows that NDL may record the year (and month, if necessary) when the person’s first material for cataloging was published as a part of this category.

While NLC records abbreviated names of the country in access points for foreign entities, only NCL records Country Associated with the Person. In this study, only data elements recorded for entities in the region where the organization is located were researched. Although Country Associated with the Person is not necessarily equal to nationality, in Japan, nationality of persons sometimes is regarded as a privacy issue, which is why there may be only a few records with this information from Japanese organizations.

Because only one organization recorded Place of Death and Language of the Person and only two organizations recorded Address of the Person, these elements are not regarded as strong elements for personal identification in the Chinese character cultural sphere.

For corporate bodies, core elements recorded by almost all organizations include Location of Conference, etc. (RDA 11.3.2), Date of Conference, etc. (RDA 11.4.2), and Number of a Conference (RDA 11.6). The exception is NDL, which do not establish conference name access points. The following are considered core elements with reservations: Date of Establishment (RDA 11.4.3), Date of Termination (RDA 11.4.4), Period of Activity of the Corporate Body (RDA 11.4.5), Associated Institution (RDA 11.5), and Other Designation Associated with the Corporate Body (RDA 11.7).

Associated Institution for conference names is recorded when “the institution’s name provides better identification than the local place name or if the local place name is unknown or cannot be readily determined” (RDA 11.5). Because place of conference recorded by each organization may be a local place name or an institution’s name, organizations might record this element for conferences. However, for corporate bodies other than conferences, only NCL recorded Associated Institution.

Location of Headquarters was recorded by all organizations. Date of Establishment and Date of Termination were recorded by five of the eight, and three of the eight organizations, respectively, though it is possible that these elements appear in the history field. Only certain data elements that the author confirmed through interviews, manuals, or samples are shown in tables, so some data elements might escape these findings.

Other Designation Associated with the Corporate Body includes Type of Corporate Body, Type of Jurisdiction, and Other Designation. Only two organizations recorded the type of corporate status, such as “company limited” or “educational corporation,” and NCL recorded the type of corporate body.

Among non-core elements, Corporate History and Field of Activity of the Corporate Body were recorded by more than half of the organizations. Two organizations recorded the URL of the webpage associated with the corporate body, while other elements such as Place Associated with the Corporate Body, Period of Activity of the Corporate Body, and Language of the Corporate Body were less likely
to be recorded. Special data elements (relative to Western countries) for corporate bodies are not recorded in the Chinese character cultural sphere, while some data elements for persons are special (such as place of ancestry).

9.5 Findings and recommendations

Core elements defined in RDA were recorded by most organizations. Among non-core elements for persons, field of study, lineage (especially in Japan), gender, place of ancestry (especially in China) were recorded by many organizations. Biographical information, which includes various data elements such as academic degree, was recorded by most organizations. Among non-core elements for corporate bodies, nature or character and history were recorded by many organizations.

As shown in Table 9-1 and Table 9-2, some organizations recorded a lot of authority data elements in the note fields, which are text-based and do not include subfield codes. Thus, even if each organization applies new fields of MARC 21/A in accordance with RDA, such as 3XX fields, it is difficult to sort these elements automatically. It is therefore desirable that each data element is recorded distinctively.

Some organizations recorded authority data elements for internal use only, likely due to privacy issues or a lack of confidence in the data itself. Verifying data often requires labor and monetary investment, so making the information private is often easier; however, this does not facilitate international sharing of authority data. Ideally, organizations should open all data elements except those involved with true privacy issues.

Each region and organization has its own views on privacy and which data elements should be publicly available. Thus, even if each organization decided to apply RDA, there will be discrepancies in the selected authority data elements due to these differing views.

In RDA 9.19.1, additions to access points for persons are defined (e.g., fuller form of name, period of activity of the person, other designation). As indicated in Table 9-1 and Table 9-2, additions to access points in Japan and China are extremely varied and many of them are included in “other designation.” For example, field of study, place of ancestry, and gender might be added; however, following RDA, data elements such as Field of Activity, including field of study (RDA 9.15), Place of Residence, including place of ancestry (RDA 9.11), and Gender (RDA 9.7) should be recorded as separate elements and should not be recorded as part of an access point. These are inadequate rules for the Chinese character cultural sphere. Because RDA aims to become internationally accepted and is a content (rather than an encoding) standard, it should not be so restrictive in mandating that an element should be recorded separately rather than as part of an access point. To remedy this, the rule should be changed and more examples of “other designations” should be provided in Sections 9.19.1.7 and 9.6.1.9.

RDA could also benefit from augmenting examples of data elements for persons such as Place of Residence and Title, as it is not explicitly stated whether place of ancestry or business title should be included, respectively.
The field of study data element, which was recorded by many organizations, is not controlled by a standard vocabulary list in Japan. Moreover, different lists of terms are used by NLC, CALIS, and two organizations in Taiwan. Because field of study is a data element frequently recorded in the Chinese character cultural sphere, creating a list of terms in each region would help unify this data element across organizations in each region.

Some data elements also could be recorded using relationships; for example, “place of ancestry” might be recorded as a “place of residence” of a related family. Whether each data element should be recorded as an attribute or a relationship, however, is decided by each organization, not by RDA, so there is a concern that each organization could record the same data element in a different fashion. This could cause confusion during international authority data sharing. A common authority data format and specific rules might avoid such confusion.

There are many persons who share the same name in China and South Korea, so providing a unique access point for each of them is difficult. While many data elements might be added to create a unique access point in China, non-unique access points are allowed in South Korea. Due to each region’s varied conditions, it is not easy to apply RDA to their current authority data.

The main tools used for this study were interviews, cataloging rules, authority formats, manuals, and limited sample records, so it is difficult to know whether or not each data element is recorded with a high frequency. While this study roughly depicted which kinds of authority data elements are recorded in authority records produced by organizations in the Chinese character cultural sphere, some data elements might not be captured in the results because they are embedded in other data elements such as history or biography. This study provided recommendations to libraries in this cultural sphere to improve their authority data to be more internationally sharable, as well as recommendations to RDA to increase its compatibility with authority data in the Chinese character cultural sphere.

Notes

1 Part of this chapter has already been reported as: Kimura, Maiko. A comparison of recorded authority data elements and the RDA framework in Chinese character cultures. Cataloging & Classification Quarterly. 2015, (accepted).
7 “CALIS 联合目录规范控制过程详细说明（更新版）”.
9 “國家圖書館人名權威檔個人專家學科領域表”．國家圖書館編目園地全球資訊網.
(accessed 2015-02-16).
(accessed 2015-02-16).
11 “團體權威整理作業手冊”．國家圖書館編目園地全球資訊網.
12 American Library Association; Canadian Library Association; Chartered Institute of Library and
2015-05-07).
Chapter 10

A proposal of a modification of FRAD model

In Chapters 5 through 8, the representations recorded in each organization in the Chinese character cultural sphere were investigated, revealing that any single type of representation is insufficient as a master key for name identification when name authority data are shared. Rather, the combination of several representations seems to be helpful for name identification. This combination, where two representations should be shown as a pair, depends on what kind of relationships the two representations have.

The representation is one of three components of authority data, as shown in Chapter 3. Since it lacks in FRAD model, this chapter proposes a modified FRAD model for names in non-Latin languages. As noted in Chapter 2, one problem with the current FRAD model is that it cannot sufficiently identify relationships among names in non-Latin languages. Thus, the ability of the model to correctly interpret authority data in non-Latin languages is highly questionable. The relationships derived by FRAD are representative only and do not aim to provide an exhaustive taxonomy of relationships, so that specific applications may select or generate relationships as needed. To prevent applications from establishing their own protocols for non-Latin languages, a modification of the FRAD model is needed in order to reliably share authority data internationally.

In this Chapter, differences between transliteration, transcription, and Romanization are clarified and the necessity of showing the parent-child relationship between an original form and its Romanization and/or transcription is advocated. Afterward, a modification of the FRAD model for recording names in non-Latin languages is proposed.

The modified FRAD model uses the Name and Controlled Access Point (CAP) entities in FRAD. Attributes and relationships shown in FRAD are adopted wherever possible, but are amended or supplemented as needed. The proposed model attempts to mimic the practice of name expression in each local region, and all names are those of authors or the creators of works. Personal and corporate names used as subjects in works are excluded, because these are more properly handled by FRSAD.

10.1 Differences among transliteration, transcription, and Romanization

The ISO 5127:2001 (Information and documentation: vocabulary) defines Romanization as the “representation of non-Roman writing systems in the Latin alphabet by means of transliteration, transcription or both.” The word transliteration means the “representation of the characters of one writing system, alphabetic or syllabic, in terms of corresponding characters of a second writing system.” Transcription is defined as the “representation of the pronunciation of a given language by the characters of a writing system or by a specially devised system of notations”. The library community, however, frequently interprets transcription as “the action or process of transcribing or copying”. Indeed, RDA 1.7 uses the word in this context. In addition, transliteration in most library
catalogs is actually a form of phonetic transcription.\(^6\) The *PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets* defines *transliteration* as the “systematic conversion of text from one script to another.”\(^7\) This definition does not clearly distinguish between *transcription* and *transliteration*. Linguistically, however, “transliteration needs to be distinguished from transcription”\(^8\) because they are two different things. For example, some languages, such as Yiddish, could transcribe or transliterate but each process would yield a different result.\(^6\) This study assumes the ISO 5127:2001 definitions of *transcription*, *transliteration*, and *Romanization*.

Generally, Chinese-Japanese-Korean (CJK) languages are Romanized by transcription rather than transliteration. Some languages, such as the Cyrillic alphabet, can be precisely transliterated letter-by-letter, and are easily converted from the original to the transliterated forms and vice versa.\(^9\) This is called *reversibility*, which is a feature of exact transliteration only.\(^6\) Since CJK languages are Romanized by pronunciation, they do not have reversibility.\(^10\)

The result of transcription or transliteration is not always evident in the Latin alphabet. If the result is in the Latin alphabet, it should be called “Romanization.” In other words, there is non-Latin transliteration and non-Latin transcription. The Korean *Hangul* and *Hanja* is an example of non-Latin transliteration relationships. As noted in Section 3.1.3.1, with few exceptions, one *Hanja* character transliterates into one *Hangul*. However, a single *Hangul* corresponds to multiple *Hanja*. While a *Hanja* name can be transliterated to *Hangul*, the reverse does not apply, because several *Hanja* candidates exist for a given *Hangul* name. Thus, Korean *Hangul* and *Hanja* possess a one-to-many transliteration relationship.

An example of non-Latin transcription is the Japanese *yomi*. Since multiple pronunciations exist in Japanese *Kanji* and *kana* (meaning *katakana* and *hiragana*), character strings called *yomi*, which represent pronunciations recorded in *katakana*, are added to the access point of bibliographic and authority records in Japan. The *yomi* serve to standardize and collocate access points. As mentioned in Section 3.1.2.1, *yomi* is very important in Japanese because, even if the original form of the name is the same, if their *yomi* is different (i.e., their pronunciation for the original form is different), they are two different persons.

As explained above, Romanization, non-Latin transliteration, and non-Latin transcription are distinguished from each other in this study. However, no distinction is made between Romanization based on transliteration and that based on transcription. Whether a Romanized form is generated by means of transliteration or transcription is hardly identifiable because some Romanization rules intermix them. Therefore, it is impractical to divide Romanization into Roman transliteration and Roman transcription.

10.2 The parent-child relationship between an original form and its Romanization and/or transcription

As depicted in Table 10-1, Romanization forms of names can be divided into two types. One is the
Table 10-1 Existence of the parent-child relationship in Romanization, transliteration, and transcription

<table>
<thead>
<tr>
<th>Type of relationship</th>
<th>Parent-child</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romanization</td>
<td>(1) “English name” in common use</td>
<td>× 孫逸仙 and “Sun, Yat-sen”</td>
</tr>
<tr>
<td></td>
<td>(2) Imposed by libraries</td>
<td>☑ 孫文 to “Sun, Wen”</td>
</tr>
<tr>
<td>Non-Latin transliteration</td>
<td>☑ 金永蕙 to 김영혜</td>
<td></td>
</tr>
<tr>
<td>Non-Latin transcription</td>
<td>× 陈云金 and 陳雲金</td>
<td></td>
</tr>
<tr>
<td></td>
<td>☑ 河野明 to コウノアキラ</td>
<td></td>
</tr>
</tbody>
</table>

name that is currently in use. For example, when a person who has a Chinese original form of their name published in an English book, his or her Romanized name is represented in the Latin alphabet. In this case, generally the Romanized name is decided by the author, and there is no need to be in accord with one kind of Romanization scheme. In other words, the author chooses his or her “English name,” which is more common in the Western country.

Another type of the Romanization form was developed because libraries in Western countries could not handle original scripts of names. This Romanization is in accord with Romanization schemes, such as the ALA/LC Romanization Tables. It is not an actual “name” in the real world, although it serves as an access point for the library community. This imposed Romanization may or may not be the same form as the author’s “English name.” Adding to the confusion, authorized access points for the LC/NACO Authority File (LCNAF) may be the former or latter. For example, although ALA/LC Romanization Tables assert that the *Hanyu pinyin* system should be used for Chinese Romanization, an authorized access point for 孫文 (one of his pseudonyms is 孫逸仙) in LCNAF is “Sun, Yat-sen,” which does not follow the *Hanyu pinyin* system, given “Sun, Yat-sen” is in general English-language reference sources. This observation is according to an alternative rule of AACR2 22.3C2 and now, an alternative rule of RDA 9.2.2.5.3. It could be said that “Sun, Yat-sen” is an “English name” that is more common in Western countries.

As for former Romanized names (i.e., currently “English names”), their corresponding original form is not always shown on materials because the material itself is in the Latin alphabet. In addition, the necessity of showing original forms seems weakened because the form of the name in the Latin alphabet could be considered an independent name used in a Western country. As for imposed Romanization, however, they are not real names in any countries and all names are derived from the original (i.e., the original form and its imposed Romanization have a parent-child relationship).

The imposed Romanization is derived from the original form, so these two access points must be shown as a pair. The reason for this is that in the case of a person who has many pseudonyms or
aliases, which Romanized form is derived from which original form of name should be clearly shown in the record, so that name identification is easily performed. Thus, the author objects to Aliprand’s account\textsuperscript{11} that pairing the original form of name and its Romanization is not needed in authority data. It is needless to say that many people share the same name, especially in China and Korea,\textsuperscript{12,13} and even if their original name forms are different, their Romanization form might be the same.\textsuperscript{14} Therefore, pairing an original form with its Romanization form is very helpful to differentiate names.

Similarly, the original form of a name and its non-Latin transcription also have a parent-child relationship and should be shown as a pair. For example, Japanese Kanji and its yomi should be handled together, because yomi is not a real name. It is derived from the original form of a name (in many cases, it is the Kanji form of a name), and to distinguish from other yomi form of names derived from other pseudonyms a pairing is needed. Imposed Romanization for Japanese names is derived from yomi, not directly from Kanji, thus there is a twofold structure. At the first step, Kanji and its yomi form a pair. Then, as a second step, yomi and its imposed Romanization form a pair. As a result, Kanji, yomi, and the imposed Romanization should be represented as a triple data point in the record.

Non-Latin transliteration and its original form, on the other hand, may or may not have the parent-child relationship. Names in Hanja can be transliterated into Hangul, thus there is a parent-child relationship. The Hanja name is useful to identify a person because several persons share the same Hangul name, thus the Hanja form of a name and Hangul form of a name must be shown as a pair, if it is sure that the Hangul name is derived from the Hanja name. However, there are some people whose Hanja name is unclear because there is only one Hangul name on the material, or people do not have a name in Hanja and only have the name in Hangul. In this case, the parent-child relationship does not come into play.

Names in traditional Chinese script also can be transliterated into simplified Chinese script. However, names in simplified script in mainland China cannot always be transliterated into traditional Chinese script accurately, because a simplified Chinese character corresponds to multiple traditional Chinese characters. For example, the simplified Chinese character “雲” corresponds two traditional Chinese characters “雲” and “云,” so it is difficult to determine which traditional Chinese script is correct for the name. Although simplified letters are essentially formed from traditional letters, for now they are two independent writing systems, one is used in mainland China and another is used in Taiwan (and Hong Kong and Macau, with minor differences in shape), and the two systems are never intermixed in one text. Thus, neither system is a parent or child, nor do they need to be handled as a pair. In another example, Traditional Mongolian, used in the Inner Mongolia Autonomous Region of mainland China, could be transliterated to Modern Mongolian, which can be represented in the Cyrillic alphabet used in Mongol and each other, although transliteration would not be perfect.\textsuperscript{15} However, because they are two independent writing systems, it is needless to handle them as a pair.

10.3 A modification of the FRAD model for recording names in non-Latin languages

The review of FRAD conducted in Chapter 2 revealed that recording names in non-Latin languages
under the FRAD model creates several difficulties. A new problem now arises in addition to the problems highlighted in Chapter 2 in that the parent-child relationship between two CAPs cannot be shown in the FRAD model.

In this section, a modified FRAD model that matches names in non-Latin languages as well as Western languages is proposed. The model is based on the FRAD model, because despite the problems, the model is accepted globally and can depict the structure of authority data comprehensibly. Given the problems in the current FRAD model, the following components were identified for a modification of the FRAD model to accommodate names in non-Latin languages:

1) *Can differentiate transliteration and transcription*

For example, *yomi* in Japanese is neither Romanization nor transliteration. It is a transcription of its original form, and should be represented in the modified model. Chinese characters used in Vietnam before 1945 and *Chữ quốc ngữ*, an official writing system of Vietnam, have the same transcription relationship. For example, the name of Nguyễn Du, a poet from Vietnam, is depicted in the Chinese character “阮攸,” and “Nguyễn Du” is a form of phonetic transcription of “阮攸.” It is not a Romanization in the precise sense because *Chữ quốc ngữ* is written in the Latin alphabet with tone letters. Without tone letters, the meaning of the alphabet would be hardly determined.

2) *Can express a Romanization scheme*

As previously suggested, displaying the Romanization scheme would assist the user. FRAD has an attribute “transliteration scheme,” which is substantially a Romanization scheme. However, transliteration and Romanization should be differentiated because there is the possibility of non-Latin transliteration. Signification of the word *transliteration* should be reconsidered.

3) *Can show the parent-child relationship between an original form and its Romanization/transcription.*

In the current FRAD model, all CAPs are equivalent. That is, the parent-child relationship among variant access points cannot be shown. However, users should understand which access point has been derived from which. This knowledge would assist in matching algorithms because the parent access points could be assigned higher priority than the derived ones.

A modified FRAD model that fulfills the above three points is illustrated in Figure 10-1. Although the entity represented in Figure 10-1 is Person, it can be switched to Corporate Body. Other entities, attributes, and relationships remain unchanged even if the entity is switched. The model is applicable to any name expressed in several writing systems, regardless of language. Table 10-2 is a comparison of attributes and relationships regarding representations of name and CAP between FRAD and the modified model.
Figure 10-1 The modified FRAD model for personal names in non-Latin languages

Note. BAP – Base Access Point.
<table>
<thead>
<tr>
<th>Attributes of a Name</th>
<th>Language of name</th>
<th>Script of name</th>
<th>Transliteration scheme of name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language of BAP</td>
<td>Language of cataloging</td>
<td>Script of BAP</td>
<td>Script of cataloging</td>
</tr>
<tr>
<td>Script of BAP</td>
<td>Transliteration scheme of BAP</td>
<td>(obsolete)</td>
<td>(obsolete)</td>
</tr>
<tr>
<td>Transliteration scheme of cataloging</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attributes of a CAP</th>
<th>Romanization scheme of BAP*</th>
<th>Romanization scheme of cataloging*</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Relationships between Names of Persons</th>
<th>Alternative linguistic form relationship</th>
<th>Alternative linguistic form relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transliteration relationship (as one of Other variant name relationships)</td>
<td>(obsolete)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationships between Names of Corporate Bodies</th>
<th>Alternative linguistic form relationship</th>
<th>Alternative linguistic form relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transliteration relationship (as one of Other variant name relationships)</td>
<td>(obsolete)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relationships between CAP</th>
<th>Parallel language relationship</th>
<th>Alternate script relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note. Newly defined attributes and relationship are shown with *.
CAP - Controlled Access Point; BAP - Base Access Point.
Names and CAPs are distinguished in this model, as in the original FRAD, which distinguishes between the name in currently in use in the real world and the access point imposed by libraries. CAPs derived from an original form are not names in the real world, but imposed access points given by libraries (i.e., an imposed Romanization is not a name, but it is a CAP).

FRAD specifies the name attributes “script of name,” “transliteration scheme of name,” and “language of name.” The term transliteration is used in reference to Romanization in FRAD, as pointed out above, and should be differentiated. Thus, the word transliteration should be avoided. Moreover, name should be in commonly used in the real world in this model, so no name has a transliteration scheme. CAP, on the other hand, will have a Romanization scheme in this model. Since a CAP with “Romanization scheme of BAP (Base Access Point)” attributes exists only when another CAP is Romanized, a CAP based directly on a name will lack this attribute.

Since CAPs are based on the name, it takes the language of the name, although the attribute “language of BAP” is not shown in Figure 10-1. CAPs also have the attributes “script of cataloging,” “Romanization scheme of cataloging,” and “language of cataloging,” which are applied to additions (supplementary to BAP in CAP). However, these attributes are also not shown in Figure 10-1 because the languages and scripts of access points are unlikely to differ between BAP and additions in countries that use non-Latin characters. For the same reason, although a CAP could be divided into BAP and additions, which could also have relationships of representations, these are not shown in Figure 10-1. Practically, additions of BAP are also recorded in various representations. When Romanized forms of BAP are recorded, additions of BAP are also Romanized. Similarly, when yomi for BAP are recorded, yomi for these additions are also recorded. Therefore, this study assumes that relationships for representations of additions could be handled with BAP, and thus, in Figure 10-1, CAP (meaning BAP plus additions) has representational relationships.

As implemented in FRAD, if two CAPs have different scripts, an “alternate script relationship” exists between them. An “alternate script relationship” is a non-Latin transcription relationship.

The new model introduces a “Romanization relationship” between CAPs. Note that only a CAP based on an imposed Romanization has this relationship; therefore, a CAP based on an “English name” does not have this relationship. Only the CAP indicated by the “Romanization relationship” arrow possesses the attribute “Romanization scheme of BAP.” The “non-Latin transcription relationship” is newly defined, because it was lacked in FRAD model. The “parallel language relationship” of the original FRAD model is used for a name in another official language, as already explained in Chapter 2. Thus, the “parallel languages relationship” still remains in the modified model.

In the modified model, all CAPs are not equal, and CAPs indicated by arrows of Romanization or transcription relationship are derived from its parent CAPs. In this implementation, when Name identification is required, two CAPs connected by these arrows should be considered as a pair. On the other hand, CAPs connected by an “alternate script relationship,” which means a non-Latin transliteration relationship in the modified model, sometimes should be handled as a pair, but sometimes this is not necessary. For example, Hanja and its Hangul should be paired only when a
Hangul name is derived from a Hanja name. A name in traditional Chinese characters and a name in simplified Chinese characters should be paired only when the simplified name is derived from the traditional name; otherwise pairing is difficult and superfluous.

Sometimes one original form of name can be Romanized to several Romanized forms because several Romanization schemes exist. In this case, several Romanization relationships can be set to one parent CAP, as Figure 10-2 shows.

Figure 10-2 Several Romanization relationships for one Controlled Access Point

Although Sun Yat-sen has many pseudonyms and derived Romanization forms, only Romanizations derived from one of his pseudonyms, “孙逸仙,” is shown in this figure. The name “Sun, Yat-sen” is his English name in common use in the Western countries, thus it is not derived from an original form. On the other hand, the access point in pinyin Romanization, “Sun, Yixian,” is derived from the name in simplified Chinese script, “孙逸仙,” and the name in traditional Chinese script, “孫逸仙.” The access point in the Wade-Giles system of Romanization, which was used in Taiwan, is also derived from the name in traditional Chinese script. Here, “孫逸仙” has two kinds of Romanization, and each of them has a different “Romanization scheme” attribute. In this manner, one CAP can have several Romanization relationships. In this case, the original form of name “孙逸仙” and each Romanization form a parent-child relationship, so each pair should be handled together. A CAP also can have several
transcription relationships, because one access point in Kanji may have several yomi.

The work of Japanese writer Haruki Murakami has been translated into many languages. Many variant access points exist in his LCNAF records, as shown in Figure 10-3, and these access points can be structurally depicted, as shown in Figure 10-4. Here, users can easily understand the access point “ムラカミ，ハルキ，1949–” is derived from “村上，春樹，1949–” in Kanji, so they form a pair. In other words, even if two persons share a name, “ムラカミ，ハルキ，” when their parent names in Kanji are different, a machine can discern that they are different persons. The access point in English “Murakami, Haruki, 1949–” and the imposed Romanization in Japanese “Murakami, Haruki, 1949–” are the same result. However, they are distinguished in this model because imposed Romanization may vary based on the Romanization scheme.

LC control no.: n 81152383
LCNN permalink: http://lccn.loc.gov/n81152383

Figure 10-3 A LCNAF record for Haruki Murakami

163
Discussion

The modified FRAD model proposed in this Chapter organizes several CAPs in non-Latin languages. Several writing systems embraced by non-Latin languages are accommodated by Romanization and transcription relationships, which are newly added to the relationships defined in FRAD. In addition, although the “alternate script relationship” of the original FRAD model includes both Roman and non-Latin transliterations, this refers only to non-Latin transliteration relationships in the modified model. Moreover, to circumvent the vague definition of transliteration in FRAD, the modified model makes the attribute “transliteration scheme” obsolete. Instead, the “Romanization scheme” is established in its place. The “Romanization scheme” attributes are assigned to the CAP based on the name, not “to” name. The entity name has a “language” attribute that is inherited by the CAP and any derived CAPs.

The original FRAD model assigns equal status to all CAPs. However, in the modified model, CAPs may arise from several writing systems of non-Latin languages; such CAPs should be treated as a child of parent entities that assist in name or person identification. In the modified model, a single CAP is assigned a single script form and a Romanization scheme (if related by the Romanization relationship). However, some traditional Chinese characters differ by region; for example, “衛” in Taiwan becomes “衞” in Hong Kong. Such cases could be accommodated by adding a region code to the script code, as implemented in the DCMI abstract model. In addition, several patterns may be
implemented by a given Romanization scheme, such as in Japanese Hepburn Romanization. This complication is best resolved by unifying the Romanization schemes. At present, however, we must individually name each Romanization scheme, such as “NDL Hepburn” (the system adopted by NDL) and “ALA-LC Hepburn.”

The modified model only amends attributes and relationships of the original FRAD model regarding non-Latin languages, thus it still can be applied to names in Western languages. All entities, attributes, and relationships in Figure 2-1 and 2-2 remain unchanged under the modified model. Because Russian is a non-Latin language, relationships in Figure 2-3 will be changed to those depicted in Figure 10-5.

A concrete authority data format based on this model is tried to be developed in Chapter 11. In 2013, the Bibliographic Framework Initiative suggested replacing MARC21/A with a new authority data framework, On BIBFRAME Authority. This move is currently under discussion. The present author suggest incorporating the tenets of the new model (i.e., adequate attributes and relationships for non-Latin languages) into the new authority data format, thus facilitating its international use. In the proposed model, there are two types of CAPs; the original CAP (i.e., the parent CAP) and its derivations (i.e., the child CAPs). Both original and derived information is needed for name or person identification. In the new format, each language, script, and Romanization scheme should be encoded, and the parent-child relationship between CAPs related through Romanization, non-Latin transcription, and alternate script (i.e., non-Latin transliteration) should be expressed.

Figure 10-5 The modified FRAD model for the Controlled Access Point “Гоголь, Николай Васильевич.”
10.5 Chapter conclusion

This Chapter proposes a modification of the FRAD model for matching personal names in non-Latin languages as well as names in Western languages. Attributes and relationships of the FRAD model are modified for clarification based on the definition of ISO 5127:2001. Because the proposed model only modifies attributes and relationships regarding representations of names in non-Latin languages, the model retains high compatibility with the original FRAD model. The ambiguous definitions of attributes and relationships in the FRAD model, which limit their applicability to names in non-Latin languages, are resolved in the modified model. This new model clearly illustrates two types of CAPs; those directly based on names and those derived from other CAPs. The latter should complement the data in their parent CAPs to assist identification. Thus, the parent and the child CAPs should be handled as a pair. The proposed model is universal in that it accommodates both names in non-Latin languages and names in Western languages. The attributes and relationships proposed in this modified FRAD model should be considered in the construction of the new authority data format.

Notes

1 Part of this chapter has already been published as Kimura, Maiko. A modification of the FRAD model for personal names in non-Roman languages. Journal of Documentation. 2015, (accepted).
10 Lee, Thomas H. The development of CJK bibliographic databases in North America and East Asia.
Chapter 11

Trial on creation of new authority data formats

This chapter aims to investigate the feasibility of the modified FRAD model proposed in Chapter 10. To implement the modified FRAD model, the author proposes two authority formats: (1) modified MARC 21 Format for Authority Data (MARC 21/A) and (2) RDF/XML. Because MARC 21/A is widely accepted not only by the Western libraries but also libraries in the Chinese character cultural sphere, revising the format based on the modified FRAD model will help libraries understand and accept the main characteristics of the model. Adopting the modified MARC 21/A is the easiest way to adopt the concept of “representation” into their authority data. In addition, a trial on the creation of sample records in RDF/XML format, which is widely recognized as newer bibliographic and authority formats, is also conducted. Although all examples are for personal names, both formats also can be used for corporate names in the same way.

11.1 Modified MARC 21/A

11.1.1 Modification of MARC 21/A

As reviewed in Chapter 2, MARC 21/A has two models that could record authority data in multiple scripts: the vernacular and transliteration model (Model A) and the simple multiscript records model (Model B). While Model A can show relationships among multiple scripts of one name using field 880s, Model B merely shows one equivalent script form of an authorized access point using the 7XX heading linking entry field.

In this study, Model A is adopted as the basis of the modified MARC 21/A format because it can link authorized access points and variant access points to each of their multiple representations. An example of authority data for Haruki Murakami ("村上春樹") in Model A is shown in Figure 11-1.

Here, the Japanese Kanji form is an authorized access point and both its corresponding yomi and Romanized forms are in field 880 with “$6100-01.” In the first 400 field (see from reference), the Russian Cyrillic form of the name is shown, and its corresponding Romanized form is in the third 880 field. The second 400 field is a Cyrillic form in Ukrainian, and its corresponding Romanized form is in the fourth 880 field. The third 400 field is in Greek. It may have a Romanized form, but in this example, it does not have a corresponding Romanized form and therefore no link with 880 is provided.

A defect of Model A, which was pointed out in Chapter 2, is that the script identification codes only distinguish limited kinds of scripts. In particular, it cannot distinguish among scripts in
Chinese-Japanese-Korean (CJK) languages. In addition, by recording authority data according to the modified FRAD model proposed in the previous chapter, MARC 21/A cannot express the difference between transliteration, transcription, Romanization relationships, and Romanization schemes. It also cannot distinguish between an English name and the Romanized form of an original form of the name imposed by libraries.

To solve these problems, the Internet Engineering Task Force (IETF) language tag defined by RFC 5646, which is adopted by the DCMI Abstract Model (DCAM), could be introduced. As reviewed in Chapter 2, a language tag consists of subtags such as language codes, script codes, and region codes. Because valid subtags registered in the Language Subtag Registry, which is maintained by the Internet Assigned Numbers Authority (IANA), covers almost all languages and scripts worldwide, adopting the language tag will increase the availability of many kinds of languages and scripts in access points, compared to using existing codes only available for Arabic, Latin, Chinese/Japanese/Korean, Cyrillic, Greek, and Hebrew.

However, the IETF language tag cannot distinguish between forms recorded in the same script, for example, names in katakana (e.g., "ウメマツ, カヲル") and its yomi (e.g., "ウメマツ, カオル"), as well as two different Romanized forms in Latin scripts. The former differences can be expressed using transcription relationships, and the latter differences can be expressed using Romanization schemes. Thus, the author defines relationship identification codes for Romanization, non-Latin transliteration, and non-Latin transcription and their parent access point, as well as codes for Romanization schemes for Romanization.

Figure 11-2 is the modified data of Figure 11-1. The subfield code $9 is newly defined to express language tags, relationship identification codes, and Romanization scheme codes. The syntax of $9 is a language tag followed by “/” with a relationship identification code followed by a hyphen and a Romanization scheme code if available. The relationship identification codes
Figure 11-2 Modified authority data of Figure 11-1

Table 11-1 Relationship identification codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>pt</td>
<td>A parent of a transliteration/transcription/Romanized form</td>
</tr>
<tr>
<td>ro</td>
<td>A Romanized form (a child)</td>
</tr>
<tr>
<td>tl</td>
<td>A non-Latin transliteration form (a child)</td>
</tr>
<tr>
<td>ts</td>
<td>A non-Latin transcription form (a child)</td>
</tr>
</tbody>
</table>

Table 11-2 Romanization scheme codes for CJK languages (provisional)

<table>
<thead>
<tr>
<th>Language (Mandarin)</th>
<th>Romanization scheme</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Hanyu pinyin</td>
<td>pinyin</td>
</tr>
<tr>
<td></td>
<td>Wade-Gales</td>
<td>wg</td>
</tr>
<tr>
<td></td>
<td>NDL</td>
<td>ndl</td>
</tr>
<tr>
<td></td>
<td>Keio University Libraries</td>
<td>keio</td>
</tr>
<tr>
<td>Korean</td>
<td>McCune-Reischauer</td>
<td>mr</td>
</tr>
<tr>
<td></td>
<td>The South Korean Ministry of Culture and Tourism (2000)</td>
<td>mct</td>
</tr>
</tbody>
</table>
are of four kinds: “ro” for Romanized forms, “tl” for non-Latin transliteration forms, “ts” for non-Latin transcription forms, and “pt” for parent forms of transliteration/transcription/Romanized forms, as shown in Table 11-1. For Romanized forms, a Romanization scheme code can be additionally recorded. There is no exhaustive list of Romanization schemes worldwide; thus, the author attempted to prepare a provisional list of Romanization schemes for CJK languages and generated Romanization scheme codes for these schemes, as shown in Table 11-2. Although there are many other Romanization schemes for CJK languages such as Cantonese, they are omitted from Table 11-2 because, to the best of the author’s knowledge, no library in CJK countries adopts such schemes for their authority data. Note that Romanization scheme codes are only available for names of “imposed Romanization” given by libraries. A common English name, which is not derived from the original form of the name, should be recorded in 4XX with no link to field 880. It should also be noted that $9 is not a mandatory element. The important point is that the information about representations can, not must, be recorded by means of this augmentation. For example, the third 400 field of Figure 11-2 does not have a relationship identification code. This is also allowed in this format if the Romanized form of Greek does not need to be recorded.

11.1.2 Examples of modified MARC 21/A records

Several examples of authority data using modified MARC 21/A format (i.e., using $9) are provided below. Some examples correspond to examples of the modified FRAD model given in Chapter 10. It looks like a very complicated format. It is worth noting again that recording the language, script, and relationship identification codes is optional, and the codes do not need to be recorded for all access points. It would be a great help for authority data sharing if even only one library recorded these codes for each access point, because it could be said that if a parent access point of one database and a normal (not using the parent-child relationship) access point of another database are identified through a matching process, the probability of consistency of two access points is higher than a pair consisting of a child access point of one database and a normal access point of another database. Libraries of each region could share the tasks of recording these codes for access points for persons and corporate bodies in their own regions. For example, NDL conducts this only for Japanese names.

Figure 11-3 is another example of authority data for Haruki Murakami, with Chinese and Korean representations.

The direction of Romanization, transliteration, and transcription should be carefully considered in this format. For example, “Cunshang, Chunshu” is a Romanization of “村上春樹” but the reverse is not true. To show that “村上春樹” is a parent access point and “Cunshang, Chunshu” is a child, “村上春樹” (the second 400 field of Figure 11-3) has the relationship identification
code “pt.” Similarly, “촌상춘수” is a transliteration of “村上春樹”; thus, “촌상춘수” has the relationship identification code “tl” and its parent form “村上春樹” (the fifth 800 field) has “pt.”

“무라카미 하루키” is a transcription of “村上春樹” based on Japanese pronunciation. In Chapter 10, the author argued that the original form of a name and its non-Latin transcription have a parent-child relationship and should be shown as a pair. However, in South Korea, Kanji (or Hanja) representations of Japanese names are rarely used; only their Hangul transcriptions such as “무라카미 하루키” are used, and Hangul representations are not considered to be derived from Kanji. Similarly, in South Korea, Chinese names are described in Hangul based on Chinese pronunciation, but Korean people do not think Hanzi (or Hanja) representations should be shown in addition to Hangul. The author believes that transcriptions of foreign names (from a Korean person’s viewpoint) should be considered as the common “Korean name”, and no relationship with Kanji or Hanzi needs to be expressed in the authority data. Thus, “무라카미 하루키” in Figure 11-3 has no link to field 880. However, the author does not preclude that an organization handles “무라카미 하루키” as a transcription of “村上春樹.”

Figure 11-3 An example of authority data for Haruki Murakami with CJK representations

There may be several Hangul transcriptions for non-Korean names. For example, “草彅剛 (Tsuyoshi Kusanagi),” who is a Japanese pop star, can be transcribed as “쿠사나기 츠토시”, “구사나기 쓰요시”, and “구사나기 쓰요시”. Although the National Institute of the Korean Language issued 외래어표기법 (Foreign word notations), which defines transcription tables of several languages into Korean Hangul based on pronunciations, it is not always used by the Korean public. The fact that the Korean version of Wikipedia and NAVER personal search adopt “쿠사나기 츠토시”, while “구사나기 쓰요시” is the correct form according
to 외래어표기법, provides evidence for this finding. For transcriptions of non-Korean names, all forms should be in 400 fields as common “Korean names”, and each “Korean name” could, but not necessarily, be linked to Japanese Kanji forms in 880 fields. This situation may also arise in data for non-Japanese names created in Japan, such as “Robert Louis Balfour Stevenson” for “スチーブンソン”, “スティーブンソン”, and “スティーヴンスン”, and others. These Japanese transcriptions of non-Japanese names are also viewed as common “Japanese names” and should be recorded in 400 fields in this study.

In Figure 11-3, the access point in the 100 field, the second 400 field, and the last 880 field look like the same representation, even though their languages are different. However, in this study, the duplication could not be eliminated, because if they were merged, as in Figure 11-4 (an errant example), it would be difficult to identify the Japanese Romanization form “Murakami, Haruki” as being based on the Japanese yomi form “ムラカミ, ハルキ”. In other words, the relationship between Japanese Kanji, yomi, and its Romanized form would be invisible in Figure 11-4.

Some Japanese people have several yomi for one Kanji name. For example, 藤原定家, who was a Japanese poet of the early Kamakura era, has several yomi such as “フジワラサダイエ”, “フジワラテイカ”, “フジワラノサダイエ”, and “フジワラノテイカ.” In this case, although his name is written in only one way in Kanji, the second (the same) Kanji form should also be recorded as a variant access point to make a set of three with its second yomi form and Romanized form of the second yomi form, as shown in Figure 11-5. If the second Kanji form is omitted and all forms are recorded in 880 fields with “$6100-01,” it would be difficult to identify which Romanized form came from which yomi form. To avoid such confusion, one Kanji form should have only one yomi form.

100  1#  $6880-01$9/pt$Sa 村上, 春樹,$d1949-
400  1#  $6880-02$9zh-Hans/pt$Sa 村上春樹, $d1949-
400  1#  $9ko-Hang$Sa 무라카미 하루키,$d1949-
880  1#  $6100-01$9ja-Kana/ts$Sa ムラカリ, ハルキ,$d1949-
880  1#  $6100-01$9ja-Latn/ro-lc2012ja$a$SaMurakami, Haruki,$d1949-
880  1#  $6100-01$9zh-Latn/ro-pinyin$SaCunshang, Chunshu,$d1949-
880  1#  $6100-01$9 ko-Hang$tl$Sa 촌상춘수,$d1949-
880  1#  $6400-02$9zh-Latn/ro-pinyin$SaCunshang, Chunshu,$d1949-

Figure 11-4 An errant authority data for Haruki Murakami
Figure 11-5 An example of data which has two *yomi* forms for one *Kanji* form

In Western libraries, authorized access points should be in Latin alphabets. Thus, the authority data in Figure 11-3 may be revised to result in those of Figure 11-6. The three access points, namely, the 100 field and the first and the second 880 fields, are triple because they are linked using $6$. Note that the second 880 field "ムラカミ, ハルキ" is not a transcription of the 100 field, but a transcription of the first 880 field because it has the relationship identification code "pt." Of course, the English name of "Murakami, Haruki" can be added to the record. If the English name were an authorized access point, the data would be as shown in Figure 11-7.

An example of a Chinese personal name is shown in Figure 11-8. Access points are partially derived from The LC/NACO Authority File (LCNAF), the HKCAN Database OPAC (HKCAN), the CALIS Union Catalog Authorities (CALIS), and the Web NDL Authorities (NDL). The fifth 400 field “Sun, Chʻing-ling Sung” is merely a different sequence version of the last 880 field, which handles her maiden name as a middle name. Obviously, “Sun, Chʻing-ling Sung” is derived from “孫宋慶齡”, but because the order of the name is different, it is assumed to be an
Figure 11-7 An example of authority data with an English name in the 100 field

100 1# $9en $aMurakami, Haruki, $d1949-
400 1# $6880-01$ja-Latn/ro-lc2012ja $aSaMurakami, Haruki, $d1949-
400 1# $6880-02$zh-Latn/ro-pinyin $aCunshang, Chunshu, $d1949-
400 1# $6880-03$zh-Latn/ro-pinyin $aCunshang, Chunshu, $d1949-
400 1# $9ko-HangSa 무라카미 하루키, $d1949-
400 1# $6880-04$9ko-Hani/ptSa 村上春樹, $d1949-
880 1# $6400-01$ja-Hani/pt $a村上, 春樹, $d1949-
880 1# $6400-01$ja-Kana/ts $aムラカミ, ハルキ, $d1949-
880 1# $6400-01$ja-Latn/ro-le2012ja $aSaMurakami, Haruki, $d1949-
880 1# $6400-02$zh-Hant/ptSa 村上春樹, $d1949-
880 1# $6400-03$zh-Hans/ptSa 村上春樹, $d1949-
880 1# $6400-04$9ko-Hang/tlSa 촌 상춘수, $d1949-

Figure 11-8 An example of authority data for Song Qingling with CJK representations

100 1# $6880-01$zh-Hant/ptSa 宋慶齡, $d1893-1981
400 1# $6880-02$zh-Hans/ptSa 宋慶齡, $d1893-1981
400 1# $6880-03$zh-Hant/ptSa 宋慶齡, $d1893-1981
400 1# $6880-04$zh-Hant/tlSa 宋慶齡, $d1893-1981
400 1# $9en$SaSun, Chʻing-ling Sung, $d1890-
400 1# $9en$SaSoong, Ching-ling, $d1893-1981
400 1# $9en$SaSun Soong, Ching-ling, $d1893-1981
400 1# $9en$SaSun, Yat-sen, ScMme, $d1893-1981
880 1# $6100-01$zh-Hant/ro-pinyin $aSong, Qingling, $d1893-1981
880 1# $6100-01$zh-Hant/ro-wg $aSong, Chʻing-ling, $d1893-1981
880 1# $6400-02$zh-Hant/ro-pinyin $aSong, Qingling, $d1893-1981
880 1# $6400-03$zh-Hant/ro-wg $aSong, Chʻing-ling, $d1893-1981
880 1# $6400-03$9ja-Hani/ptSa 宋慶齡, $d1893-1981
880 1# $6400-03$9ja-Latn/ro-ngdlSo $aKeirei, $d1893-1981
880 1# $6400-04$9ko-Hani/ptSa 宋慶齡, $d1893-1981
880 1# $6400-05$9zh-Hant/ro-pinyin $aSong, Qingling, $d1893-1981
880 1# $6400-05$9zh-Hant/ro-wg $aSong, Chʻing-ling, $d1893-1981
English name here. The first 880 field “Song, Qingling” is the authorized access point for her in LCNAF. It is possible to exchange the first 880 field for the 100 field by changing $6. The triple set of “宋慶齡”, “Song, Qingling,” and “Sung, Chʻing-ling” still remains because $6 links them, and the relationship identification code still shows “宋慶齡” as a parent among them. It can be said that there is a transliteration relationship between the 100 field and the first 400 field “宋慶齡”; however, the relationship does not need to be shown because traditional and simplified Chinese scripts are two independent writing systems.

Examples of Korean personal names are shown in Figures 11-9 and 11-10. Access points in Figures 11-9 and 11-10 are derived from LCNAF, HKCAN, and NACSIS-CAT. Because “이광수” is a transliteration of “李光洙”, the latter has the relationship identification code “pt.” Korean people, however, may feel “이광수” is a more important access point. Thus, “이광수” is an authorized access point in Figure 11-9. Again, “I, Gwangsu” or “Yi, Kwang-su” can also be authorized access points, with “李光洙” as a parent access point. If a Chinese character form “李光洙” is unknown, “이광수” will be a parent access point instead.

100 1# $6880-01$ko-Hang/tl$a 이광수,$d1892-1950
400 1# $6880-02$ja-Hani/pt$a 李光洙,$d1892-1950
400 1# $6880-03$ja-Hani/pt$a 李光洙,$d1892-1950
400 1# $6880-04$ko-Hang/tl$a 리광수,$d1892-1950
400 1# $9en$a Lee, Kwang Soo
400 1# $6880-05$zh-Hani/pt$a 李光洙,$d1892-1950
880 1# $6100-01$ko-Hani/pt$a 李光洙,$d1892-1950
880 1# $6100-01$ko-Latn/ro-mct$sal, Gwangsu,$d1892-1950
880 1# $6100-01$ko-Latn/ro-mr$sa Yi, Kwang-su,$d1892-1950
880 1# $6400-02$ja-Kana/pt$t$sal, 및, 코우시,$d1892-1950
880 1# $6400-02$ja-Latn/ro-ndl$sal, Guansu,$d1892-1950
880 1# $6400-03$ja-Kana/pt$t$sal, り, コウシュ,$d1892-1950
880 1# $6400-03$ja-Latn/ro-ndl$sal, Koshu,$d1892-1950
880 1# $6400-04$en$a Li, Guangzhu,$d1892-1950
880 1# $6400-05$zh-Latn/ro-pinyin$sal, Guangzhu,$d1892-1950

Figure 11-9 An example of authority data for Yi Kwang-su with CJK representations

“이광수” is another transliteration of “李光洙” used in North Korea, where the beginning-sound rules are not applied. For “이광수,” another parent access point “李光洙” is needed because it is similar to the relationship of Kanji and yomi, one Hanja parent could have
only one _Hangul_ form, although a _yomi_ or a _Hangul_ form could have several Romanized forms as differences of Romanization schemes. If one parent _Hanja_ form had various _Hangul_ forms, shown as Figure 11-10 (an errant example), it would be unclear whether the Romanization form “I, Gwangsu” is derived from “이광수” or “리광수.” In fact, it is obviously derived from “이광수.” To confirm this, it is important to recall the principle that one parent should have only one non-Latin transliteration/transcription form, even though the number of Romanized forms that a parent could have is unlimited.

Figure 11-10 An errant authority data for Yi Kwang-su (1)

Figure 11-11 An errant authority data for Yi Kwang-su (2)

“이, 구언스” is a transcription of “李, 光洙” based on Korean pronunciation, while “리, 코우슈” is a transcription of “李, 光洙” based on Japanese pronunciation. Recently, transcriptions based on Korean pronunciation are preferred, especially by Korean people, while Japanese people do not understand that “이, 구언스” is derived from “李, 光洙” without _furigana_ representations on the materials to be cataloged. Therefore, whether “이, 구언스” and “李, 光洙” should be shown as a pair is a point to consider. Similar to “무라카미 하루키”, “이, 구언스” might be considered a common “Japanese name” that should be shown independently from its Hanja representation. In Figure 11-9, however, both transcriptions are linked to their parents “李, 光洙” because linking “이, 구언스” and “李, 光洙” might be helpful in making it clear to Japanese users that “이, 구언스” is derived from “李, 光洙.” Again, each transcription should be linked to separate parent access points, because each parent access point can have only one non-Latin transliteration/transcription. If the parent access point
were shared, as shown in Figure 11-11 (an errant example,) it would be difficult to know which Romanized form came from which transcription (yomi) form.

Figure 11-12 is an example of a Korean name that does not have representations in Hanja. In this case, the Hangul form “김하늘” is a parent access point in Korean representations. Because the name is not represented in Hanja, a Kanji representation in Japanese does not exist. Instead, a katakana form (a Japanese transcription of “김하늘”) provides a parent access point. “金荷娜” is a Chinese transcription of “김하늘.” It is her “Chinese name” and not an original Hanja form of “김하늘.”

An example of a Vietnamese personal name is shown in Figure 11-13. According to the IANA Language Subtag Registry,1 suppress-script of Vietnamese is “Latn,” which means most Vietnamese texts are written in Latin alphabets; thus, the script code does not need to be recorded in the language tag.6 However, the possibility of using Chinese characters in Vietnamese still remains; therefore, “vi-Latn” and “vi-Hani” are used for each access point in Vietnamese.

\[
\begin{align*}
100 & 1# \ 6880-01\$ko-Hang/ptSa \ 김하늘, \$d1978-
400 & 1# \ 6880-02\$ja-Kana/ptSa \ キム, \ハヌル, \$d1978-
400 & 1# \ 6880-03\$zh-Hant/ptSa \ 金荷娜
880 & 1# \ 6100-01\$ko-Latn/ro-mctSaKim, \ Haneul, \$d1978-
880 & 1# \ 6100-01\$ko-Latn/ro-mrSaKim, \ Ha-nūl, \$d1978-
880 & 1# \ 6400-02\$ja-Latn/ro-ndlSaKimu, \ Hanuru, \$d1978-
880 & 1# \ 6400-03\$zh-Latn/ro-pinyinSaJin, \ Hena
\end{align*}
\]

Figure 11-12 An example of authority data for Kim Ha-nūl with CJK representations

Whether Chữ quốc ngữ is “Latn” should be considered. Chữ quốc ngữ is written in the Latin alphabet with diacritics. Many languages, such as French and German, are written in Latin alphabets with diacritics and their scripts are equal to “Latn.” However, in Vietnamese, diacritics appear much more frequently than in French or German, and, without diacritics, the name might be completely different. In this study, the author uses the tag “Latn” for Chữ quốc ngữ representations, based on precedent, but the relationship between the 100 field and the first 880 field “阮攸” is assumed to be a transcription, not a Romanization, because Romanization suggests an imposition by libraries in this study, which is not the case.

As “グエン, ズー” is a Japanese transcription of “阮, 攸”, they can also be handled as a pair. However, in Figure 11-13, “グエン, ズー” is assumed to be a separate access point from “阮, 攸,” because some materials in Japanese do not show Vietnamese names with Chinese
characters. Because “阮攸” is shared by both traditional and simplified Chinese characters, the script code of the third 400 field is “Hani (Chinese characters),” not “Hant (traditional Chinese characters)” nor “Hans (simplified Chinese characters).” “응우옌주” is a Korean transcription of “Nguyễn, Du” but assumed to be a “Korean name” that is distinguished from other representations.

Figure 11-14 is an example of a more current Vietnamese name. It is not expressed in Chinese characters, whereas it is represented as “楊秋香” in Taiwan. Whether “杜兰, Thu Hurlong” is in fact derived from “楊秋香” in Vietnamese is indeterminable. A difference is evident between small “オ” and “オ” in two katakana names in Japanese, which, though minor, results in different Romanization; thus, they are assumed to be two different “Japanese names.”

In this section, examples of authority data of names in the Chinese character cultural sphere are shown in the format of the modified MARC 21/A. Names are recorded in the format using IETF language tags and relationship identification codes, which are newly defined by the author. The modified MARC 21/A format successfully shows the parent-child relationship between access points regardless of whether it is an authorized or variant access point. The format also leaves a margin of choice for each library as to whether it should be in field 4XX or 880 for some access points, i.e., an independent access point or a child access point of a parent access point. For example, whether “イ，グアンス” (the transcription form of “李，光洙”) should be in field 400 or 880 could be decided by each library. The grounds for this choice may be whether showing the parent-child relationship of these access points is useful for the user of the library; it may differ by each region and country.

100 1# $6880-01$vi-Latn/ tsSaNguyễn, Du,$d1765-1820
400 1# $6880-02$ja-Hani/ptSa 阮, 攸,d1765-1820
400 1# $6880-03$ja-Kana/ptSa グエン, ズー,$d1765-1820
400 1# $6880-04$zh-Latn/ro-pinyin$aRuan, You,$d1765-1820
400 1# $6880-05$ko-Hani/ptSa 완유,$d1765-1820
880 1# $9ko-Hang$s응우옌주,$d1765-1820
880 1# $6100-01$vi-Hani/ptSa 阮攸, $d1765-1820
880 1# $6400-02$ja-Kana/tsSa ゲン, ユウ,$d1765-1820
880 1# $6400-03$ja-Latn/ro-ndl$sGen, Yu,$d1765-1820
880 1# $6400-04$zh-Latn/ro-pinyin$sRuan, You,$d1765-1820
880 1# $6400-05$k-Hani/ptSa 阮攸,$d1765-1820

Figure 11-13 An example of authority data for Nguyễn Du with CJKV representations

179
The parent-child relationship helps author identification by machines, because if an authority record includes the parent-child relationship, the machine could determine that the parent access point is more important, and if the parent access point matches another access point from another data source, the possibility that these two access points indicate the same entity is pretty high, while matching a child access point may decrease the possibility.

The language and script tags also increase the accuracy of matching if both the object and subject data source for matching records these tags. Tags of Romanization schemes also have this function. Besides, Romanization schemes indicate from which sources the access point originates. This information is more meaningful for people than machines, especially when many Romanization forms for one name exist. Thus, the availability of a Romanization scheme enables people to judge which Romanization form is more suitable for use in reference lists, websites, and others.

11.2 The Resource Description Framework (RDF)

Current discussions involve the search for a new format to replace MARC. BIBFRAME initiated by LC provides “a foundation for the future of bibliographic description” and is expected to replace the MARC 21 Bibliographic Format. BIBFRAME also includes BIBFRAME Authorities (currently in draft form). However, BIBFRAME Authorities “are not designed to replace or compete with existing authorities but rather to provide a common abstraction layer, or wrapper, around them.” BIBFRAME Authorities attempt to describe authority data using RDF. Therefore, supplying authority information using other ontology such as MADS/RDF is allowed in BIBFRAME Authorities. In other words, describing authority data in RDF is nearly equal to describing authority data in the context of BIBFRAME Authorities.
The author therefore tried to describe authority data in RDF/XML format based on the modified FRAD model, using the definitions of new vocabularies regarding “representation” since no other vocabularies including BIBFRAME successfully describe representations of names or access points.

11.2.1 Class and property newly defined

RDF/XML format is already adopted by Web NDL Authorities as well as RDF/Turtle and JSON formats. In Web NDL Authorities, vocabularies such as SKOS-XL, SKOS, DC-NDL are used. For example, Figure 12 is an authority record in RDF/XML format derived from Web NDL Authorities.

In Web NDL Authorities, an authorized access point is represented by using “skosxl:prefLabel,” and variant access points are represented by “skosxl:altLabel.” These terms are defined as extensions to the Simple Knowledge Organization System, called the SKOS eXtension for Labels (SKOS-XL). The original forms of names are expressed using “skosxl:literalForm,” and to express yomi of Japanese, “dcndl:transcription,” which is one of the properties of the DC-NDL vocabulary, is adopted. The preferred and alternative labels consist of a string of Unicode characters and an optional language tag defined by RFC 5646.

Using an RDF visualization tool called MR3, the RDF graph for the example in Figure 11-15 shown in Figure 11-16 could be described. In this RDF graph, only parts of authority data regarding representations are visualized and data elements such as date of birth are omitted.

Although language tags are not shown in Figure 11-16, plain literal data such as “村上，春樹, 1949-” and “Murakami, Haruki, 1949-” have a language tag as shown in Figure 11-15 with the “xml:lang” attribute. In Figure 11-16, we can see that “ndl:transcription” combines transliteration, transcription, and Romanization into one “ndl:transcription” property, but it is not enough to express the modified FRAD model proposed in this study. Therefore, this study develops new properties that express non-Latin transliteration, non-Latin transcription, and Romanization, respectively, and also proposes a way to show a Romanization scheme for the “Romanization” property.

First, two new classes, namely, “ex:Representation” and “ex:RomanizationScheme” are defined. “ex:Representation” is a class that represents variations of notational representations. It is a subclass of “rdfs:Literal.” On the other hand, “ex:RomanizationScheme” is a class representing Romanization schemes. Secondly, new properties of the class “ex:Representation” are defined, as shown in Table 11-3. Table 11-3 was designed in accordance with the simplified Description Set Profile (DSP), which was defined in the guideline of sharing metadata information published by the Metadata Information Infrastructure Initiative in 2011.
Figure 11-15 An authority data of Haruki Murakami derived from Web NDL Authorities

182
Figure 11-16 The RDF graph of NDL's representations of Haruki Murakami
<table>
<thead>
<tr>
<th>Statement name</th>
<th>Property name</th>
<th>Minimum Occurrence Count</th>
<th>Maximum Occurrence Count</th>
<th>Value type</th>
<th>Value constraint</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transliteration</td>
<td>ex:hasTransliteration</td>
<td>0 -</td>
<td></td>
<td>Strings</td>
<td></td>
<td>Has non–Latin transliterated representation.</td>
</tr>
<tr>
<td>Transcription</td>
<td>ex:hasTranscription</td>
<td>0 -</td>
<td></td>
<td>Strings</td>
<td></td>
<td>Has non–Latin transcripted representation.</td>
</tr>
<tr>
<td>Romanization</td>
<td>ex:hasRomanization</td>
<td>0 -</td>
<td></td>
<td>Structure</td>
<td></td>
<td>Has Romanized representation.</td>
</tr>
<tr>
<td>Representations</td>
<td>ex:hasRepresentation</td>
<td>0 -</td>
<td></td>
<td>Structure</td>
<td>ex:hasParent ex:hasTransliteration ex:hasTranscription ex:hasRomanization</td>
<td>Has representation(s) other than the authorized access point (use when the instance of “xl:literal” from “xl:prefLabel” is not a parent access point.)</td>
</tr>
<tr>
<td>Parent</td>
<td>ex:hasParent</td>
<td>0 -</td>
<td></td>
<td>Strings</td>
<td></td>
<td>Has a parent representation.</td>
</tr>
<tr>
<td>[Structured Romanization]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romanized strings</td>
<td>ex:hasRomanizedLiteral</td>
<td>0 -</td>
<td></td>
<td>Strings</td>
<td></td>
<td>Actual character strings which are Romanized from other non–Latin character strings. Subproperty of “ex:hasRomanization.”</td>
</tr>
<tr>
<td>Romanization scheme</td>
<td>ex:hasRomanizationScheme</td>
<td>0 -</td>
<td></td>
<td>Reference value</td>
<td>Value defined in the table of Romanization scheme*</td>
<td>Romanization Scheme. Subproperty of “ex:hasRomanization.”</td>
</tr>
</tbody>
</table>

Note. *Currently does not exist, examples are shown in Table 11-2.
As a result of adopting new classes and properties in Table 11-3, the RDF graph shown in Figure 11-16 could be revised to that in Figure 11-17. To express Romanization schemes in RDF graph, one more blank node should be inserted between the arc “ex:hasRomanization” and “ex:hasRomanizationScheme.” The Romanization form “Murakami, Haruki, 1949-” is recorded using the property “ex:hasRomanizedLiteral,” while the Romanization scheme is recorded using “ex:hasRomanizationScheme.” As “ムラカミ，ハルキ，1949-” is a non-Latin transcription form of “村上，春樹，1949-,” the property “ex:hasTranscription” is used.

Authority data of Figure 11-17 in RDF/XML format is shown in Figure 11-18. Again, data included in Figure 11-18 are only representations of access points. In this modified RDF/XML format, access points that have “xl:literalForm” are considered to be parent access points, and thus indicators such as “pt,” which are used in modified MARC 21/A, are not added. As seen in Figure 11-18, modified authority data based on modified FRAD model could be expressed in RDF format. Therefore, it could be recorded and used in the BIBFRAME environment.

Figure 11-17 The modified RDF graph of representations for Haruki Murakami
Figure 11-18 Modified authority data for Haruki Murakami in RDF/XML
11.2.2 Example of RDF/XML records

In this section, authority data examples in the RDF/XML format are shown using newly defined properties. The examples correspond to those in the modified MARC 21/A format in Figures 11-3, 11-5–11-9, and 11-12–11-14 of this chapter.

Figure 11-19 is an example of authority data for Haruki Murakami (expressed in MARC 21/A format in Figure 11-3) in RDF format, with CJK representations.

The RDF graph for Figure 11-19 is shown in Figure 11-20. In general, in this RDF/XML format, an access point with “xl:prefLabel” is an authorized access point, and access points with “xl:altLabel” are variant access points. However, as shown in Figure 11-20, an arc of “xl:prefLabel” is connected to a blank node, and this node is connected to several arcs connected by a literal or another blank node. This means that there is a group of authorized access points in one RDF graph because literals indirectly connected by an arc “xl:prefLabel” are varied, such as “村上，春樹,” “ムラカミ，ハルキ,” and “Murakami, Haruki.” Among these, only one access point with “xl:literalForm” is a parent of the group of the authorized access point. Other access points are children of the parent access point. In other words, there is one “authorized” parent access point and several “authorized” child access points. A parent access point is sometimes unequal to an exclusive authorized access point, such as field 100 of MARC 21/A format, especially in Western libraries. In this RDF format, only one exclusive authorized access point is not designated. Therefore, the RDF graph of authority data in MARC 21/A format shown in Figure 11-6 is similar to Figure 11-20.

If a name in the Latin alphabet is an “English name,” it could only be one authorized access point in the RDF graph, as in Figure 11-21, which shows the same authority data as in Figure 11-7. The authority data in RDF/XML format is as shown in Figure 11-22.

Figures 11-23–11-27 are authority data of Chinese, Korean, and Vietnamese in RDF/XML format corresponding to data in MARC 21/A format in Figures 11-8–11-14 (except Figures 11-10 and 11-11), respectively.

From the examples shown in this section, it could be said that the modified FRAD model proposed by the author can be expressed in RDF/XML format, and thus it can be used in the BIBFRAME environment. However, three problems need to be resolved.

Firstly, the Uniform Resource Identifier (URI) for each Romanization scheme should be defined. In the examples shown in this section, provisional URIs such as “http://examples.com/pinyin” (shortened to “ex:pinyin” in Figures) are used. Once the URI for each scheme is defined, many organizations can use it, and these data can be linked to each other easily. Therefore, URIs as well as the table of Romanization schemes and their corresponding codes and URIs should be developed.

Secondly, the RDF graphs shown in this study were more hierarchized and complicated than
the RDF graphs of DC-NDL, because Romanization schemes need to be expressed. This complication may cause difficulties in matching various authority data sets. There is room to consider a simpler structure for implementation in the future.

Thirdly, in the proposed format, only one exclusive authorized access point is not defined if an authorized access point has many representations. For Western libraries, however, only one exclusive authorized access point is needed. This issue could be solved in the BIBFRAME environment, because only one exclusive authorized access point can be recorded by using a term “bf:authorizedAccessPoint” in BIBFRAME Authority.

<?xml version="1.0"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:ex="http://example.com/
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:xl="http://www.w3.org/2008/05/skos-xl#"
 xmlns:Description rdf:about="http://id.ndl.go.jp/auth/ndl/00104237">
  <xl:prefLabel rdf:parseType="Resource">
    <xl:literalForm xml:lang="ja-Hani">村上春樹, 1949-</xl:literalForm>
    <ex:hasTranscription xml:lang="ja-Kana">ムラカミ, ハルキ, 1949-</ex:hasTranscription>
    <ex:hasRomanization rdf:parseType="Resource">
      <ex:hasRomanizedLiteral xml:lang="ja-Latn">Murakami, Haruki, 1949-</ex:hasRomanizedLiteral>
      <ex:hasRomanizationScheme rdf:resource="http://example.com/lc2012ja"/>
    </ex:hasRomanization>
  </xl:prefLabel>

  <xl:altLabel rdf:parseType="Resource">
    <xl:literalForm xml:lang="zh-Hant">村上春樹, 1949-</xl:literalForm>
    <ex:hasRomanization rdf:parseType="Resource">
      <ex:hasRomanizedLiteral xml:lang="zh-Latn">Cunshang, Chunshu, 1949-</ex:hasRomanizedLiteral>
      <ex:hasRomanizationScheme rdf:resource="http://example.com/pinyin"/>
    </ex:hasRomanization>
  </xl:altLabel>

  <xl:altLabel rdf:parseType="Resource">
    <xl:literalForm xml:lang="zh-Hans">村上春树, 1949-</xl:literalForm>
    <ex:hasRomanization rdf:parseType="Resource">
      <ex:hasRomanizedLiteral xml:lang="zh-Latn">Cunshang Chunshu, 1949-</ex:hasRomanizedLiteral>
      <ex:hasRomanizationScheme rdf:resource="http://example.com/pinyin"/>
    </ex:hasRomanization>
  </xl:altLabel>

  <xl:altLabel rdf:parseType="Resource">
    <xl:literalForm xml:lang="ko-Hani">村上春樹, 1949-</xl:literalForm>
    <ex:hasTransliteration xml:lang="ko-Hang">촌상춘수, 1949-</ex:hasTransliteration>
  </xl:altLabel>

  <xl:altLabel rdf:parseType="Resource">
    <xl:literalForm xml:lang="ko-Hang">무라카미 하루키, 1949-</xl:literalForm>
  </xl:altLabel>
</rdf:RDF>

Figure 11-19 An example of RDF/XML authority data for Haruki Murakami with CJK representations
Figure 11-20 The RDF graph of authority data in Figure 11-19
Figure 11-21 The RDF graph of the authority data in Figure 11-7
Figure 11-22 RDF/XML authority data of Figure 11-21
Figure 11-23 RDF/XML authority data of Song Qingling in Figure 11-8
<?xml version="1.0">  
<rdf:RDF  
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
xmlns:ex="http://example.com/"  
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  
xmlns:xl="http://www.w3.org/2008/05/skos-xl#">  
<rdf:Description>  
<xl:prefLabel rdf:parseType="Resource">  
<xl:literalForm xml:lang="ko-Hani">이광수, 1892-1950</xl:literalForm>  
</xl:prefLabel>  
<xl:altLabel rdf:parseType="Resource">  
<xl:literalForm xml:lang="en">Lee, Kwang Soo</xl:literalForm>  
</xl:altLabel>  
<xl:altLabel rdf:parseType="Resource">  
<xl:literalForm xml:lang="ja-Hani">イグアンス, 1892-1950</xl:literalForm>  
</xl:altLabel>  
<xl:altLabel rdf:parseType="Resource">  
<xl:literalForm xml:lang="zh-Hani">李光洙, 1892-1950</xl:literalForm>  
</xl:altLabel>  
</rdf:Description>  
</rdf:RDF>  

Figure 11-24 RDF/XML authority data of Yi Kwang-su in Figure 11-9
Figure 11-25 RDF/XML authority data of Kim Ha-nûl in Figure 11-12
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://example.com/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xl="http://www.w3.org/2008/05/skos-xl#">
  <rdf:Description>
    <xl:prefLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="vi-Hani">阮攸, 1765-1820</xl:literalForm>
      <ex:hasRomanization rdf:parseType="Resource">
        <ex:hasRomanizedLiteral xml:lang="vi-Latn">Nguyễn, Du, 1765-1820</ex:hasRomanizedLiteral>
      </ex:hasRomanization>
    </xl:prefLabel>
    <xl:altLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="ja-Hani">阮攸, 1765-1820</xl:literalForm>
      <ex:hasRomanization rdf:parseType="Resource">
        <ex:hasRomanizedLiteral xml:lang="ja-Latn">Gen, Yu, 1765-1820</ex:hasRomanizedLiteral>
      </ex:hasRomanization>
    </xl:altLabel>
    <xl:altLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="zh-Hani">阮攸, 1765-1820</xl:literalForm>
      <ex:hasRomanization rdf:parseType="Resource">
        <ex:hasRomanizedLiteral xml:lang="zh-Latn">Ruan, You, 1765-1820</ex:hasRomanizedLiteral>
      </ex:hasRomanization>
    </xl:altLabel>
    <xl:altLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="ja-Kana">ゲン, イウ, 1765-1820</xl:literalForm>
      <ex:hasRomanization rdf:parseType="Resource">
        <ex:hasRomanizedLiteral xml:lang="ja-Latn">Guen, Zu, 1765-1820</ex:hasRomanizedLiteral>
      </ex:hasRomanization>
    </xl:altLabel>
    <xl:altLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="ko-Hani">완유, 1765-1820</xl:literalForm>
      <ex:hasTransliteration rdf:parseType="Resource">
        <ex:hasTransliteratedLiteral xml:lang="ko-Hang">Won-you, 1765-1820</ex:hasTransliteratedLiteral>
      </ex:hasTransliteration>
    </xl:altLabel>
    <xl:altLabel rdf:parseType="Resource">
      <xl:literalForm xml:lang="ko-Hang">응우옌주, 1765-1820</xl:literalForm>
    </xl:altLabel>
  </rdf:Description>
</rdf:RDF>

Figure 11-26 RDF/XML authority data of Nguyễn Du in Figure 11-13
Figure 11-27 RDF/XML authority data of Dương Thu Hương in Figure 11-14
Notes

3 “쿠사나기 츠토시”. 위키백과. 2015-01-28. http://ko.wikipedia.org/wiki/%EC%BF%A0%EC%82%AC%EB%82%98%EA%B8%B0_%EC%8B%A0%EC%9A%94%EC%8B%9C, (accessed 2015-03-31).
4 “쿠사나기 츠토시”. NAVER 인물검색. http://people.search.naver.com/search.naver?where=nexearch&sm=tab_ppn&query=%EC%BF%A0%EC%82%AC%EB%82%98%EA%B8%B0%EC%8B%A0%EC%9A%94%EC%8B%9C&os=96255&ie=utf8&key=PeopleService, (accessed 2015-03-31). NAVER is one of the most popular search engines in Korea.
5 「TRC MARC 人名典拠録」編集部編. 17 人のスティーヴンソン. 図書館流通センター, 1985, 144p.
Chapter 12

Conclusion

The results and findings from this research are as follows:

a) Based on the characteristics of personal names in the Chinese character cultural sphere, three components of name authority data, namely, representations, data elements, and structures, were defined.

b) Representations of name authority data recorded by organizations in Mainland China, Hong Kong, Taiwan, Japan, and South Korea have been examined and clarified. The results can be summarized in the following eight statements:

b-1) Chinese character forms are recorded in letter types, used by each region where each organization is located. This means Chinese character forms are not always an “accurate” form used by a person or corporate body in their native country. For example, a Japanese personal name is not always recorded in Japanese Kanji form in regions outside of Japan.

b-2) Romanized forms for Chinese names are recorded using Hanyu pinyin in all organizations studied, except organizations in South Korea. However, umlaut marks are handled differently depending on the organization, and it may be an obstacle to string matching based on Romanized forms of Chinese names.

b-3) Despite the importance of yomi for Japanese names, it is not recorded by organizations outside of Japan.

b-4) Romanized forms of Japanese names might be different depending on the organization, because the Hepburn Romanization system adopted by each organization is slightly different. It may be an obstacle to identifying Japanese names using Romanized forms among organizations.

b-5) Korean name Romanization systems adopted by organizations in South Korea and in other countries are different. Therefore, identification of Korean names using Romanized forms among organizations is difficult.

b-6) Organizations outside South Korea do not record Hangul forms for Korean names as a mandatory element, excepting one organization in Japan. It may preclude the possibility of identifying Korean names using Hangul forms among organizations.

b-7) In Vietnam, name authority control for author names is not conducted. Although it is desirable that Chinese character forms and Vietnamese forms are recorded together in authority records, such an authority database does not exist.

b-8) In summary, any single type of representation is insufficient as a master key for name identification when name authority data are shared. Rather, the combination of several representations seems to be helpful for name identification.
c) Authority data elements recorded by organizations in the Chinese character cultural sphere are compared to data elements defined by RDA. It was established that core elements defined in RDA were recorded by most organizations. Among non-core elements, field of study, lineage (especially in Japan), gender, place of ancestry (especially in China), nature or character, and history were recorded by many organizations. Some organizations recorded a lot of authority data elements in the note fields, but these elements are not available for data identification. RDA stipulates that some data elements such as Field of Activity, Place of Residence, and Gender should be recorded separately from access points. These elements are, however, recorded as additions to access points in Japan and China.

d) The modified FRAD model, which can represent many kinds of representations of names in non-Latin countries, was proposed. Three kinds of representations were defined: namely, non-Latin transliteration, non-Latin transcription, and Romanization. The necessity of introducing the parent-child relationship into Controlled Access Points was explained.

e) Based on the modified FRAD model, authority data formats, which can adopt the model, were proposed.

Based on the above findings, the author makes the following suggestions:

1) As noted in Chapter 1, many attempts have been made to use identifiers rather than name strings for name authority control. However, for authors already recorded in various authority databases, name identification using name strings is still inevitable because the unification of several authority records should be conducted before assigning a uniform identifier. As this study has shown, representations of names in non-Latin languages, especially in Chinese-Japanese-Korean (CJK) languages, are complicated and, therefore, mechanical name identification cannot achieve a 100% success rate. Constructing more detailed authority data including relationships of representations will contribute to enhancing the accuracy of such mechanical name identification. Libraries should take the central role in constructing rich and valuable authority data, which can be used for various kinds of databases worldwide. In other words, libraries cannot escape such manual work.

2) For people in Western countries, inputting scripts of non-Latin languages is tedious and error-prone. Therefore, they have to spend much time recording Romanized forms for names in non-Latin languages as authorized access points. However, this method is not suitable for the cultures of non-Latin countries. A Romanized form of a name should be expedient for all times; it is not another “name” for the person or the corporate body, but merely a “representation” of the name. In addition, each original form is addressed equally in authority data created by Western countries. In fact, there are relationships between the original forms, and these relationships are useful for name identification. The modified FRAD model proposed by the
present author conveys such relationships and differentiates a Romanized form of a name, which is imposed by libraries, from a real “English” name. Adopting the modified FRAD model for authority and name identification work will facilitate name authority control more precisely.

3) As is the case under BIBFRAME Authority, linking to VIAF appears to be the main trend in authority control. However, as this study shows, VIAF identification is not perfect. The present author is concerned that many applications including library catalogs assume that their authority control has been conducted as long as it links to VIAF. If everyone follows this belief, it is unlikely that anyone will construct precise authority data because authority work is costly. However, to achieve ideal authority control, which fulfills the user’s needs, a precise authority database is needed to which every application should be linked. Although VIAF has the potential to become a central hub for authority data sharing on a global level, manual validations and corrections of authority data will be needed to achieve this.

4) RDF, which was adopted by BIBFRAME Authority, presupposes linking to other resources automatically, using URI references. Some defects will occur during the linking process, because linking is conducted without human confirmation, but these defects are permitted under this technology. Whether this technology is suited to authority control, which should seek 100% precision, is an issue that should be investigated in a future study.