It is commonly believed that the quality of original cataloging on OCLC (Online Computer Library Center) and RLIN (Research Libraries Information Network) significantly differs. Past stereotypes reflected the fact that OCLC traditionally emphasized increasing the size of the database, while RLIN ascribed top priority to cataloging quality. The literature is mixed regarding whether significant differences do exist. Focusing on Russian name headings, this project investigates whether the quality of member-contributed cataloging in this Slavic area does significantly differ on OCLC and RLIN. The study draws member-contributed records containing Russian name headings from both the OCLC and RLIN databases and compares the name headings with respect to correct transliteration, punctuation, and tagging, as well as conformance to AACR2 (Anglo American Cataloging Rules 2) and Library of Congress rule interpretations. (Contains 12 references.) (Author/JLB)
RUSSIAN CORPORATE AND PERSONAL NAME HEADINGS
ON OCLC AND RLIN: A COMPARISON STUDY

A Master’s Research Paper submitted to the
Kent State University School of Library Science
in partial fulfillment of the requirements
for the degree Master of Library Science

by
Melinda Saveleva
November, 1990

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Rosemary Du Mont
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)"
RUSSIAN CORPORATE AND PERSONAL NAME HEADINGS
ON OCLC AND RLIN: A COMPARISON STUDY

It is commonly believed that the quality of original cataloging on OCLC and RLIN significantly differs. Past stereotypes reflected the fact that OCLC traditionally emphasized increasing the size of the database, while RLIN ascribed top priority to cataloging quality. The literature is mixed regarding whether significant differences do exist. Focusing on Russian name headings, this project will investigate whether the quality of member-contributed cataloging in this Slavic area does significantly differ on OCLC and RLIN. Member-contributed records containing Russian name headings will be drawn from both the OCLC and RLIN databases, and a comparison of the name headings will be made with respect to correct transliteration, punctuation, and tagging, as well as conformance to AACR2 and Library of Congress rule interpretations.
Master's Research Paper by
Melinda B. Saveleva
B.A., Ohio State University, 1980
M.A., Ohio State University, 1983
M.L.S., Kent State University, 1990

Approved by
Advisor [Signature] Date [Date]
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of Findings</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>Research Objectives</td>
<td>7</td>
</tr>
<tr>
<td>Areas of Study</td>
<td>8</td>
</tr>
<tr>
<td>Methodology</td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td>9</td>
</tr>
<tr>
<td>Error Schedule</td>
<td>13</td>
</tr>
<tr>
<td>Identification of Errors</td>
<td>14</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>14</td>
</tr>
<tr>
<td>Findings</td>
<td>15</td>
</tr>
<tr>
<td>Conclusions</td>
<td>17</td>
</tr>
<tr>
<td>Bibliography</td>
<td>20</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1 - Occurrence of errors
Table 2 - OCLC Errors by MARC Field
Table 3 - RLIN Errors by MARC Field
PREFACE

I gratefully acknowledge the professional guidance given to me by my advisor, Dr. Ann Allan. She freely volunteered both her expertise and time, and it was a pleasure working with her. I also wish to thank Dr. Lubomyr Wynar for his guidance in the field of ethnic librarianship. Heartfelt thanks are also due Mrs. Marge Markewicz, secretary of the Kent State University School of Library Science, Columbus, and Kui Qui, KSU graduate laboratory assistant, for their support and encouragement.
Summary of Findings

The present study examined 181 matched pairs of catalog records from OCLC and RLIN in an attempt to discover whether any differences exist in the quality of member-contributed cataloging on both networks. Differences in most name heading areas were negligible, with the exception of errors in conformance to AACR2 and Library of Congress rule interpretations: member-contributed cataloging on both databases contained numerous errors in respect to the latter. RLIN records contained 19 name heading errors whereas OCLC records contained 14.

A total of 63 name heading errors of all types were found, an average of .17 errors per record. The 181 OCLC records contained a total of 26 errors, or .14 errors per record, while the 181 RLIN records possessed a total of 37 errors, or .20 errors per record. The majority of errors occurred in the application of AACR2 and LCRIs: 14 errors in OCLC and 19 errors in RLIN. MARC tagging followed in total number of errors (17). OCLC records contained 7 errors in MARC tagging and RLIN 10.

The hypothesis that RLIN records are more accurate than OCLC records was not upheld, at least in terms of name headings: Of the 63 total errors, 37 were attributable to RLIN member-input records. The small sample size, however, renders additional testing necessary to confirm the above results.
Introduction

In the present era of online information retrieval, accurate entry of name headings is a necessity for efficient searching of online bibliographic records. Typos, inaccurate transliteration, inaccuracies in name structure with respect to Library of Congress rule interpretations can all impede the retrieval process.

Accurate and consistent entry of name headings online is perhaps even more crucial than in a manual system, because name headings entered inaccurately will often not be retrievable at all online. In a manual system, slight inaccuracies in spacing, minor typos, etc. will often present no problems as far as retrieval. Errors in non-English language name headings may present special problems (such as transliteration error for languages in non-Roman alphabets). The decision to document name entry errors in Russian records specifically was occasioned by the researcher's experience in the OCLC Retrospective Conversion area and in the Eastern European Studies Department of the Ohio State University Main Library. In the course of daily online searching, both at OCLC and at OSU (using LCS), the author noticed incorrect or inconsistent Russian name headings, inaccuracies that would often have made proper retrieval of records by the inexperienced searcher impossible.

Since both OCLC and RLIN serve as major resources for shared cataloging in the United States and abroad, it is proposed that a comparison of name headings on the above
databases with respect to Russian, in order to systematically
document the quantity and type of errors occurring, would
contribute to improving the quality of Russian records
derived from the two largest sources of shared cataloging in
the U.S. Name heading entries were chosen because of their
significance as a major access point to the online record, in
addition to title and numeric search keys.

In the early years of both OCLC (Online Computer Library
Center) and RLIN (Research Libraries Information Network)
Library of Congress machine-readable cataloging (LC-MARC
records) formed the major portion of both these databases.
(Interne 1989, 3). By 1989, however, the situation had
altered somewhat and member-contributed copy comprised the
greater portion of each database. Quality control became
increasingly important as the proportion of member-
contributed copy vs. LC copy increased. The present study
will primarily focus on and draw its sample from the last 1.5
million records input on OCLC and the corresponding matching
RLIN records, the majority of these items being published in
1989.

In its early years, OCLC was primarily concerned with
building its database, and awarded each institution that
entered a record for the first time a small financial reward,
irrespective of the fullness and accuracy of the record.
Neither entry of a K-level record (one with only the minimum
number of fields) nor entry of an I-level record (a record
containing all fields specified by AACR2 and the network) was
charged if the above records were entered for the first time. OCLC did publish bibliographic standards, but quality control was largely on a voluntary basis. Errors could be corrected by members submitting a report to the network. OCLC cataloging, unlike RLIN, has only one master record for each item in its database. Only the entering library was permitted to alter an I-level record, and was able to do so only if no other institution had attached holdings symbols. LC copy would replace member-contributed copy, as would copy from "Enhance" libraries (libraries with consistently top-notch cataloging that are permitted to correct or improve other institutions' records) (Inter 1989, 4).

RLIN was dependent on voluntary compliance to its cataloging standards as well, but its pricing scale varied with the completeness and accuracy of the record; quality control was a high priority from its inception. The majority of RLIN member libraries were academic and research institutions, dedicated to providing materials for graduate and post-graduate study. Complete and accurate catalog records were particularly important in serving these research needs. OCLC member libraries included many colleges, smaller universities, and public libraries. The users of these libraries were not primarily researchers and placed less importance on complete and accurate catalog records.

Quality control is particularly important with respect to foreign language cataloging, especially in dealing with materials written in non-Roman alphabets. This study will
focus on monographs in Russian, with a view toward improving quality control of name headings on member-contributed records.

There are several factors that one must take into account in researching Russian name headings on the two major bibliographic databases:

1. Does one consider LC copy, member-contributed copy, or both?
2. What errors in name entry are significant, i.e., affect retrieval?
3. How are these errors measured?
4. What implications will such a study have for the library community?

Literature Review

Existing research pertinent to the proposed study falls into three categories: OCLC-RLIN comparison studies, treatment of problems in transliteration, and discussions of AACR2 in relation to Slavic name headings. A search of existing literature did not reveal any previous studies dealing specifically with Russian-langauges name headings on OCLC and/or RLIN.

OCLC-RLIN comparison studies generally focus on one aspect of the services these bibliographic networks provide. Studies comparing OCLC vs. RLIN's usefulness as a reference tool exist. Studies on the comparative cost effectiveness of the above networks, such as "RLIN/OCLC, A Cataloging Cost Study in the Health Sciences Library" (Dailey, Jaroff, and Gray 1982) and hit rate studies such as "Chasing MARC: Searching in Bibliofile, Dialog, OCLC, and RLIN" (Allan 1990)
are also numerous. More germane to the present paper, however, are those comparison studies specifically treating the question of cataloging quality on OCLC and/or RLIN.

Two outstanding studies should be cited in this connection: "Quality in Bibliographic Databases: An Analysis of Member-Contributed Cataloging in OCLC and RLIN" (Intner 1989) and "Accuracy of LC Copy: A Comparison between Copy that Began as CIP and Other LC Cataloging" (Taylor and Simpson 1986). Intner analyzed a group of 215 matched pairs of catalog records contributed by member libraries to OCLC and RLIN in from 1983 to 1989, and concluded that the widespread notion of RLIN's preeminence in cataloging quality was without basis in fact. The Taylor study, although it dealt exclusively with OCLC, was useful from a methodological standpoint: it provided the researcher with guidelines in sampling technique for this type of comparison study.

Little discussion of transliteration problems relevant to East Slavic or Russian was found in the literature, but one study did prove especially relevant. "Establishing Slavic Headings under AACR2" discusses several pertinent transliteration problems (Markiw 1984). For example, one must be careful to differentiate Ukrainian and Russian personal names, particularly in cases when the Ukrainian author's work is published in Russian. One might tend to establish a heading for Russian language material written by a Ukrainian as "Mikhailov, Igor" (using the LC transliteration table for Russian) if one was unaware of the
author's Ukrainian nationality. However, correct establishment of the name should be "Mykhailov, Ihor" (using the LC transliteration table for Ukrainian).

One must also consider conformance to AACR2 in establishing correct Russian language headings. Markiw (1884) discusses this in some detail in the above article, delineating changes in the structure of corporate name headings and changes in the manner in which personal names are established under AACR2.

**Research Objectives**

This study addressed the commonly held belief that the quality of original cataloging on OCLC and RLIN differs. The present study's purpose was two-fold: First, error rate estimates for types of Russian language name heading errors on OCLC and RLIN were developed. Secondly, an analysis was conducted in order to discover what types of name heading errors in Russian occur more frequently on both databases, and whether there is a difference in type and level of error by database.

The first analysis, testing the hypothesis that quality control of Russian language name headings is superior on the RLIN database in terms of fewer errors, could lead to an investigation of quality control practices regarding member-contributed foreign language copy at both OCLC and RLIN. The second analysis, that of developing a typology of errors occurring most frequently on both databases, and a comparison of name heading error patterns on each database,
could perhaps point toward specific solutions (such as workshops or handbooks) to specific quality control problems. As outlined above, the study both quantified and categorized errors. On both databases, errors were to be quantified with respect to kind (transliteration error, punctuation error, etc.) and type of name heading in which they occurred (personal name or corporate name). Random OCLC samples were drawn by the OCLC Office of Research, while pairing of OCLC sample records with RLIN records was done by the researcher. Control procedures built into the study included random selection, exclusion of LC copy (which is likely to be the same for both bibliographic networks), and limitation of the study to one format, books. Microform records were also eliminated. In addition, only Russian language materials published in the Russian republic (fixed field country code "rur") were chosen, in order to eliminate from consideration non-Russian authors who may have been published in Russian at one time.

Areas of Study

Areas of study included name headings for main and added entries. Name headings used as subject entries were excluded, as were title added entries (field 740). Unfortunately, no conference headings appeared on any of the records sampled. "Name headings" referred to in subsequent areas of this study are to be defined as above.

In all name heading fields treated in the present paper data will be quantified as to type of error: name heading
punctuation error, name heading transliteration/spelling error, name heading tagging error, and name heading AACR2 and LCRI error.

Name heading punctuation errors can be defined for the purposes of this study as the misplacement, omission, or incorrect usage of marks of punctuation (such as a period where a comma would be appropriate). Name heading punctuation errors will also include errors in capitalization.

Transliteration and spelling errors are grouped together because it is virtually impossible for a researcher to determine whether the inputter was confused regarding correct transliteration of the Cyrillic or was simply careless in typing. Transliteration/spelling errors usually occur as a "typo", and are to be distinguished from name headings that are completely inconsistent with NAF (Name-Authority File) forms.

Name heading tagging errors include incorrect or absent tagging of MARC fields, subfields, and indicators. AACR2 and LCRI errors will primarily represent headings whose forms are inconsistent with the Library of Congress Name-Authority File. Records with incorrect tagging (such as a 6xx that should be entered as 7xx) will be listed as tagging errors rather than AACR2 and LCRI rule errors.

Methodology

The study consists of four parts: data collection, construction of an error schedule, identification of errors,
and data analysis.

Data Collection  The final sample of paired bibliographic records drawn from OCLC and RLIN was chosen according to the four basic parameters described below:

1. No cataloging for the item from LC or any other national library such as NLM or NAL existed.
2. The item was present on both databases.
3. The item represented full cataloging: I-level on OCLC or 9114,9115,9116,9117, or 9118 on RLIN. RLIN cataloging category had to be "b" or "c".
4. Data was limited to the last 1.5 million records cataloged by OCLC and their matching RLIN pairs.

The population under study was originally designed to include Russian-language books published 1983-1989, but later modified. The final sample was chosen from the last 1.5 million records cataloged by OCLC, and the majority of the sample items have a publication date of 1989. The author hopes to publish the results of this study in the future, using a larger sample size (240 record pairs) and a wider time span (1983-1989).

It is hypothesized that altering the time frame from which the final sample was drawn could have engendered problems in matching. The initial OCLC mini-sample was drawn from the years 1983-1989, and it is from the mini-sample that the estimate of RLIN hits relative to OCLC items was taken. In addition, altering the time frame may have altered the proportion of matching records available on RLIN in the final sample: Only 152 matching RLIN records were located from the 700 records selected from OCLC (27 matched pairs were used.
from the mini-sample, for a sample total of 181 matched pairs). Because the final OCLC sample was drawn from books that were quite recently cataloged, this may have contributed to the numerous unusable "in process" records on RLIN that were not usable as matches for the purpose of the present study.

The initial 1983 date chosen by Intner, two years from the date LC began implementing AACR2 cataloging, allowed for several years to elapse during which errors attributable to the change in cataloging rule might occur. Intner noted that "training for experienced catalogers might take time, and newly graduated catalogers might have been taught earlier rules up to the end of the 1981 academic year" (Intner 1989, 6). This paper, however, will consider OCLC and RLIN paired records for Russian language titles contributed by members primarily, but not exclusively, in 1989, and does in fact yield results that differ somewhat from the Intner study.

A preliminary mini-sample was drawn, in order to anticipate potential problems in the sampling and pairing procedure. Preliminary sampling was conducted by the OCLC Development Division. The preliminary OCLC sample and RLIN matching procedure not only aided in avoiding methodological and statistical problems; it also assisted in predicting how large the actual OCLC sample needed to be in order to locate a sufficient number of paired records in the RLIN database.

Matching RLIN records were located, using any search key necessary to locate the appropriate items. Personal and
corporate name headings were obviously the last choice for search key, ISBN numbers being the first choice. The RLIN primary cluster member was used as the basis for comparison with the corresponding OCLC record. Pairing of OCLC/RLIN records was to be determined according to the following criteria:

1. Both items were member-contributed, full cataloging
2. Title: Exact match to shortest string (field 245)
3. Edition: Matches on number or name (field 250)
4. Publisher: Matches on two words (field 260 subfield b)
5. Dates of publication (date 1 in fixed field or 260 subfield c) may differ by one year
6. Pagination: Matches largest Arabic number within 10 pages (field 300 subfield a)

The final OCLC sample (to be paired with RLIN records, an RLIN hit determined by examining the main cluster record) was to have been drawn according to the following parameters:

**Fixed Field**

- **Type:** a
- **Source:** d
- **Bib lvl:** m
- **Lang:** rus
- **Enc lvl:** I or L
- **Ctry:** rur
- **dates:** 1983 through 1989 for date 1

**Variable Fields**

- **040:** DLC, NLM, and NAL must be absent from this field
- **041:** record should not include 041 field
- **1xx, 7xx:** Either 1xx, 7xx, or both must be present in record
In actuality, the final OCLC sample was drawn using solely the fixed field parameters described above. All OCLC records were examined manually, and those records lacking name fields or that were obviously translations were excluded from consideration. After OCLC records were drawn and paired with RLIN records, the study examined name headings in the following fields: 100, 110, 700, and 710.

**Error Schedule** An error schedule was constructed using the following categories:

1. Name heading punctuation error
2. Name heading transliteration/spelling error
3. Name heading tagging error (such as tagging for personal name when corporate name would be appropriate)
4. AACR2 rule error or LCRI error (such as the pre-AACR2 practice of entering East European corporate names under place, then under name. Current AACR2 practice is to enter East European name heading directly, i.e. "Gosudarstvennyi Ermitazh (Soviet Union)" rather than "Leningrad. Ermitazh"). Headings will be verified in the LC Name-Authority File when possible, and discrepancies with the NAF will be considered LCRI errors.

The validity of the instrument was determined by having the instrument reviewed by a panel of experts chosen from the cataloging field.

**Identification of Errors** Four types of name heading errors were defined and distinguished above. These name heading errors will also be quantified as to field in which they occur.

In addition, one should note that since the items catalogued were not in hand, coding for AACR2/LCRI errors was
limited to those errors that could be determined by examining the catalog record and the LC Name-Authority File. In the case of discrepancies between RLIN and OCLC name headings that could not be resolved by examining the catalog record or the NAF, the pair was discarded.

**Data Analysis** Errors in name heading entry for both OCLC and RLIN follow in table format. Errors are described by type (transliteration error, AACR2 error, etc.) and by place of occurrence (corporate name or personal name). Percentages for type and place of error were calculated for both OCLC and RLIN. Statistical packages available through Kent State University School of Library Science were used to tabulate data, and statistical experts reviewed all statistical procedures for appropriateness.

Each record pair was first assigned a control number to identify it as a single unit, and each pair was further identified by listing the appropriate OCLC number and RLIN record number. A coding sheet for the pair was developed, and included all the error types listed above (i.e., name heading punctuation error, name heading transliteration/spelling error, name heading tagging error, and name heading AACR2/LCRI error). These errors were also categorized as to field in which they occurred. Data was tabulated using Lotus.

**Findings**

Analysis of the data indicates that numerous errors in
name heading entry occurred on both databases, a total of 63
errors in the 181 matched pairs (i.e., 362 catalog records).
Of these 63 errors, 26 were name heading errors of OCLC
member libraries, while the remaining 37 errors were made by
RLIN member libraries.

The greatest number of errors in both databases occurred
in the application of AACR2 and LCRIIs, the most common error
type in this category being inconsistency with LC's Name-
Authority File. Errors in MARC tagging occurred somewhat
less frequently, with 7 errors on OCLC and 10 on RLIN.
Errors in Library of Congress rule interpretations were
considered to be potentially serious because they might
affect retrieval of the catalog record. Spelling and
transliteration error occurred infrequently on both
databases, with one error on OCLC and four errors on RLIN.
Punctuation errors were not considered to be significant, as
they almost never affect retrieval. Four punctuation errors
occurred on OCLC and an equal number on RLIN. For a summary
of the study's findings in table form, see page 16.
### TABLE 1. Occurrence of Errors

<table>
<thead>
<tr>
<th>Error Type</th>
<th>OCLC #(%)</th>
<th>RLIN #(%)</th>
<th>Total #(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuation</td>
<td>4(6.3)</td>
<td>4(6.3)</td>
<td>8(12.6)</td>
</tr>
<tr>
<td>Trans./Spell</td>
<td>1(1.6)</td>
<td>4(6.3)</td>
<td>5(7.9)</td>
</tr>
<tr>
<td>Tagging</td>
<td>7(11.1)</td>
<td>10(15.9)</td>
<td>17(27.0)</td>
</tr>
<tr>
<td>AACR2/LCRI</td>
<td>14(22.2)</td>
<td>19(30.2)</td>
<td>33(52.4)</td>
</tr>
</tbody>
</table>

| Total errors | 26(41.3) | 37(58.7) | 63(100.0) |

### Table 2. OCLC Errors by MARC Field

<table>
<thead>
<tr>
<th>Error Type</th>
<th>1xx</th>
<th>7xx</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuation</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Trans./Spell</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tagging</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>AACR2/LCRI</td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
</tbody>
</table>

| Total errors | 7   | 19  | 26    |
Table 3. RLIN Errors by MARC Field

<table>
<thead>
<tr>
<th>Error Type</th>
<th>1xx</th>
<th>7xx</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuation</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Trans./Spell</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Tagging</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>AACR2/LCRI</td>
<td>11</td>
<td>8</td>
<td>19</td>
</tr>
</tbody>
</table>

Total errors 7 19 37

Conclusions

Keeping in mind that the study's results may be skewed due to sample size and a predominance of record publication dates in the late 1980's, one can nevertheless draw several tentative conclusions from the above analysis.

Contrary to the prevailing notion, RLIN records possessed more total name heading errors than the corresponding OCLC records. This appears to support Intner's conclusion that the quality of member-contributed cataloging on OCLC is not significantly poorer than that of RLIN. One can only draw tentative conclusions at best using data from only two fields, but the study's findings at least lend credence to the notion that the quality of member-contributed original cataloging on OCLC and RLIN is similar.

Both databases contain numerous errors with respect to Library of Congress rule interpretations; these are mostly
discrepancies with the NAF. In some cases, the NAF record was added after the catalog record was input. In other cases, the correct form of entry might have been made had the cataloger understood the Russian language (For example, if one sees the title "doktor biologicheskikh nauk" in reference to a name heading in the Name-Authority File, one may conclude that the person in question has a Ph.D. in biology, and was likely to write books and articles on that subject. In addition, the number of errors in MARC tagging on both databases indicate the need for further staff training in both OCLC and RLIN member libraries.

The present study serves to confirm the sense of all of Intner’s recommendations regarding the training of catalogers, namely:

1. Correct punctuation should be emphasized.
2. Greater emphasis should be placed on the encoding of data for the MARC formats
3. Greater emphasis should be placed on the application of AACR2 according to the Library of Congress’ policies and practices

Further research in areas related to this study might include the following: a) An analysis similar to the present study using the 1983-1989 time span to obtain 240 matched pairs b) A study on the scale of Intner’s, analyzing all aspects of the catalog record (i.e., fixed field data and all variable field data such as subject headings, title, imprint, and collation c) The application of a similar methodology to other languages or error types d) The classification of member-input records as to quantity and type of error.
Possibilities are numerous.
BIBLIOGRAPHY


Whitby, Thomas J. and Tanja Lorkovic. 1979. Introduction to
Soviet national bibliography. Littleton, Colo.: Libraries Unlimited.