

## Description of Holdings and Circulation Data

### *I. Circulation data*

#### **UCLA: Taos**

**Total transactions:** 7,704,079

**Circulation transactions between:** 9/27/1999 and 7/6/2004

**Unique bibliographic record ids:** 923,625

**Processing.** This dataset was delivered as a large text file, one record per line, fields not delimited. A Java routine "chunked" data out by position according to the data dictionary, and assigned values to fields. The routine then generated a tab-delimited load file appropriate for ingest into MySQL. Values present in each field were checked against the original lines of text, and checked against the data dictionary to ensure that they matched expected values.

**Record contents.** The records provided for each transaction included:

- Minimal bibliographic information (isbn, lccn, oclc number, call number classification, and material type).
- Minimal patron information (anonymized patron id, patron rank, privilege code and institution code).
- Circulation details (item barcode, due date, transaction code and date, and item location).

#### **UCLA: Voyager**

These data pick up where the Taos System records left off in July 2004.

**Total transactions:** 1,662,896

**Circulation transactions between:** 7/1/2004 and 6/30/2005

**Unique bibliographic record ids:** 360,820

**Processing.** The dataset was delivered in multiple related files, in tab-delimited format suitable for loading directly into MySQL tables.

**Record contents.** The records provided for each transaction included:

- Minimal bibliographic information (isbn, lccn, oclc number).
- Minimal patron information (anonymized patron id, patron group, block code).
- Circulation details (item barcode, due date, reserve status, transaction type and date, and item location).

**UC Berkeley:**

**Total transactions:** 11,693,189 (excluding institutional transactions)

**Circulation transactions between:** 7/1/1998 and 5/31/2005

**Unique bibliographic record ids:** 1,155,035 (excluding institutional transactions)

**Processing.** This dataset was delivered as a large text file, one record per line, tab delimited. A Java routine assigned values to fields then generated a load file appropriate for ingest into mysql (\\N instead of blanks to indicate null values, sequence numbers added as primary key). Values present in each field were checked against the original lines of text, and checked against the data dictionary to ensure that they matched expected values.

**Record contents.** The records provided for each transaction included:

- Bibliographic information (isbn, lccn, oclc number, call number category and class, country of publication, language, format, record type).
- Patron information (major department affiliation, college code, loan privilege, patron type, gateway privilege, charge count).
- Circulation details (item barcode, due date, reserve status, transaction type, date and location, and item location, loan period).

**General notes:**

There are no transactions from April - June 2002 in this set.

This record set originally included transactions that were not made by individual patrons; for example, removing a book from circulation or transferring it to the storage facility. Institutional transactions were excluded by filtering based on patron\_id, selecting only those with patron\_id="Forbidden" (which indicates individual transactions, as these are overwritten for privacy protection).

## ***II. Holdings data***

### **OCLC**

**Total records:** 7,097,924 million (by OCLC number)

**Processing.** Records were delivered as tab-delimited text in a single flat file, suitable for loading directly into a MySQL database.

**Record contents.** Holdings counts (by OCLC number) for all WorldCat records that have UCLA OR UCB holdings. Each record consists of a separate count for each of the 10 UC campuses, and a "worldwide" holdings count" from WorldCat.

### **RLG**

**Total records:** 9,886,979 (by ISBN)

**Record Contents.** Each record consisted of one ISBN/holdings count pair. Because of the method of generation, ISBNs were not unique; a single ISBN could occur multiple times in the data set.

**Processing.** Records were delivered as series of 108 comma-delimited text files. Before delivery, RLG staff removed the initial line in the file (SQL column names) and removed alphabetic strings in the ISBN data, (e.g. 'PBK' and 'CASED'). Once received at CDL, a Java routine read and split incoming lines, validated the ISBN and tracked ISBN/holdings count pairs. When a repeat ISBN was encountered, the holdings number was checked; if it matched, the record was skipped as a duplicate. If it did not match, then an error message was generated and the ISBN was removed from the set.

#### **General Notes:**

Approximately 15% of the data were being discarded due to the issue of ISBNs with differing counts.