Workflow for Contributing Internet Archive-Digitized Volumes Into the HathiTrust: A Guide for the UC Campus Library

Prepared by the California Digital Library
# Table of Contents

Internet Archive to HathiTrust Workflow .......................................................... 3

Background........................................................................................................... 4

Overview of Roles ................................................................................................. 4
  Campus Library ................................................................................................. 4
  Internet Archive ............................................................................................... 5
  CDL ..................................................................................................................... 5

Overview of Deliverables ..................................................................................... 5
  Campus Library ................................................................................................. 5
  CDL ..................................................................................................................... 6
  Internet Archive ............................................................................................... 6

Notification/Communications .............................................................................. 6

HathiTrust Ingest Eligibility Requirements ......................................................... 7

IA Requirements .................................................................................................. 7

Picklists .................................................................................................................. 8

Annotated Picklists ............................................................................................... 9

Bibliographic Records ......................................................................................... 9

Tracking And Quality Control ............................................................................ 11

Project Checklist ................................................................................................. 13
  Before Scanning ............................................................................................... 13
  During Scanning ............................................................................................... 13
  After Scanning ................................................................................................. 13

Appendix: Technical Requirements .................................................................. 14
  Section A.1: Metaxml file .............................................................................. 14
  Section A.2: ARK Minter .............................................................................. 15
  Section A.3: Example: OCLC Number Documentation .................................... 15
  Section A.4: Picklist format ........................................................................... 17
  Section A.5: Rejection Codes ........................................................................ 18
  Section A.6: Example MARC record with required fields ......................... 19
Internet Archive to HathiTrust Workflow

Internet Archive

- Digitization
- Images + OCR
- Metadata
- IA Web Server

Campus Library

- Picklist
- QA & Tracking
- Annotated Picklist
- MARC Records w/ OCLC Num.

CDL

- Campus ILS
- CDL FTP Server
- MDID
- Processing/ Prep
- Ingest/ Validation
- Hathi Web Server

Users
Background

The University of California has long benefited from the Internet Archive’s (IA’s) expertise in digitizing books and book-like volumes, and historically this work has been brought together by the California Digital Library (CDL) for the benefit of the UC Libraries. From 2006-2009, CDL had a contractual relationship with the Internet Archive and coordinated IA’s digitization of books across the libraries. The Internet Archive deployed clusters of Scribe scanning stations at NRLF and SRLF, where ongoing operations were conducted on site. Through this relationship, we were collectively able to digitize nearly 200,000 books from a number of campuses.

In the process, CDL learned that consistency is key in making digitization projects work together cohesively and in realizing their benefits on a UC-wide scale. The hardest problem to solve was inconsistency in metadata and digitization practices, and this experience has led CDL to recommend more stringent metadata creation and formatting specifications for future UC Internet Archive scanning projects.

As of this writing, the Internet Archive and CDL do not have a direct, contractual relationship. However, CDL seeks to facilitate the efforts of UC libraries to digitize collections with the Internet Archive. While the campus library is responsible for the physical materials and operations staff involved in digitization, CDL oversees preservation of and access to digitized objects via HathiTrust.

This guide provides the campus library a summary of the process of working with the Internet Archive, along with the standards and technical details necessary to enable CDL to submit digitized objects to HathiTrust.

Overview of Roles

The following items are critical to each participant’s role in the overall workflow.

Campus Library

The contributing library will:

- Coordinate and consult with CDL before beginning the project
- Identify a dedicated project manager to serve as the point of contact for communications
- Negotiate with the Internet Archive to set the terms of the scanning agreement, including the requirements set out by CDL in this document
- Identify the collections to be digitized
- Make the candidate collections available for scanning
- Perform Quality Assurance on metadata received from the Internet Archive before providing it to CDL
Internet Archive

The Internet Archive will:

- Provide scanning equipment and staff to operate it
- Scan physical materials into a set of digital files (images and OCR)
- Make these digital files available to CDL.
- Create and make available metadata for each digitized item
- Return the physical volume to the partner library

CDL

CDL will:

- Provide standards for metadata creation and delivery
- Provide tracking of baseline information about UC Libraries’ high-volume scanning projects
- Provide coordination and management for ingesting digitized objects into HathiTrust (The ingest itself is performed by staff at the University of Michigan.)
- Track digitized assets deposited into HathiTrust for preservation purposes

Overview of Deliverables

Campus Library

1. Picklist in .tab format (for use by Internet Archive and CDL)
2. Annotated Picklist in .tab format (for use by CDL)
3. Item level bibliographic records in MARC-8 or MARCXML format (for use by Internet Archive, CDL and HathiTrust)
4. Samples of all files prior to the beginning of digitization (for use by CDL)

---

1 It is understood that Internet Archive will retain a copy of the digital item, as they have in the past.
**CDL**

1. Provide FTP server\(^2\) for campus library file delivery to CDL; user accounts and directories

**Internet Archive**

1. Digitization output will consist of:
   i. JPEG-2000 images
   ii. OCR files
   iii. Metadata files (see step 2 of section “Tracking And Quality Control” below)

2. Provide samples of digitized book images to contributing library prior to the beginning of full-scale digitization.

3. Provide samples of metadata to contributing library prior to the beginning of full-scale digitization.

**Notification/Communications**

The contributing library must coordinate with CDL before commencing scanning operations with the Internet Archive in order to determine operational timelines and deliverables.

**Steps:**

1. The contributing library will contact massdig@cdlib.org to notify CDL of the planned commencement date of scanning, at least 60 days before the date.

2. The contributing library and CDL will confer about the nature of the collections to be digitized to ensure that they are good candidates for inclusion in HathiTrust.

3. CDL and the contributing library will together determine a timeline for delivery of Picklists, Annotated Picklists, and MARC files to CDL.

4. When the scanning project completes, the contributing library will notify massdig@cdlib.org.

Note: Campus libraries can contact Robert Miller at the Internet Archive [robert@archive.org] for more information.

---
\(^2\) The CDL FTP Server is ftpei.cdlib.org and requires an FTPS connection. Accounts, login information and usage requirements will be communicated during the initial consultation with CDL.
HathiTrust Ingest Eligibility Requirements

All items to be ingested into HathiTrust must have OCLC numbers in their metadata and be book-like in format. Having OCLC numbers in bibliographic metadata permits CDL to include digitized books in downstream services, most importantly their linking in WorldCat and ingest into HathiTrust. At present, the HathiTrust workflow does not support non-book materials, although these items may be considered in the future.

Steps:

1. The contributing library will consult with CDL regarding the candidate collection(s) and feasibility for digitization.
2. The contributing library will ensure that all items submitted for ingest are in bound, book-like format. They cannot be standalone maps, archival papers, images, postcard collections, audio or video.
3. The contributing library will confirm all records associated with items for submission contain an OCLC number consistently recorded in a specific field. (See section A.3 for details)

IA Requirements

Before commencing scanning, the contributing library will negotiate with the Internet Archive to ensure that files produced from scanning operations will be available to UC, and are consistent with other UC scanning partners. The campus library must confirm metadata is created and captured according to the following specifications. (If a contract is being produced between the Internet Archive and the contributing library, these negotiations are best codified there.)

Key Points:
It is the campus library’s responsibility to convey the following to the Internet Archive, and to ensure that the Internet Archive complies:

1. The Internet Archive must provide a meta.xml file for each item. This metadata in this file must be structured in a way consistent with other UC partners (see appendix A.1).
2. ARK: The Internet Archive must create ARK identifiers using the CDL minter and insert them into the meta.xml file for each item. (See Appendix A.2). ARKs must be recorded within an <identifier-ark> tag. In order to be valid, each ARK must be generated according to the rules of the minter included in the Appendix.
3. Collection Tags: All materials must be tagged as with collection tags as below:
   <collection>cdl</collection>
   <collection>uc-local-[campusID]</collection> (This is a newly
added collection tag)
    ([campusID] must be one of the following: ucb, ucd, uci, ucla, ucm, ucr, ucsb, ucsc, ucsd, ucsf, nrlf, srlf)

4. SysID/BibID: IA must extract both the bibliographic record ID and system ID (if different from bib record ID) for each item from the Picklist and insert them into the meta.xml file. They should be recorded using an <identifier-bib> tag for the bibliographic record ID and an <identifier-sys> tag for the system ID. The <identifier-sys> tag is a new requirement.

5. Volume Data: Volume metadata must be stored in the <volume> tag in the meta.xml file and copied exactly from the Picklist.

6. MARC file retrieval: IA must retrieve MARC files from the campus library’s Z39.50 server, based on instructions provided by the campus library.

---

**Picklists**

All items to be digitized must be documented on a Picklist, in tab-delimited format. The Picklist helps both CDL and the contributing library to track the progress of scanning. Picklists should be delivered to CDL at the start of the scanning project.

Note: See Appendix A.4 for an example Picklist.

**Steps:**

1. Use the local OPAC to output a tab-delimited file with the records for the books to be digitized. Name it according to the following convention: [owning campus code]-[owning library code]-[picklist number].tab (Example: ucla-srlf-04.tab). The file extension and UTF-8 encoding is required.

2. Populate the file with the following fields, in this order:
   - System ID
   - Bibliographic ID (if different from System ID)
   - Bibliographic Item ID (if applicable)
   - Call Number
   - Barcode (if applicable)
   - Volume (directly copied from item data in record)
   - Author
   - Title
   - Publication Date (formatted as YYYY if possible) – CDL prefers this to come from the MARC 008 field.
3. Upload the file to the CDL FTP Server (this information will be supplied by CDL).

Each new project will require its own Picklist.

**Annotated Picklists**

Annotated Picklists are based on the original Picklists produced to support the scanning workflow, but include additional metadata about scanning outcomes. The Annotated Picklist is a critical tool for tracking the progress of scanning. CDL requires Annotated Picklists from contributing libraries in order to keep track of the scanning or rejection of individual items, and to record the assignment of identifiers. Having the Picklists will ensure our ability to surface these items in discovery tools (via identifiers) and for the UC Libraries to be able to make collaborative and/or UC-wide collections decisions.

The contributing library will provide updates and/or final Annotated Picklists to CDL on a regular schedule as agreed. It is often easiest to annotate Picklists in Excel and then export as tab-delimited upon completion.

Steps:

1. Import the original Picklist into Excel, or your chosen tool. Name it according to this convention: [original picklist name]_yyyymmd.d.

   **Example:** ucla-srlf-04_20081023

2. Include all the original Picklist columns, and add the following (in order from left to right):

   - Sent date; formatted as YYYYMMDD
   - Rejection codes (must conform to controlled vocabulary in Appendix A.5)
   - Internet Archive ID (IA ID)
   - ARK

3. Save the file in .tab format

4. Deliver Annotated Picklists to CDL within 12 weeks of completion of scanning.

**Bibliographic Records**
Item level MARC records must exist for each physical item to be digitized, and must be delivered to CDL with the Annotated Picklist (at the completion of digitization activities). The minimum requirements for content, formatting, and documentation are listed below. Having a minimum standard for records, along with adequate documentation, will permit CDL (and HathiTrust) to ingest the records more or less automatically and independently and to surface the digital volumes within discovery services. Fuller bibliographic records are always useful, but not required. The campus library ILS is expected to be Z39.50 enabled.

Requirements:

1. Item level records: Each physical item must have a discrete bibliographic record. Boundwiths and boxed sets should have one record for each item included. Serials should have one record for each item that will receive an Internet Archive identifier, as in Appendix A.2.

2. Each record must contain the following fields:
   • Bibliographic Record ID
   • Bibliographic Item ID (if applicable)
   • System ID (if different from Bib Record ID)
   • OCLC number
   • Barcode (if exists)
   • Title
   • Author
   • ISBN/ISSN (if applicable)
   • LCCN (if applicable)
   • Publication year; formatted as YYYY if possible
   • Item data (volume, etc.) - split to be as granular as possible
   • Internet Archive identifier (IA ID): 995$p
   • ARK identifier: 995$q

3. Each record must also have the Internet Archive identifier (also referred to as “IA ID”) and ARK identifiers present in the 995$p field and 995$q field, respectively.

4. Formatting: Records must be formatted as MARC or MARCXML files with UTF-8 encoding.

5. Documentation: The contributing library must create a document describing the location and formatting of identifiers in its bibliographic records.

   The locations of the following identifiers must be noted:
   • SystemID
• Bibliographic Record ID
• Bibliographic Item ID (if applicable)
• OCLC number
• Item data (enum/chron)
• Barcode (if applicable)

Formatting of identifiers must also be documented. Elements such as prefixes, suffixes, structure, length, and legal characters should be noted, with samples. See Appendix A.3 for an example.

### Tracking And Quality Control

When the Internet Archive returns metadata for digitized items to the contributing library, the library must perform QA to verify that Internet Archive is abiding by the formatting requirements agreed upon prior to the commencement of scanning. This is necessary to prevent small omissions or variations from becoming large problems by catching them early in the process. The contributing campus library should work out any inconsistencies in the metadata with the Internet Archive before submission of any files to CDL.

**Steps:**

1. Verify that each object’s meta.xml file contains the following tags:
   - ARK: <identifier_ark>
   - Internet Archive identifier: <identifier>
   - Collection tags:
     - <collection>cdl</collection>
     - <collection>uc-local-[campusID]</collection>
   - System/Bibliographic identifier(s):
     - <identifier-bib>XXXXXXXX</identifier-bib>
     - <identifier-sys>XXXXXXXX</identifier-sys>
   - Volume data: <volume>

   See Appendix, sections A.1 and A.2 as needed.

2. Ensure that the following files are available from the Internet Archive on their web server. This set comprises the minimum requirements for HathiTrust ingest (see preferred package documentation at [http://www.hathitrust.org/ingest](http://www.hathitrust.org/ingest)).
   - jp2
• djvu.xml
• djvu.txt
• marc.xml
• meta.xml
• scandata.xml
Project Checklist

Before Scanning

1. Verify all items are in book-like in format
2. CDL to provide login/password for the CDL FTP Server
3. Confirm and document the Internet Archive’s metadata requirements
4. Provide instructions to the Internet Archive regarding how to access campus library’s Z39.50 server to retrieve MARC files
5. Generate Picklist
6. Check bibliographic records for required fields
7. Format bibliographic records as MARC or MARCXML
8. Confirm all bibliographic records have an OCLC number consistently recorded in a specific field
9. Upload Picklist to the CDL FTP server
10. Document location and formatting of identifiers in records and submit to CDL

During Scanning

11. Perform QA on IA Metadata
12. Verify existence of required files (as listed in step 2 of “Tracking And Quality Control” above)
13. Perform ongoing updates of Annotated Picklists with rejection reasons, or with date sent, ARK and IA ID

After Scanning

14. Complete Annotated Picklist and deliver to CDL
15. Export and deliver MARC records for all scanned items to CDL
Appendix: Technical Requirements

Section A.1: Meta.xml file

Below is an example of a properly encoded meta.xml file from the Internet Archive with all required fields:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<metadata>
  <title>The adolescent girl; a study from the psychoanalytic viewpoint</title>
  <creator>Blanchard, Phyllis Mary, 1895-</creator>
  <subject>Girls</subject>
  <subject>Adolescence</subject>
  <subject>Psychoanalysis</subject>
  <subject>Adolescent psychotherapy</subject>
  <description>Bibliography at end of each chapter</description>
  <publisher>New York : Moffat, Yard</publisher>
  <date>1920</date>
  <language>eng</language>
  <copyright-evidence-operator>Alyson-Wieczorek</copyright-evidence-operator>
  <possible-copyright-status>NOT_IN_COPYRIGHT</possible-copyright-status>
  <copyright-region>US</copyright-region>
  <copyright-evidence>Evidence reported by Alyson-Wieczorek for item adolescentgirlst00blan on August 20, 2008: visible notice of copyright; stated date is 1920.</copyright-evidence>
  <sponsor>MSN</sponsor>
  <contributor>University of California Libraries</contributor>
  <scanningcenter>la</scanningcenter>
  <mediatype>texts</mediatype>
  <collection>americana</collection>
  <collection>cdl</collection>
  <collection>uc-local-srlf</collection>
  <call_number>SRLF_UCLA:LAGE-607890</call_number>
  <collection-library>SRLF_UCLA</collection-library>
  <identifier-bib>LAGE-607890</identifier-bib>
  <identifier-sys>LAGE-607890</identifier-sys>
  <volume/>
  <updatedate>2008-08-20 20:40:48</updatedate>
  <updater>Alyson-Wieczorek</updater>
  <identifier>adolescentgirlst00blan</identifier>
  <uploader>scanner-alyson-wieczorek@archive.org</uploader>
  <addeddate>2008-08-20 20:40:50</addeddate>
  <publicdate>2008-08-20 20:40:53</publicdate>
  <ppi>400</ppi>
  <camera>Canon 5D</camera>
  <operator>scanner-deanna-flegal@archive.org</operator>
  <scanner>scribe2.la.archive.org</scanner>
</metadata>
```
Section A.2: ARK Minter

The Internet Archive uses the CDL NOID Minter to create ARKs for digitized UC books.

A test NOID minter is available at: http://dot.ucop.edu/nd/noidu_fk8?mint=1.

Standard practice is:

a. IA to prepend "ark:" to the minted ids that the minter generates. For example, when the minter mints “id: 13960/fk3qv3c71j”, IA will produce “ark:/13960/fk3qv3c71j”

b. IA to place the ARK in the following MODS/MARC21 tag and describe as follows:
   • <identifier type="ark"> ark:/13960/fk3qv3c71j</identifier>

c. IA to place the access URI of the resource in the following MODS/MARC21 tag and describe as follows:
   • <identifier type="uri"> http://www.archive.org/details/triumphoflightk00keelrich</identifier>

The minter is completely documented at http://www.cdlib.org/inside/diglib/ark/noid.pdf

Section A.3: Example: OCLC Number Documentation

OCLC numbers are required for all objects to be ingested into HathiTrust. Documenting
the format of OCLC numbers in its bibliographic records, a campus library might provide CDL with an explanation similar to this one:

**OCLC Numbers Format (Example Documentation)**

If MARC field: **079$a**
- If exists, this is the authoritative OCLC number.
- OCLC number is preceded by "(OCoLC)", "(OCoLC)ocm", "(OCoLC)ocn", "ocm", or "ocn"

**EXAMPLE:** (OCoLC)26485580

Else, MARC field **039$a** or **035$a**
- OCLC number preceded by " (OCoLC) " , "(OCoLC)ocm", "(OCoLC)ocn", "ocm", or "ocn"

**Fields may contain extraneous data (usually trailing the OCLC number). If both fields exist, always trust the 039$a, but only if it has an "(OCoLC)" prefix.**

**EXAMPLE:** (OCoLC)26485580
### Section A.4: Picklist format

<table>
<thead>
<tr>
<th>System ID</th>
<th>Bibliographic ID</th>
<th>Bibliographic Call Number</th>
<th>Barcode</th>
<th>Vol</th>
<th>Author</th>
<th>Title</th>
<th>Publication Date</th>
<th>Sent Date</th>
<th>Rejection IA II</th>
</tr>
</thead>
<tbody>
<tr>
<td>b167284964</td>
<td>VM140.F9 C6</td>
<td>$B 16 439</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b167287576</td>
<td>LA707 .G7</td>
<td>$B 16 699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b167287606</td>
<td>LC1045 .P3</td>
<td>$B 17 891</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16419579</td>
<td>AC8 .C5</td>
<td>$B 18 380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13776737</td>
<td>HV523.W6 G</td>
<td>$B 20 335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13977463</td>
<td>HV5293 .W5</td>
<td>$B 20 443</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15767360</td>
<td>920.A881.a</td>
<td>$B 27 390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2337425</td>
<td>984s.R784.s</td>
<td>$B 28 089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4034</td>
<td>2706206</td>
<td>$B 28 834</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17427014</td>
<td>906s.M97</td>
<td>$B 31 437</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19736204</td>
<td>TX645 .B3</td>
<td>$B 31 746</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13757056</td>
<td>TX791.A73</td>
<td>$B 31 761</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32040</td>
<td>5687256</td>
<td>$B 31 762</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>866245</td>
<td>604047</td>
<td>$B 31 764</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3258303</td>
<td>TX303 .W5</td>
<td>$B 31 900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17436782</td>
<td>TX145 .B39</td>
<td>$B 32 175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17324485</td>
<td>TX145 .L8</td>
<td>$B 32 178</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6185722</td>
<td>BX5133.S649 1 B 3</td>
<td>334 786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17494399</td>
<td>BX5133.S649 1 B 3</td>
<td>334 787</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17519200</td>
<td>BX5133.S649 1 B 3</td>
<td>334 788</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1750656</td>
<td>PS88.M5</td>
<td>$B 335 133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7389541</td>
<td>PS88.d5</td>
<td>$B 335 134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17436782</td>
<td>JN118 .L95g 1917</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Section A.5: Rejection Codes

<table>
<thead>
<tr>
<th>CODE</th>
<th>REJECTION REASON</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>fragile or no binding (includes items in clam shells or phase boxes)</td>
</tr>
<tr>
<td>CAT</td>
<td>cataloging error</td>
</tr>
<tr>
<td>DAM</td>
<td>damaged</td>
</tr>
<tr>
<td>DAT</td>
<td>still in copyright</td>
</tr>
<tr>
<td>DUP</td>
<td>exact duplicate of another on list</td>
</tr>
<tr>
<td>FO</td>
<td>foldouts</td>
</tr>
<tr>
<td>FOR</td>
<td>non-book format</td>
</tr>
<tr>
<td>LAN</td>
<td>outside language parameters</td>
</tr>
<tr>
<td>LG</td>
<td>too large</td>
</tr>
<tr>
<td>LINK</td>
<td>unsuccessful link to metadata</td>
</tr>
<tr>
<td>LIST</td>
<td>picklist error</td>
</tr>
<tr>
<td>MAR</td>
<td>margins too tight</td>
</tr>
<tr>
<td>MIS</td>
<td>missing pages</td>
</tr>
<tr>
<td>MUL</td>
<td>multiple titles bounds together</td>
</tr>
<tr>
<td>NA</td>
<td>not available</td>
</tr>
<tr>
<td>NOS</td>
<td>not on shelf – missing/lost</td>
</tr>
<tr>
<td>OTH</td>
<td>unknown, non – standard codes</td>
</tr>
<tr>
<td>OUT</td>
<td>not on shelf – checked out or billed to patron</td>
</tr>
<tr>
<td>PAG</td>
<td>pagination problems: section(s) bound out of order or upside down</td>
</tr>
<tr>
<td>PAM</td>
<td>pamphlets</td>
</tr>
<tr>
<td>PAP</td>
<td>brittle paper, tissue paper</td>
</tr>
<tr>
<td>SKW</td>
<td>skewed text – to point of being unreadable</td>
</tr>
<tr>
<td>SM</td>
<td>too small</td>
</tr>
<tr>
<td>SPH</td>
<td>requires special handling</td>
</tr>
<tr>
<td>UNC</td>
<td>uncut pages (more than 5)</td>
</tr>
<tr>
<td>VEL</td>
<td>vellum</td>
</tr>
<tr>
<td>WD</td>
<td>withdrawn</td>
</tr>
</tbody>
</table>
Section A.6: Example MARC record with required fields

Below is an example of a MARC-XML record with all required fields:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<record>
  <leader>01189cam a2200337i 4500</leader>
  <!-- Required: Bibliographic ID -->
  <controlfield tag="001">607890</controlfield>
  <controlfield tag="005">20071231105928.0</controlfield>
  <controlfield tag="008">790108s1920 nyu b 000 0 eng|</controlfield>
  <datafield tag="010" ind1=" " ind2=" ">
    <subfield code="a">20008047</subfield>
    <subfield code="d">607890</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)04531075</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="9">07-ABG-2031</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">uc0clc4531075</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">607890</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">IAR</subfield>
    <subfield code="d">IAR</subfield>
    <subfield code="d">CHS</subfield>
    <subfield code="d">CLU</subfield>
  </datafield>
  <datafield tag="050" ind1=" " ind2=" ">
    <subfield code="a">HQ35</subfield>
    <subfield code="b">.B6 1920</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Blanchard, Phyllis Mary</subfield>
    <subfield code="d">1895-</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="4">
    <subfield code="a">The adolescent girl</subfield>
    <subfield code="b">a study from the psychoanalytic viewpoint</subfield>
    <subfield code="c">by Phyllis Blanchard ... with a preface by Dr. G. Stanley Hall</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" ">
    <subfield code="a">New York :Moffat, Yard</subfield>
    <subfield code="c">1920</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">xiv, 242 p.</subfield>
    <subfield code="c">20 cm.</subfield>
  </datafield>
</record>
```
Bibliography at end of each chapter.

Girls.

Adolescence.

Psychoanalysis.

Adolescent psychotherapy.

rcp

MC2805401

MARS

vol. 15 no. 1

AA0010476364

7585719

adolescentgirlst00blan

ark:/13960/t9b56rj52

<record>