PAPR Data Standards and Cataloging Principles

The **Print Archives and Preservation Registry (PAPR)** system is designed to support archiving and management of serials collections by providing comprehensive information about titles, holdings, and the terms and conditions of archiving of the major print archiving and shared print programs.

This document outlines the data standards and cataloging principles which CRL follows, and describes the validation work CRL performs, to encourage responsible growth of the PAPR database. CRL is keenly aware that it is nearly impossible to have a flawless database of serials holdings when aggregating data from many sources. However, by following industry standards, validating data to the best of our ability, working with all contributors of PAPR to improve their data, we can, provide the community with the best possible data about serials held by print archives or shared print programs.

Bibliographic Data Elements in PAPR Records
To support aggregation of data sources, collection comparisons, and the synchronization of data across the library community, PAPR uses key elements of library bibliographic data. They include: OCLC#, ISSN, title and publisher. Key to automated aggregation, discovery and comparison is the OCLC number and ISSN. Records need only one of those number to ingest into PAPR, but CRL strongly encourages contributors to include both, when possible.
The ISSN is a unique eight digit number used to identify serials. ISSNs are assigned by ISSN National Centres in coordination with the International ISSN Centre. The ISSN is the primary identifier used by the PAPR database. All holdings records containing an ISSN will attach to the PAPR record associated with that ISSN.

The OCLC number is a unique number assigned to a single bibliographic record contributed to OCLC's Worldcat database. OCLC numbers is are used to identify a unique bibliographic record representing and describing a resource. In the absence of an ISSN, the OCLC number becomes the primary identifier for the PAPR database. All holdings records with no ISSN, but having an OCLC number, will attach to the PAPR record with the matching OCLC number. It is common to find multiple OCLC numbers and records representing a serial title. Without a matching ISSN, a title with more than one OCLC number, will create duplicate records in PAPR.

Title and publisher are displayed in PAPR, but not used as a match point. Title and publisher are taken from the first instance of a record for that title in PAPR.

Cataloging Principles
In keeping with current cataloging standards, PAPR follows successive entry cataloging rules for describing serial resources. Over the course of its publication life, a serial resource often undergoes changes (such as in title, publisher, etc.) that may require a new record. Submitting records, which follow outdated or local cataloging standards, can cause holdings to attach to the wrong title record or to be separated from the correct one.

Following current cataloging standards ensures that PAPR can accurately map holdings to the correct title record, aggregate holdings from multiple sources, and report on the number of copies held. Rules for successive entry cataloging can be found in the *CONSER Cataloging Manual CCM: Module 16 When to create a new description*.

Holdings Disclosure in PAPR Records
To support the disclosure of the extent of holdings and the condition of holdings in repositories of print archives and shared print collections, PAPR uses guidelines established in the OCLC Shared Print Archives Pilot Project (2010-2012).

Detailed metadata guidelines can be found at https://www.oclc.org/en/services/shared-print-management/metadata-guidelines.html

A key element of the metadata is the 583 field. The PAPR database uses some elements of the 583 for display or linking records to program and library records. These include:
- subfield 3 Materials specified - Display
- subfield a Action - Display
- subfield d Action interval - Display
- subfield f Authorization (Archiving program) – Display and record linking
- subfield i Method of Action (Validation level) - Display
- subfield l Status - Display
- subfield z Public Note - Display
-
Bibliographic and holdings data can be downloaded. In addition to the fields listed above, subfield j Holdings Library Code is available in the downloaded record.

Please see PAPR Record Use Policy http://papr.crl.edu/policy and PAPR Terms of Service http://papr.crl.edu/terms under which use of the data is subject.

Validation
CRL performs extensive review and validation on all data submitted for ingest into PAPR. Reports, with the results of the review, are returned to all data contributors. CRL makes every attempt to be as accurate as possible in its data validation. However, without being able to verify the data with a check of the physical materials, some validation questions cannot be resolved by CRL. CRL depends on the cooperation of the contributing libraries to review the reports and make the changes indicated in those reports. Correcting data errors and discrepancies at the source makes the PAPR database more accurate, and reduces the risk of loss of resources due to inaccurate reporting.

CRL uses the ISSN database and OCLC's Worldcat database to validate the following data elements:

- Record is a serial:
    - if a record is not a serial, it will not be ingested into PAPR
    - check MARC Bibliographic Level (Leader position 07)
    - check MARC Serial Type (008 position 21)
- Record is for print format:
    - if a record is not print format, it will not be ingested into PAPR
    - check MARC Form of Item (008 position 23)
- OCLC number is accurate:

- is found in WorldCat
- and is in a MARC 001 field there
- ISSN is associated with the OCLC number provided:
  - in WorldCat, this ISSN appears in a MARC 022 subfield a
  - and that 022 is in a record where the MARC 001 contains this OCLC number
- ISSN is in the ISSN database
  - ISSN database title matches title provided by the contributing library
  - ISSN publication dates are congruent with the library holdings
- Holdings from contributing library fall within the valid publication dates for the title:
  - dates in the holdings are between Date1 and Date2 in MARC field 008 (years)
  - where MARC field 362 data is available, dates in the holdings must be between the specific dates indicated (months/days)

There are frequently discrepancies between the OCLC record and the ISSN record. In those cases, CRL will provide the contributing library with copies of both records for the contributor's use in determining which is correct. Most of the time, the ISSN database record is preferred, but there are exceptions.

Validation Reports
PAPR validation reports are in the form of a multi spreadsheet workbook. They provide a confirmation of number of records submitted for ingest, a count of items to review, detailed information about each record needing review, and a list of records with no problems.
A sample report is attached.

Excluding records from ingest into PAPR
CRL will exclude from ingest records for anything other than a print serial record. A print serial records is identified by its associated OCLC#, ISSN. Holdings that refer to electronic or microform, despite having a print OCLC# or ISSN, will be excluded.
Records with errors of formatting or substance (like using an ISSN for more than one title or including holdings outside of the publication range), will be loaded into PAPR if submitting programs prefer, but participants must realize that loading data with errors negatively affects true aggregation of the data and compromises CRL's ability to report on the status of print archiving as disclosed in PAPR.