OCLC Multi Match List Page

Automatically generates a daily Excel report of all OCLC multi-match bibliographic records from the previous day’s NZ import jobs.

At a glance

- **Status:** Active
- **Environment / Tags:** Network Zone
- **Applies consortium-wide?:** Yes *(The workflow operates in the Alma Network Zone and reports on Network Zone import jobs.)*
- **Runs on:** Alma Network Zone (NZ)
- **Trigger:** Scheduled — runs **daily at 10:30 PM**
- **Primary outcome:** Generates a spreadsheet listing all **OCLC multi-match bibliographic records** from the previous day’s import jobs.
- **Who receives results:** No direct notification; results are written to a shared Excel workbook.

Why this exists

When OCLC import jobs encounter **multi-matches**, Alma cannot automatically determine which existing record should be updated. These records require review.

This workflow creates a structured, review-ready spreadsheet listing all multi-match records from the previous day’s OCLC import jobs, including:

- OCLC number
- Network ID
- Institution holdings
- Import profile type
- Title

This supports consortium-level quality control and faster resolution of ambiguous matches.

What it does

- Runs nightly.
- Identifies specific OCLC import jobs by job ID.
- Retrieves completed job instances from the previous day.

- Excludes FTP-related job variants.
- Retrieves all **multi-match records** from those jobs.
- Splits comma-separated Network IDs into one row per record.
- Retrieves full NZ bib records for each match.
- Parses MARC XML to extract:
  - First 035$a OCLC number
  - All AVA fields (institution holdings)
- Produces one row per holding per bib.
- Removes duplicate institution holdings.
- Creates a new dated worksheet in Excel.
- Appends formatted rows to a structured table.

Where it runs

- **Alma IZ(s):**
  - Alma **Network Zone**
- **Systems touched:**
  - Alma APIs (read-only for job data and matches; read/write for bib retrieval)
  - Microsoft Excel (Office 365)
- **Reports / queries used:** None (operates directly on Alma job APIs)

How it works

Logic overview

1. Schedule & Date Setup

- Runs daily at **10:30 PM**.
- Calculates **yesterday's date**.
- Uses that date to retrieve completed job instances.

---

2. Identify Relevant Import Jobs

- Uses a hard-coded list of OCLC import job IDs.
- Retrieves job instances submitted yesterday.
- Keeps only those with status **COMPLETED_SUCCESS**.
- Filters out instances whose name contains **"w/FTP:"**.

⚠ If import profile IDs change, this workflow must be updated.

---

3. Retrieve Multi-Match Records

- Calls Alma's job matches endpoint with population = `MULTI_MATCHES`.

- Extracts:
  - Incoming record ID
  - MMS Ids (comma-separated list)
  - Job name (used to determine import type)
  - Submission date
- Determines **Import Type** (Deleted, Merged, New, Updated) using pattern matching on job name.

⚠ If naming conventions change, import type detection may break.

---

4. Separate Network IDs

- Splits comma-separated MMS IDs into **one row per Network ID**.
- This is required for downstream bib-level processing.

---

5. Retrieve and Parse Bib Records

For each Network ID:

- Retrieves the full NZ bib record.
- Parses MARC XML using regex to extract:
  - First 035$a beginning with `(OCoLC)`
  - All AVA fields:
    - subfield 0 → MMS Id (institution-level)
    - subfield a → Institution code

Returns one row per AVA field (institution holding).

⚠ Assumptions:

- MARC structure remains consistent.
- Only the first `(OCoLC)` number is used.
- Regex-based MARC parsing may fail if XML structure changes.

---

6. Normalize and Deduplicate

- Merges import metadata with parsed MARC data.
- Removes duplicate rows based on MMS Id (institution holding).
- Prepares structured fields for Excel.

---

7. Excel Output

For each run:

1. Creates a **new worksheet** named with yesterday's date (`yyyy-MM-dd`).
2. Waits 15 seconds (ensures Excel sheet creation completes).
3. Appends a header row.
4. Creates a table.
5. Appends all processed rows.

Columns written:

- Network Id
- OCLC Number in Alma
- Incoming Record Id
- Title
- MMS Id
- Institution
- Import Profile
- Import Date

Network IDs and MMS IDs are written as text to prevent numeric reformatting.

If results exist

- A new worksheet is created for the date.
- Each multi-match record appears as one row per institution holding.
- Spreadsheet becomes a permanent audit record for review.

If no results

- A new worksheet is still created.
- Only the header row is present.
- No data rows are appended.

Artifacts produced

- **Excel workbook:**
  OCLC Multi Match List
- One worksheet per day.
- Structured table with normalized fields.

No emails or additional reports are generated.

Revision #1
Created 2026-02-19 20:12:43 UTC by Jackie Saavedra
Updated 2026-02-19 20:15:54 UTC by Jackie Saavedra