



AGUA Development Phase 5

Mid-Phase Report

Summary

Planning for the fifth AGUA Development Phase began in the summer of 2020, with project ideas submitted by the WEST project team, the AGUA technical team, WEST Operations and Collections Council, and WEST membership. From the long list of possible enhancements, the OCC identified a short list of development priorities. The project team crafted a development plan based on OCC priorities and AGUA technical team capacity. The final development list and project plan were endorsed by the WEST Executive Committee in October 2020. The start date for AGUA Development Phase 5 was shifted to begin slightly later than past Development Phases to accommodate the work the AGUA technical team was engaged in in support of the [CDL, CRL, HathiTrust Shared Print Collaboration](#) to develop an [open collection comparison tool](#) for CRL's Print Archives Preservation Registry. Development work for Phase 5 began in late January 2021, after the launch of the [PAPR comparison tool](#), and is anticipated to run through the end of September 2021.

WEST originally prioritized 10 projects for development during Phase 5. Of these, five projects are completed and one has been deprioritized for development; 2 projects are in progress; and 2 are still pending. At this time the team appears on track to complete all prioritized projects ahead of schedule. Any time remaining in the development period may be used to develop features in Phase 5 projects that had been deferred, develop projects that had not been prioritized during Phase 5 planning, or for other WEST priorities as appropriate.

In addition to the prioritized projects, the AGUA technical team also designed and implemented a project to update the AGUA database to support the migration of retained holdings off of Archivers' shared print second symbols and onto their primary OCLC symbols. This work took place while they were also building the PAPR comparison tool and beginning development work for Phase 5 projects.

AGUA Development Phase 5 project team

Lucy Liu - Programmer Analyst
Jing Jiang - Programmer Analyst
Dana Jemison - Metadata Analyst
Judy Dobry - Technical Team Manager
Alison Wohlers - WEST Program Manager
Anna Striker - WEST Collections Analyst

Summary of Development Projects with Current Status

Ref #	Brief description	Foundational / Innovative	Priority Group	Estimated / Actual Time	Status
9	Surface historically unanalyzed collections	Foundational	Priority 1	2-4 weeks / 3 weeks	Complete
2	Enhanced Metadata Validation	Foundational / Innovative	Priority 1	>4 weeks / 7 weeks	Complete
4	Backend enhancements - Improvements to Disclosure-to-Bibliographic Data Matching and Storage	Foundational	n/a	>4 weeks / 9 weeks	Complete
12	Comparing volumes held against published run	Innovative	Priority 1	>4 weeks / 6 weeks	Complete
6	Add countdown of Builder budget to Review Proposed Journals page	Innovative	Priority 2	1-2 weeks / 2 weeks	Complete
1	Local discovery of WEST collections	Innovative	Priority 1	>4 weeks / tbd	In progress
8	Add unarchived holdings to On Demand Collection Comparison report	Innovative	Priority 2	>4 weeks / tbd	In progress
17	Support for Builder 'Calls for Holdings'	Innovative	Priority 1	>4 weeks / tbd	Pending
16	Support for voluntary archiving	Innovative	Priority 2	>4 weeks / tbd	Pending
15	Expanding archiving of high-risk journals (operationalizing Ref 9)	Innovative	Priority 2	2-4 weeks / n/a	Deprioritized

Completed development projects

Ref 9: Surface historically unanalyzed collections

- **Brief summary of project:** Identify high-risk (limited electronic availability) journals in WEST member collections that are not held by Archive Builders and so would not be eligible for archiving proposal under the normal collection model. Analyze findings and assess the value of building out processes in AGUA to detect and propose these journals as part of the regular biennial collections analysis.
- **Development activities:** The AGUA team generated a report that identified unarchived holdings that meet criteria for Title Categories 3, 4, or 5 (Silver and Gold) but are *not* held by any of the Builder or Builder-affiliated institutions.
- **Outcomes:** The report identified 4,065 journal families that, based on electronic availability, should be categorized as Title Category 3, 4, or 5, (moderate and high risk) but are ineligible for archiving proposal under the current collections model because they are not held by any of the Archive Builders. These journal families had very low average duplication across the group, indicating that these are scarce titles that would not provide significant opportunities for space reclamation. The project team does not recommend pursuing development of processes in AGUA to regularly identify and propose these titles in the course of the regular collections [analysis \(Ref 15\)](#), but instead to re-generate this report after the Cycles 12/13 unarchived holdings files have been ingested and conduct a one-time project to secure retention commitments for these scarce and at-risk materials.
- **Development time:** Initial estimate - 2-4 weeks; real time - 3 weeks total.
- **Project status:** Complete.

Ref 2: Enhanced Metadata Validation

- **Brief summary of project:** Enhance metadata validation practices to align more closely with CRL practices and to include more points in the archiving cycle where validation is performed and reported. This enhancement offers members support for ongoing metadata cleanup efforts and intercepts certain metadata issues early in order to correct and improve ahead of disclosure to outside organizations, streamlining those workflows.
- **Development activities:** CRL generously provided the code for their updated validation script to support aligning metadata validation practices between the two organizations and to help identify metadata errors earlier in the collection analysis/disclosure process. The AGUA technical team worked to review the code and analyze the output files and concluded that this is a very robust tool that achieves all of the validation requirements WEST had been considering, plus a few that had not been under consideration. The Metadata Analyst has provided feedback to CRL based on her analysis of the tool and its output reports. The tech team is not able to modify the code locally, but sees an opportunity to partner with CRL on future development work and implementation of their code.
- **Outcomes:** WEST will use the CRL validator on both unarchived holdings and disclosure records, which will help streamline analysis of member records. WEST will provide reports to members about specific problems categories:
 - Holdings out of range

- OCLC number change
- ISSN mismatch

This information will be provided in addition to the automated reports developed in Phase 4 and automatically sent to members by AGUA during file ingest.

- **Development time:** Initial estimate - more than 4 weeks; real time - 7 weeks (mostly analysis).
- **Project status:** Complete.

Ref 4: Backend enhancements - Improvements to Disclosure-to-Bibliographic Data Matching and Storage

- **Brief summary of project:** Update AGUA processing of MARC disclosure records and related bibliographic title data to improve database workflows and data that haven't been enhanced or improved since our initial implementation in 2011.
- **Development activities:** The AGUA technical team undertook development work to support several interdependent enhancements to the AGUA database:
 - Remove duplicate title records from AGUA's internal "warehouse" of bibliographic data
 - Merge holdings from duplicative titles
 - Update the match process to use OCLC number as the primary search key in OCLC where needed
 - Link disclosure and matching bibliographic entries to AGUA's OCLC cross-reference table, so that the full "set" of OCLC numbers (primary and cross-reference) are associated with each unique title
 - Fully refresh bibliographic title data in AGUA's bibliographic data warehouse
 - Create a script that can be run yearly to re-refresh bibliographic title data
 - Capture additional bibliographic data points to support other development projects for this, and future development cycles
- **Outcomes:** This project made a significant improvement to internal WEST processes and data that haven't been enhanced or improved since our initial implementation in 2011, and has laid the groundwork for future, potential development projects. This development work supports a number of enhancements that will impact AGUA's analytical and matching abilities and improve internal practices. This work was undertaken with an eye to the future as well as current needs, and WEST has the option to continue leveraging this development work in future projects. Specific outcomes of this project include:
 - One-to-one matching on OCLC number rather than ISSN ensures that the constituent's disclosure data is matching with the correct title data, that is, the same as what is reflected by their local catalog
 - Bibliographic title, publisher, control number data that is mapped to reports will be fresh and correct, and data will be refreshed every year prior to disclosure ingest
 - Duplicate title entries in reports will be eliminated
 - Diacritics will display correctly
 - Internal processing of disclosure and bibliographic will be more correct and efficient, building a strong foundation for other, current and future development processes
 - Data has been captured to support other projects in the current development phase, and in anticipation of future, possible development projects
- **Development time:** Initial estimate - more than 4 weeks; real time - 9 weeks.
- **Project status:** Complete.

Ref 12: Comparing volumes held against published run

- **Brief summary of project:** Compare the number of held volumes at an institution against the depth of the published backfile and calculate the estimated 'completeness' of the held run.
- **Development activities:** The tech team developed methods for calculating the estimated depth of a publication run both at the journal family level and at the title (primary OCLC number) level, and then compares this against the holding institution's calculated "depth of backfile" to generate an estimated completeness of the local holdings compared to the published run of the journal. This development leveraged enhancements made to 008 date normalization and storage developed in the [Ref 4](#) backend enhancements.
- **Outcomes:** This project was originally conceived to support the collections analysis and making archiving proposals, the development to calculate the depth of the published backfile at the title level has applications beyond the project team analysis. The project team determined that this information would be of value to all members in the All Archived Titles report and the On Demand Collection Comparison report, as this provides valuable information about the relative completeness of the archived copy when making local collection management and deselection decisions. Additionally, the project team solicited feedback from the Archive Builders on the value of seeing the relative completeness of their local holdings for journals WEST is proposing they commit to retain, which was found to have significant support from both new and continuing Builders. The new data points were added to the following AGUA reports:
 - Include journal family-level calculation in Summary report to support prioritizing 'complete' backfiles for archiving proposal
 - Include title-level (primary OCLC number) calculation in reports for Archivers' local holdings for proposed journals to facilitate identifying 'complete' backfiles when making commitment selections (also included in reports for Archivers' local holdings for committed journals)
 - Include title-level (primary OCLC number) calculation in the All Archive Titles report and the On Demand Comparison report to facilitate members' identification of complete archived backfiles to support local collection management decisions
- **Possible additional work:** The project team identified a number of additional AGUA reports where this information could be useful, but did not develop as they were determined to be out of scope for the project requirements. The team has identified the following opportunities for future enhancements:
 - Add title-level calculation to "WEST Holdings for Proposed/Committed Journals" reports (so-called "others" reports) to support Builders' ability to identify potential contributors for Silver/Gold archive building
 - Add 008 date_1/date_2 information to All Archived Titles and On Demand Comparison reports to support transparency and data sharing
 - Add title-level calculation to yearly Collection Comparison reports to support local deselection decisions
- **Development time:** Initial estimate - more than 4 weeks; real time - 6 weeks.
- **Project status:** Complete.

Ref 6: Add countdown of Builder budget to Review Proposed Journals page

- **Brief summary of project:** Include a dynamic Archive Builder Budget display on the Review Proposed Journals pages for Title Categories 3, 4, and 5 to help facilitate Builder workflows selecting archiving commitments while staying within their allotted Builder budgets without having to switch between screens in AGUA.

- **Development activities:** AGUA calculates Builders' archive creation subsidy encumbrance to help Builders track how much of their archiving subsidy their commitments will use and make adjustments as necessary ahead of submitting retention commitment selections (tracked on the [Archive Builder Budget Review](#) page in AGUA). The tech team brought this information onto the Review Proposed Journals page to support Builder subsidy tracking and to improve visibility of the impacts of retention commitment selections on the Builder budget. Additionally, the Archive Builder Budget Review page now calculates Builder Subsidies using proposed journals and saved commitment selections, improving the transparency of the initial archiving proposals and the impacts on the Builder budget to support Builder workflows identifying where they have been 'overproposed.'
- **Outcomes:** This development brings this information onto the Review Proposed Journals pages for Title Categories 3, 4, and 5, listing the following information
 - *Archive Creation Subsidy:* Calculated using the estimated volumes committed for the Title Category currently being viewed, multiplied by the per-volume payment amount set by WEST (\$4 for Silver volumes, \$7 for Gold volumes), e.g.,
 - Title Category: 4
 - Estimated Volumes Committed: 7,447 (updated dynamically [previous development])
 - Archive Creation Subsidy: \$52,129 (updated dynamically)
 - *Archive Builder Budget Remaining:* Calculates the total subsidy for all proposed and committed TC3, 4, and 5 journals and compares against the total budget set for the Archive Builder, e.g.,
 - Archive Builder Budget Remaining: \$xx,xxx (updated dynamically) / \$100,000.00 total budget

Additionally, the Archive Builder Budget Review page has been updated to display the subsidy calculation based on archiving proposals as well as saved and submitted commitment selections. Previously, this page only displayed the subsidy calculation for saved and submitted commitment selections, requiring Builders to begin the selection process before knowing the extent to which their budgets were overproposed (it is WEST's usual practice to over-proposed so Builders have some flexibility in selecting which journals to commit to retain). This enhancement will improve Builders' understanding of how WEST is proposing these journals and the impacts on their archive creation budgets ahead of any journal review and selection.
- **Next steps:** Due to perceived limits on Builder bandwidth during the development period and given the time until Builders will be able to use this feature, user testing was deferred. This and other AGUA development projects will be presented to Builders at a future Archive Builder Forum, and any feedback will be included in the AGUA technical backlog for development at a later date.
- **Development time:** Initial estimate - 1-2 weeks; real time - 2 weeks.
- **Project status:** Complete.

In progress development projects

Ref 1: Local discovery of WEST collections

- **Brief summary of project:** Enhance the visibility and discoverability of WEST retained collections by library staff and end users at member institutions. Close the gap between WEST's stated policy that all WEST members have access to all WEST-retained materials, and the lack of local discovery options. Users (library staff and patrons alike) must know about and search outside databases in order to find what titles they have access to and what has been retained by partner institutions on behalf of the group. This development project supports the "shared" in shared print by bringing these retained titles into local discovery systems so library staff and end

users are able to see what titles are available to them through ILL with minimal requirements for prior knowledge or additional search steps on their part. This project makes this data available to member institutions for display in local systems, which can then connect to local resource sharing systems.

- **Development activities:** Following OCC's decision to pursue local development of a MARC record file that can be distributed to members, the team
 - has written detailed technical specifications for building the records ([Ref 4](#) enhancements support creation of bibliographic records),
 - is working to define how they will query the AGUA database to gather the necessary data and how records will be output from the system,
 - is working to identify bad datasets that will need to be resolved before implementation, and
 - is collaborating on identifying outstanding questions, both for the AGUA team to answer and which need member input.
- **Additional activities:** The project team spent a significant amount of time investigating member needs, capacities, and preferences to help inform development decisions, and to engage in conversations with a major library systems vendor to discuss ways shared print can be more fully integrated into their native system architecture.
 - Initiated discussions with Ex Libris on how they could support integrating shared print collections into Primo/Summon. Our latest discussion took place on 03/30/2021 with representatives from the Ex Libris Product and Systems Architecture teams. The project team came out of this meeting feeling positively; Ex Libris has conducted internal discussions about how they might address this within their own architecture and is currently working with the project team to schedule a follow up meeting.
 - Distributed a survey to WEST Primary Contacts to gather more information about their local discovery and technical needs and preferences. The findings support taking a combined approach to closing the discovery gap between the stated WEST Access Policy and technical challenges that limit easy and convenient discovery of shared print materials in local systems.
- **Outcomes:** Tbd.
- **Development time:** Initial estimate - more than 4 weeks; real time - tbd.
- **Project status:** In progress.

Ref 8: Add unarchived holdings to the On Demand Collection Comparison tool

- **Brief summary of project:** Expand the functionality of the On Demand Collection Comparison tool to return results for matches to WEST members' unarchived holdings in addition to archived holdings to help facilitate identification of scarcely held titles and support formal or informal retention of these high-risk materials.
- **Development activities:** The tech team is currently preparing to begin development on this project. The tech team is reviewing the functional requirements and will begin development shortly.
- **Outcomes:** Tbd.
- **Development time:** Initial estimate - more than 4 weeks; real time - tbd.
- **Project status:** In progress.

Pending projects

Ref 17: Support for Builder 'Calls for Holdings'

- **Brief summary of project:** Streamline Builders' 'calls for holdings' workflows by providing an interface for holdings comparison, contributor selection, and communication management.
- **Development activities:** The formal development of this project has not yet begun. However, the project team has engaged the Archive Builders to review and provide feedback on the requirements for this project, and is working to incorporate their feedback and priorities into the final requirements that will be the basis for the development team's work.
- **Development time:** Initial estimate - more than 4 weeks; real time - tbd.
- **Project status:** Pending.

Ref 16: Support for voluntary archiving

- **Brief summary of project:** Integrate the title nominations process into AGUA by providing an interface where member institutions could submit 1) journals they want to contribute to the archive, or 2) journals they cannot contribute but think would be important/valuable retentions to secure.
- **Development time:** Initial estimate - more than 4 weeks; real time - tbd.
- **Project status:** Pending.

Deprioritized project

Ref 15: Expanding archiving of high-risk journals

- **Brief summary of project:** Develop processes in AGUA to routinely identify and propose for archiving high-risk journals in the WEST unarchived collections that are not held at any of the Archive Builder institutions.
- **Development activities & outcomes:** Based on the findings from [Ref 9](#), this development project was recommended for deprioritization, which OCC endorsed. WEST will not build out functionality in AGUA to routinely identify and propose high-risk journals held only by non-Builder institutions in this development phase. WEST will reuse the code developed for Ref 9 to re-generate this report after the Cycles 12/13 unarchived holdings ingest is complete to identify and secure retention commitments for scarce and high-risk journals at non-Builder members and secure retention as a one-time project.
- **Development time:** None.
- **Project status:** Deprioritized.
 - Development of routine processes is shelved, but the project team will likely employ the code developed in Ref 9 for one-time projects in the near future.