

Against the Grain Article, November 2010 issue

Heading West: Circling the Wagons to Ensure Preservation and Access

by **Emily Stambaugh** (Shared Print Manager, California Digital Library, WEST Assistant Project Manager)
<emily.stambaugh@ucop.edu>

Research libraries have inherited a legacy of print duplication; duplication that made sense in its time to ensure institutional competitiveness. But a network-wide shortage of storage space requires us to reduce the physical footprint of retrospective collections. Research libraries seek ways to make informed decisions about what to preserve and what to withdraw. The recent growth in last copy agreements suggest there is real momentum in the community to find collaborative solutions.¹ But taken together, these efforts do not reach the scale that is needed to address the systemic and long-term shortage of space to house physical collections. Among the factors that have hampered such efforts are: the absence of business models, organizational structures, collection decision-making models, disclosure systems and incentives to create and sustain trusted archives. Large scale collection consolidation has real operational costs that surpass existing consortial capabilities. A network level (regional, national, international) solution is required. Research libraries and consortia in the western United States have prepared a business model and operational structure for a Western Regional Storage Trust (WEST) which is designed to support network level archive creation services to preserve the scholarly record, provide access, when needed, and manage reallocation of space.

About Aggregate Print Journal Collections

Print journal archives are ideal candidates for space reclamation for reasons that are well-known; large amounts of shelf space can be reclaimed with a relatively small number of titles (and decisions about those titles). To put the size of the aggregate print journal collections in perspective, there are about 4.18 million print serials in WorldCat and the average number of libraries that hold a title is about nine. At the high end of the duplication spectrum are roughly 10,000 titles in Portico and JSTOR with average holdings of 250 and 600 libraries, respectively.² While titles in Portico and JSTOR are the usual suspects for collaboration, there is clearly a need for collaboration on other electronically held titles and on titles published only in print. As much as 40% of the refereed scholarly journal literature is not available in electronic format. Some 56% of peer-reviewed history journals are published in print-only format. By contrast, almost 80% of the refereed medical journal literature is available online.³ There is an economic sweet spot for consolidating print collections, and it can be found where duplication is highest and where holdings can be compared in semi-automated ways for ready decision-making. The extent of possible candidates may be great enough to remedy library and storage facility space problems without dipping into more costly monograph deselection projects or more risky restrictions on collection growth.

¹ The Center for Research Libraries has inventoried recent last copy and shared print agreements <<http://archivereg.crl.edu/project/index>>. Most are focused on journals, some on government documents, but none extend in scale to the scope of the aggregate collections that require attention.

² OCLC Research. 9/16/2008.

³ Ibid.

In the western region of the United States, an initial analysis of print journals held by thirteen research libraries and their storage facilities revealed at least 60,000 commonly held journal “families” (current + previous titles of a journal). About 30,000 are held by 3 or more institutions in the region and about 17,000 by 5 or more (up to 21 copies). These duplication rates are probably understated at the title level, as a significant number of records supplied for analysis could not be meaningfully compared due to lack of match points (ISSNs). Further analysis is underway to compare regional rates of overlap network (national, international) level overlap.

Table 1: Levels of Print Duplication within WEST Planning Libraries

Duplication Level	# Copies	# Journal Families (current + previous titles)	%
High	5 or more	17,233	28%
Moderate	3 or 4	13,381	22%
Low	1 or 2	29,966	49%
Total		60,580	

On Collaboration Scale

The scale of collaboration requires careful consideration: state, regional, national? Creating archives at a certain pace has real operational costs and requires dedicated staff trained in project management and validation. In 2009, the University of California Libraries considered going it alone with a consortial archiving service that would serve the ten UC campuses. Experiments were conducted with low level (issue) validation and different organizational models (campus distributed effort and storage facility based services.) We found that on a per-unit and gross productivity basis, the most effective model was to concentrate this work at (and move materials to) storage facilities. A proposal was prepared for a lightweight service to consolidate UC holdings at its storage facilities. It was immediately recognized that the resulting archives would benefit a broader library constituency and that economies of scale could be gained if partners were cultivated beyond the consortia. Parallel conversations with other libraries in the state suggested there was a real desire to support shared preservation and archiving commitments. It was also acknowledged that a rich history of gifts and exchange of physical materials between libraries in different states might serve as a useful model for completing physical collections and could enable a partnership beyond a single state. Furthermore, it was felt that a broader partnership with inter-institutional dependencies on shared archives would ensure sustainability of the service and the archives and create a fabric of trust and operational capabilities that could be leveraged in future collaborations.

All of these factors combined suggested that real tangible and intangible benefits could be gained with a regional partnership. In terms of cost-benefit, the Western Regional Storage Trust, as proposed, will achieve similar results to the earlier consortial proposal at less than one-tenth the cost to the University of California. Other WEST partners will experience similar economies of scale. What partner libraries forgo to gain this benefit is sole discretion over title selections. Group priorities will outweigh local preferences for archiving. The collection model for WEST is designed to balance these sometimes competing needs.

Nuts and Bolts of the Western Regional Storage Trust

In Fall 2009, with support from the Andrew W. Mellon Foundation, an initial set of research libraries and consortia were identified to create a plan for a distributed retrospective print journal archiving service (WEST). Guided by Lizanne Payne, WEST's project consultant, and a core planning team including Ivy Anderson (CDL), Sherrie Schmidt (ASU), Brian Schottlaender (UCSD), Lizanne Payne and me (CDL) and supported by several functional working groups, the Trust has been designed to scale. It includes new organizational and business models and new modes of collection decision-making and disclosure.

The long-term goals for the Trust are to preserve the scholarly print record at the lowest possible cost through a coordinated system of persistent archives and network level disclosure. An additional goal is to create significant opportunity for space reclamation in libraries and storage facilities. The model can be replicated and supports reciprocity with other regional efforts. These goals will be achieved through low cost archiving of a single print copy of titles that are also available and preserved electronically. At the same time, Trust participants will invest effort in proactively building and validating archives for print only journals with moderate to high duplication in the region. Among the 13 planning institutions, approximately 8,000 journal families (275,000 volumes) were selected for archiving, providing the potential to deselect an estimated one million duplicate volumes in libraries and storage facilities and freeing up the equivalent space of one mid-sized ARL library.

WEST planning partners agreed that the service needed to provide avenues of participation by diverse partners with different institutional motivations for collaboration. Some institutions would seek to secure access to backfiles, when needed, without having to maintain archives on site (needs based access). Some would have already divested many print holdings but would value access to titles never previously held (extension of breadth) or to support value-added services (digital access). Others might seek operational support for ongoing archiving commitments (stewardship). And there will always be free riders. To satisfy these diverse needs and achieve greater buy-in and therefore sustainability, the Trust has been designed to work on multiple categories of titles in parallel and provide avenues for both content and financial contributions.

From Storage to Archiving

The Trust involves a transition from storage to archiving, which is as much a shift in mindset as in operational approach. Trust partners proactively select, build, and store a set of print journal backfiles in designated facilities focusing on titles that can provide substantial benefit to the majority of partners. Titles identified for archive creation and retention are aligned with specific storage facilities and libraries (archive providers) based on existing depth of holdings. This data-driven approach to aligning backfiles with archive locations effectively transforms storage facilities (and some libraries) from passive receivers of uncoordinated, incomplete deposits to sites where archives are actively created and curated.

Business Model

The business model provides avenues for large and small libraries to participate in different capacities and distributes costs equitably across a broad partnership. The model also includes mechanisms to compensate archive providers (storage facilities and some libraries) for archive creation services for higher risk titles. The initial membership term is for five years with 12 months notice for withdrawal. And archive providers agree to a 25-year retention period (through 2035), a commitment that survives membership.

Membership fees support only those costs that a single institution or consortium cannot support on its own, including validation of a planned number of volumes each year and project management. Trust members support all other costs in kind (deselection and access services, transfer of materials to archive providers, etc.) Membership levels are determined by collection size, and archive providers receive a discount based on the size of archive held as an incentive to participate and indirect compensation for ongoing storage costs.

Archive providers for higher risk titles are directly compensated to hire staff to process archives, thereby ensuring a certain pace of archive consolidation. This direct compensation not only provides incentives to serve as a provider but also supports other members' needs for a rapid timeline to make informed collection management decisions.

New Approaches to Collection Decision-Making

The collection model for WEST allows partners to make collection decisions for large classes of material and to balance efforts on different classes. Titles are categorized based on risk, using risk management principles, such that low risk titles can be archived with the lightest weight methods and higher risk titles receive more attention. The collection model is informed by Ithaka S+R's optimal copies research⁴; Ithaka S+R's recommendations for what to withdraw⁵; and an initial analysis of print journal titles held by WEST storage facilities and libraries.

Risk is defined as the likelihood of loss of content, loss of access, or a stewardship failure in the region as deselection occurs for print journal backfiles. A print title that is electronically available, digitally preserved, and widely duplicated in print in the region is at the lowest risk on all three counts. A

⁴ Yano, Candace, et. al. *Optimizing the Number of Copies for Print Preservation of Research Journals*. University of California, Berkeley. October, 2008. Advance copy. Submitted to *Interfaces*.

⁵ Schonfeld, Roger and Ross Housewright. *What to Withdraw: Print Collections Management in the Wake of Digitization*. Ithaka S+R, September 29, 2009.

title that is only available in print and is moderately duplicated in the region may be at higher risk. Some factors that mitigate risk for an individual title include electronic availability of the backfile, post-cancellation access permissions to the electronic backfile, level of duplication within WEST, level of duplication beyond WEST, presence of an existing, validated print archive and access to a validated print archive.⁶

The Trust has identified six categories of risk or “title categories.” Titles are categorized by their format of publication and digital preservation status. Within each category, candidate titles are selected based on various additional criteria (e.g. scholarly/academic, years of publication, subject) but most importantly based on the current print duplication level within the region. Uniquely held titles are not candidates for the Trust; presumably these will be retained by the institution regardless of a cooperative effort.

The Trust has also defined several archive types analogous to the Olympic medal theme (e.g. Bronze, Silver, Gold); archive types explicitly define the level of effort to be placed on archive creation. Bronze is intended for low risk categories. Very little effort is placed on these archives; holdings are disclosed, but not validated or moved to storage. Silver is intended for moderate risk titles and includes an organized call for holdings, volume level validation for the completeness of a run and disclosure of holdings and gaps. Gold is for higher risk titles and includes an organized call for holdings, issue-level validation for completeness and condition, and disclosure of holdings, gaps and conditions. Storage facilities are preferred (or required) for Silver and Gold archives.

The relationship between title category and archive type ensures predictability and transparency across the Trust; partners know what level of effort will be placed on a title with certain characteristics, and it keeps decision-making overhead low. Archive providers will work on multiple title categories in parallel each year to gain experience with the operational requirements associated with each.

Table 2: Title Categories and Archive Types

Title Category	Duplication Level in region sought in candidate titles	Risk Level	Archive Type	Number of Archive Copies ⁷
1 Print and Electronic plus Digital Preservation (CLOCKSS, Portico)	High	Low	Bronze	1
2 Print and Electronic, no Digital Preservation, publisher e-journal packages	High	Low	Bronze	1

⁶ Initially, the level of duplication beyond WEST and image density and quality will not be taken into consideration in WEST’s collection decisions but may be incorporated later in the Title Category definitions as metadata becomes available for those aspects.

⁷ Given historical use rates, one copy is viewed as sufficient to meet regional demand. Silver and Gold WEST archives will be validated, and as such, will be eligible for contribution to a broader network of optimal copies..

3	Print only with selected full-text access through aggregator databases	High	Moderate	Silver	1
4	Print only with electronic abstracting and indexing	Moderate - High	High	Gold	1
5	Print only, no electronic access points	Moderate - High	High	Gold	1
6	JSTOR ⁸	N/A	Moderate	Silver	1

The Trust has also outlined standards for issue and volume-level validation to be applied to each archive type. The standards explicitly define “a reasonable level of effort” to be placed on verifying completeness and condition of a print backfile.⁹ The standards provide uniformity and transparency as multiple archive providers engage in validation work. They also temper our tendencies to seek perfection, which is unnecessary in an optimal copies environment.

Disclosure and Collection Analysis

Disclosure is critical in a networked collection management environment. One region’s commitment to retain a print journal backfile might facilitate another’s collection management decisions to duplicate or not. WEST is planning to use existing OCLC WorldCat capabilities to disclose archival commitments.

Disclosure includes several activities: the registration of an archival commitment for a title, the explicit declaration of preservation actions taken to verify completeness and condition (i.e., the level of validation) and the identification of specific holdings, gaps and conditions in the backfile.

Decisions to build or declare an archive are made in the context of aggregate print holdings and existing shared print archives. During the planning phase for the Western Regional Storage Trust, the California Digital Library built a proof-of-concept prototype collection analysis system. Over a million records were supplied by thirteen institutions (libraries and storage facilities) and shared print initiatives. Data from Ulrich’s was used to normalize and enhance library-supplied data and to trace title histories. And finally, normalized holdings were compared at the title level (not item level) to identify overlap and refine lists for each title category.

The collection analysis and disclosure requirements for collaborative archiving are non-trivial and require systems infrastructure and standards of practice. The Center for Research Libraries is planning to develop a production-level collection analysis and disclosure system to support the efforts of WEST and other consortial archiving initiatives.

Access

⁸ The Trust will grandfather in existing built archives including the Orbis Cascade Alliance’s Distributed Print Repository (DPR) and the University of California’s CoreSTOR and Institute of Electrical and Electronics Engineers (IEEE) archives. WEST priorities for JSTOR are to complete gaps in existing shared collections.

⁹ The standards are informed by Ithaka S+R’s optimal copies research, the University of California’s experience with the JSTOR Shared Print archive (which includes a form of issue-level validation in preparation for page validation) and experiments with issue-level validation for two shared print projects: the IEEE and CoreSTOR projects.

WEST planning partners acknowledged that print journal backfiles are declining in use. Indeed, in UC's experience with the JSTOR Shared Print Archive, an assembled print backfile is far more likely to be used for digitization/redigitization than for direct access by researchers. In this context, WEST partners agreed not to restrict use to member institutions, which would require additional investment in systems development. WEST archives will be made discoverable and accessible to researchers through existing interlibrary services and protocols. Whenever possible, digital scans or photocopies are provided before physical volumes to reduce wear and tear and avoid re-validation.

Areas for Future Research

The Western Regional Storage Trust aims to move into production in January 2011. As backfiles are consolidated and the regional model is possibly replicated elsewhere, some new areas of research might emerge. Future lines of inquiry might include an evaluation of components of the WEST model that might be applicable to print monographs, exploration of the network effects of one region's retention commitments on another, stewardship expectations from both user and university administration perspective, the value of assembled backfiles to publishers, aggregators and other digitization partners¹⁰, and refinement of the optimal copies framework in the absence of a page validated archive¹¹. In the long term, shared print efforts will probably focus on collaborative prospective collection development for journals, monographs and other forms of publication. The landscape for collaboration and print publishing will have shifted by then, offering a bright and interesting new future for cooperative collection development.

¹⁰ Publishers often do not maintain complete backfiles of their publications and may find a complete resource valuable and worthy of support and/or partnership. IEEE has showed interest in UC's archive consolidation effort to fill in gaps in its digital backfile.

¹¹ Ithaka S+R and Candace Yano are planning to refine the optimal copies research conducted in 2008. UC Libraries and others will supply data about levels of validation, and disclosed conditions and gaps to facilitate that research.