

Constructing a Plan for a JSTOR Paper Repository

Overview

As scholars and students become increasingly comfortable using electronic materials, the need for paper repositories to stand behind digitized versions of converted documents has become more important. We are in a time of transition and the methodologies for archiving electronic documents are not yet in place, nor are the resources to accomplish the task. Paper is a stable medium, and methods for preserving paper documents are well understood and considered reliable. So systematically establishing paper repositories for digital materials is a timely endeavor, and would complement JSTOR's digital archiving efforts.

From the earliest days of JSTOR, the need for a paper archive to back up the electronic collections was articulated and framed as an organizational goal. In one of the first JSTOR planning documents, it states, "JSTOR should play a proactive role in helping to fund and establish regional depositories for the paper materials to reassure librarians that the documents will not be lost for all time." Libraries continue to seek ways to reduce costs and some are beginning to discard backfiles of journals when JSTOR is available at their institutions. There are indeed savings to be realized, but care must be taken to be sure a systematic solution has been implemented so that a sufficient number of copies remain protected and available. Similar movements to reformat scholarly materials (newspapers to microfilm) have led to the loss of some originals that has generated considerable controversy. We should not repeat past mistakes.

I. The Mission:

To build a distributed and worldwide system of failsafe paper repositories designed to protect, preserve, and provide access to complete collections of the journal runs available in the JSTOR digital archive.

II. Agenda

A. Defining a System of Dark Archives

JSTOR intends to provide for the safekeeping of the journals in our collections and extend their useful life for scholars and researchers in the future. Our first course of action will be to devise a plan for the establishment of a "dark" archive that is designed to preserve a set number of JSTOR journals stored in a chosen number of locations. We are defining a dark archive as a closed, protected set of JSTOR journals designed strictly for preservation. This archive will not be opened to the academic community unless every other means of access to the original content or artifact should completely fail. By collecting and storing these volumes in facilities that are designed for their long-term preservation, JSTOR can provide a measure of assurance to the academic community that the original copies will be available, should there one day be a need.

The task before us on September 4 will be to address the issues surrounding dark archives and to create a plan that can be implemented in a relatively short time frame. Below are most of the issues that we will need to address. We believe that some of these topics have a straightforward resolution and can be addressed prior to our meeting by explaining the issue and what we see as a logical solution. Other issues are more complex and will, therefore, require a more thoughtful discussion by our group. It is our hope that this document can frame the issues we need to discuss and focus our efforts for a productive meeting.

1. Validation of the Complete Set of JSTOR Journals

The creation of dark archive leads to the critical question of what exactly will be “kept.” Defining a truly full run of JSTOR journals and validating its completeness is of very high importance. JSTOR has considerable experience here, as we have had to define what volumes to digitize, reaching back to Volume 1, Issue 1. For the past seven years, the JSTOR staff has carefully worked through content issues -- considering advertisements, special supplements, and additional materials -- with the goal of providing a comprehensive set of volumes for each title. Our goal in this section is to outline the process we use for assembling complete runs and solicit your input as to whether this validation process is acceptable for the creation of this dark archive.

The first step JSTOR takes is to conduct a title history for the individual journal. Staff librarians research potential issues that are associated with the journal name – such as title changes, supplements, etc. – to determine all the parameters of the title. This is done through the use of academic libraries' online catalogs, in order to trace the publication history of a specific title. There are also indicators in the catalog that provide information about other titles that might be closely related to the title and should be included. This information also lists the journal's publication years, the frequency, and special supplements. The title history searches are documented and this becomes the map or standard used for acquiring the JSTOR journals. This step gives JSTOR a sense of how frequently the title is published and lays the foundation for further inquiry.

The next step takes place after the paper journals begin arriving at JSTOR. The production staff begins to get a sense of the structure of the run as they note the patterns found in the volumes and issue sequencing. Special care is given to the way a journal's volumes are numbered, in order to spot potential gaps in issues. The staff also use library catalog holding statements, which is a listing of what is found on the library's shelves, when they think a volume might be missing. If the issue in question appears on a library holding statement, they know it exists and should, therefore, be included in the JSTOR journals.

One of the last steps in the verification process is for the staff to go through journals, page by page. They look for consistency and no gaps in the numeration of the pages. They also look for notices that would indicate a special supplement was included in a particular volume. For example, if the production staff finds an editorial statement that comments on a supplemental issue that accompanies that volume, JSTOR will go out and track down the supplement to add to the content of that volume. For example, a user once contacted JSTOR and said that she noticed a special issue was missing from the volumes that were digitized for that particular year. JSTOR staff researched and acquired the special issue for its collection.

Even after this series of validation checks, there are still some pieces of volumes that we don't have. Occasionally, a JSTOR volume will be missing its cover. While JSTOR makes great effort to offer volumes with complete content, but we do not make efforts to acquire missing covers. In the process of accessioning journals, librarians often remove covers and advertisement pages. We try to find copies with those pages, but they don't track them down if they are the only missing piece in an otherwise comprehensive volume.

In conclusion, JSTOR's method for validating complete backruns journals attempts to compile a complete set and locate any pieces that are missing. Our goal in this piece of our discussion is to determine if this process is satisfactory for this project.

An additional item worth mentioning is the issue of journals that are difficult to locate. Most journals that are used for scanning are either donated or sold to JSTOR. When a journal cannot

be acquired in this manner, JSTOR has been able to borrow the volume for a period of time. While this scenario is infrequent, the possibility of rare or small-circulation volumes being hard to acquire does exist. When these instances occur, the volume is loaned and the staff continues searching for the volumes after the scanning is complete and the journal is returned. With regard to the paper repository collection, we would, of course, make every effort to find the missing volumes, but some issues may take longer to locate and acquire than the majority of the collection. Therefore, gaps may exist while the search for the missing volume continues. We realize that this may prove to be a significant challenge in assembling the dark archive and we may need to address this issue if we are unable to acquire the hard-to-find volumes.

2. Ownership

After researching a number of off-site shelving facilities across North America, it is clear that most depositing institutions generally retain ownership of all materials that are deposited for storage. While the responsibility for maintenance is transferred to the facility and materials are accessioned into the site's system, the depositing organization keeps all rights and continues to list the materials in its own catalog system or in shared systems, when materials are jointly owned, and reference the new, off-site location. We recognize that volume count for ARL statistics continue to be important to libraries and are a factor in libraries' decisions regarding collection management.

However, we expect that JSTOR can create this archive primarily through discarded and deaccessioned volumes. One issue that might warrant discussion in this area would be to determine if we need to keep track of each depositing institution for recording the provenance of each donated journal.

3. Characteristics of a Provider: The Facility Environment

Preservation of information is an area of critical concern for both librarians and scholars in the 21st century. In the past 200 years, preservation primarily occurred at the item level, and was generally reactive. It is only within the past two decades that libraries have been exposed to research that can help their preservation efforts become more proactive and extend the useful lives of the materials in their collections. Building off-site storage facilities with strictly controlled climate conditions is one way that librarians are taking action. Part of their decision-making has been informed by the emerging research in the materials sciences that addressed the variables affecting preservation. This research includes the work done by Don Sebera, Library of Congress, through his "Isoperm Theory,"¹ which has proven that temperature and relative humidity are intricately tied to the long-term survival of organic (paper) materials. Further research by the Image Permanence Institute at Rochester University corroborated his findings and argues for consistently cool, dry conditions as regularly as possible. The Institute created a Preservation Index (PI) that expresses how temperature and relative humidity affect the chemical decay rate of collections.²

Other conditions that affect the life of organic materials include proper lighting (as free from UV light as possible), protection from pollutants and particles, protection from fire and natural disasters, back-up and emergency control-systems, and procedures for proper care and handling. Yale University and other academic institutions considered these conditions as they sought to build their own library shelving facilities. Paul Conway, the Yale University Librarian at the time, details the important elements of this effort in his chapter, "The Preservation Environment" taken

¹ "The Preservation Environment" by Paul Conway, a chapter in Library Off-Site Shelving: Guide for High-Density Facilities, edited by Danuta A. Nitecki and Curtis L. Kendrick, 2001.

² Ibid.

from the book, Library Off-Site Shelving: Guide for High-Density Facilities, edited by Danuta A. Nitecki and Curtis L. Kendrick.

The conditions we would like this group to consider are – at minimum – temperature and relative humidity. The current standards for paper preservation given by the National Information Standards Organization (NISO)³ recommend a maximum temperature of 65 degrees Fahrenheit with a daily fluctuation of +/- two degrees and an optimum relative humidity between 35 and 65 percent, with a daily fluctuation of +/- three percent. These two variables have a noticeable effect on the life of paper. In addition, we have included the preservation chapter from the Nitecki/Kendrick book as reference material because it represents many of the best practices used for creating a preservation environment in a storage facility. As we consider the type of facility desired for JSTOR's dark archive, we believe these variables should be part of our discussion.

4. Number of facilities needed and locations

The discussion in this section will focus on number of facilities needed for the dark archive only. As mentioned in our first email, we have decided to section this project into manageable tasks. This is only the first piece of the repository system that JSTOR plans to build, with our next step being the development of a series of accessible archives. In preparation for this meeting, we have had numerous informal conversations with librarians in the academic community about the topic of “how many” dark archives would be needed to secure the survival of these journals. The response has varied from one well-secured collection to comments as vague as “several.” We believe that the number of copies for a truly dark archive will be few, but the final tally is – at this point – not clear. Questions that arise for discussion are: How many copies are “enough” in order to establish a baseline preservation and, if needed, future access? If the answer is one copy, does the geographical location matter? If two or more copies are needed, should they be located in separate parts of the country? Or is location a factor that does not need our consideration? Are the locations of dark archives tied to location of possible “light” archives?

5. Collecting the titles

The starting point for the creation of the JSTOR paper repository (whether dark or light) is to amass the complete collection. We must therefore create a plan for collecting JSTOR journals as they are discarded or become available and welcome your suggestions and input. Facilitating communication between institutions within the community to learn where these journals are and which libraries might donate the journals in the near future will be an important first step to building this repository. What we would essentially be doing is building a net with which to catch JSTOR journals as they become available. The goal would be to capture as many available copies as possible to fill the various collections we expect to establish. The more journals we have on hand, the easier it will be to take the next step of organizing a series of accessible regional depositories nationally and around the globe.

Collecting journals is not a new task for JSTOR. Our staff has experience in locating and obtaining the best possible copies of the journals. We believe that your advice and input are valuable and we would like to brainstorm with you to find the best methods for contacting libraries and learning of their current and future plans for JSTOR journals. This may involve discussions about how we get libraries to contribute volumes and how to motivate them to act after making decisions to release journals. We might also need to discuss logistics. In addition to these queries, we have listed a few additional questions we would like you to consider:

³Taken from NISO TR01-1995, Environmental Guidelines for the Storage of Paper Records by William K. Wilson. <http://www.niso.org>

1. What are the best means of informing libraries of our plans for these journals?
2. Should the journals be moved to a centralized location for validation and processing or should we pursue a regional approach?
3. Who should bear the costs for transporting the journals?

We look forward to hearing your thoughts on this subject.

6. Legal agreements and logistics

Working through the details of a formal agreement with an off-site shelving facility will require guidance from our legal staff and consultation with possible shelving facilities. Some of the issues that we will need to address in the terms of an agreement would include facility standards, content to be deposited, shelving, rights of ownership, liabilities, costs, shipping and handling agreements, processing agreements, contact information, forms of communication, and terms of renewal. A provision might need to be inserted that would include a sampling of content at the end of the contract to verify proper preservation of the paper copies. We anticipate formulating a legal agreement that would span a period between two and five years.

The costs associated with archiving this collection vary according to facilities and the environmental conditions established. As expected, the better the conditions, the higher the cost of storage. The estimates also vary according to additional costs. Some estimates make provisions for access to the archive. None of the estimates include the upfront costs of preparing the volumes for storage (cleaning, bar-coding, boxing, and shipping). The costs for storage prior to archiving are also not included in these estimates. We will provide a sheet that details some of the anticipated costs and the associated variables on September 4.

B. Development of Future Accessible Repository Systems

1. Defining Accessible Archives

Should time permit, we would like to move forward and begin discussing what would be needed to develop an accessible, or "light" system of archives in order to complete a comprehensive JSTOR paper repository. We define a light archive as one that is stored in a protected, preservation environment and made available to the academic community through a well-defined set of policies and procedures.

2. Trigger Events

We have purposely shifted this agenda item from the dark archive discussion to the light archive section because we believe the two archives will be linked, in some manner, through the establishment of "trigger events" associated with levels of access. Therefore, addressing the trigger events in the context of levels of access within a complete system of JSTOR paper repositories might be a good way to begin.

III. Wrap Up: Next steps

We anticipate that there will be a need for follow-up after our initial meeting. We have set aside a period of time to conclude the discussion and set forth the next steps for our group.