

# Total Digital Access to the League of Nations Archives: Digitization, Digitalization, and Analog Concerns

Colin M. Wells, United Nations Office at Geneva; Geneva Switzerland

## Abstract

*The terms digitization and digitalization have often been used interchangeably, confounded, and misunderstood [1]. Case studies on archives digitization, which can be understood to be the transformation of analog resources to digital formats, are plentiful, but there appears to be less discussion of digitization as but one aspect of the digitalization of archives, here meant to include the broader digital transformation of business models and services. The example of the Total Digital Access to the League of Nations Archives project can be used to examine how comprehensive digitalization, including digitization alongside traditional analog concerns, have shaped the definition of project workflows and operations within the more limited sphere of its digitization operations.*

## Introduction

In 2017, with the support of a generous grant from a private Genevan foundation, the United Nations Library Geneva, Institutional Memory Section (IMS) launched a massive five-year project to digitize the entire League of Nations Archives, estimated to contain some 15 million pages of content. This project, called Total Digital Access to the League of Nations Archives Project (LONTAD), will allow for the comprehensive study of this rich and unique collection and will help better illuminate the diplomacy of the interwar period.

The LONTAD project will go beyond a “simple” digitization project, however, and will also encompass physical preservation, digital preservation, and the provision of online access to digitized content. The goals of the project are to provide true global access to the collection, while ensuring physical and digital preservation of the original and digitized materials, and to provide new pathways for research into the collection via digital humanities and data analytics methodologies. Thus, the project aims not just to digitize, but rather to *digitalize*, the archives. We use the term *digitalize* here to mean the broad transformation of the existing archives activities towards digital business processes, which include everything from the provision of online finding aids and reference services, to less visible, and perhaps less considered, aspects such as integrated physical inventory and loans management, a process which had already been started before the project began.

This digitalization impetus is also expected to be integrated with the ever present analog concerns that all archives must handle, and here we refer to issues such as physical preservation, environmental control of storage areas, as well as the physical risk inherent to the digitization process. In addition to the typical archives activities of providing access to materials, analog concerns for UNOG also address a range of communication activities that include internally and externally hosted exhibitions and a range of activities that will celebrate the centenary of the founding of the League and the establishment of modern multilateralism under the umbrella of the UNOG program “100 Years of Multilateralism in Geneva

1919-2019”. This program will cover a range of event types over some 18 months in 2019 and 2020.

The confluence of all of the concerns and organizational contexts mentioned above has greatly affected the design and implementation of the LONTAD project workflows. This paper will address how these issues have been incorporated into the project workflows.

## The League of Nations Archives

The League of Nations archives are the definitive record of the League, which existed from 1919-1946 and was created as part of the Treaty of Versailles as the first global intergovernmental organization, with the essential objective of preventing war while promoting international cooperation and establishing systems of international law. While the onset of the Second World War led to an initial historical evaluation and public opinion of the League as a political failure, the League is nevertheless now seen by contemporary historiography as having a critical influence on the structure and activities of the United Nations, and it broadly shaped much of the work of undertaken by the international community and the United Nations System today. The importance of the League and its archives was recognized by UNESCO with their inscription on the Memory of the World Register in 2009 [2].

When the League of Nations finally and officially closed its doors in 1946, its assets were transferred to the newly created United Nations, who has been the custodian of the League’s records ever since. Today, the League archives are managed by IMS, which provides traditional archives services for the League archives, as well as historical archives produced by the United Nations Office at Geneva and records management for current staff.

The League archives consist of approximately three linear kilometers of essentially paper-based archives, arranged organizationally, either by administrative section (e.g., Political Section, Economic and Finance Section, Health Section), or by League offices and entities that existed outside of its headquarters in Geneva. Documents within the archives are extremely heterogeneous, varying widely in size, text formats and languages, and quality of paper. There are four categories of materials:

- Registry files: the core of the archives, these files were precisely organized and maintained during the time of the League in a comprehensive and detailed registry system for which various indexes and other finding aids exist
- Section files: the day-to-day working files of the various sections, they consist of less well organized materials that vary widely in contents and arrangement
- Commission files: the files of the external offices and temporary commissions created by the League, the also vary considerably in content and arrangement

- Official documents: the parliamentary documents produced by the League, including the minutes, official documents, and publications of the League Council, Assembly, and all other technical organs and commissions.

Additional types of materials do exist and will be addressed during the course of the project, but are not necessary to mention here.

It is also worth noting that the archives are essentially closed and static, with no further accessions or arrangement to be made.

### **The Organizational Context**

As is true of all archives, every digitization project is unique and reflects the specific context of the digitizing institution, including its overall objectives and its user communities and audiences. Foremost among the many aspects that have influenced the implementation of the LONTAD project has been the comprehensive scope of the work. To state it plainly, *all* of the League archives materials will be digitized. As a result of its comprehensiveness, and the fact that research into the League archives represents 80-90% of its reference activities, the project clearly must take into consideration the broader context of the IMS and its activities, as the project will influence the work of IMS for decades to come.

Several points about the development of IMS services are worth mentioning in the context of LONTAD. First, is the long, ongoing work to create a comprehensive online finding aid of the League archives (<https://biblio-archiv.unog.ch>). Previous indexes to the registry files were created as part of this system's administration during the time of the League, but these can be problematic for contemporary users, both due to the indexes' intended function (i.e. administration rather than research) as well as to their complexity, and the fact that they cover only the registry files. During the 1960s and 1970s two valuable syntheses of the League-era indexes were produced: the *Répertoire Général*, which provides essentially a series level index to the registry, section, and commission files, and the *Repertoire of League of Nations Serial Documents, 1919-1947*, which provides a general index to the various series of official documents produced by the League. While these resources were a great step forward in providing intellectual access to the League archives and are still used on a daily basis, they can be problematic for researchers. They provide access at a relatively high level of the collection and provide limited points of access, which can result in the need to review extremely large quantities of materials to find a desired file or document. Additionally, description for series can be simplistic and haphazardly mixes the use of French and English. As a result, in 2002, IMS began the process to describe the entirety of the archives to the file level using ISAD(G) and acquired an archives management system (AMS) to manage and provide online access for this work. This has meant that significant portions of the collection, at least some 30%, were already completely described to the file level, and higher level description is in place for the remainder. This represents IMS earliest steps towards the digitalization of its operations, and are steps that many archives are now familiar with. LONTAD could therefore take advantage of this work, but would also need to integrate its workflows into existing processes and description decisions.

Second, while some of the reference services conducted by IMS have been moving into digital modalities through the

provision of online finding aids and electronic communications, other aspects remain stubbornly analog, such as the management of physical storage spaces, inventories, and reading room loans. While AMS systems are capable of managing these processes, they have not been implemented due to competing priorities and limited resources. LONTAD, however, brings the opportunity to contribute to just such an implementation, as taking the additional steps necessary in the AMS becomes more efficient due the comprehensive nature of the project.

Third, physical preservation issues related to the League archives have been a long-standing concern for IMS. The archives are stored within the physical spaces of the United Nations Library Geneva, housed in the Palais des Nations, which was constructed in the late 1920s and 1930s to serve as the headquarters of the League. Library storage spaces lack environmental controls beyond simple heating, and large portions of the collection are still housed in registry boxes that date back to the time of the League and that are acidic. Portions of the archive have been heavily used, resulting in wear and tear from physical handling, which is compounded by the fact that onionskin paper is very prevalent throughout the collection. While considering physical preservation within the context of digitization projects is nothing new, specific tasks were integrated into the LONTAD workflows that, while not necessary for the digitization itself, contribute to addressing these issues [3]. In addition, LONTAD will fund the installation of climate control in the portion of the Library's storage area where the collection is housed.

### **A Word on Large-Scale Archives Digitization**

Large-scale archives and special collections projects have become increasingly common, and digitization project methodologies have similarly followed suit, moving away from early "boutique" digitization, which often involved small, carefully selected portions of collections that were treated at the item level, to more comprehensive large-scale and even mass digitization of archival collections [4]. Archives and special collections are increasingly abandoning item-level treatment of digitization as a practical means to digitize more materials, provide improved access to those materials, and to reach previously undiscovered audiences [5]. IMS's own experiences have mirrored this development; a series of early digitization projects completed between 2001 and 2009 treated materials on an item level, but subsequent projects have moved to at least file-level treatment. This process has been complimentary to the highly influential More Product, Less Process movement (MPLP) within archives [6]. It has even been suggested that digitization alone can serve as a sufficient method to replace archival processing and description while still providing sufficient access to collections [7]. Large-scale archives and special collections digitization often try to leverage existing finding aids and higher level descriptions to provide sufficient intellectual access to collections for research purposes, while reducing the incredible resource demands required for more detailed item-level description and handling of digitized files.

It is clear that a project like LONTAD is a large-scale project, if not a mass digitization project, and the challenges inherent to a project of this scale. Many of these challenges, however, no longer relate to the more basic aspects of archives digitization, such as the selection and configuration of file formats, image resolution, and color profiles, which have now been promulgated in a number of standards and recommendations at

both the international and national level [8]. IMS has also increasingly been adapting sometime unintentionally a MPLP approach to digitization, often simply borne out of necessity due to limited resources. What does present a significant challenge, however, is the organization of the range of complex workflows and processes, which include but are not limited to scanning operations, to allow for the successful achievement of such a comprehensive project while meeting the overall digitalization goals of IMS.

## The Total Digital Access to the League of Nations Archives Project

We will now turn to look at how the concerns and contexts mentioned above influenced the creation of project workflow. This examination will focus largely on what we consider to be the operations workflows, necessary to complete the digitization aspects of the project; it will not specifically address other project activities, such as the procurement and implementation of various software systems.

### Initial Project Planning

From the earliest discussions about the possibility of such a project as LONTAD, IMS established a set of broad objectives. These objectives integrated existing objectives and long-standing needs of the organization. In the broadest terms, the project aims to:

- Digitize and provide online access to the entirety of the League of Nations Archives
- Improve the long-term physical preservation of the collection
- Provide for the long-term digital preservation of the products produced by the project.

Each of these objectives, however, are clearly complex undertakings. Further consideration of how to implement any one of these also continually brought us back to previous and ongoing work of IMS as well its general mission and objectives (cite IMS mission?). The project clearly has the potential to be transformative to the nature of IMS, but how could we ensure that objectives that were outside of the strict scope of the project would be sustained? The answer was that the LONTAD project would have to integrate certain activities into its workflows that contribute to these broader objectives. A detailed look at these workflows will now illustrate how we have attempted to achieve this.

### Overall Operations Workflow

An overall workflow, largely driven by experiences on previous small- and medium-scale projects, was initially envisioned. This overall workflow attempts to permit the fastest achievement of the project objectives, reducing the project risks and production bottlenecks, while also identifying “release valves” or processes that could be postponed or delayed if necessary. The workflow design also had to allow the continuation of in-person and online reference services with a minimum of disruptions, seen as critical due to heightened research interest fueled by the centenary of the League’s founding.

Three primary operations teams were established, with their general activities identified below:

- Pre-digitization: physical preparation for scanning, preservation and conservation treatment, and AMS work to prepare for digitization

- Digitization: digital capture (scanning), format conversions, optical character recognition (OCR)
- Post-digitization: quality control (QC) of digitization, metadata creation (description and indexing), and electronic file management and publishing

It is important to note that the digitization operation have been outsourced to a third-party service provider, although the work is done on-site. Because of this, however, the transfer points between pre- and post-digitization and the digitization team is more formalized, and also requires an additional QC phase and a clearly defined product acceptance process.

The establishment of a general project timeline of five years meant that it would be impossible to run each step of the project in series and that they instead would run concurrently, although with staggered start and end points. This stagger was planned as a means to ensure that sufficient material would be prepared by pre-digitization before the digitization commenced.

Initial analysis, as well as previous experience, identified a significant project risk at the point of transfer between pre-digitization and digitization; it is essential that the digitization team always has sufficient materials to scan, both to meet the project timeline as well as to ensure compliance with the service provider contract. We were also able to identify means to mitigate this risk, however, by planning and allowing for personnel to move from pre- to post-digitization if and when necessary.

The decision was made to approach the project by digitizing materials by the League’s sections (again, such as Political Section, etc.). This provides a number of advantages: it allows the project to produce and communicate results in smaller increments, it corresponds to existing finding aids and indexes, including the ongoing description work, and perhaps most importantly, it lets us establish a processing order that corresponds to research interest and other usage demands, including exhibitions and centenary events. This is critical, because a possible “release valve” would be to postpone metadata creation and publishing of seldom-used portions of the collection to the end of the project, or even, if necessary, to leave this work to regular IMS staff after the closure of the project. This decision was not the most obvious or only possibility however, as the collection boxes are only very roughly organized by section and this decision to proceed by section thus requires us to pull out rather randomly grouped sets of boxes. It would likely have simpler, from a pure digitization perspective, to simply start at “Box 1” and proceed through the collection. As suggested, the decisions on how to prioritize the order of League section also must take into consideration a range of other factors outside of the project and is in itself a reflection of other analog and digitalization concerns.

We will now take a closer look at how these issues and concern have shaped the workflows and activities of the operations teams.

### Pre-Digitization

Although some level of physical processing and preparation is usually necessary when archives materials are being digitized, the LONTAD pre-digitization team goes beyond the minimal steps necessary to allow for safe digitization. These additional actions are taken with the long-term preservation of the materials in mind. In addition to the

removal of a multitude of fasteners and unfolding of documents that would typically occur prior to any digitization, the pre-digitization team may also perform basic repairs and stabilization of torn paper, place materials into preservation enclosures to ensure their longer-term stabilization, and isolate any photographs with photo-safe preservation paper (photographs are generally kept in the files and not removed). Significantly, all old boxes are removed and replaced with modern, preservation quality archives boxes. This has entailed substantial efforts and costs, not just in terms of the purchase of the boxes themselves, but also in terms of the time and resources to perform the additional necessary tasks of creating, printing and affixing labels. When necessary, spacers are used to better support materials within boxes, or annex boxes are created to alleviate over-filled containers. These steps clearly have effects that will last well beyond the project life cycle and address broader IMS concerns.

After these physical treatments, files are sorted in appropriate order according to pre-existing registry or other reference numbers, and checked against any existing entries in the AMS and corrected if necessary, which improves the overall accuracy of the existing metadata. Where no entries exist, very basic metadata (a unique reference code and physical characteristics) are entered, which permits a digitization file cover sheet to be generated and also allows subsequent metadata work to be accomplished more rapidly and accurately.

There was significant amount of discussion in early planning stages about the extent to which these processes would be industrialized, that is, how finely to split up individual tasks in order to improve efficiency. It theoretically would be possible, for example, to have one or more staff specialize in preservation work, with others managing the replacement of boxes and box labels, and perhaps another to verify and enter data in the AMS. Some limited attempts at this have found, however, that while short-term time gains can be realized, overall quality suffers in the long-term and can create further issues downstream. As a result, we have settled on a process where pre-digitization staff are assigned multiple boxes at a time, and they perform all the pre-digitization tasks for any given box, although they typically industrialize the process to some extent as well, performing a step for a group of boxes before moving on to the next step in the process. This also makes it easier to address any quality issues, as all tasks related to a box were performed by only one person.

The initial planning had intended for the pre-digitization team to also implement the steps necessary to digitalize the archives physical inventory by codifying and entering storage locations and shelf information into the AMS. This was quickly postponed, however, as it became clear that all the staff allocated would be needed to prepare materials in order to avoid the production bottleneck, as well as unexpected changes to the overall size of the collection as a result of re-boxing and the additional of annex boxes. It therefore made more sense to wait until all materials have been processed and established in their final containers before implementing this system. This will be more efficient and again will also improve the accuracy and digitalization of later stacks management and loans.

### **Digitization**

Several choices relating to the digitization phase were also determined with broader contexts in mind. Overhead scanners designed for cultural heritage digitization were specified in the procurement process as the only acceptable scanners. Image

formats and specifications were, in part, selected based on the desire to provide preservation quality images that could serve as surrogates for consultation of originals, while access copies are generated with the intention to provide the widest possible use. Two sets of delivery files are produced, a single-page master file (JPEG-2000, 300ppi, 24-bit color), as well as a multipage access PDF which has also been processed with OCR.

Previous digitization projects with League of Nations materials provided evidence that high-quality full-text renditions of League materials based on OCR are, in most cases, difficult to achieve due to the abundance of hand-and type-written documents, as well as poor quality mimeograph text, and a large number of languages. Despite these shortcomings, OCR was seen as a potential benefit to full-text searching and online visibility of the documents, particularly as description would only be done at the file level. In addition, a fairly robust technical metadata schema was requested to be provided, with an eye toward the digital preservation aspects of the project. The digitization service provider was also requested to generate checksums for image files, influence by digital preservation concerns, but also as a means of ensuring the authenticity of materials once online access is provided.

The delivery of digitized images was requested on a relatively frequent basis, every two weeks, largely, from the IMS perspective, to ensure the continued access to reference services to original materials. Once boxes are pulled and entered into the pre-digitization process, they are considered unavailable to researchers until after scanning as a means to ensure their readiness and reduce the risk of researchers unintentionally re-ordering files or pages. A short delivery cycle ensures materials are available for traditional reference needs as quickly as possible.

### **Post-Digitization**

Quality control is an essential component to any digitization project, particularly when external service providers perform the scanning work. LONTAD is no exception. Initially in the project the post-digitization team was performing a 100% check on scanned images, largely due to a recurring error that was eventually corrected, and a 10% sampling is now performed. In addition to this QC, a physical QC check is also performed on a sample of materials, again with physical preservation in mind. Staff here are looking to ensure that materials are returned to boxes in good order and physical condition, an important check in view of the possibility that the originals will remain un-consulted for some time after digitization. It should also be noted that some decisions, such as the generation of checksums at digitization, have complicated QC, as many issues that could be fairly easily addressed by post-digitization staff must go back to the service provider in order to generate new checksums.

Metadata creation is a key component of post-digitization activity, but also the one that is perhaps most tied to pre- and post-project activities. The descriptive metadata created by the project will serve as a key component to providing access to its results. As mentioned previously, the LONTAD project has also had the benefit of being able to readily take advantage of existing description and indexing work that has been underway for a number of years, although with some caveats. As description has been underway for some time, a certain amount of drift in description quality can be observed in the AMS metadata of the League archives, due to changes in personnel as



well as changes in recommended practices and processes that were not universally adapted or retroactively enforced. As a result, significant efforts have been made to standardize and document description processes, and existing metadata is thoroughly checked.

In addition to the descriptive metadata, post-digitization also provides a link within the AMS to the physical box or container. This is completely unnecessary for the digitization process, but is critical to eventually allowing management of loan and storage spaces with the AMS. As LONTAD will require work on each file-level entry in the AMS, however, it became clear that the most efficient means to do this processing was at the within he post-digitization processing.

## Conclusion

The presentation of this case study has not necessarily been intended to be an exhaustive look at how the LONTAD project has been implemented, but rather to illustrate how various considerations external to the project itself can influence the design and establishment of its workflows, namely the organizational contexts of the institution implementing archives digitization projects. Specifically, within IMS, that has meant the necessity for LONTAD to work in collaborative pursuit of IMS's broader objectives, including the overall *digitalization* of archives services, while still respecting the traditional analog concerns of physical preservation, in-person reference services, and other uses of materials.

## References

- [1] J. Bloomberg, "Digitization, Digitalization, and Digital Transformation: Confuse Them at Your Peril" *Forbes* (April 29, 2018).
- [2] UNESCO, "Memory of the World: League of Nations Archives 1919-1946", <http://www.unesco.org/new/en/communication-and-information/memory-of-the-world/register/full-list-of-registered-heritage/registered-heritage-page-5/league-of-nations-archives-1919-1946/> (accessed 14 March 2019).
- [3] P. Conway, "Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas," *Library Quarterly* 80, 1 (2010); J.F. Dean "Digital Imaging and Conservation: Model Guidelines," *Library Trends* 52, 1 (2003); J. Paris, "Asking the Right Questions: The Role of the Conservator in Digital Projects," *Liber Quarterly* 18, 2 (2008); and Oya Y. Reiger, *Preservation in the Age of Large-Scale Digitization* (CLIR, Washington, DC, 2008).
- [4] C. Lampert, "Ramping Up: Evaluating Large-Scale Digitization Potential with Small-Scale Resources," *Digital Library Perspectives* 34, 1 (2018).
- [5] Rick Erway, *Rapid Capture: Faster Throughput in Digitization of Special Collections* (OCLC, Dublin, OH, 2011).
- [6] M.A. Greene and D. Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, 2 (2005); and S.C. Sutton, "Balancing Boutique-Level Quality and Large-Scale Production: The Impact of 'More Product, Less Process' on Digitization in Archives and Special Collections," *RMB, J of Rare Books Manuscripts and Cultural Heritage* 13, 1 (2012).
- [7] L.K. Miller, "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *The American Archivist* 76, 2 (2013).
- [8] See, e.g., UK National Archives, *Digitising at the National Archives* (2015), <http://www.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf> (accessed 14 March 2019); US National Archives and Records Administration, "Guidelines for Digitizing Archival Materials for Electronic Access", <https://www.archives.gov/preservation/technical/guidelines.html> (accessed 14 March 2019); and National Archives of Australia, "Preservation Digitisation Standards" (2018), <http://naa.gov.au/about-us/organisation/accountability/operations-and-preservation/preservation-digitisation-standards.aspx> (accessed 14 March 2019).

## Author Biography

*Colin M. Wells received his BA in history from Union College (1997) and his MS in library and information science from the Pratt Institute (2006). Since 2010 he has worked as an archivist at the United Nations Office a Geneva, first as Chief of the Archives Management Unit and since 2017 as Project Manager for the Total Digital Access to the League of Nations Archives Project.*