Learning from the past and looking to the future: The necessity of reevaluating digitization practices and procedures

Brie Grey-Noble
Library and Archives Service, International Telecommunication Union, Geneva, Switzerland.
E-mail address: brie.grey-noble@itu.int

Abstract:

Since 2007, the International Telecommunication Union (ITU) Library and Archives Service has scanned over 600,000 pages of historical publications and documents produced by ITU. In 2010, the History of ITU Portal (www.itu.int/history) was launched providing free, public access to digitized texts as well as a selection of complementary materials. The Portal tells the story of the Union since its founding in 1865 by giving access to key documents and providing the context for discovering more about ITU's history as one of the oldest existing international organizations.
While the digitization of library and archival materials at ITU began as a simple scanning project, over time it has become a central part of our operations. Our efforts have evolved from simply scanning publications to fill user requests to a more methodical approach. Before scanning, we research and assess each collection thoroughly to ensure that what we present on the Portal is more than just scanned pages. By placing the documents in context, we provide our users with a comprehensive view of the Union's work and history.

In 2013, we embarked on a new project: digitizing the Telecommunication Journal, one of the earliest publications produced by ITU. First published in 1869 as the Journal télégraphique, it is a resource with enormous historical value as it provides a record of ITU's work giving a unique international perspective on the growth and evolution of telecommunications technology, policy, and industry over time.

After studying the extent and complexity of the Telecommunication Journal collection, it was clear that we would need to review our digitization procedures. This paper will discuss why changes in our approach were necessary, as well as explore the challenges faced and lessons learned throughout the process of selecting, preparing, digitizing, and providing access to this collection.

Keywords: Digitization, Libraries, International Organizations, Institutional History, Procedures.

Introduction

The International Telecommunication Union (ITU) relies on its in-house library and archives to manage many of the resources that it produces. Part of the mission of the ITU Library and Archives Service (hereafter referred to as the Library and Archives) is to acquire, preserve, and provide the best possible access to all ITU publications and official printed documents that result from its activities. These publications and documents include the outcomes of ITU conferences, some of which are treaties, as well as conference and meeting documents, regulatory publications, statistics, a journal, and standards. In addition, it maintains historic archives that document the activities and administration of ITU since it was founded in 1865. By preserving and storing these important records the Library and Archives ensures that access to the organization's work and history will be available for many years to come.

Digitization has provided the opportunity for the Library and Archives to expand its mission by increasing access to many of the materials in its collections on a global level by making digitized versions of ITU's key publications and documents available online. Since 2007, the Library and Archives has scanned over 600,000 pages of historical publications and documents produced by ITU. In 2010, the digitized texts as well as a selection of complementary materials were made publicly available, for free, with the launch of the History of ITU Portal web site[1]. The aim of the web site is to tell the story of the Union since its founding by giving access to key documents, but also by providing the context for discovering more about ITU's history as one of the oldest existing international organizations. Making the collections of historical publications and documents available online has not only increased access to some of the organization's most important historical records, but also provided an opportunity to highlight its contributions to the international community over time.

To ensure the effective management of each digitization project, it has been crucial to develop and maintain a standard workflow and set of procedures which, with the programme's expansion, have been streamlined for efficiency. Occasionally, minor modifications of our working methods have been made in order to address the specific requirements of each individual project within the digitization programme. With this in mind, each project begins with a thorough assessment of the materials in order to ensure that the procedures in place will result in high-quality digital representations of the source publications and documents.

In 2013, we embarked on a new project: digitizing the *Telecommunication Journal* (hereafter referred to as the Journal), one of the earliest publications produced by ITU. After studying the extent and complexity of the Journal collection, it was clear that our current digitization procedures would have to be revised in order to successfully launch and complete the new project. This paper will discuss why changes in our approach were necessary, as well as explore the challenges faced and lessons learned throughout the process of selecting, preparing, digitizing, and providing access to our collections.

---

[1] See: *History of ITU Portal* – http://www.itu.int/history

Why digitize?

The rapid growth of the Internet has changed the way people seek information. As Erway & Schaffner (2007) suggest there is a sense that we are living "[in] a world where it is increasingly felt that if it's not online it doesn't exist" (p.2). Consequently, many of the materials in library and archives collections may be overlooked and undervalued because they are inaccessible and perhaps even unknown to researchers who are not able to visit the physical locations and consult items from the shelves. In order to prevent valuable materials from becoming hidden in physical collections, many institutions are making an effort to digitize at least portions of their collections as a way to expose them to a larger research community by making them available online. Digitization can increase the likelihood of discovery and help our users uncover the wealth of information found in our collections.

As the guardians of ITU's intellectual and administrative output, the Library and Archives is responsible for making these materials available to both external and internal researchers through requests for information. Digitization gives us the opportunity to further increase access to many of ITU's key publications and documents by providing electronic versions online. In addition, the process enables us to take an active role in telling ITU's story by researching and selecting documents to scan and share on our web site.

From project to programme

Our first digitization project began in response to frequent requests for materials in the collection of publications and documents resulting from ITU's treaty-making conferences. Consequently, we made the decision to digitize all of the official publications and documents in this greatly used, and immensely useful, collection. As work progressed on the conferences project, the value of taking what we learned during our first experience with digitization and applying it to other significant series of documents in the Library and Archives' collections became apparent. Additionally, we saw the potential to use the publications and documents in our collections to tell the story of the Union's growth and development over its nearly 150-year history. Upon reflection, it was clear that incorporating digitization into our day-to-day operations would enhance our mission and benefit our user community. As a result, we expanded our first project into a full-scale, on-going digitization programme with two main goals:

- to produce high-quality, authoritative digitized versions of key ITU publications and documents; and
- to tell the story of the International Telecommunication Union.

Since completing our first scanning project in 2009, the number of digitized collections accessible through the History of ITU Portal has grown from one to six. In addition to the conference series, other projects have included digitizing ITU's annual reports, administrative regulations, and the series historical statistics. We have also carefully selected individual publications and articles that provide an historical overview of ITU's work.

When we began our first digitization project there were no specific procedures in place. We were mindful of this and by the time we completed the project, we had established a basic set of guidelines. As the digitization programme expanded, we created a workflow and set of procedures based on our experiences and resources. In addition, we studied the procedures at similar institutions to see how we could integrate established digitization standards and best

practices into our work. Finally, we collaborated with a team of consultants to evaluate our processes, to seek advice for improving our work, and to provide a realistic assessment of the resources available to us. To streamline our workflow and improve the team's efficiency, we implemented many of the consultants' recommendations. Consequently, we developed a standard set of procedures that addresses each step of our digitization workflow: selection and assessment, cataloguing, scanning, quality control, storage, and delivery.

Three phases of the digitization programme

The digitization programme at ITU can be divided into three phases. What began as a modest project has grown steadily and is now a significant part of our daily operations.

Phase One

In 2007, in response to frequent requests from researchers, the Library and Archives decided to begin scanning publications and documents in the ITU conferences collection. Prior to this, each time a new request was made, we fulfilled it by photocopying a specific section of a publication or set of documents. To improve efficiency and to eliminate unnecessary duplication of effort, we decided to scan the entire item once and save the file for future requests. The digital copies allowed us to fill requests more quickly and it saved time in not having to create new copies each time we received a new request for the same item.

During phase one, scanning was done somewhat sporadically. Documents were simply taken off the library shelves and, with little assessment of the contents or condition, sent to the ITU Reprographic Service for scanning. Following scanning, we performed only a minimal amount of quality control on the digital files. The assumption was that since the publications and documents were on the library shelves, they were accurate and authoritative. However, upon closer inspection, it became clear that this was not the case. For example, in some cases, we discovered that some of the items in the library collection were actually photocopies of the originals. In other cases, we determined that some translations were unofficial and, therefore, not necessarily reliable representations of the source publications. It became clear that in order to ensure that the materials being digitized were authoritative and of high-quality, greater quality control was necessary before scanning.

Phase Two

In addition to the lack of standardized procedures, it was also clear that more time and resources would be needed to properly identify, assess, catalogue, prepare, and scan documents. During phase one, there was only enough time for the Library and Archives staff to meet once a week to work on the digitization project. Consequently, the work proceeded very slowly. Therefore, in 2009, a project librarian was hired to complete the digitization of the conferences collection. The addition of a staff member to focus specifically on digitization enabled the project to move forward quicker and, perhaps more importantly, we could finally develop and document a workflow and standard set of procedures.

One of the most significant distinctions between phases one and two of our digitization programme was that we became much more critical during the entire process. According to Dahlström (2012), one of the characteristics of critical digitization is that it implements the many links in the long digitization chain in a manual, intellectual, and critical way by giving one the opportunity to make choices, select, leave out, and interpret the materials being

digitized (p. 462). As a result of approaching our project critically, many key decisions were made during this period and a workflow was developed:

- **Selection** - Selection decisions would be made at the collection level in order to enable access to a comprehensive and complete series of materials;
- **Assessment** - All items would be brought together from the Library and Archives collections to be assessed simultaneously;
- **Cataloguing** - All items, including unofficial translations, would be catalogued and bibliographic records would be created for both the print and the electronic versions;
- **Metadata** - Metadata relating to the digitization programme would be centralized by storing it in an exclusive database;
- **Scanning** - Items would be scanned on dedicated equipment acquired and operated by Library and Archives staff;
- **Post-scanning** - Quality control would be performed on the digital files to ensure completeness and readability; then, a derivative consultation copy of the master file would be created for access and optical character recognition (OCR) would be applied;
- **Storage** - Master and consultation files would be stored in the ITU document management system (DMS) and on a local server; and
- **Delivery** - Access to the consultation files would be provided via a dedicated web portal.

In order to ensure that work would be completed efficiently and consistently, procedures were created for each step in the digitization workflow. Overall, these procedures have been sufficient in meeting the needs of the digitization programme. Occasionally, minor modifications were made in order to address exceptional requirements for each collection.

Establishing a workflow and creating standard procedures were fundamental to expanding our first digitization project into an on-going digitization programme. Another important step in moving our digitization efforts forward was widening our original goal from simply increasing access to our collections to using our digital collections as a way to tell the story of ITU. Initially, access to the digital files was provided through web pages based on document type. However, during phase two we made the decision to provide an historical context for each collection of materials in order to facilitate an understanding of their content, function, and significance in relation to the Union's work and history. The result was the creation of the History of ITU Portal, a web portal that functions as a platform for providing access to our digital collections as well as a growing collection of additional information and resources about the Union's history.

Since its launch in 2010, the History of ITU Portal has garnered positive feedback from colleagues within ITU, as well as external researchers. Putting the collections of historical publications and documents online and making them widely available at no charge to the user has not only increased access to some of the organization's most important historical records, but also provided an opportunity to highlight its contributions to the international community since the late nineteenth-century.

Phase Three

Until recently, identifying collections to digitize was based largely on analyzing requests for information from both internal and external researchers. For our next digitization project, we decided to change our approach and select a collection that we felt had enormous potential as a historical resource. In 2013, we began working on one our most ambitious projects to date: the digitization of the Journal. The Journal is one of ITU's oldest publications. When the International Telegraph Conference (Vienna, 1868) [2] decided to establish a centralized secretariat for the International Telegraph Union, it gave the secretariat five responsibilities, including the preparation and publication of a telegraph journal. First published in 1869 as the *Journal télégraphique*, it has been published regularly ever since. It is a resource with enormous historical value and high research potential. It not only traces the institutional history of ITU, but also gives a unique international perspective on the growth and evolution of telecommunications technology, policy, and industry.

In his discussion of corporate journals, Heller (2008) identifies them as resources that possess "an almost seamless and coherent organization discourse, [and that] these journals act as the first step into the organization's past" (p. 179). Similarly, the Journal provides an accessible and comprehensive view of ITU's institutional history. It has been a line of communication between the secretariat of the Union, its Member States, and other key stakeholders since its early days. Providing online access to the Journal, a key historic and technical publication, will undoubtedly supply the research community with greater insight into the Union's rich history.

Once the decision was made to scan the Journal collection, the first step was to evaluate its scope and begin planning for the assessment of the collection. It became quickly apparent that we had to reevaluate certain aspects of our current procedures to apply them to this project. In addition, due to the volume and requirements of the project we decided to outsource the scanning portion of the project. In view of this decision, we determined that it would be necessary to undertake a much more rigorous and thorough assessment of the original materials than with previous projects. In principle, thoroughly evaluating the collection before scanning would contribute to the creation of high-quality, complete digital files that would require a minimal amount of post-scanning quality control.

Unlike materials produced by and related to ITU conferences, which could be processed at the event level, the Journal collection is a continuous publication series that spans nearly 150 years. We had to modify the assessment procedure to provide a systematic review of the entire collection of the Journal from its first, then subsequent issues. As a result, we have spent significantly more time assessing and preparing the Journal collection for scanning than previous collections.

As in phase two, we brought together the items from the Library and Archives to perform a comprehensive review of all of the copies in our collections. During the assessment, we discovered a number of inconsistencies – missing pages, loose inserts, and large format foldouts – amongst the copies from the various collections. Due to the volume of items, we added the additional step of documenting all of the missing and unique elements in a detailed item-level inventory to the assessment procedure. By keeping track of everything, we were

---

[2] See: *International Telegraph Conference (Vienna, 1868)*
- http://www.itu.int/en/history/Pages/PlenipotentiaryConferences.aspx?conf=2&dms=S0201000002

able to reconstruct a complete set of the Journal by compiling the best examples of each issue and replacing missing and unique elements as required. In addition, to prepare for scanning, we ensured that all loose inserts and large format foldout pages were flagged and special instructions were drawn up to ensure that each issue would be properly scanned.

To further complicate matters, during the assessment we discovered that during certain periods, the Journal included a number of supplement publications. The supplements were included in some issues the Journal, but were also published and distributed separately as multiple publication series. Therefore, we decided to remove the supplements from the Journal collection and create separate digital collections for each series. Each of the supplement series would undergo the same treatment as the materials in the Journal collection. Thus, the Journal project expanded from one publication series to four separate publication series of varying extents and complexity.

During phase two, we produced two PDF files: a master file, which underwent a page-by-page evaluation to verify that the contents were complete and in correct order; and an access file, a copy of the master, which was also reviewed page-by-page to check that the image quality was acceptable. Library and Archives staff used a set of guidelines to assess the image quality; however, ultimately it was up to the staff member to decide, subjectively, whether or not the image was acceptable. In phase three we decided to outsource the scanning portion of the Journal project. In addition to scanning, the terms of reference for the project specified that the vendor was responsible for ensuring the overall quality control of each digital image. Consequently, since some aspects of the quality control process would be completed by the vendor, it was necessary to develop a modified set of quality control procedures specifically for the Journal project. This included creating a quality control checklist with very specific criteria that would allow us to be more objective when evaluating the digital images. In addition, the page-by-page review of each digital file was eliminated and replaced by the systematic review of a predetermined selection of individual image files. However, as required, a representative set of digital image files, such as a complete journal issue, would be identified and evaluated from cover-to-cover to ensure completeness and that the pages are in correct sequential order.

Another significant difference between phase three and the first two phases was the decision to change the master file format from PDF to TIFF. Previously, our main goal was providing access to our digital files and, therefore, PDF was an ideal format for this purpose. As we entered phase three, we questioned whether this procedure was going to be effective in the long term. Sustainable file formats, such as TIFF, are necessary to ensure the continued retention and accessibility of digitally archived materials. In addition, the TIFF format offers the ability to embed metadata in the actual files, something we had not done previously. In addition, creating TIFF master files would also increase the opportunity to share our digital collections with other institutions in the future. For these reasons, we made the decision to create master files in TIFF format and continue to create derivative files in PDF format for access.

Another important part of the project is providing access to the digital files. As with our other digital collections, access to the Journal collection will be provided through the History of ITU Portal. However, unlike the other collections for which we have included contextual information, the Journal will be presented as a standalone collection with no additional background materials. Essentially, its format and contents will allow it to speak for itself.

Furthermore, due to the importance and extent of the Journal collection it will be featured more prominently than some of the other collections.

Finding a way to facilitate access to and delivery of the Journal digital collection continues to be a challenge. In making plans to develop a web presentation exclusively this digital collection, we have identified a number of limitations that will affect how we present and provide access to this digital collection. For example, metadata will only be captured at the collection and issue levels. We do not have plans to index the collection at the article level. This means that our users will only be able to access the collection by browsing issues year-by-year or month-by-month. In addition, it will be possible to perform full-text searches across the entire collection. However, once the user has found what they are searching for, they will only have the option to download the entire Journal issue in PDF format. Unfortunately, due to limited technical resources, we do not have the facilities similar to those used in Gallica's "Presse et revues[3]" and Library of Congress' "Chronicling America[4]" and, therefore, it will not be possible to offer features such as article extraction or multiple format options for downloading.

The need to review and modify our procedures has been significant part of phase three of the digitization programme. With regards to the Journal project specifically, another challenge that we encountered was how to handle the recent issues of the Journal. Today, the Journal continues to be published as *ITU News*, which is currently available in print by subscription and electronically through a dedicated web site[5]. When exploring options for access to and delivery of the digitized collection, this presented the additional challenge of determining the moment when a current issue becomes "historic." Furthermore, one of the goals of this project is to create a centralized repository for all issues of the Journal including current and future issues. Consequently, it will be crucial to establish a procedure to facilitate the integration of new and current issues into the historical collection. As a starting point, since the Journal is currently produced electronically, we made the decision to scan the issues in the Library and Archives collection up until the point when reliable digital versions were available. After establishing this, Library and Archives staff created copies of the digital files and stored them on a local server. As new issues are published, a copy of each digital file will be transferred to the Library and Archives to ensure that the publication series is complete.

Providing access to the wealth of information available in the Journal collection will undoubtedly generate new research opportunities not only for our users, but also for Library and Archives staff. In order to provide the contextual information included on the History of ITU Portal, we have spent a great deal of time searching for relevant articles in the print collection of the Journal. The articles that we have discovered are only a small example of what is available in this valuable collection. Once the digitization project is complete, the ability to perform full-text searches across the entire collection will enable a more comprehensive selection of results than what we have been able to find by leafing through each issue page by page.

Conclusion

Many lessons have been learned since embarking on our first digitization project and, undoubtedly, as we move forward with new projects we will continue to find areas for

---

[3] See: *Presse et revues (Gallica)* - http://gallica.bnf.fr/html/presse-et-revues/presse-et-revues
[4] See: *Chronicling America: Historic American Newspapers* - http://chroniclingamerica.loc.gov/
[5] See: *ITU News Magazine*: https://itunews.itu.int/en/

improvement in our workflow and procedures. As our digitization programme evolves, we will continue to document and track our decisions and respond to changes by reevaluating and modifying our policies and procedures. Being open to change by accepting that what has worked well for past projects may not be applicable in the future is crucial to maintaining a sustainable and productive digitization programme.

The digitization programme at ITU has enriched the mission of the Library and Archives. It facilitates the searchability of and access to the Union's materials, it reduces the unnecessary duplication of labour by Library and Archives staff to retrieve and process materials, and it establishes the policies for a sustainable digital collection. Furthermore, by understanding and demonstrating the value of the materials in our collections we are providing an authoritative history of its past and ensuring the safeguarding of its institutional memory.

References

Dahlström, M., Hansson, J., & Kjellman, U. (2012). 'As we may digitize' - institutions and documents reconfigured. *Liber Quarterly*, 21 (3/4), 455-474. doi:http://persistentidentifier.nl/?identifier=URN:NBN:NL:UI:10-1-113640

Erway, R., & Schaffner, J. (2007). *Shifting Gears: Gearing up to get into the flow*. OCLC Programs and Research. Retrieved from: http://www.oclc.org/research/activities/capture.html

Heller, M. (2008). Company magazines 1880-1940: An overview. *Management & Organizational History*, 3 (3-4), 179-196. doi:10.1177/1744935908094084

Union télégraphique internationale. (1868). *Documents de la conférence télégraphique internationale de Vienne*. Vienne : Imprimerie impériale et royale de la Cour et de l'Etat.