About the Digitization Programme & History of ITU Portal

Purpose - Vision

In order to promote ITU's history and to make unique ITU records widely available to researchers, ITU staff, and ITU delegates , the ITU Library and Archives Service has been working on a digitization programme for several years now. The purpose of the programme is to organize, catalogue, digitize and publish on the ITU website a selection of historical documents.

Priorities & Selections : Critical digitization

The ITU Library and Archives Service only digitizes ITU materials in order to avoid any copyright issues. Every scanning project is preceded by thorough research and appraisal of our collections. Our priorities are quality and uniqueness, rather than quantity and completeness. Our choices are driven by usage and demand.

Milestones: from project to programme

What began as a simple scanning project has developed into a full-scale digitization programme and into a web portal with a wealth of documents and information that trace the development of ITU as an international organization. Today, ITU Library and Archives are taking an active role in telling the story of ITU.

❖ **2005 – 2007: A slow start: response to requests**

By 2005, the combined reference desk of the Library and Archives showed a high frequency of requests for information contained in ITU's plenipotentiary conferences.

In response to these frequent requests, the ITU Library and Archives took the proactive decision to digitize all of the editions in this highly used, and highly useful, series of documents. A professional book scanner was installed. This was the beginning of our first digitization project.

❖ **2007 – 2009 : The learning curve & first web portal**

Simply scanning documents from the shelves without any real attention to the contents or the condition, proved not the right procedure. It became clear that in order to produce high-quality and authoritative digital copies, more time and resources were needed to properly identify, assess, catalogue, select, prepare, and scan the original documents.

During this period, a key decision was made: we would group the documents together in relation to the conferences that produced them. By bringing the documents together in their context we enable users to understand their content.

Once the documents were selected, assessed, and scanned we needed a way to deliver them to our users. We decided to create a web portal. This first phase of the programme resulted in the launch of the History of ITU Portal in February 2010. For the first time, scanned copies of key documents and publications related to the ITU plenipotentiary conferences, 1865-present, as well as administrative telegraph and telephone conferences, 1879-1988, were made available and freely accessible to anyone with internet access, wherever they live or work.

❖ **2010 : Radio**

In 2010 – the second phase – we added key documents and publications from the early international radiotelegraph conferences, 1903-1947, and the more recent world administrative radiocommunication conferences (WARC) and world radiocommunication conferences (WRC), 1979-2007. The "Radiocommunication Collection" of the History Portal was officially launched on 6 December 2010 during the World Radiocommunication Seminar (WRS-10).

New staff and new equipment allowed us to more than double our output.
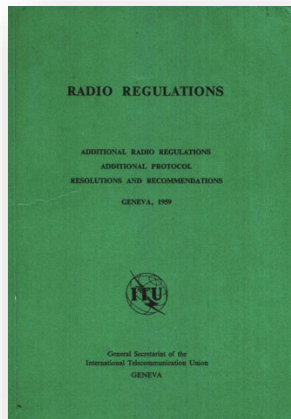
❖ **2011 : Regulations**

The third phase of the programme, undertaken in 2011, treated the historic series of Radio Regulations, 1906-present. The story of the radio regulations is complex and tracing the different editions of these regulations is often requested both by outside researchers and ITU staff and delegates. Again, the selection of materials to be scanned was driven by the demands of our users. The official launch of this collection took place on 18 October 2011 during the 2011 session of the ITU Council.

❖ **2012 : Re-organizing and re-designing**

In the fourth phase, the programme focused to a number of smaller series of key publications including ITU's Annual Report and various historical statistical publications (telegraph, telephone, radio). More importantly, the architecture and back-end of the History Portal was entirely reviewed and redesigned to improve the Portal's services and to provide new functionalities.

❖ **2013 & further**

In 2015, ITU will celebrate its 150th anniversary. In anticipation of this milestone, we will be adding more information pages and expanding our digital collections. Digitizing the ITU Telecommunication Journal is high on our priorities list. We are also planning to explore new ways of presenting historical information such as creating maps and timelines.

Staffing

Over the years, the ITU historic digitization programme has gradually grown from occasional scanning activities with very limited staff resources to a full blown programme supported by a real management team. Today, the digitization team consists of 6 people (not full time equivalent): two librarians and one archivist make selections and decisions and provide intellectual control (assessment, cataloguing, metadata, context, web site pages to host the information).Three technical assistants are responsible for the preparation, scanning, quality control, and administration of the project. This team represents 75% of our total library and archives staff.

In 2012 the new web portal was developed with contract staff during 7 months.

Equipment

As of today, scanning was done entirely in-house. This is the logical result of our choice for critical digitization that requires exhaustive intellectual control to accompany the scanning . The Library makes use of a variety of digital capture devices, including a book scanner (b&w/grayscale), a flat bed scanner (b&w/grayscale/color), an ADF scanner (b&w/grayscale/color) and a film/slide scanner (negative film/diapositives).
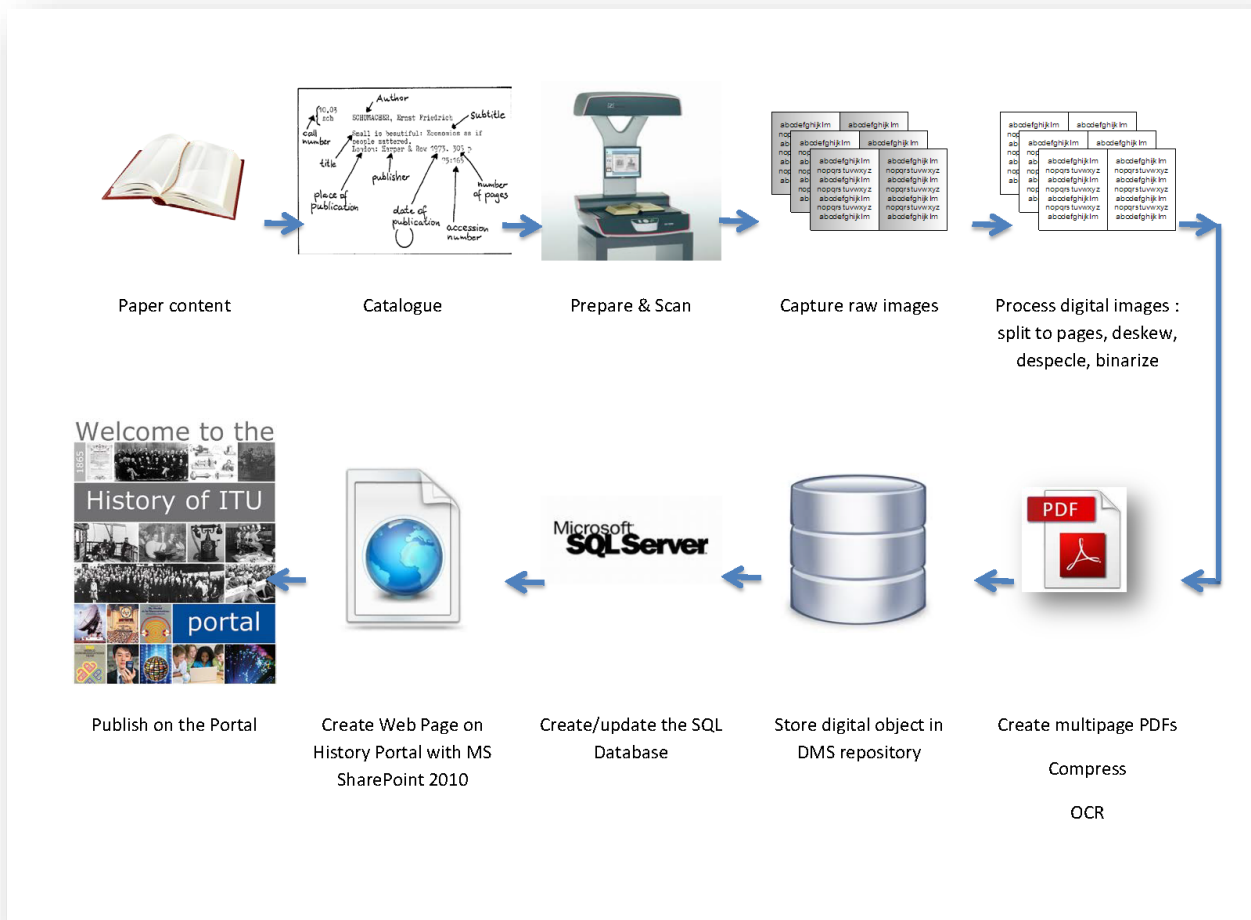
Before 2007, we used a variety of office equipment for simple scanning, such as photocopiers and multi-function office machines. We also received support from the Reprography Service that is equipped with professional scanning machines.

As the programme evolved, the need for more sophisticated post scanning processing and OCR prompted the acquisition a professional book scanner (Zeutschel) in 2007. Quality of our scans improved tremendously. A fast ADF scanner with flat bed (Kodak) was added to our equipment in mid-2010: scanning output doubled. Today, the majority of our digitized copies (<80%) is produced by the ADF scanner.

The nature of the material to be digitized, particularly its condition and construction, determines the type of equipment used for digital capture. All collections are surveyed prior to digitization and minimal handling during the digital capture process is preferred. The following general guidelines apply to the selection of the digital capture device.

- ❖ Most documents are digitized in black & white. If photos or graphic images are included on the pages, we may scan them in grayscale and occasionally in color.
- ❖ Bound volumes, very large formats or particularly fragile documents are scanned on the book scanner.
- ❖ Where possible –if we have multiple copies of the original— we guillotine the volumes and scan loose pages with the ADF scanner.
- ❖ Strip films, negatives, mounted slides are digitized with the negative scanner.

Workflow

Paper content	Catalogue	Prepare & Scan	Capture raw images	Process digital images : split to pages, deskew, despecle, binarize
Publish on the Portal	Create Web Page on History Portal with MS SharePoint 2010	Create/update the SQL Database	Store digital object in DMS repository	Create multipage PDFs Compress OCR

❖ Selection: managers select the materials for digitization
❖ Bibliographic description: all materials are described in the library catalogue (Voyager / Marc21 / AACR2)
❖ Assessment and preparation: materials are checked for their condition, fold-outs, graphic images that require grayscale/color scanning; staples are removed.
❖ Digitization and image processing: publications/documents are scanned with a resolution of 400 dpi. The images are captured and processed by the imaging software (Kodak Capture Pro/Omniscan): images are split, flattened, de-skewed, and de-speckled ; the image is then binarized to a bi-tonal page]. Finally the images are combined into a PDF.
❖ Quality Control: The resulting PDFs are checked for completeness and accuracy. Further image processing may be done to enhance the quality of the image for better readability. Both master copies and consultation copies are multipage PDFs (we don't save the TIFF images that are created during the digitization process).

❖ Compression and OCR: Consultation copies are compressed for faster delivery; they are OCRed and a sheet is inserted to indicate the provenance of the digital copy. These digital objects are checked into ITU's corporate repository.

❖ File upload into the Corporate Repository (DMS): all digital objects (master + consultation copies) are checked into ITU's corporate repository with some basic metadata.

❖ Web display: the digitized collections are made available through the History Portal that is built with SharePoint. Web pages are created dynamically and all page contents are stored in a database.

Quality Assurance is undertaken at several stages in this process: during the image capture, the compression and OCR of the PDFs, the file upload, and when the images are made available online.

Architecture

The architecture of digital objects, metadata and web content is quite complex. Objects and data are captured and stored on several platforms and in different systems.
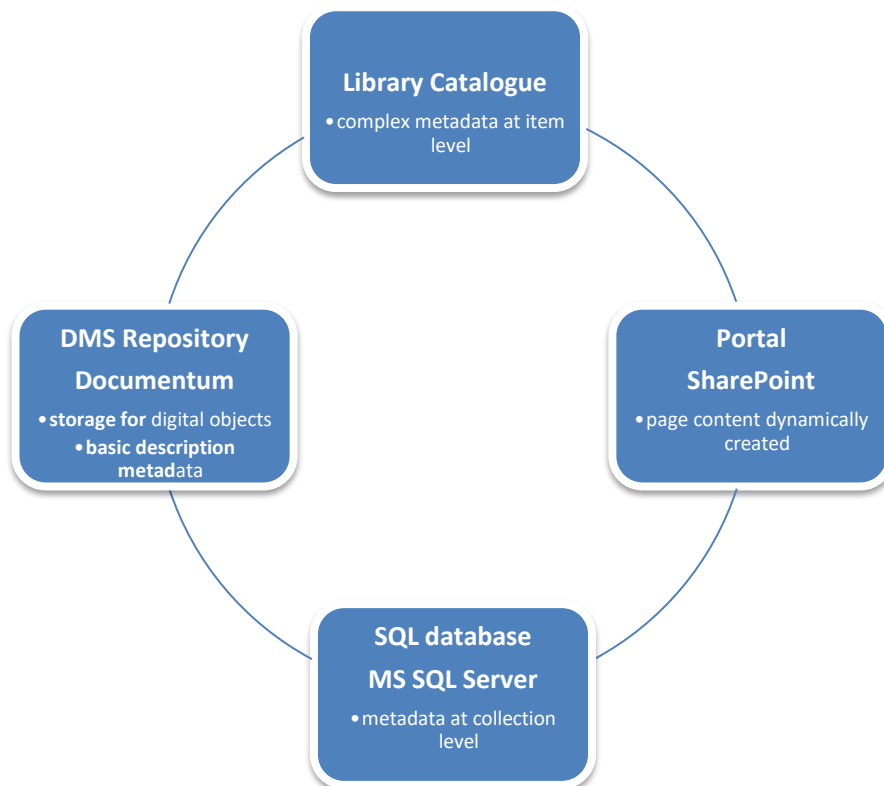

Complex descriptions of all materials (digitized or not) at item level (one description per document) are captured in the library system, built with Voyager. The descriptions are linked by authority records of conference names, authors names, and subjects.

The digital objects are uploaded into ITU's corporate repository, that is built on a Documentum platform. The objects receive some very basic metadata when checked into the repository.

ITU's web site is developed with SharePoint Server 2010. The History Portal web pages are for the largest part built dynamically, based on an underlying database. This allows us to create pages with complex lists that can be sorted or filtered by the end user. The SQL database contains metadata that collection level.

Some Figures
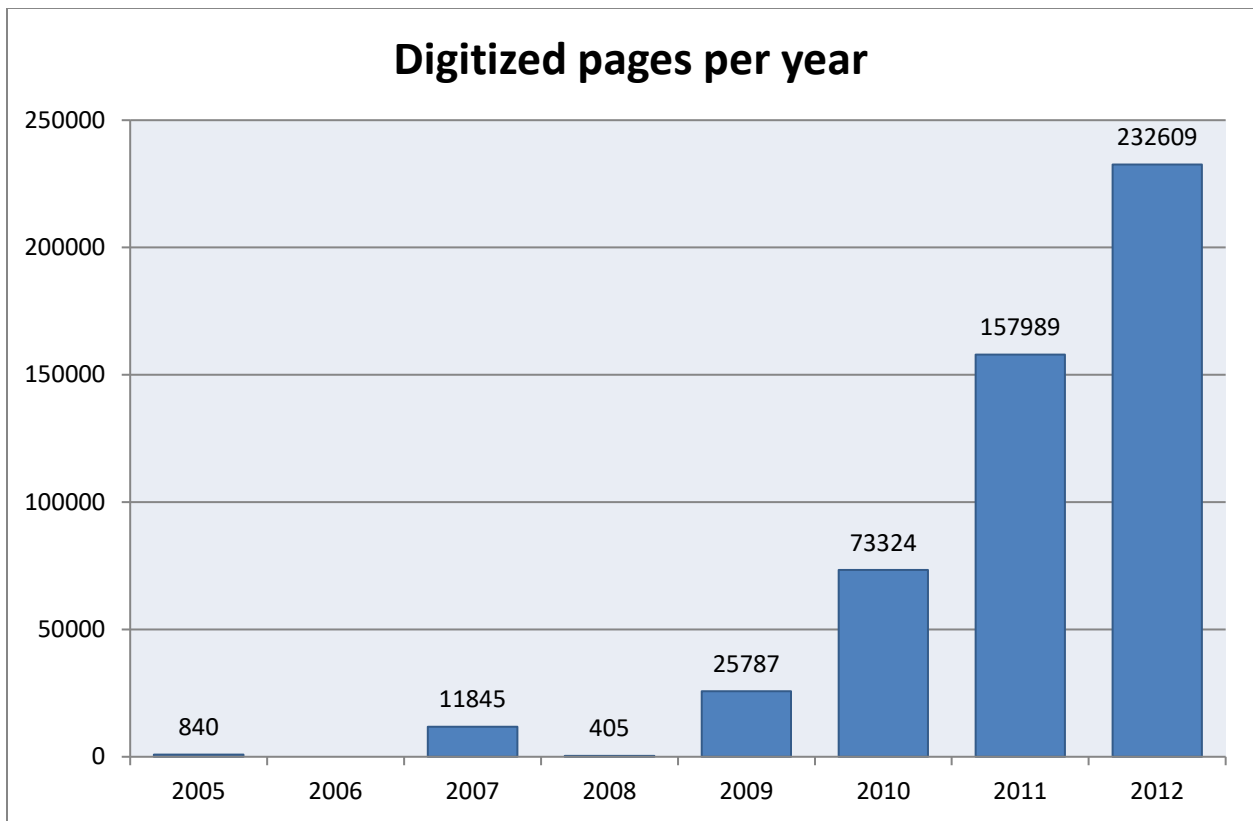
As of today (November 2012), we digitized about 500,000 pages, distributed over different collections. Half of the digitized documents are from the Radiocommunications collection.

Consultation

The first version of the History Portal was launched in February 2010. Since that date, the number of page views has grown steadily every month – from about 400 pages views per day in February 2010 to almost 5,000 page views per day in September 2012. We expect that the new version of the History Portal, launched in November 2012 will improve the user experience considerably and increase consultation of the web site accordingly.

History Portal usage : # page views

Page Views   Av. Page Views/Day
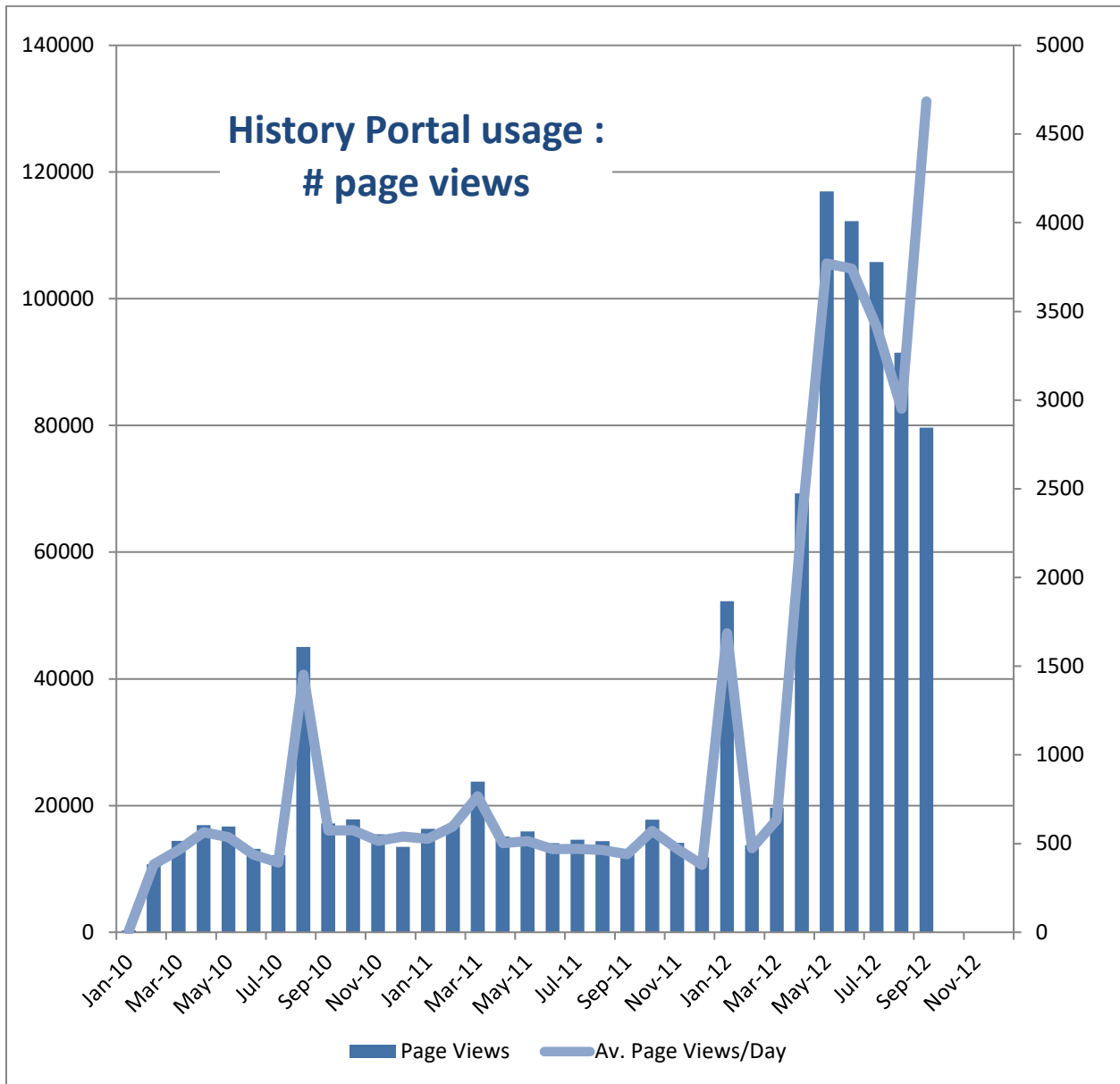
Prepared by the ITU Library and Archives Service
October 2012