

## **Case Studies**

**Digitisation Projects of**



## **CASE STUDY 1: Digitisation of FAO Bibliographical Records:**

### **Background:**

The FAO is a specialised agency of the United Nations that leads international efforts to defeat hunger. Serving both developed and developing countries, FAO acts as a neutral forum where all nations meet as equals to negotiate agreements and debate policy. FAO is also a source of knowledge and information, and helps developing countries and countries in transition modernise and improve agriculture, forestry and fisheries practices, ensuring good nutrition and food security for all. Its Latin motto, *fiat panis*, translates into English as "let there be bread!".

### **The FAO is Headquartered in**

Viale delle Terme di Caracalla  
00153 Rome, Italy

FAO's most important publications present comprehensive and objective information and analysis on the current global state of food and agriculture, fisheries and aquaculture, forests, agricultural commodity markets and hunger. These titles are issued regularly, to inform public debate and policy-making at national and international levels.

FAO maintains a Library catalogue as FAOBIB which is a multilingual, on-line catalogue of documents and publications produced by FAO since 1945, books added to the library collections since 1976, and serials held in the FAO library. FAO mentions that Books which are post 1998 are available in digital format and other documents which predate 1998 are available in Print and Microfiche format.

The FAO Bibliographical record is a list of all Papers and documents written in many different languages gathered from all of its partnering countries across the world.

### **Requirement:**

Publications are central to FAO's work as a knowledge organisation. More than 300 titles per year are published -- usually in multiple language versions -- on topics such as hunger and food security, commodity markets, climate change, nutrition, fisheries, forests, rural livelihoods and much more.

FAO in order to achieve its objectives of making available its vast pool of agricultural research available to everybody was keen to convert the same into Digital format.

FAO has then as a preamble to the digitisation process of all its existing data held in print format has decided to digitise its existing Bibliographical Records. Make them available in databases in XML besides PDF & Text to create an Index of existing records.

### **Solution:**

In order to ensure that the Partner it chose would bring out the best results, FAO had selected CMU, who is presently undertaking the Million Book Project amongst other prestigious digitisation initiatives across the world, was best suited to deliver on the same. The time period for completion of the project in all respects was set for one month.

---

Carnegie Mellon University, in turn shared these requirements with Thrinaina Informatics limited, India who has been its Indian Partner in the Million Book Project, Digitisation of Archives for the Qatar Foundation amongst others. Thrinaina was engaged by CMU for its proven track record of handling similar project and technology implementations and ability to provide face-to-face assistance at all stages.

Thrinaina with its vast pool of resources and existing workflows was up to the challenge.

The client's needs and all specifications were captured through the initial documentation which includes a brief questionnaire. By answering the same the client allows the capture of all of their requirements such as:

1. Physical document details
2. Digitisation specifications
3. Indexing and Tagging specifications
4. Rendering Specifications
5. File and Folder structures/nomenclatures

into a specifications sheet.

The specification sheet is then sent for vetting by the client. Once vetted and signed off by the client this becomes the standard on which the Digitisation sequence follows. Screens will then be designed to capture the required Meta and Structural data details.

Thrinaina uses its patented Digiflo Software Suite to set up the production line at its State-of-the-art Digitisation centers. Each Digitisation center is equipped with Scanners, Software Computer systems installed with Thrinaina's Digiflo Suite and requisite Data Capture tools. The process is then initiated by setting up the operations to follow criteria as per the specifications for finished document delivery (Scanned, Cleaned, OCRd, converted to PDF and Text formats at original size and ready for print). The Digiflo Suite with its various modules is set up to ensure that the document being processed is at all times adhering to the delivery specifications through its multiple quality check points and at final delivery too.

In addition to the Digiflo Software, Thrinaina used OCR engine to capture the data and to enable full text searching. It should take special mention here that each page of the Bibliographical record would consist of 3 to 11 different languages. The Thrinaina team was able to do the necessary capture and correction of this with utmost effectiveness.

Thrinaina was able to deliver the work as per the client's requirement and on time too.

Not to mention, that when we went to Rome to deliver the work Mr. ' was very much impressed and to quote his own words "....."

### **Learning:**

1. The Bibliographical Record was a record made up of entries in as many as 15 different languages, with some pages having listings of articles of almost 11 different languages. Here we can definitely suggest that care should be taken to note down the languages of which data is to be captured and if required take all the assistance of the client to help recognize those languages.
  2. The need to fine tune and build up assets to take on work catering to different languages and the challenges that accrue out of it were definitely a great learning experience and will definitely hold the company at stead in the future for all such kinds of work.
  3. Digitisation of Bibliographical records has to be done by enabling and preplanning for the need for OCRing multiple languages at one go.
-

4. The delivery of the output in PDF formats have to follow correct alignment and proper look to match the original as these form the basis for the next step in the digitisation of the individual articles/books listed out in them.
  5. The delivery of the output in text format should be made available in a manner that it is easy to convert the same into any other format such as CSV as required by the client in future by ensuring that proper care is taken during the conversion and saving of Text format by not introducing any blank spaces and removal of unwanted characters.
  6. The Project has once again proved that by having standardized workflows like Digiflo Suite, projects can always be planned to achieve consistency in quality and reliability at delivery regardless of the format and condition of availability of the Physical data.
  7. Addition of delimiters to programmatically convert the text files into databases as SQL & XML for further digital cataloguing of documents.
-