

Case Studies

Digitisation Projects of



CASE STUDY 2: The Economic and Political Weekly:

The *Economic and Political Weekly*, published from Mumbai, is an Indian institution which enjoys a global reputation for excellence in independent scholarship and critical inquiry.

First published in 1949 as the Economic Weekly and since 1966 as the Economic and Political Weekly, EPW, as the journal is popularly known, occupies a special place in the intellectual history of independent India. For more than five decades EPW has remained a unique forum that week after week has brought together academics, researchers, policy makers, independent thinkers, members of non-governmental organisations and political activists for debates straddling economics, politics, sociology, culture, the environment and numerous other disciplines.

EPW is also unique because it is the one forum where there is an exchange of ideas across the social science disciplines - political scientists debate with economists, sociologists read what political scientists have to say, historians study what economists have to say and so on.

From the beginning EPW was well aware of the Treasure trove of information it was holding in its printed journal form. EPW formulated a business plan and vision to create an environment wherein it could provide access to its journal to its subscribers over the internet. For this the EPW portal provides a convenient means for its subscribers to view online all of its journals in digital format.

Requirement:

To achieve this The *Economic and Political Weekly*, had expressed interest in digitising and archiving its old issues by converting all the Journals into individual article PDF's which are then searchable by Name of the Journal, Year and week of publication, author and Article by using a Search and Retrieval mechanism and to make them available over the internet for its subscribers.

A clear cut budget and timeline was allocated by EPW for this purpose. EPW had then approached Thrinaina Informatics Limited to provide the solution and the processes to achieve the necessary delivery formats.

Solution:

To understand fully the Client's requirements TIL provides the client with specifications questionnaire sheet which enables the client to clearly specify their overall needs. This sheet will capture all the details pertaining to:

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

Based on the requirements specified by the client Thrinaina generates a clear and comprehensive specifications document incorporating all phases from Scanning to delivery as per the Client's requirement. The document will also be able to report the number of physical pages that are mutilated and other pertinent information (in case of old Journals). This document will then be submitted to the client for approval and signing off.

The approved specifications along with the Physical documents will be sent to the Digitisation centre. The Digitisation centre is equipped with State-of-the-art Scanners, Software Computer systems installed with Thrinaina's Digiflo application. Thrinaina's Digiflo is then configured for the process with all of its predefined specifications. Screens will then be configured to capture the Meta and Structural data of the Journals. The Digiflo will then be able to track the progress of any individual document through its various stages such as Scanning, Cropping, and Processing. Reporting adherence and non-adherence to set specifications, any discrepancies with the specifications at all stages will be filtered out through the built in Quality Check modules of the Digiflo Suite. All deviations are corrected before being made available to the next stage of operation. This ensures that errors do not move forward in the flow but are corrected.

Content capture from the scanned images was done through OCR engine to enable full text searching with settings to ensure the highest quality levels.

The captured content was then converted into PDF versions of Journals and Articles ready for delivery. Metadata of the Journals captured through the Digiflo application was then used to develop a 'Search and Retrieval' application to be used by EPW. HTML version of the Books was also delivered. XML tagged version based on data schema defined by the client was delivered.

Learning:

1. Journals such as EPW which hold important data and are fragile (age of the book) have to more necessarily be scanned through a non-contact process using Book scanners. This will protect the Books from further deterioration during the scanning process.
 2. Specifications and delivery schedules have to be rigorously met since these books are meant for public usage and the client expects the finished product to be used for revenue generation. Hence, quality is of prime importance.
 3. It is far more important to clearly define the Metadata fields to be captured in order to create a database of the articles information for usage in the Search and Retrieval applications.
 4. It is also necessary for the Client to be properly informed and advised if necessary to how and what formats the output will have to be standardized before he can host the data on to the internet. This will have to be done during making up of the specifications document. This will eliminate the necessity of any reworking during the implementation process.
 5. Samples of completed work are sent to the client to ensure that the delivery is following their expected specifications prior to production delivery.
 6. The 'Search and Retrieval' application developed for such processes can become the basis for design of such applications as and when required by other clients. The design process and testing being successfully done for this project.
 7. This allows TIL to showcase its capabilities as a end to end specialist in Digitisation and Archiving Solutions provider with different delivery models to suit individual client's requirements.
 8. The Project was able to prove once again the objectives of designing workflows such as Digiflo Suite to ensure the achieving of Consistency, Quality and Reliability are a necessity for all such Digitisation Projects whatever may be the format of availability of the Physical Data.
-