# The PAGE Format

There is a plethora of established and proposed document representation formats but none that can adequately support individual stages within an entire sequence of document image analysis methods (from document image enhancement to layout analysis to OCR) and their evaluation. This paper describes PAGE, a new XML-based page image representation framework that records information on image characteristics (image borders, geometric distortions and corresponding corrections, binarisation etc.) in addition to layout structure and page content.

The suitability of the framework to the evaluation of entire workflows as well as individual stages has been extensively validated by using it in high-profile applications such as in public contemporary and historical ground-truthed datasets and in the ICDAR Page Segmentation competition series.

|  |  |  |
| --- | --- | --- |
| Column 1 | Column 2 | Column 3 |
| Cell 1 | Cell 2 | Cell 3 |
| Cell 4 | Cell 5 | Cell 6 |