

Page 1: Text-Based Content

This page contains regular text content with embedded text layer.
It demonstrates a proper text layer in the document.

Born-digital content like this doesn't require OCR.

Page 2 will contain scanned/image content that requires OCR.

This hybrid approach is common in document management.

The text on this page is fully extractable and searchable.

Page 2: Scanned Content

This page is rendered as image pixels.

No text layer exists - OCR is required.

This hybrid approach is common for documents that mix born-digital and scanned pages.