

RiverDocs Converter – Starter Document

If you wish, you can use this PDF document for your first test run with the Converter. It contains multicolumn text, images and tables and provides a test example for RiverDocs PDF-to-HTML conversion, together with some information on web accessibility aspects of conversion.

Layout and accessibility

One of the first things you will notice is that the Converter renders multicolumn text in single-column format. This is recommended for web accessibility, and makes the text easier to read on mobile phones and PDAs.

Documents converted using RiverDocs software enable Government and corporate organisations to conform to W3C standards and better comply with their legal obligations in terms of Web accessibility for documents.

The original PDF format, which was designed primarily for print, had a number of drawbacks with regard to web accessibility. Based on the PostScript language, it was essentially a description of a page containing scanned images of the document's contents, and lacked the digitised representation required by the assistive technologies that people with a wide range of disabilities use.

It is now technically possible to create "accessible" PDF documents using an extension to the PDF format. However, this requires active commitment from authors and publishers. In practice, most online PDF documents, recent or otherwise, still lack basic accessible features such as textual alternatives for graphical elements and the separation of document structure from visual presentation.

For people with disabilities, this traditional print-style PDF is little better than the results of scanning and optical character reading (OCR). Even when best practice has been observed, and published PDF documents are designed for accessibility, barriers remain:

- users need compatible reader software that is integrated with their individual assistive technologies
- they need to be capable of using the new system

Progress has been made in developing accessible reader software and integrating it with assistive technologies. However, free screenreaders such as Thunder do not support PDF, because the complexity of the format makes this too costly. More advanced commercially available screenreaders that do support PDF are too expensive for many users.

For some users with cognitive deficits, adopting the latest solutions may also present a learning problem.

In fact, accessibility is achievable using standard, established Web technologies recommended by the World Wide Web Consortium. (W3C) – Extensible Hypertext Markup Language (XHTML) and Cascading Style Sheets (CSS). It therefore seems convenient and logical to provide documents in these formats in order to conform with Web Content Accessibility Guidelines (WCAG 1.0), which state

Use W3C technologies when they are available and appropriate for a task, and use the latest versions when supported.

WCAG 1.0, checkpoint 11.1 (priority 2)

Features

RiverDocs Converter's automatic conversion and accessibility process does a great deal of the work of converting PDFs or Word DOCs to accessible (X)HTML.

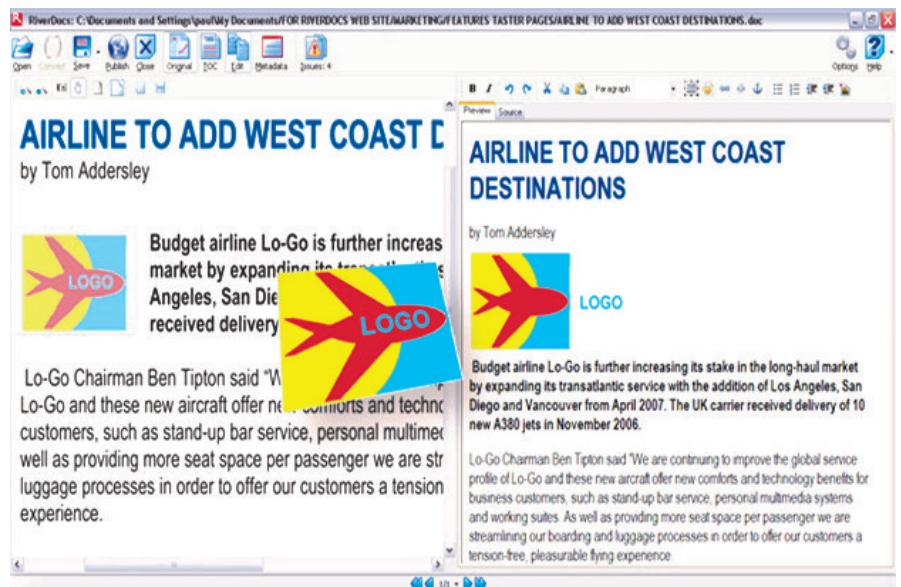
A further editing stage is required for full compliance, as resolving some accessibility issues involves human assessment and manual input. There may also be unsupported graphics that need fixing with screen capture (see below).

The Converter provides all the editing tools you need for this stage, making it easy to fix issues such as missing image alt texts and table summaries (the Converter inserts summaries automatically in some tables, but will not succeed with all types of table).

The CSS editing features are especially useful for changing the overall style of elements such as headings.

RiverDocs Converter – Features
Conversion
★ Convert PDF, DOC and RTF document formats to web-accessible, editable (X)HTML ready for publication
★ Select single or multi-page structure for converted document
★ Editable table of contents generated automatically
★ Built-in WC3-approved accessibility checker
★ CSS stylesheet is generated automatically or can be imported
Editing
★ Built-in WYSIWYG editing of previewed HTML output and source code
★ CSS Editor
★ Standard text editing controls and HTML formatting
★ Multipane editing – compare original with output, edit TOC
★ Copy and insert items from original document display using screen capture
★ Insert and edit images, tables and links

RiverDocs Converter – Advantages	
Editability	Converted output is fully editable, including CSS style control via automatically generated or imported stylesheet
	Automatic highlighting of accessibility issues speeds editing process
	TOC: Editable table of contents based on the page divisions or section headings of the original document
Usability	No additional plug-in reader required
	Multipage originals can be published and read as single web pages, saving the user clicking from page to page
	File size of RiverDocs CHM is usually smaller than the original PDF
Compliance	Enables compliance with accessibility guidelines (WCAG 1.0) and social inclusion legislation



Screen capture

The Converter's screen capture feature (illustrated) enables you to quickly replace or relocate items that have not converted satisfactorily, which can happen occasionally for various reasons, including

- unsupported image formats such as EPS or EMF
- graphical-text combinations (e.g. labels on maps) becoming split
- divisions in composite images causing a 'shredded' effect

In the example shown, the logo text, shown in pale blue in the Editing pane on the right, was not part of the image and has separated from it during conversion. It takes only a few seconds to copy and paste the complete logo image into the Editing pane.

Performance

The Converter's automatic conversion phase is rapid. Because of the great variations in content and formatting between documents, the rate of conversion does not relate directly to the number of pages.

For example, using a computer equipped with an AMD Athlon™ 64 1.8 GHz processor, the European Constitution, an online PDF document of 485 pages with few images or tables but a large number of headings, took 1 minute 6 seconds to convert, a rate of 438 pages per minute. However, if we combine this example with the others shown in the table, which were converted on the same system, the average rate is 367 pages per minute.

While the automatic conversion process is fast, the significant time and cost savings obtainable with the Converter derive from the combination of fast conversion to (X)HTML with speedy identification, display and editing of residual accessibility issues, enabling faster processing of documents than can be achieved by manual searching and correcting with conventional editors.

Examples of Conversion Times

PDF Document	Length (pages)	Conversion time (seconds)
2005-018 EICSA Web.pdf	9	3
Exist-seus-whitepaper-rev2.pdf	11	7
psb-overview-v1.pdf	23	7
UMAAI05preprint.pdf	30	8
AgencyPSBConnectorSpecification-v1.0	39	4
Eaccessibility(eu)report.pdf	84	16
AERF_apm_best_practices.pdf	84	17
Nrdctnch1992.pdf	123	17