Antenna House PDFXML Conversion Library V2.0

Transform collections of legacy PDFs to XML!

Legacy PDFs → Conversion Library →
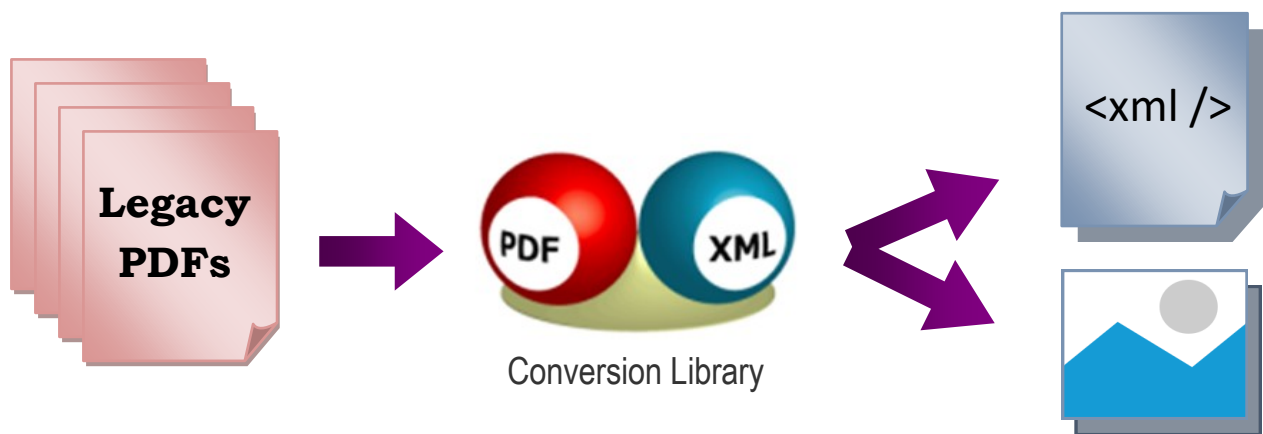
Unlock the content in your PDFs

If you want to reuse content from old PDFs, you no longer need to retype or go through the trouble of reconstructing your documents' content from the PDF binary format. With the Antenna House PDFXML Conversion Library, it is now possible to extract text, tables and images from PDFs and convert them to an XML format which we call "AHPDFXML". The data can then be transformed to any desired output by applying XSLT stylesheets. Benefits and uses for XML include:

- **Content Reusability**
- **Improved Searchability**
- **Good for Accessibility**
- **Promotes Interoperability and Data Integration**
- **Platform Independent**
- **Vendor Independent**

Antenna House PDFXML Conversion Library

The Antenna House PDFXML Conversion Library allows you to unlock the content from your legacy PDFs. It is designed for those organizations that need to convert large volumes of PDFs into XML, HTML5, XSL-FO, DocBook, or any other file formats.

Easily handles PDF content

The library generates a richly structured XML document from the PDFs by using Antenna House's PDF Analyzer Technology.

Reuse PDF content anywhere

The resulting XML can then be transformed with XSLT to the desired output, whether it is EPUB, XML, HTML, or simply text. With the Antenna House PDFXML Conversion Library, you now have the means to take advantage of PDF content for a wide range of environments. Transforming PDF content to XML makes it much easier to reuse, transform, manipulate, and search for data. By applying an XSLT stylesheet, there is more flexibility to processing data depending on how it's being used.

Quickly generates structured content from a PDF

The XML format outputted by this conversion library is called AHPDFXML. It is created by converting the contents in a PDF into XML expressions for text, tables, and images. Using the text frame element (includes line, paragraph and column), table frame element and image frame element, you can convert them to any format that displays the document structure such as DocBook, HTML5, or XSL-FO.

Extract text from a PDF

You can easily extract the text in an arbitrary range in any PDF document with the text element output to AHPDFXML.

PDF Support

- PDF 1.3—1.7
- PDFs compliant with ISO 32000-1: 2008
- PDFs created by Antenna House software

System Support

- Windows Server 2016
- Windows Server 2012 R2
- Windows Server 2012
- Windows Server 2008 R2
- Windows Server 2008 (32bit/64bit)
- Windows 10 (32bit/64bit)
- Windows 8.1 (32bit/64bit)
- Windows 7 (32bit/64bit)
- Linux Red Hat Enterprise series
- Linux CentOS series
- Linux Fedora series

For more Antenna House products, please visit www.antennahouse.com
AH Formatter ● AH Regression Testing System ● Sever Based Converter ● i18n Index Library

A Data Usability Company
ANTENNA HOUSE