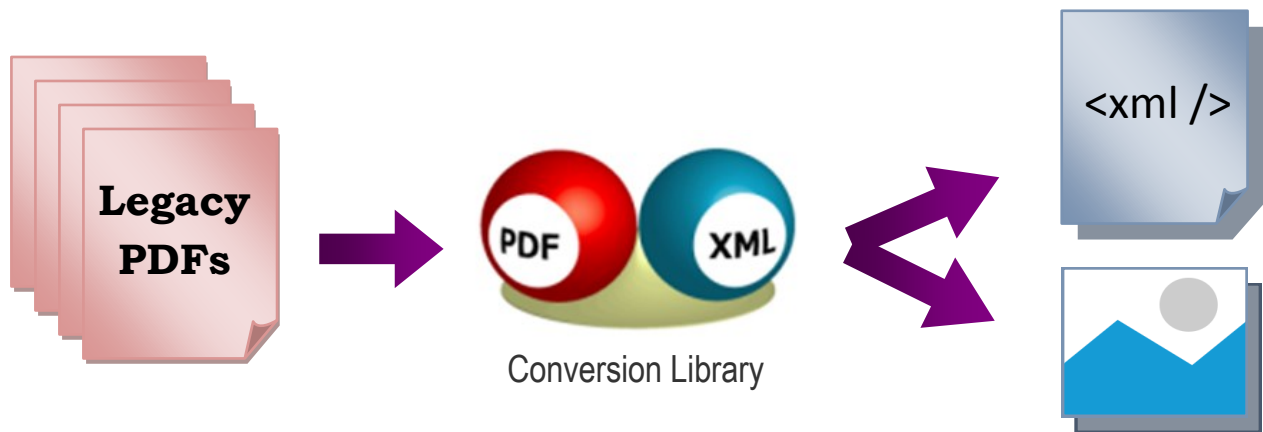


# AHPDFXML Conversion Library

Transform collections of legacy PDFs to XML!



## Unlock the content in your PDFs

If you need to reuse content from old PDFs, there's no longer any need to retype or take the difficult route of reconstructing your documents' content from the PDF binary format. With the AHPDFXML 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". By then applying XSLT stylesheets the data can be transformed to the desired output. Benefits and uses for XML include:

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

# AHPDFXML Conversion Library

Antenna House's PDF to XML Conversion Library (AHPDFXML) now makes the content of your legacy PDFs available. AHPDFXML 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 AHPDFXML, you now have the means to take advantage of PDF content for a wide range of environments. By transforming PDF content to XML, it's easy 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 that conforms to ISO 32000-1
- PDFs created with Antenna House products

## System Support

- Windows Server 2012/2012 R2 (64bit)
- Windows Server 2008 R2 (64bit)
- Windows Server 2008 (32bit/64bit)
- Windows 8.1 (32bit/64bit)
- Windows 8 (32bit/64bit)
- Windows 7 (32bit/64bit)
- Linux (32bit/64bit) and has been built with GCC 4.1