

(/)

BLOG  
(/BLOG)

## APACHE PDFBOX

[Overview \(/index.html\)](#)[License](#)<https://www.apache.org/licenses>[Downloads \(/download.cgi\)](#)

## COMMUNITY

[Support \(/support.html\)](#)[Mailing Lists](#)[\(/mailinglists.html\)](/mailinglists.html)[Issue Tracker](#)<https://issues.apache.org/jira/browse/PDFBOX>[Project Team \(/team.html\)](#)

## DOCUMENTATION

[2.0](#)[1.8](#)[Architecture](#)[\(/1.8/architecture.html\)](/1.8/architecture.html)[Dependencies](#)[\(/1.8/dependencies.html\)](/1.8/dependencies.html)[Cookbook](#)[Document Creation](#)[\(/1.8/cookbook/documentcreation.html\)](/1.8/cookbook/documentcreation.html)[Text Extraction](#)[\(/1.8/cookbook/textextraction.html\)](/1.8/cookbook/textextraction.html)[Document Encryption](#)[\(/1.8/cookbook/encryption.html\)](/1.8/cookbook/encryption.html)[Working with Fonts](#)[\(/1.8/cookbook/workingwithfonts.html\)](/1.8/cookbook/workingwithfonts.html)[Working with Metadata](#)[\(/1.8/cookbook/workingwithmetadata.html\)](/1.8/cookbook/workingwithmetadata.html)[Working with](#)[Attachments](#)[\(/1.8/cookbook/workingwithattachments.html\)](/1.8/cookbook/workingwithattachments.html)[PDF/A Validation](#)[\(/1.8/cookbook/pdfavalidation.html\)](/1.8/cookbook/pdfavalidation.html)[Creating a PDF/A](#)[Document](#)[\(/1.8/cookbook/pdfacreation.html\)](/1.8/cookbook/pdfacreation.html)[Command-Line Tools](#)[\(/1.8/commandline.html\)](/1.8/commandline.html)[API Docs](#)[\(/docs/1.8.13/javadocs/\)](/docs/1.8.13/javadocs/)[FAQ \(/1.8/faq.html\)](#)

## DEVELOPMENT

[Building from Source](#)[\(/building.html\)](/building.html)[Coding Conventions](#)[\(/codingconventions.html\)](/codingconventions.html)

# Working with Metadata

## Introduction

PDF documents can contain information describing the document itself or certain objects within the document such as the author of the document or it's creation date. Basic information can be set and retrieved using the `PDDocumentInformation` object.

In addition to that more metadata can be retrieved using the XML metadata as described below. Getting basic Metadata

To set or retrieve basic information about the document the `PDDocumentInformation` object provides a high API to that information:

```
PDDocumentInformation info = document.getDocumentInformation();
System.out.println( "Page Count=" + document.getNumberOfPages() );
System.out.println( "Title=" + info.getTitle() );
System.out.println( "Author=" + info.getAuthor() );
System.out.println( "Subject=" + info.getSubject() );
System.out.println( "Keywords=" + info.getKeywords() );
System.out.println( "Creator=" + info.getCreator() );
System.out.println( "Producer=" + info.getProducer() );
System.out.println( "Creation Date=" + info.getCreationDate() );
System.out.println( "Modification Date=" + info.getModificationDate());
System.out.println( "Trapped=" + info.getTrapped() );
```

## Accessing PDF Metadata

See class: `org.apache.pdfbox.pdmodel.common.PDMetadata`

See example: `AddMetadataFromDocInfo`

See Adobe Documentation: `XMP Specification`

PDF documents can have XML metadata associated with certain objects within a PDF document. For example following PD Model objects have the ability to contain metadata:

```
PDDocumentCatalog
PDPage
PDXObject
PDICCBased
PDStream
```

The metadata that is stored in PDF objects conforms to the XMP specification, it is recommended that you review that specification. Currently there is no high level API for managing the XML metadata, PDFBox uses standard `InputStream/OutputStream` to retrieve or set the XML metadata.

```
PDDocument doc = PDDocument.load( ... );
PDDocumentCatalog catalog = doc.getDocumentCatalog();
PDMetadata metadata = catalog.getMetadata();

// To read the XML metadata
InputStream xmlInputStream = metadata.createInputStream();

// Or to write new XML metadata
InputStream newXMPData = ...;
PDMetadata newMetadata = new PDMetadata(doc, newXMLData, false );
catalog.setMetadata( newMetadata );
```

Update the Website

(/siteupdate.html)

Jenkins

(<https://builds.apache.org/view/All/job/PDFBox-trunk/>)

Travis CI (<https://travis-ci.org/apache/pdfbox>)

SonarCloud

(<https://sonarcloud.io/project/issues?id=pdfbox-reactor>)

JIRA

(<https://issues.apache.org/jira/browse/PDFBOX>)

Ideas (/ideas.html)

External Links

(/references.html)

#### APACHE SOFTWARE FOUNDATION

Apache Software Foundation

(<https://www.apache.org/>)

ASF Sponsors

(<https://www.apache.org/foundation/thanks.html>)

Security

(<https://www.apache.org/security/>)



(<https://www.apache.org/foundation/contributing.html>)

---

Copyright © 2009–2019 The Apache Software Foundation (<https://www.apache.org/>). Licensed under the Apache License, Version 2.0 (<https://www.apache.org/licenses/LICENSE-2.0>).

Apache PDFBox, PDFBox, Apache, the Apache feather logo and the Apache PDFBox project logos are trademarks of The Apache Software Foundation.