# PDF Reference

## fourth edition

## Adobe® Portable Document Format

**Version 1.5**

Adobe Systems Incorporated

PDF Reference, fourth edition: Adobe Portable Document Format version 1.5.

# Contents

# Preface

THE ORIGINS OF THE Portable Document Format and the Adobe® Acrobat® product family date to early 1990. At that time, the PostScript® page description language was rapidly becoming the worldwide standard for the production of the printed page. PDF builds on the PostScript page description language by layering a document structure and interactive navigation features on PostScript's underlying imaging model, providing a convenient, efficient mechanism enabling documents to be reliably viewed and printed anywhere.

The PDF specification was first published at the same time the first Acrobat products were introduced in 1993. Since then, updated versions of the specification have been and continue to be available from Adobe via the World Wide Web. This book is the third professionally published edition of the specification. Like its predecessor, it is completely self-contained, including the precise documentation of the underlying imaging model from PostScript along with the PDF-specific features that are combined in version 1.5 of the PDF standard.

Over the past ten years, aided by the explosive growth of the Internet, PDF has become the *de facto* standard for the electronic exchange of documents. Well over 500 million copies of the free Adobe Reader® application have been distributed around the world, facilitating efficient sharing of digital content. In addition, PDF is now the industry standard for the intermediate representation of printed material in electronic prepress systems for conventional printing applications. As major corporations, government agencies, and educational institutions streamline their operations by replacing paper-based workflow with electronic exchange of information, the impact and opportunity for the application of PDF will continue to grow at a rapid pace.

PDF is the file format that underlies Adobe ePaper® Solutions, a family of products supporting Adobe's vision for Network Publishing—the process of creating, managing, and accessing digital content on diverse platforms and devices. ePaper

fulfills a set of requirements related to business process needs for the global desktop user, including:

- Preservation of document fidelity across the enterprise, independently of the device, platform, and software
- Merging of content from diverse sources—Web sites, word processing and spreadsheet programs, scanned documents, photos, and graphics—into one self-contained document while maintaining the integrity of all original source documents
- Real-time collaborative editing of documents from multiple locations or platforms
- Digital signatures to certify authenticity
- Security and permissions to allow the creator to retain control of the document and associated rights
- Accessibility of content to those with disabilities
- Extraction and reuse of content using other file formats and applications

A significant number of third-party developers and systems integrators offer customized enhancements and extensions to Adobe's core family of products. Adobe publishes the PDF specification in order to encourage the development of such third-party applications.

The emergence of PDF as a standard for electronic information exchange is the result of concerted effort by many individuals in both the private and public sectors. Without the dedication of Adobe employees, our industry partners, and our customers, the widespread acceptance of PDF could not have been achieved. We thank all of you for your continuing support and creative contributions to the success of PDF.

*Chuck Geschke and John Warnock*
*May 2001*

CHAPTER 1

# Introduction

THE ADOBE PORTABLE DOCUMENT FORMAT (PDF) is the native file format of the Adobe® Acrobat® family of products. The goal of these products is to enable users to exchange and view electronic documents easily and reliably, independently of the environment in which they were created. PDF relies on the same imaging model as the PostScript® page description language to describe text and graphics in a device-independent and resolution-independent manner. To improve performance for interactive viewing, PDF defines a more structured format than that used by most PostScript language programs. PDF also includes objects, such as annotations and hypertext links, that are not part of the page itself but are useful for interactive viewing and document interchange.

## 1.1 About This Book

This book provides a description of the PDF file format and is intended primarily for application developers wishing to develop *PDF producer* applications that create PDF files directly. It also contains enough information to allow developers to write *PDF consumer* applications that read existing PDF files and interpret or modify their contents.

Although the PDF specification is independent of any particular software implementation, some PDF features are best explained by describing the way they are processed by a typical application program. In such cases, this book uses the Adobe Acrobat family of PDF viewer applications as its model. (The prototypical viewer is the fully capable Acrobat product, not the limited Adobe Reader® product.) Similarly, Appendix C discusses some implementation limits in the Acrobat viewer applications, even though these limits are not part of the file format itself. To provide guidance to implementors of PDF producer and consumer applications, compatibility and implementation notes in Appendix H describe the be-

havior of Acrobat viewer applications when they encounter newer features they do not understand, as well as areas in which the Acrobat products diverge from the specification presented in this book.

This fourth edition of the *PDF Reference* describes version 1.5 of PDF. (See implementation note 1 in Appendix H.) Throughout the book, information specific to particular versions of PDF is marked as such—for example, with indicators like *(PDF 1.3)* or *(PDF 1.4)*. Features so marked may be new in the indicated version or may have been substantially redefined in that version. Features designated *(PDF 1.0)* have generally been superseded in later versions; unless otherwise stated, features identified as specific to other versions are understood to be available in later versions as well. (PDF viewer applications designed for a specific PDF version generally ignore newer features they do not recognize; implementation notes in Appendix H point out exceptions.)

The rest of the book is organized as follows:

- Chapter 2, "Overview," briefly introduces the overall architecture of PDF and the design considerations behind it, compares it with the PostScript language, and describes the underlying imaging model that they share.
- Chapter 3, "Syntax," presents the syntax of PDF at the object, file, and document level. It sets the stage for subsequent chapters, which describe how that information is interpreted as page descriptions, interactive navigational aids, and application-level logical structure.
- Chapter 4, "Graphics," describes the graphics operators used to describe the appearance of pages in a PDF document.
- Chapter 5, "Text," discusses PDF's special facilities for presenting text in the form of character shapes, or *glyphs*, defined by fonts.
- Chapter 6, "Rendering," considers how device-independent content descriptions are matched to the characteristics of a particular output device.
- Chapter 7, "Transparency," discusses the operation of the *transparent imaging model*, introduced in PDF 1.4, in which objects can be painted with varying degrees of opacity, allowing the previous contents of the page to show through.
- Chapter 8, "Interactive Features," describes those features of PDF that allow a user to interact with a document on the screen, using the mouse and keyboard.
- Chapter 9, "Multimedia Features," describes those features of PDF that support embedding and playing multimedia content.

- Chapter 10, "Document Interchange," shows how PDF documents can incorporate higher-level information that is useful for the interchange of documents among applications.

  ***Note:*** *Chapter 9 is new in this edition; it contains information on the new multimedia features (see Section 1.2, "Introduction to PDF 1.5 Features" below), as well as some material that was previously in Chapter 8. The material that was in Chapter 9 of the previous edition is now in Chapter 10.*

The appendices contain useful tables and other auxiliary information.

- Appendix A, "Operator Summary," lists all the operators used in describing the visual content of a PDF document.
- Appendix B, "Operators in Type 4 Functions," summarizes the PostScript operators that can be used in PostScript calculator functions, which contain code written in a small subset of the PostScript language.
- Appendix C, "Implementation Limits," describes typical size and quantity limits imposed by the Acrobat viewer applications.
- Appendix D, "Character Sets and Encodings," lists the character sets and encodings that are assumed to be predefined in any PDF viewer application.
- Appendix E, "PDF Name Registry," discusses a registry, maintained for developers by Adobe Systems, that contains private names and formats used by PDF producers or Acrobat plug-in extensions.
- Appendix F, "Linearized PDF," describes a special form of PDF file organization designed to work efficiently in network environments.
- Appendix G, "Example PDF Files," presents several examples showing the structure of actual PDF files, ranging from one containing a minimal one-page document to one showing how the structure of a PDF file evolves over the course of several revisions.
- Appendix H, "Compatibility and Implementation Notes," provides details on the behavior of Acrobat viewer applications and describes how viewer applications should handle PDF files containing features that they do not recognize.
- Appendix I, "Computation of Object Digests," describes in detail an algorithm for calculating an object digest (discussed in Section 8.7, "Digital Signatures").

A color plate section provides illustrations of some of PDF's color-related features. References in the text of the form "see Plate 1" refer to the contents of this section.

The book concludes with a Bibliography and an Index.

## 1.2 Introduction to PDF 1.5 Features

Several features have been introduced or modified in PDF 1.5. The following is a list of the most significant additions, along with the primary sections where the material is discussed:

- The ability to display images using JPEG2000 compression (Section 3.3.8, "JPXDecode Filter"), and to allow 16-bit images (Section 4.8.4, "Image Dictionaries")
- Additional options for the encryption of documents (Section 3.5, "Encryption"). Major new features include crypt filters (Section 3.3.9, "Crypt Filter" and "Section 3.5.4, "Crypt Filters") and syntax for public-key security handlers (Section 3.5.3, "Public-Key Security Handlers", which contains information introduced in PDF 1.3 but not documented in the *PDF Reference* until this edition)
- An extension to the use of streams to allow greater compression of PDF files (Section 3.4.6, "Object Streams" and Section 3.4.7, "Cross-Reference Streams")
- The ability to selectively view or hide content in a PDF document (Section 4.10, "Optional Content" and "Set-OCG-State Actions" on page 608)
- New predefined CMaps and character collections ("Predefined CMaps" on page 404)
- Enhancements to interactive presentations (Section 8.3.3, "Presentations"), including navigation between pages ("Sub-page Navigation" on page 555) and a new action type ("Transition Actions" on page 611)
- Additional annotation types (Section 8.4.5, "Annotation Types") and other enhancements to annotations (Section 8.4, "Annotations")
- Miscellaneous enhancements to interactive forms (Section 8.6, "Interactive Forms"), including support for forms based on Adobe's XML Forms Architecture (Section 8.6.7, "XFA Forms") and the ability to use styled text in form fields and markup annotations ("Rich Text Strings" on page 620)

- Enhancements related to digital signatures and signature fields, including the ability to compute object signatures ("Signature Fields" on page 636, Section 8.7, "Digital Signatures," and Appendix I, "Computation of Object Digests")
- Greatly enhanced support for embedding and playback of multimedia (Section 9.1, "Multimedia", "Screen Annotations" on page 588 and "Rendition Actions" on page 609)
- The ability to allow the display of a PDF file as a slideshow (Section 9.4, "Alternate Presentations"). (This feature is considered part of PDF 1.4, although it was not previously documented.)
- Enhancements to Tagged PDF (Section 10.7, "Tagged PDF") and accessibility features (Section 10.8, "Accessibility Support"). Related updated information is found in Section 5.7, "Font Descriptors" and Section 5.9, "Extraction of Text Content."

## 1.3 Related Publications

PDF and the PostScript page description language share the same underlying Adobe imaging model. A document can be converted straightforwardly between PDF and the PostScript language; the two representations produce the same output when printed. However, PostScript includes a general-purpose programming language framework not present in PDF. The *PostScript Language Reference* is the comprehensive reference for the PostScript language and its imaging model.

PDF and PostScript support several standard formats for font programs, including Adobe Type 1, CFF (Compact Font Format), TrueType, and CID-keyed fonts. The PDF manifestations of these fonts are documented in this book. However, the specifications for the font files themselves are published separately, because they are highly specialized and are of interest to a different user community. A variety of Adobe publications are available on the subject of font formats, most notably the following:

- *Adobe Type 1 Font Format* and Adobe Technical Note #5015, *Type 1 Font Format Supplement*
- Adobe Technical Note #5176, *The Compact Font Format Specification*
- Adobe Technical Note #5177, *The Type 2 Charstring Format*
- Adobe Technical Note #5014, *Adobe CMap and CID Font Files Specification*

See the Bibliography for additional publications related to PDF and the contents of this book.

## 1.4 Intellectual Property

The general idea of using an interchange format for electronic documents is in the public domain. Anyone is free to devise a set of unique data structures and operators that define an interchange format for electronic documents. However, Adobe Systems Incorporated owns the copyright for the particular data structures and operators and the written specification constituting the interchange format called the Portable Document Format. Thus, these elements of the Portable Document Format may not be copied without Adobe's permission.

Adobe will enforce its copyright. Adobe's intention is to maintain the integrity of the Portable Document Format standard. This enables the public to distinguish between the Portable Document Format and other interchange formats for electronic documents. However, Adobe desires to promote the use of the Portable Document Format for information interchange among diverse products and applications. Accordingly, Adobe gives anyone copyright permission, subject to the conditions stated below, to:

- Prepare files whose content conforms to the Portable Document Format
- Write drivers and applications that produce output represented in the Portable Document Format
- Write software that accepts input in the form of the Portable Document Format and displays, prints, or otherwise interprets the contents
- Copy Adobe's copyrighted list of data structures and operators, as well as the example code and PostScript language function definitions in the written specification, to the extent necessary to use the Portable Document Format for the purposes above

The conditions of such copyright permission are:

- Authors of software that accepts input in the form of the Portable Document Format must make reasonable efforts to ensure that the software they create re-

spects the access permissions and permissions controls listed in Table 3.20 of this specification, to the extent that they are used in any particular document.

- Anyone who uses the copyrighted list of data structures and operators, as stated above, must include an appropriate copyright notice.
- Accessing the document in ways not permitted by the document's access permissions is a violation of the document author's copyright.

This limited right to use the copyrighted list of data structures and operators does not include the right to copy this book, other copyrighted material from Adobe, or the software in any of Adobe's products that use the Portable Document Format, in whole or in part, nor does it include the right to use any Adobe patents, except as may be permitted by an official Adobe Patent Clarification Notice (see the Bibliography).