BEST PRACTICES

PORTABLE DOCUMENT FORMAT

HEALTHCARE

A BEST PRACTICES GUIDE

Approval date:
February 11, 2008

AIIM/ASTM BP01-2008

ASTM INTERNATIONAL

aiim® The ECM Association

AIIM/ASTM BP-01-2008

1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910-5603 USA
Telephone: 301.587.8202
Fax: 301.597.2711
E-mail: aiim@aiim.org
Website: http://www.aiim.org

100 Barr Harbor Drive
West Conshohocken, PA 19428
Telephone: 610.832.9500

E-mail: service@astm.org
Website: www.astm.org

ISBN # 0-89258-418-1

Printed in the United States of America.

AIIM/ASTM BP-01-2008

Portable Document Format Healthcare

A Best Practices Guide

An AIIM Best Practices Guide
Prepared by AIIM
Co- sponsored by ASTM International

Abstract

This *Portable Document Format Healthcare Best Practices Guide* describes the features and functions of the proposed, voluntary, and industry-wide use of the Portable Document Format (PDF) for the healthcare industry. As such, the guide is intended to be used as a reference tool for defining PDF as an electronic container by which healthcare information can be captured, exchanged, preserved, and protected for consumers, care providers, and other stakeholders within the healthcare delivery system.

The guide does not describe a normative file format in the same manner as PDF/A. Instead the guide provides education on the use of PDF to support eXtensible Markup Language (XML) standards in the healthcare ecosystem and enable longer-term retention of PDF healthcare documents.

Note: This document is not intended to address compliance with any applicable state and federal regulations that might apply to this information, including Public Law 104-191 (1996), the Health Insurance Portability and Accountability Act (HIPAA).

Information about PDF/A can be found at:
http://www.aiim.org/documents/standards/19005-1_FAQ.pdf

The *Implementation Guide for PDF Healthcare* is published as a separate document by AIIM and can be found at: http://www.aiim.org/pdfh/ig

TABLE OF CONTENTS

FOREWORD

Contributors to the PDF Healthcare Best Practices Guide

Adobe Systems Incorporated
AIIM
American Academy of Family Physicians
American Academy of Pediatrics
ASTM International
CapMed
Dak Systems Consulting
eClinicalWorks LLC
Generator LLC
Good Health Network
Intel Corporation
Massachusetts Medical Society
Medirex Systems Inc.
NextGen
Northrop-Grumman
Schering-Plough
Solventus
Voice of the Physician
YourTimeMatters.com

Supporters of the PDF Healthcare Best Practices Guide

American Academy of Pediatrics
Cerner Corp.
Epocrates
Harvard University
MedCommons Inc.
MEDecisions
MinuteClinic
Northern Illinois Physicians for Connectivity
SureScripts

INTRODUCTION

Portable Document Format (PDF) is a digital file format that provides a method for presenting information that is independent of the application software, hardware, and operating system used to create the information and of the output device used to display or print the information. The independent nature of PDF facilitates the process of creating, managing, securing, collecting, and exchanging digital content on diverse platforms and devices. As such, the use of PDF provides the basis for information portability and interoperability. The migration of multiple medical record types to a universal digital format would be enabled by implementation of an easily adopted document encapsulation practice. This practice would contain specifications for portability, interoperability, and security and would promote the exchange of healthcare information.

Today's processes for collecting, maintaining, and accessing healthcare information inhibit effective and efficient communication between and among caregivers, consumers, and other stakeholders in the healthcare delivery system. Because the current processes are primarily paper-driven, resultant inefficiencies often lead to medical errors, cost redundancies, and sub-optimal patient care.

Most consumers of healthcare do not have access to their health information. Even when copies of medical records and documents are provided to patients they are often disjointed, either because they are in paper form or they are provided by various provider organizations in disparate electronic formats. Consequently, important clinical information can be overlooked, misread, or omitted.

Allowing consumers to access and manage their health information significantly enhances communications between patients, their providers, and other healthcare industry stakeholders—and helps to improve the quality of care. Fortunately, with the advent of the Internet, a new paradigm in healthcare communication emerged, allowing consumers to become more active participants in healthcare information management and exchange.

PDF is a de facto standard for information exchange among major corporations, government agencies, and educational institutions. Because of its wide adoption and acceptance, PDF can support the capture of healthcare information that typically includes structured data, text, graphics, images, and multimedia. Also, PDF is an open industry tool that allows any healthcare information developer to create software tools that, in turn, create or consume documents conforming to the format rules.

PDF Healthcare includes a *Best Practices Guide* and an *Implementation Guide* that describe how to use PDF as a trusted means by which healthcare information can be captured, exchanged, preserved, and protected between and among all stakeholders in the healthcare delivery system. As long as healthcare legal, regulatory, and operational requirements are satisfied, it is believed that PDF's usability, portability, simplicity, and availability will garner the wide use of this *Best Practices Guide in Healthcare*.

This *PDF Healthcare Best Practices Guide* focuses on PDF as a generic technology. Because all Adobe Systems Incorporated PDF specifications referenced in this guide are freely available, any healthcare information developer can implement the functionality and solutions described in this guide. Furthermore, as of January 2007, Adobe released the full *PDF Reference* (version 1.7) to ISO (the International Organization for Standardization) for issuance as an international standard. Consequently, this guide references features that are implementable by any vendor or organization that wishes to do so. There are currently many commercial vendors, organizations, and individual software developers that implement the PDF functionality described in this document.

There are many distinct use cases for *PDF Healthcare*. Some may be fulfilled by the traditional role of PDF in providing readability on a wide scale, while others may support interoperability with XML standards through the use of more advanced PDF features. One of the main goals of *PDF in Healthcare* is to promote and accelerate adoption of a digital healthcare record model using eXtensible Markup Language (XML)-based standards. PDF is not the only technology that can be used to support XML-based healthcare standards, but is one of a number of technologies that may be used to provide this functionality. This guide does not preclude or discourage the use of other standards or technologies in healthcare, but rather focuses on the uses and value of PDF in healthcare. The patient and the healthcare provider are the focus of this important model.

A sample implementation of the *PDF Healthcare Best Practices Guide* is included in a separate document. The *Implementation Guide for PDF in Healthcare* is available on the AIIM website at http://www.aiim.org/pdfh/ig. The *Implementation Guide* serves as one example and does not preclude development or implementation of any other features or implementations of *PDF Healthcare*. The *Best Practices Guide* references the *Implementation Guide* throughout, and in particular, in places where topics are outside of the scope of the *Best Practices Guide*.

PATENT POLICY

The ASTM and ANSI patent policies apply to technical standards that "require" the use of a particular patented technology in order to use the standard or be in compliance with the standard. This document does not include any technical requirements or references to necessary patented technologies. All technical references are used solely as non-mandatory, descriptive or illustrative examples.

SCOPE

The *PDF Healthcare Best Practices Guide* describes PDF features useful in healthcare and documents points to consider for these features. As such, users of this document can decide what features are important to them under their specific circumstances. The *PDF Healthcare Best Practices Guide* does not describe normative requirements, nor is it a language specification. For detailed language issues, it references the PDF and XFA Specifications. (See the Terminology and Reference Documents sections for more information on these specifications). For implementation specific guidance, it references an accompanying *Implementation Guide*.

TERMINOLOGY

Term	Definition
ADA	Americans with Disabilities Act
AIIM	The Enterprise Content Management Association. Enterprise Content Management (ECM) is the technologies, tools, and methods used to CAPTURE, MANAGE, STORE, PRESERVE, and DELIVER information, content, and documents related to organizational processes. ECM enables four key business drivers: Continuity, Collaboration, Compliance, and Costs. *Source:* http://www.aiim.org
ASTM	ASTM International (www.astm.org) is an international voluntary standards organization. Formerly called American Society for Testing and Materials.
CCR	Continuity of Care Record See: ASTM E2369-05, *Standard Specification for Continuity of Care Record (CCR)* *Source:* http://www.astm.org/cgi-bin/SoftCart.exe/STORE/filtrexx40.cgi?U+mystore+ildv3343+-L+CCR+/usr6/htdocs/astm.org/DATABASE.CART/REDLINE_PAGES/E2369.htm
CDA	Clinical Document Architecture Approved as an ANSI Standard (ANSI/HL7 CDA) November 2000. The CDA, which was until recently known as the Patient Record Architecture (PRA), provides an exchange model for clinical documents (such as discharge summaries and progress notes)—and brings the healthcare industry closer to the realization of an electronic medical record. By leveraging the use of XML, the HL7 Reference Information Model (RIM) and coded vocabularies, the CDA makes documents both machine-readable—so they are easily parsed and processed electronically—and human-readable—so they can be easily retrieved and used by the people who need them. CDA documents can be displayed using XML-aware Web browsers or wireless applications such as cell phones. Source: http://www.hl7.org/
cryptography	The discipline which embodies principles, means, and methods for the transformation of data in order to hide its information content, prevent its undetected modification, or prevent its unauthorized use. *Source:* http://www.ca0.net/Glossary.aspx
DICOM	Digital Imaging and Communications in Medicine The industry standard for transferral of radiologic images and other medical information between computers. Patterned after the Open System Interconnection of the International Standards Organization, DICOM enables digital communication between diagnostic and therapeutic equipment and systems from various manufacturers. *Source:* http://www.rsna.org/Technology/DICOM/
Dublin Core (DCMI)	Dublin Core metadata is used to supplement existing methods for searching and indexing Web-based metadata, regardless of whether the corresponding resource is an electronic document or a "real" physical object. The Dublin Core Metadata Element Set (DCMES) provides a vocabulary for describing the "core" information properties, such as "Description" and "Creator" and "Date." Dublin Core metadata provides card catalog-like definitions for defining the properties of objects for Web-based resource discovery systems. The Dublin Core Metadata Element Set is a set of 15 descriptive definitions. It represents a core set of elements likely to be useful across a broad range of vertical industries and disciplines of study. Managed by the Dublin Core Metadata Initiative. *Source:* http://libraries.mit.edu/guides/subjects/metadata/standards/dublincore.html

Term	Definition
HIPAA	HIPAA is the United States Health Insurance Portability and Accountability Act of 1996. For additional information see: http://www.hhs.gov/ocr/hipaa/
HL7	Health Level Seven (www.HL7.org) HL7 refers to Health Level Seven, Inc., an all-volunteer not for profit organization involved in development of international healthcare messaging standards. HL7 is also used to refer to some of the specific standards created by the organization (i.e. HL7 v2.x, v3.0, HL7 RIM etc.).
HSM	Hardware Security Module
IHE	Integrating the Healthcare Enterprise (www.ihe.org)
ISO 17090	ISO 17090 (in three parts) defines the basic concepts of a healthcare public key infrastructure (PKI) and provides a scheme of interoperability requirements to establish a PKI enabled secure communication of health information. It also identifies the major stakeholders who are communicating in health, as well as the main security services required for health communication where PKI may be required More information can be found at: http://www.iso.org/
NCPDP	National Council on Prescription Drug Programs (http://www.ncpdp.org/)
OAIS	Open Archival Information System (ISO 14721:2003)
OTP	One Time Password Generator
PACS	Picture Archiving and Communication Systems
Implementation Guide for PDF Healthcare	The *Implementation Guide for PDF Healthcare*, published as a separate document by AIIM, contains useful implementation samples and more detailed information. This document can be found at: http://www.aiim.org/standards.asp?ID=31832
RDF XML	Resource Description Framework XML Syntax The RDF is a general-purpose language for representing information in the Web. RDF/XML defines XML syntax for RDF. *Source:* http://www.w3.org/TR/rdf-syntax-grammar/
RFC 3881	Information regarding RFC 3881 can be found in the memo which describes the Security Audit and Access Accountability Message XML Data Definitions for Healthcare Applications. Source: http://www.faqs.org/rfcs/rfc3881.html
TPM	Trusted Platform Module TPM is both the name of a published specification detailing a microcontroller that can store secured information, as well as the general name of implementations of that specification.
W3C	World Wide Web Consortium An international consortium of companies involved with the Internet and the Web. The W3C was founded in 1994 by Tim Berners-Lee, the original architect of the World Wide Web. The organization's purpose is to develop open standards so that the Web evolves in a single direction rather than being splintered among competing factions. The W3C is the chief standards body for HTTP and HTML.
XFA	XML Forms Architecture *Source:* http://partners.adobe.com/public/developer/en/xml/xfa_spec_2_4.pdf
XML	eXtensible Markup Language XML is a W3C initiative that allows information and services to be encoded with meaningful structure and semantics that computers and humans can understand. XML is great for information exchange, and can easily be extended to include user-specified and industry-specified tags. *Source:* http://www.orafaq.com/glossary/faqglosx.htm
XMP	eXtensible Metadata Platform

REFERENCE DOCUMENTS

Americans With Disabilities Act (42 U.S.C. 12101 et seq.), Public 101-336.

ANSI/HL7 CDA-R2 2005, *HL7 Clinical Document Architecture, Release 2.0*. Health Level 7, April 2005.

ANSI/NISO Z39.85-2007, *Dublin Core Metadata Element Set*. National Information Standards Organization, May 2007. <http://www.niso.org/standards/standard_detail.cfm?std_id=725>

ASTM E2369-05, *Standard Specification for Continuity of Care Record (CCR)*. ASTM International, 2005.

ISO 14971, *Medical Devices – Application of risk management to medical devices.* International Organization for Standardization, 2007.

ISO 19005-1, *Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)*. International Organization for Standardization, 2005.

ISO/IEC 7810, *Identification cards -- Physical characteristics.* International Organization for Standardization, 2003.

ISO/IEC 7816-1, *Identification cards -- Integrated circuit(s) cards with contacts -- Part 1: Physical characteristics*. International Organization for Standardization, 1998, amended 2003.

ISO 17090, *Health informatics -- Public key infrastructure*. International Organization for Standardization, 2008. Three parts: Part 1: *Ooverview of digital certficate services*; Part 2: *Certificate profile*; Part 3: *Policy management of certification authority*.

ISO/WD 19005-2, *Document management – Electronic document file format for long-term preservation, Part 2: Application of PDF 1.6.* International Organization for Standardization, committee draft; not yet published.

PDF Reference: Adobe Portable Document Format, Version 1.7, 6th Edition. Adobe Systems Incorporated Inc., November 2006.
Latest version available from: <http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference.pdf>

RFC 3161, *Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)*. IETF, August 2001. <http://www.ietf.org/rfc/rfc3161.txt>

RFC 3280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL)* IETF, April 2002. (Updated by RFC 4325 and RFC 4630) <http://www.ietf.org/rfc/rfc3280.txt>

RFC 3881, *Security Audit and Access Accountability Message XML Data Definitions for Healthcare Applications*. IETF, September 2004. <http://www.faqs.org/rfcs/rfc3881.html>

SCRIPT Standard Implementation Guide Version 8.1. National Council for Prescription Drug Programs, October 2005.

Section 508 of the Rehabilitation Act (29 U.S.C. 794d), P.L. 105-220.
<http://www.section508.gov/index.cfm?FuseAction=Content&ID=14>

Web Content Accessibility Guidelines 1.0. W3C Recommendation, May 5, 1999. <http://www.w3.org/TR/WCAG10/>

XML Forms Architecture (XFA) Specification, Version 2.4. Adobe Systems Incorporated Inc., September 11, 2006. <http://partners.adobe.com/public/developer/en/xml/xfa_spec_2_4.pdf>

XMP Specification. Adobe Systems Incorporated Inc., September 2005. <http://partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf

PART I: GENERAL CONCEPTS

WHY PDF?

The main benefit of PDF is that PDF provides a high level of accessibility while preserving Extensive Markup Language (XML) standards of interoperability. Also, PDF provides a "single window" for authorized stakeholders to securely access healthcare information. PDF provides interactive forms that are accessible to parties without electronic healthcare systems. People are familiar with PDF!

WHY NOT PDF/A?

This document assumes that the reader is familiar with PDF/A . Information about PDF/A can be found at: http://www.aiim.org/documents/standards/19005-1_FAQ.pdf

While the PDF/A work was taken into consideration and leveraged where applicable for this guide, the goal of *PDF Healthcare* was to not duplicate what had been accomplished with PDF/A. In addition, PDF/A does not support XML / XML Forms Architecture (XFA) forms, making PDF/A less than ideal for XML standards-based supporting roles. Finally, PDF/A does not permit the use of those security features in PDF that may be required to protect confidential information in healthcare documents. There will be cases where PDF healthcare documents will need to be formally archived (long term) at the end of their active roles. However, the requirements for those documents are already outlined by ISO for PDF/A. The *PDF Healthcare Best Practices Guide* is targeted at live, transactional documents, whose requirements go well beyond the scope and functionality of PDF/A.

TYPES OF PDF

There are various types of PDF files and different ways to create them. The most common types are those that are created from paper—from a scanner or bitmap image of a document, and those that are created directly from other electronic files—from office or publishing applications, electronic forms, or illustration software. Determining which type of PDF file to use varies by process.

The following section describes the various types of PDF files and methods to create them. It also highlights additional PDF features and functionality that will be useful for healthcare processes.

PDF Files Generated from Electronic Sources

PDF files generated from an original electronic source, such as an electronic file, generally have smaller file sizes and greater fidelity to the original than PDF files that are created from secondary sources such as a printed or rasterized copy of that original. Examples of electronic sources include: word processor documents, digital radiographs, databases, and publishing software. PDF files generated from electronic sources usually are text-searchable and allow the user to select and copy text.

There are many types of PDF converters for electronic files. The most basic types of converters intercept the information that an application would send to a printer when physically printing a document and then convert the information to PDF, optimizing for size and efficiency in the process. More advanced converters capture additional information from the authoring

application and convey it to the PDF. Some applications can create PDF files directly without using a separate converter.

Additional information from the authoring application, beyond the strictly visual, often is captured and used to increase the usability or capability of the final PDF. For example, information about a document's logical structure is captured and used in the PDF to provide better accessibility to the content, especially for use by individuals with disabilities. Metadata is also captured from the original source. Metadata may include information such as the document title, author, creation date, and software type. Finally, more complex applications might require capture information regarding a table of contents, hyperlinks, or document layers.

Points to Consider: *It is generally best to utilize the most advanced PDF generation tools available, as the resulting PDFs will have the potential to provide more value. In many cases, depending on the file formats employed, PDF generation can capture significant metadata, logical information for accessibility, layers, bookmarks, links, tables of contents, or other items. Capturing these items directly from the authoring application cuts down on redundant work, errors, or future usability issues. Whenever possible, users should generate PDF files from electronic files because of the size and accuracy advantages as well as the available feature set. Note that "…to utilize the most advanced PDF generation tools…" does not necessarily mean using the most recent version of the PDF specification. Each version of the PDF specification includes new and revised features that may or may not be relevant to a particular PDF usage or software application. See also the section below,* PDF Versions and Features.

PDF Files Generated from Scanned Documents or Image Files

While PDF files that are generated directly from electronic files are generally the most efficient, often access to electronic originals is not possible. In many cases, only paper or bitmap image files (such as TIFF) are available. There are many benefits to converting these files to PDF, but users should be aware of their conversion constraints.

Image-Only PDF Files

The most basic type of image-based PDF file is that which is scanned or directly imported. These types of PDF files are images contained in a PDF "wrapper" that may have compression applied to reduce the file size. Generally, these types of PDF files will be significantly larger than those that were created from electronic files. They may contain metadata captured from a scanner or original image, but, by default, image-only PDF files do not contain selectable, searchable text. Using Optical Character Recognition (OCR) technology, some PDF applications have the ability to convert image-only files into fully text searchable PDFs. In these cases, there are two types of resultant PDFs—Searchable Image and Fully Converted.

Points to Consider: *The PDF specification supports diagnostic quality medical images in PDF documents provided that the images are processed in an appropriate manner during PDF conversion, and provided that appropriate steps are taken to capture appropriate metadata during conversion. Diagnostic quality image standards stipulate requirements for resolution, compression, metadata, and downsampling / subsampling that can be controlled and maintained during PDF generation to preserve those image aspects that make the images suitable for diagnostic purposes. There are many use cases where diagnostic images may also be presented in PDF at lower resolutions or without diagnostic-specific metadata. Implementations should carefully consider these requirements and note the differentiation*

between diagnostic and non-diagnostic images in PDF. The Implementation Guide includes samples demonstrating the use of DICOM images in PDF. See also the section below on Image Files.

Searchable Image PDF Files

Formerly known as “image-over-text” or “image + hidden text” PDF files, searchable image PDF files are created by processing an image or scanned PDF with an OCR application. The original image is preserved and the "captured" text is then added to an invisible layer in the PDF document. As such, users of the document can then search and select text. This type of PDF file is especially useful in preservation or archival applications because it can provide for both archival quality of an original image as well as the added value of searchable, extractable text.

Points to Consider: *In general the size of these PDF files is slightly larger than an image-only PDF file. Therefore these files should be utilized instead of image-only PDF files when there is text in the image. Human verification of any automated OCR process is highly recommended.*

Fully Converted PDF Files

Like searchable image PDF files, fully converted PDF files (a.k.a. formatted text and graphics files) also begin as image only or scanned PDF files. Also, when completed, these files end up with selectable text. However, they are processed in a different manner.

Fully converted PDF files are processed with OCR and only those sections of the document that are actual images will be retained. Other image information will be discarded for any areas of the document that are identified as text by the OCR process. The OCR process will attempt to identify the fonts used in the image and will re-create them in the document. This results in a dramatically smaller file size that may even approach that of electronic-format based PDF files; however, often at the expense of a degree of precision. Consequently, fully converted PDF files may not precisely capture the original look and feel of a document, though this varies based on the resolution of the original image.

Points to Consider: *Fully-converted PDF files might not be suitable for archival or records purposes since they may not accurately preserve the original look and feel of a document. Implementations using fully-converted PDF files may want to consider keeping the original files and providing a reference to them, if archival is important. However, fully-converted PDF files provide an option for portability and file-size reduction.*

PDF Forms

PDF is widely used for electronic forms. PDF forms may be interactive for completion by end-users or they may be merged directly with data by the creating application. Examples include patient data collection, insurance forms, and structured reports. Forms are usually best created in an application specifically designed for PDF forms development.

While it is possible to layer form fields over an existing PDF document, usability is somewhat limited to simple fill-and-save / fill-and-print scenarios. Properly designed PDF forms can be used in conjunction with healthcare industry XML standards such as HL7’s CDA and ASTM’s CCR to provide automatic data validation and a high degree of process automation. Also, applications specifically designed for PDF forms development lend themselves well to the precision layout requirements of forms. Finally, forms design tools usually produce smaller and

more re-usable PDF files. (For additional information on PDF Forms, see the separate document, *Implementation Guide for PDF Healthcare*.)

HEALTHCARE XML-BASED STANDARDS

PDF Healthcare applications are intended to accelerate XML-based healthcare standards by providing a widely supported format for visual presentation, security, and portability for XML-based healthcare data. While there are many ways to build applications around XML standards, PDF documents provide the ability to contain both XML data and a visual representation of that data within a single file format. The use of XML standards-based PDF does not impair data interoperability with other XML-based systems and it provides a method of presenting, transferring, interacting with, and securing healthcare information.

PDF and XML Forms Architecture (XFA) provide a platform for XML standards in healthcare. PDF can be used simultaneously with XML to provide greater security, portability, and consumer-friendly features. XFA can be used to import and export standardized XML content. Implementation examples of PDF and XFA are included in the *Implementation Guide for PDF Healthcare* document. Additional information regarding XFA can be found later in this document in the section Part III: PDF Forms.

Alternatively, an XML document can be embedded inside a PDF file as an attachment. This method does not bind a presentation layer to the XML as can be the case with XFA. Regardless, the XML can be extracted from the PDF as an XML document. Implementation examples of this technique are included in the *Implementation Guide for PDF Healthcare* document.

Healthcare has begun a migration to the use of XML-based standards. Healthcare users benefit using PDF as a platform for XML because PDF provides a familiar user interface for documents while maintaining the support for XML data standards. PDF has always provided wide scale readability of documents, and when incorporated with XML-based standards, it adds no impairment to interoperability for that XML data. (In other words, if the XML provides for interoperability, the inclusion of the XML in the PDF will not impair that interoperability.)

Two examples of existing XML-based standards in healthcare are the ASTM *Standard Specification for the Continuity of Care Record (CCR)* and the Health Level Seven (HL7) *Clinical Document Architecture (CDA)*. Both of these standards represent collections of clinical data as either a dataset (CCR) or as a document (CDA), which can be used in an XFA-based PDF. Also, the National Council on Prescription Drug Programs (NCPDP) has released an XML-based implementation of their SCRIPT 8.1 standard. The SCRIPT standard represents the data of a medication prescription. Such health IT standard examples indicate the likelihood that the healthcare industry will continue a migration to XML-based standards.

This guide does not endorse or promote the CDA, CCR, or any single XML-based healthcare standard. The use of any appropriate XML-based healthcare standards can allow wide-scale interoperability of healthcare data in a PDF document. Therefore, this guide only discusses how PDF used in healthcare can leverage these existing or future standards.

PART II: PDF FEATURES

PDF VERSIONS AND FEATURES

In general, new features added to newer versions of PDF are not regarded as incompatibilities. Therefore, it is important to understand the native PDF features available in the various, major versions of PDF.

The *PDF Reference* v1.5 addresses the issue of features across versions. In addition, this guide focuses primarily on *PDF Reference* v1.7. The details of which features exist in individual PDF versions is beyond the scope of this document. This information can be found in the *PDF Reference*. Additionally, software tools to determine PDF version compatibility are available. But because *PDF Healthcare* is a Best Practices Guide and not a proposed standard, thoughtful consideration should be given to the version of PDF implemented relative to the native feature set.

***Points to Consider**: Implementers should evaluate the efficacy of keeping PDF readers and PDF tools up to date.*

IMAGE FILES

Image files can be directly converted to PDF. Also they can be attached to PDF files in their original format.

Image files such as digital photos and digital diagnostic images can be converted to PDF even at the very high resolutions required by some formats. Conversion of image files to PDF is ideal in a PDF-centric environment because not only does the conversion allow users to work in a single application, but many PDF applications provide advanced image viewing functionality, such as zoom, navigation, and measuring tools. Additionally, PDF files offer support for commenting and annotations. As such, users can mark-up images with notes, comments, arrows, lines, and shapes to indicate areas of interest or comment. It is important to consider lossless vs. lossy conversions. Some tools will convert all images to lossy JPEG.

***Points to Consider:** There are many tools available to convert image files to PDF, but users should be aware of their image resolution requirements (which may be quite high for medical images) and be sure that their conversion tools support the appropriate resolutions.*

Implementers should consider whether to preserve the original resolution of the bitmapped files and the implications of lossy format conversions. See also the section on Image-Only PDF Files *that discusses further issues regarding images in PDF and their use in healthcare.*

VECTOR GRAPHIC FILES

PDF directly supports vector graphic files. Unlike bitmapped (or raster) graphic files, which describe fixed-resolution images by the color of each individual pixel, vector graphic files are described by coordinates and mathematical formulas for lines, shapes, positions, and colors. Vector graphics may be scaled or re-sized without the loss of image quality that may occur when scaling bitmapped graphics. Vector graphics are used for graphs, charts, and waveforms. Also, they are used in files describing maps, architectural plans, engineering blueprints, and other complex diagrams.

Note: The loss of raster image quality with scaling is a complex topic. It is outside the scope of this guide to address this issue.

Integrating the Healthcare Enterprise (IHE) endorses PDF as the format for electrocardiogram (ECG) waveforms and requires that these waveforms be generated using vector graphics, not images. Because vector graphics are directly supported in PDF, users are able to take advantage of a plethora of PDF and PDF application capabilities, such as commenting, zoom, navigation, and measuring tools.

Points to Consider: *When a source document contains vector graphics, those vector graphics generally should not be rasterized (converted to bitmap images) during conversion to PDF. To avoid inadvertent rasterization, use tools that are capable of preserving vector data during PDF conversion. Vector graphics should be used where precision, scaling, and/or measurement are important, and when aliasing effects are undesirable. In many cases, vector graphics tend to produce smaller file sizes than their rasterized bitmap equivalents. Therefore, when file size is a consideration, vector graphics should be retained. However, vector graphics are not generally suitable for photos or photo-realistic images, in which case bitmap images are preferred.*

METADATA

Metadata in PDF is used to describe a whole document. Information that is usually contained in metadata includes concepts such as a document's author, creation/modification date, subject, title, and keywords.

PDF supports a very flexible format for metadata known as XMP (eXtensible Metadata Platform). XMP is not exclusive to PDF; it is based on open standard RDF XML, as well as metadata vocabulary standards like Dublin Core (DCMI). XMP is also used as a common metadata format in other file formats. Information like document identification numbers may be stored in XMP, though it is not intended to store information such as form data. For encrypted or protected files, metadata may be either openly accessible or itself encrypted, depending on encryption parameters.

Points to Consider: *Consider the implications of storing confidential or personal information in metadata/XMP/document properties. Also consider whether information that may not seem confidential could be if it were accessible to an indexing search engine. Non-encrypted PDF metadata is readily accessible by search engines and PDF viewers and should not be used to store confidential information that would not otherwise be stored in the viewable document itself. Non-encrypted metadata can be used to store general information about documents that may be helpful for search and retrieval. If encryption is used, settings that encrypt metadata will prevent proper indexing by search engines.*

ACCESSIBILITY, TAGS, AND LOGICAL STRUCTURE

PDF has features that facilitate access for visually and hearing impaired users, e.g., captioning of multimedia. This capability, often referred to as Tagged PDF, provides access to the content and logical structure of a document for use by assistive technologies such as screen readers and Braille printers. Some PDF viewers also include a built-in capability to read documents aloud.

In addition to support for assistive technology, the use of Tagged PDF also provides some benefits for all users. Tagged PDF can be re-flowed to provide easier reading for small screen devices such as PDAs or cell phones, and can be used to temporarily re-format multi-column

documents into a single continuous format. The use of tagged PDF may be required to fully comply with mandates like *Section 508 of the Rehabilitation Act*, the *ADA (Americans with Disabilities Act)*, or the *W3C Accessibility Guidelines (WCAG)*.

Many PDF generating applications have the capability to create tagged PDF automatically from electronically-originated content or electronic forms. Some PDF generating applications also allow tags to be added to image-based PDF files that have been processed with OCR software. Tagged PDF can increase file size.

Points to Consider: *Consider use of PDF software that supports tagged PDF, and create tagged PDF from electronic originals whenever possible. The use of tagged PDF may be impractical for image-only PDF files without OCR text or for files with severe size constraints, but the implications of using inaccessible content should be considered, especially for all public-facing PDFs.*

ANNOTATIONS

A PDF Annotation associates an object with a PDF page. Annotations are often added after the PDF has been created. PDF defines a variety of standard annotation types such as comments, links, sounds, and movies. For a full list of annotation types see the *PDF Reference*, 8.4.5.

One clinical use case may involve using annotations to point out a pathological structure when a specialist forwards records to a referring physician.

For a sample implementation, see the separate *Implementation Guide for PDF Healthcare*.

Annotation Type: Comments and Markup

There are many types of comment and markup annotations. Some are in-line with document text, such as highlights, strikethroughs, and edits, while others such as lines, notes, callout, cloud, or shapes exist anywhere on the page. Comment annotations may have text notes attached to them that become visible when a user hovers or clicks on the associated annotation, depending on the PDF viewing application.

Comments can be used in a review workflow, where multiple people would like to review and comment on a document without modifying the original content. Comments allow for replies, so commenting threads can be created. PDF Viewers may provide different printing options for comments.

Points to Consider: *Users can use comments for adding information to a PDF document without modifying the original content of the PDF page. When it is not desirable to change the original content of a PDF document, use comments to add information to the existing PDF.*

Use of annotations should be avoided for storing information that has form fields provided; e.g., reactions to medications should not be stored as an annotation to a medication entry.

Annotation Type: Links

Links in PDF, also known as link annotations, are much like hyperlinks in an HTML page. Links may be directly associated with text, or may be in the form of a visible or invisible box. Clicking a link in a PDF viewing application will generally redirect the users view to the target of the link. Note that links are similar to bookmarks (described later in this document) in that they facilitate

navigation throughout a PDF document, to other PDF documents, or to locations on the Internet/Web.

Points to Consider: *Some PDF converters allow for the automatic creation of links from existing hyperlinks in the native document format. Some PDF tools can also add links after the PDF has already been created. Links should be used where the author would like to facilitate navigation of a document by providing hyperlinks within the document's page.*

Links to external data (via the Web) or other documents that require specific applications could impede the host document's use in a disconnected environment.

MULTIMEDIA

Various forms of multimedia (audio and video) may be embedded in PDF files or stored remotely and linked to annotations. The playback of embedded multimedia is generally dependant on a compatible media viewer being present on the viewing computer. Both audio and video may be controlled by the PDF annotation, and video may be displayed directly inside a PDF document. A PDF file may contain multiple renditions of a single multimedia object to accommodate different formats and bandwidth constraints.

There are benefits to including multimedia files in PDFs—primarily the ability to create compound documents that let users work in a single application. Other possibilities include interactive instructional manuals or comprehensive reports that include embedded video.

Multimedia in PDF files also benefit from PDF's robust security features—if the PDF is encrypted, any media content embedded in that PDF will also be protected, and, if the PDF is digitally signed, any media content embedded in that PDF will be included as part of the signed content. One such clinical use case of multimedia in a PDF could include a video of a cardiology study in a physician's referral package.

Points to Consider: *Annotations of most types are suitable for use in many PDF applications, though they are not used for forms-type data or core document content. Care should be used in the distribution of PDFs with annotations in that they may not be fully supported by all PDF viewers. Multimedia annotations may also have dependencies on external media players that may not be present on all computers. Links are generally suitable for all documents other than those with strict archival requirements.*

Comments are generally used for collaboration, and their contents may be extracted from a PDF in data format. Visual comments may be especially beneficial for medical images or other image-type documents. Embedded multimedia should be used with caution unless all interested parties are known to have access to the appropriate multimedia player(s). The use of remote multimedia can be very beneficial with regard to file size, but requires additional care to maintain the availability of the remote content over the usable life of the document.

3D

PDF includes the ability to embed collections of three dimensional objects, such as those created by Computer Aided Design (CAD) software. The 3D facilities in PDF are intended to allow authors to transmit complex 3D information such as engineering diagrams in a ubiquitous manner. These embedded 3D collections are referred to as 3D artwork. 3D artwork can be included within a PDF page. Individual views of the rendered 3D artwork may also be printed. This same technology may be used in a clinical context to transmit 3D medical images.

A PDF file may include multiple instances of 3D artwork and may also include specific labeled views of that artwork. Views are based on the relationship between the camera and the 3D artwork. Some PDF viewers may provide the ability to rotate and move the 3D artwork, which allows the user to examine that artwork from any angle or position.

JavaScripts can also be included to manipulate the 3D artwork in some PDF viewers, including the animation, separation, or showing and hiding of the individual components that make up the 3D artwork. (See the discussion on use of JavaScript in the Scripting section of this document.)

Since this 3D information is contained within a PDF, those documents can take advantage of many of PDF's other features and attributes including security and electronic signatures.

Points to Consider: *Implementers should ensure that all targeted viewers support 3D. The inclusion of 3D in PDF has allowed users who previously had to transmit many 2D copies or slices of a model to send a single PDF. 3D in PDF supports the standard U3D format as specified at:* http://www.ecma-international.org/publications/standards/Ecma-363.htm

ATTACHMENTS

Any type of file can be embedded in a PDF as an attachment. Attachments can be other PDF files, other document types, multimedia, images—any file type. Attachments may or may not be associated with an annotation that is visually present on a page in a PDF document. Some PDF viewers do restrict the embedding of executable attachments for security purposes. When PDF files are digitally signed or encrypted, all attachments are included in the process.

A PDF signature verifies the integrity of all attachments, and an encrypted PDF restricts access to attachments in accordance with the settings of the PDF container. PDF does have the capability to encrypt only attachments, allowing it to function as a "secure envelope."

For a sample implementation, see the separate *Implementation Guide for PDF Healthcare*.

Points to Consider: *While many file types can be attached to a PDF, PDF is the preferred format for several reasons. Using only PDF attachments helps ensure that the user will have access to all embedded content. Otherwise, a user may not be able to view a particular embedded file type without the originating application. Another benefit is that PDF attachments can be searched from within the container PDF, allowing a user to perform a single search across multiple files. PDF attachments can employ compartmentalized security—allowing each attachment to have individual security settings, encryption levels, or digital signatures.*

One main use case for non-PDF attachments is to embed an original file inside of the PDF that was created from it. This allows for the preservation of the original file, while providing a universally viewable PDF. Otherwise, non-PDF attachments should follow the same rule as multimedia annotations—they should be avoided in favor of PDF versions if there is no guarantee that all intended users of the document will have access to the appropriate viewing software.

FONTS

In PDF, fonts can be embedded in a document if the positioning and look of the document is important. Most PDF viewers will attempt to match the intent if the font is not embedded. However, PDF viewers can not guarantee complete fidelity if document fonts are not embedded or otherwise available to the

viewer. Depending on the fonts used, the lack of access to document fonts may even render text unreadable in a PDF. To ensure fidelity to the original content, fonts must be embedded in a PDF document or must be made available in the viewing environment. It is incumbent on the developer of the document systems to ensure that required fonts are made available in the appropriate manner. Note that font embedding may cause an increase in the file size which will vary depending upon the complexity of the font.

Some situations, particularly those that rely on the base 14 fonts (see *PDF Reference*) may not necessitate embedding. Full embedding is appropriate for forms, while subset embedding is appropriate for final form documents. Non-embedded fonts for base 14 can be used in certain situations.

> ***Points to Consider:*** *The font subsetting option may be used if the document will no longer be edited. Subsetting should not be used if it is intended to be an interactive form. Fonts can be embedded as per PDF Reference 1.6 and ISO/WD 19005-2. This helps to ensure a self-contained record that better preserves future rendering by offering a minimal recommended practice.* ***The repercussions of not embedding fonts should be carefully considered, especially if the fonts are symbol or non-standard fonts.***

BOOKMARKS

Bookmarks are not visible on a page; rather they are usually displayed in a separate "Bookmarks" tab in a PDF viewing application. Bookmarks may be logically nested and may be used to replicate a document's table of contents. Note that bookmarks are similar to links (described earlier in this document).

Points to Consider: *Some PDF converters allow for the automatic creation of bookmarks from a table of contents. Some PDF tools can be used to add bookmarks once the PDF has already been created. Bookmarks should be used where the author would like to facilitate navigation of a document by providing an overall outline.*

For a sample implementation, see the separate *Implementation Guide for PDF Healthcare*. This feature can be used to help users navigate through extensive medical records in the same PDF file.

LAYERS

PDF documents may include layers which are sections of content in a PDF document that can be selectively viewed or hidden by authors or consumers. These may be captured from original files during generation. Layers can be used to support multiple languages. They enable a single PDF with multiple languages available for viewing.

Layers can be used to preserve a greater degree of detail and functionality for some types of content. For example, images may contain multiple layers that are hidden at lower magnification levels for clarity, but become visible at higher magnification levels to provide detail. Layers are most valuable for detailed visual content, but may not be fully supported by all PDF viewers.

Points to Consider: *Consider avoiding the use of dynamic layers to convey critical content except in situations where all users can be expected to use software that fully supports them.*

ACTIONS

PDF specifies a number of predefined *Action* types for interactive documents. Actions can include: running JavaScript, moving to a particular destination in the current document (like a hyperlink), playing a sound or movie, hiding or showing an annotation or comment, submitting a form, or moving to a particular 3D view.

PDF Viewers which support Actions will carry them out as a result of a trigger event. PDF defines a number of trigger events. Some trigger events include opening a PDF page, activating a field, typing into or clicking on a field, before and after saving or printing a document, among others. A full list of supported Actions and Triggers is documented in *PDF Reference*, section 8.5.

Points to Consider: *Actions and triggers can be used to provide a consistent initial view for users. When doing so, implementers should confirm that targeted PDF viewers support Actions and Triggers.*

SCRIPTING

PDF files may contain JavaScript. JavaScript can be used to make PDF documents more interactive, it can also be used to manipulate documents, customize the viewing environment, perform calculations or other actions on form fields, send or retrieve information from Web services, or provide detailed information about documents.

Implementers of JavaScript are encouraged to follow organizational security guidelines commensurate with the use of active content and scripting in their environment. This guide does not make a recommendation either way, rather, implementers should balance the effectiveness of using this technology in regard to the benefits derived versus the potential risks.

It is worthwhile to note that JavaScript in PDF is not the same as JavaScript in the browser. The security aspects are much different. In PDF, JavaScript can be used to provide increased interactivity and can be authenticated and protected using digital signatures. Implementers are reminded that scripting has value in some situations but it may not be appropriate for all documents.

Points to Consider: *Scripting can be very valuable in PDF, provided that certain considerations are taken. Since not all PDF viewers support JavaScript, implementers should ensure that PDF documents that depend on JavaScript for critical content or functionality will be used only with PDF viewers that support it. The software, service, or organizational security model is outside the scope of this document. However, implementers should thoroughly assess the impact of any feature, including scripting, on their individual security and risk standards. Scripting functionality should be evaluated for risk and should only be used in accordance with the overall security framework for the overall application.*

REDACTION

Redaction is the process of removing personal or confidential information from a document, usually prior to generalized distribution or publication. The redaction of paper documents generally involves crossing out words or sections of text with a wide black pen.

Redaction of electronic documents is somewhat more complicated, and when performed incorrectly can result in the inadvertent disclosure of the information that was presumed to be removed.

Electronic redaction in PDF must be performed correctly to avoid inadvertent disclosure of information that should be "blacked out." It is not sufficient to highlight text in a PDF or source document in black to obscure the text below. This method does not remove the actual text, which may still be extracted or exported. Some PDF tools have built in support for actual redaction. One example of redaction in a clinical use case could be to redact a patient's identity information when sending a teaching file to another institution.

Points to Consider: *Redaction of PDF documents should only be performed with tools specifically designed for it. Failure to use redaction-specific PDF tools may result in the exposure of confidential information.*

For a sample implementation, see the separate *Implementation Guide for PDF Healthcare*.

WATERMARKS

Some PDF generation applications can create PDF documents with watermarks. Watermarks can appear as page headers, footers, or behind the body of a document. Watermarks may be fully visible or may be invisible on-screen and visible only on printed copies of the document.

Points to Consider: *Watermarks can be useful for discouraging unauthorized use of files. They can also be useful to denote additional information about a document such as its status or origin (e.g., Draft, Confidential, or Complete).*

READER CONFIGURATION ISSUES

Initial View

During the creation and/or post configuration of a PDF file, it is best to ensure the file has a consistent initial view upon opening. If no option is selected the default saved view is used which may lead to variability between users. More importantly, offering users quick access to PDF stored and structured content on initial open is a simple and welcomed solution.

Points to Consider: *Actions and Triggers can allow a document to be more interactive and can provide for navigation and help. Users hoping to provide for interactive features without the desire to write JavaScript may want to examine Actions and Triggers. Implementers should ensure that all targeted PDF Viewers support Actions and Trigger Events.*

PART III: PDF FORMS

PDF documents support electronic forms. Electronic forms can be used interactively to collect information from users, and/or as templates to allow the merging of forms and data for display to users. In some processes, these features are combined to allow for the partial pre-population of forms before they are completed by users. PDF forms are used in many workflows, from small, decentralized processes in which users send and receive forms via email, to large enterprise systems that integrate with services and databases. PDF forms can be divided into two distinct types—AcroForms and XFA (XML Forms Architecture). See also the section above on PDF Forms and Healthcare XML-Based Standards.

ACROFORMS

AcroForms is the legacy model for PDF forms. AcroForms was largely superseded by the introduction of XFA in PDF 1.5, but remains supported by many PDF viewers and is sometimes still appropriate for simple forms. AcroForms is based on simple form field annotations that are typically created within a PDF-based application. Form field annotations can be text boxes, buttons, lists, or other basic form objects. Some PDF viewers or applications allow data to be imported, collected, saved, or exported from AcroForms PDF forms.

AcroForms allows the import or export of form data in a very basic type of flat data file that may be serialized as delimited text or FDF (Forms Data Format) or XFDF (XML-based FDF). Unlike XFA, the data extracted from AcroForms-based documents generally needs to be transformed before it can comply with any industry data standards or services.

Points to Consider: *AcroForms may have a role in simple, ad-hoc, and small office use. They may also be suitable for use in an environment where compatibility with older viewers is required. They are generally not intended for use with XML standards, Web services, or enterprise systems. Not all viewers support XFA. Implementers should examine their environment for suitability.*

XML FORMS ARCHITECTURE (XFA)

XFA was added to the *PDF reference* 1.5 as an alternative for AcroForms. Information on the XFA Specification is included in the Terminology and the Reference Documents sections of this document. XFA forms can be anything from simple static PDF forms that mirror their paper counterparts, to highly interactive and dynamic forms with flowable content and direct support for industry XML standards. One of the main benefits of XFA is its support for direct integration with XML standards like the CCR. Integration with XML standards means that the visual elements of XFA forms can be mapped to the data structures of an XML document.

From an XFA perspective, mapping individual documents is implementation-specific and is outside the scope of this *Best Practices* guide. Mapping need not be technology specific, rather standard layouts can be mapped to XML data standards in an abstracted manner.

This mapping allows the XFA/PDF forms to provide both human-readability through PDF and machine-readability through XML. XFA-based PDF forms that are mapped to an XML schema can support the import or export of XML that conforms to the schema, and can be used to validate interactive user input against the rules and data types of that schema. XFA-based

PDF forms can also be designed to be dynamic, with flowable content based on user interaction or the contents of a dataset.

Points to Consider: *XFA forms provide a much greater level of functionality over AcroForms and should be used in nearly all situations where PDF 1.5 or later is in use. XFA forms may contain scripts and dynamic content, and should follow the same points to consider for scripting as in other PDF documents. Ensure use of a compatible viewer.*

BARCODES

PDF supports the use of barcodes, as images or barcode fonts. AcroForms support simple one-dimensional (1-D) and two-dimensional (2-D) barcodes. XFA forms in PDF support the inclusion of dynamic 1D and 2D barcodes. Barcodes encode information in a line and block-based image format designed to be printed and then reliably read by hardware scanners. Barcodes may consist of a series of bars of varying widths or points or glyphs on a grid.

1-D barcodes have one line of bars, whereas 2-D barcodes may have multiple lines or may rely on grid-type systems. As a result, 2-D barcodes can generally encode a greater quantity and/or more varied information than 1-D barcodes. There are many different types of barcodes. Some are specified formally by standards organizations and others are nothing more than conventions. XFA specifies identifiers for some of the most commonly used barcodes, while AcroForms only allows for 3 types.

Some PDF viewers may allow 1-D barcodes to accept typed user information, much as a text field does. The user can click on the barcode, type a value, and when they commit that value, the field will display the value encoded as a barcode. 2-D barcodes are usually not directly interactive, and have their value set by script or calculation based on a collection or series of other field values. All barcodes include some level of redundancy to support error correction when scanning.

Some barcodes allow the amount of error correction redundancy to be explicitly specified, since more error correction means more space on the page. Some barcodes allow data to be compressed, which can reduce the size of the barcode on the page for large amounts of data.

Points to Consider: *Barcodes can be used to reliably link paper documents to electronic workflow. Users can print or fax a PDF form that contains a barcode. That printed or faxed paper form can be transmitted to another party, who can then scan in the barcode and obtain the information back into electronic format.*

Healthcare includes many workflows which rely on paper and fax. Barcodes provide the ability to incorporate these paper workflows into more accurate and automated electronic workflows. Simple, static barcodes place few if any requirements on an end user. Implementers of more complex 2-D or dynamic barcode based systems should ensure that all PDF viewers support the required fonts and forms capabilities needed in these scenarios. Barcodes can contain a limited amount of data. It is recommended that that data be compressed when attempting to encode large amounts of data.

PART IV: SECURITY AND PRIVACY

SECURITY AND PRIVACY INTRODUCTION

PDF documents may be secured using a variety of existing methodologies including passwords, digital certificates, using a public key infrastructure (PKI), data server technologies, certificates and authentication, and personal cryptographic device hardware security, depending on usage and level of desired privacy.

Note: *This guide outlines native PDF security features that are freely available and can be implemented in a variety of ways. Discussions of security device types are outside the scope of this document. It is incumbent upon each implementer to apply security and privacy controls on associated hardware and transfer devices commensurate with institutional policies. Policy formation is beyond the scope of this document. Implementers should thoroughly assess the impact of any feature on their security and risk policies and standards. Any functionality should be evaluated with regard to policy and risk and should be only used in accord with the overall security framework for the overall application. Implementers can consult other resources and standards that provide additional information about risk and security including: ISO 14971, Medical Devices – Application of risk management to medical devices, and the work of ISO Technical Committee (TC) 215, Working Group 4, Health Informatics – Security. Some examples of security related features are included in the Implementation Guide.*

While the use of a secured single merged Personal Health Records (PHR) document is possible, a more likely scenario will be one of a PDF "envelope" containing all personal health record items, utilizing various forms of security for each document (including merged documents), depending on the PHR owners' needs and wishes.

Existing security methodologies involve three basic “factors”:

- something the user knows (e.g., password, or personal identification number (PIN);
- something the user has (e.g., personal memory device, smart card); and
- something the user is (e.g., biometric characteristic, such as a fingerprint).

Security methods that depend on more than one factor are more difficult to compromise than single-factor methods.[1]

The success of a particular security method depends on more than the technology. It also depends on appropriate policies, procedures, and controls. An effective security method should have customer acceptance, reliable performance, scalability to accommodate growth, and interoperability with existing systems and future plans.

A combination of factors may be utilized to secure the privacy of PDF documents depending on the desired level of accessibility. For example, a patient's emergency data document may be open to all PDF readers, while the patient's medical data summary may be protected using

[1] Source (modified): Albinson, Scott M., *Interagency Guidance on Authentication in an Internet Banking Environment*. Department of the Treasury, Office of Thrift Supervision, CEO Letter #228, October 12, 2005. <http://www.ffiec.gov/ffiecinfobase/resources/info_sec/2006/ots-ceo-ltr-228.pdf>

both a password and the hardware security of a personal memory device (such as a USB flash drive).

One probable scenario will utilize servers, with access to PDF documents granted to owners/patients using both passwords and personal memory devices, and also to whomever owners/patients designate using passwords and public/private keys. Data synchronization with a patient's personal memory device can be implemented, as well as current server security technologies.

Several security factors are outside the scope of this document, such as password loss and recovery (including password escrow), the use of tokens (smart cards, memory devices, etc.), policy and server-based technology integration, and complex audit trails.

TYPES OF DOCUMENT SECURITY

Document security in PDF can be divided into two main concepts (though there is some overlap). The first concept is document control. Document control features restrict who has the ability to view a document, who may manipulate a document, and what manipulations or actions are allowed. These control features may be role-based, with a degree of built-in flexibility around the interplay of roles and actions. Document control features include encryption, modification prevention, and document rights management. The second security concept is that of authenticity. Authenticity applies not only to establishing the origin and integrity of documents, but also to non-repudiation and accountability. Authenticity features include digital signatures, certification, tamper-evidence, modification detection, versioning, and audit-trails.

Document Control

Authentication

Document control features require some level of authentication to establish the identity of a user. In some cases, user authentication may take the form of a simple password. In others, it may be a digital certificate or biometric device. The level of authentication employed should be commensurate with the level of security or privacy of the document; though considerations need to be made for practical access to technology, e.g., many individual users may not have access to digital certificates.

Passwords

Passwords may be used to directly control access to a PDF document. Passwords are also commonly used in conjunction with other forms of authentication such as digital certificates and biometrics. A PDF document may have two passwords. The first type of password restricts the opening of the document and must be entered each time the document is opened.

The second type of password controls access to settings that determine what users may do with the document. Examples of restrictions include limits on extracting/copying text, printing, or modifying the document.

Passwords typically restrict access to PDF files and can encrypt the entire document including metadata. However, encrypting metadata or populating metadata with personal or high-value information is not recommended. Depending on process or usage requirements, passwords may be directly selected by users. In other situations, passwords may be automatically created

by document generation systems on a document-by-document basis and/or transmitted to recipients via out-of-band methods.

Other situations that overlap the direct use of passwords in PDF include password use on portals used to distribute documents or directly on devices that may contain documents. Individually attached documents in the PDF envelope may be password protected and encrypted based on need.

Passwords alone are now generally considered to be only a rudimentary form of authentication, and are not typically relied on as a sole form of authentication for high-value or private information.

A stand alone password is considered "single-factor authentication" in that there is only one item required to authenticate a user—in this case something that the user knows. Emerging laws and regulations require higher levels of authentication for securing or authenticating high-value or confidential information. Implementers are urged to familiarize themselves with any applicable regulations.

Certificates & Public/Private Keys

PDF files can also directly employ digital certificates and public/private key cryptography (commonly referred to as PKI or certificate-based security) for both restricting access to and electronically signing documents. While the details of public/private key cryptography are outside the scope of this document, PKI systems employ a public-private key pair, with the public part stored in a digital certificate. Each key forms part of a pair that functions mathematically in a single direction (potentially allowing the user of a digital certificate to use its keys for encryption/decryption or digital signing/verification).

For more information refer to ISO 17090, the standard for healthcare PKI.

Digital certificates may be stand-alone (self-signed) or integrated into a larger network of linked certificates—known as a Public Key Infrastructure (PKI). Self-signed certificates have a disadvantage in that they do not provide any third-party source of trust and should not be used as a sole source of trust for digital signatures. A full PKI provides verifiable links between end-user certificates and a central certificate authority. When issued at appropriate assurance levels, PKI certificates are generally more suitable for digital signature use.

Certificates may be stored on a user's local computer (software certificates), on a device such as a smart card or USB token (hardware certificates), or remotely on a server (roaming certificates). Passwords are usually used to further restrict the use of a certificate to a single system or individual.

Role-Based Authentication

Access to PDF documents may be based on authentication by individual users or by groups of users. Password-based authentication provides two very simple levels: a document open password, and a password that restricts user permissions. This allows a document creator to give a different password to different users based on these predetermined levels of access. More advanced role-based authentication can be implemented through certificate-based PDF security.

Certificate security allows users to set different levels of security based on individual users or groups of users as defined by organizational directories. Some groups might be only permitted to view a document, while others may have form-filling capability. Server-based document control systems can extend this a step further, and offer the ability to centralize management of groups and policies and may provide real-time modification of document access permissions.

Access Recovery

Document encryption has some potential for information loss. Individuals may forget or misplace their passwords. Smartcards and tokens that contain user's private keys may be lost or stolen. Without proper precautions, the loss of a password or key can mean the loss of any information that may have been encrypted using it.

With certificate-based encryption, some organizations employ a system of "key-escrow" where duplicates of user's private keys are centrally stored to reduce these issues. If a user's certificate or token is lost, the stored keys can be used to decrypt any encrypted information. Password recovery can be more difficult, or in some cases impossible.

In all cases, implementations should take appropriate precautions to mitigate this potential for information loss. Users should be aware that a lost password or lost tokens may mean complete information loss, and should take appropriate precautions to back-up or co-locate encrypted information in a secure, but alternately protected manner such as on a password-protected or encrypted storage device on a home PC, or remotely in a secure web-based service.

Access recovery methods are beyond the scope of this document but should be considered by the implementation.

Biometrics and Other Forms of Authentication

PDF documents may be protected indirectly by other means of authentication. Other methods of authentication may be single-factor, two-factor, or greater (e.g., password + token + biometric). Some applications provide the ability to incorporate biometric or other authentication methods. Generally though, those applications abstract or build upon the basics of PDF encryption as described above.

Authenticity

Authenticity involves establishing the origin and integrity of documents, but also non-repudiation and accountability of the users of those documents. Authenticity features include digital signatures, modification detection, versioning, and audit-trails.

Most of the authenticity features in PDF are a function of certificate-based signatures. Self-signed certificates are generally not suitable for authenticity except in combination with other forms of authentication. Infrastructure based (PKI) certificates of appropriate assurance levels are generally more appropriate for authenticity type roles, particularly in document of record or legal situations.

Digital Signatures

Digital signatures lay the foundation for the authenticity features of PDF. Digital signatures provide a specific type of electronic signature. They use the same public/private key systems and certificates that are used for certificate based encryption in PDF. The difference is digital

signatures, instead of using the keys to encrypt and decrypt the document, use the keys to encrypt and decrypt a "digest" of the document.

This process binds the exact state of the document to the individual key of a user, creating a unique and verifiable "signature." Digital signatures can be combined with other types of electronic signatures—such as signature tablets or biometrics—to provide additional levels or methods of authentication.

Non-Repudiation

Non repudiation—establishing that an individual accepted or approved a document—can be provided with certificate-based digital signatures in PDF. The effective level of non-repudiation varies largely depending on the assurance level of the certificate. The assurance level also determines the relative certainty that a specific individual used the certificate for signing purposes.

When using digital signatures to establish non-repudiation, higher assurance (multi-factor, token-based) certificates provide more effective evidence that a particular individual executed an electronic approval or acceptance. In most countries, comprehensive technical standards currently exist in determining the legally-binding use of digital signatures for non-repudiation.

While such determinations are outside the scope of this document, when using digital signatures in a stand-alone legal capacity, implementers should perform their legal due diligence when implementing a given solution.

Timestamps and Revocation

As an adjunct to basic non-repudiation, other information can be collected and stored in a PDF at the time of signing. This information—time stamps and revocation information—can provide additional evidence for the verification of a PDF signature. Time stamps are standards-based in PDF (RFC 3161) and provide independent third-party verification of the time and date when a signature was applied. This can be important from a non-repudiation perspective, since without a remote time-stamp, it's possible that the time and date obtained from a signer's computer could be either inadvertently or deliberately incorrect. PDF files support the embedding of secure time stamps into a digital signature to provide this additional assurance.

The capturing and embedding of revocation information at the time of signing provides additional non-repudiation capability for a signature. The digital certificates used to create digital signatures are not typically perpetual; they generally expire after a set period of time. There are also cases where certificates may be explicitly revoked—such as an employee leaving a company, or evidence of fraud.

It is important to consider the implications of expired certificates and signature validity. In general, outside the presence of overt fraud, as long as a certificate was valid at the time it was used to create a signature, that signature can be considered valid. Since signature verification typically includes certificate revocation checking, this presents a problem for verification of signatures that were created with certificates that may have expired.

A signature may be valid and the certificate may have been valid at the time of signing, but it may have since expired. This has the potential to introduce a margin of ambiguity into the signature verification process, especially over very long time-frames.

PDF includes a built-in mechanism to help reduce the impact of these scenarios. PDF has the ability to store revocation information—captured exactly at the time of signing—within a signed document for later use in verification. This means that a signed PDF can contain signed evidence of the validity of the certificates used to sign it and results in a self-contained and independent record of its own validity.

Versioning and Audit Trails

The final elements in establishing authenticity are the capability to maintain the integrity of a document and to provide definitive information (audit-trails) about what changes may have occurred to a document over its life-cycle. Integrity and audit trails are closely related in PDF. PDF documents support what is known as "incremental updating."

When changes are made to an incrementally updated PDF—anything from the completion of a form to the change of a single pixel constitutes a change—those changes are not saved in-line with the rest of the document. Instead, changes are saved at the end of a file, separate from the main body of the PDF. When the PDF is rendered, the modified sections are referenced from the end of the file, and replace the original in-line sections.

Incremental updates offer the ability go beyond simply establishing the integrity of a document. When a digital signature is applied, it creates a reference-point in the document's lifecycle. The signature applies to the version of the document that was signed. Any changes made after the application of a signature are appended to the end of a file. Incremental updates are built-in to the PDF signature process, and maintain both the integrity of multiple versions as well as a secure audit trail of document events.

Certification

Certification or (author signatures) is a special type of PDF signature. It must be applied as the first signature to a PDF, and it provides some additional features to the document creator or author. Certification allows an author to specify in advance the type of modifications that are permissible by other users. Those restrictions are not in the form of encryption-based control, but are user rights defined by the parameters of the signature. Examples include allowing users to complete forms, sign a document, or add annotations.

Certification is generally used by the author of a document prior to sending it to an unknown or outside party. In the case of forms, the authors can be confident that they are receiving back the same (not spoofed or modified) form that they sent out. The recipients can also be confident that the form they are receiving is the form the author intended (very important in phishing type attacks). It is also important to note that while certification-based restrictions may not always be honored explicitly by a PDF viewing application, any modifications will always be detectable by the certifying application.

Certification allows the author of a document to authenticate it before distribution by signing the layout, data bindings, logic, and scripts inside a document. Certification signatures give recipients assurance that a document was approved by the author and has arrived unaltered. They ensure that when the sender receives a completed form or signed document at the completion of a workflow, that none of their scripts, bindings, or logic has been altered by recipients.

Certification prevents recipients or other parties from inserting malicious scripts that might alter calculations or values, modifying the text of a contract, or attempting to hide text in small or page-colored fonts.

Certification parameters can be customized by the author to selectively allow form-filling, scripting, or interactivity with the document. Certification also provides a mechanism to notify recipients of any scripts or interactivity contained inside a document. Recipients are automatically alerted if a certified document allows form-filling actions or if it contains any scripts.

Certification allows authors to bind scripts to the certification status of a document. Certain security-sensitive scripts can only be run in certified documents where the recipient has explicitly trusted the author. Author Certification is highly recommended for any documents that are expected to be signed by recipients, documents that may become legally binding, or documents of record. It protects both authors and recipients, ensuring that content can not be maliciously or accidentally altered over the lifecycle of the document. In the context of the sample CCR, certification provides this functionality.

Instead of requiring post process validation of these sections manually, PDF provides the ability for authors to authenticate the document from the start, protect the content throughout the entire process, and re-authenticate the certification upon completion to establish validity.

Tamper-Evident Seal

PDF includes robust support for certificate-based digital signatures. With respect to modification detection, digital signatures in PDF are synonymous with the concepts of "Tamper Evident Sealing." They provide proof that a document has not been modified since a signature was applied, and they can uniquely bind a signed document to a particular digital identity.

Each signature in a PDF provides a tamper evident seal on the document. Digital signatures in PDF are standards-based (X.509, PKCS), and can provide robust authenticity for documents. They also support secure time-stamping, full path validation and revocation checking for certificates. PDF can fully support the SAFE (Signatures and Authentication For Everyone) standard.

AUDIT TRAIL

Secure audit trails have two fundamental requirements:

- Audit Data – Recorded information about the date/time and nature of an auditable event.
- Audit Authenticity – Assurance of the integrity of audit information of a document.

Audit Data

Since audit data is generally domain-specific—i.e. it concerns events that have particular meaning in a certain context, "signed," "populated," "tamper-sealed"—audit data is generally a member of a particular XML vocabulary. Support for industry-specific audit data in PDF is covered through PDF support for XML data standards. A PDF can contain any audit data as part of its underlying XML instance.

Audit Authenticity

Audit authenticity can take two forms: it can be provided by a system-specific mechanism, or through the built in capability of PDF to produce a secure audit trail using digital signatures. PDF files support the use of multiple and sequential digital signatures to ensure document authenticity.

Each signature on a PDF document produces an auditable event—and can include information like secure timestamps and certificate revocation responses—to establish exact event chronologies. The structure of PDF is based on an incremental update capability which lends itself well to multiple signature workflows. As a PDF is modified, the modifications to the document are added to the end of the file. When the document is rendered, the modifications are integrated into the current rendering—preserving the previous document states.

This incremental updating is transparent to the person viewing the document, but when combined with PDF digital signatures, allows for the detection and audit of modifications to the file. PDF applications that support digital signatures may allow users to view the state of a document at the time each signature was applied. Advanced PDF tools may use the audit information to visually identify and report on document modifications.

For more information see RFC 3881, an industry consensus document adopted by DICOM and ISO as a standard.

DOCUMENT INTEGRITY

Document integrity in PDF is a function of many of the previous sections—particularly those that deal with digital signatures, with some consideration toward understanding dynamic PDF capabilities and scripting. PDF digital signatures provide a strong, PKI-based integrity mechanism for documents.

Modifications to a digitally signed PDF are intrinsically detectable, and auditable. A document author may choose to limit what levels of interactivity or form filling capability are acceptable for a particular workflow with Certification signatures. Certification signatures in particular are critical to document integrity. Certification signatures authenticate a document before users interact with it. They can eliminate the need for additional data validation by ensuring document integrity throughout a workflow. Certification signatures also enable better multi-signature workflows since they reduce the need for individual signature validation.

SERVER-BASED DOCUMENT CONTROLS

Authentication and document control may also incorporate server-based systems that incorporate PDF encryption with organizational directories (LDAP, Active Directory), and server-based permission systems (FileOpen, Adobe LiveCycle Policy Server). These systems allow more real-time authentication and permissions management and are built on general PDF encryption principles with server based key systems.

SUMMARY OF PDF DOCUMENT SECURITY

Security Function	Comments
Password to open document	Consider allowing emergency information to be readily available without a password.
Password to restrict editing	Consider restricting editing of core documents to authorized users, though medical providers may need open access to some documents. NOTE: The use of this setting relies on the viewing application to enforce those permissions. Since not all PDF viewers may enforce permissions appropriately, users should not rely on edit permissions alone to protect highly confidential information.
Encrypt all document contents	This option makes finding an individual's records more difficult. It does not significantly increase security unless metadata contains personal information.
Encrypt all document contents except metadata	Use when encrypting documents. Available in PDF v1.5 and later. Search engines can access metadata. NOTE: XFA information, which includes the XML Form Data, is not metadata and will be encrypted as part of the document.
Encrypt only file attachments	Available in PDF 1.6 and later. Can be used for secure container type of PDF files.
Restricted printing	Some PDF viewing software will allow a document creator to restrict the document users' ability to print the document. Some viewers prevent printing, while others permit printing but with noticeable poor quality. Restricted printing should not be considered secure in untrusted environments. Strong user authentication generally provides a significantly greater level of security than restricted printing.
Restrict changes	Set level as appropriate on a per-document basis.
Enable copying of text, images, and other content	Enabling copying allows full copy-and-paste access to content.
Enable text access for accessibility	This setting may not be honored by all PDF viewers and should be used carefully. If misused, it equates to the "Enable copying…" setting above.