Department of Veterans Affairs

Product Development (PD)

Veterans Benefit Management System (VBMS)

PDF Specification Document

Agreement # VA118-10-IA-0001-027

Version 1.0
April 2012

Space and Naval Warfare Systems Center Atlantic
P.O. Box 190022
North Charleston, SC 29419-9022

Revision History

Date	Version	Description	Author
4 APR 2012	Draft	Initial Version	VBMS Architecture
13 APR 2012	1.0	Release to VA	VBMS Architecture

Table of Contents

List of Figures

List of Tables

This page intentionally left blank.

1. Veterans Benefits Management System Mission

The mission of the Department of Veterans Affairs (VA), Office of Information & Technology (OIT), and the Office of Product Development (PD) is to provide benefits and services to veterans of the United States. In meeting these goals, OIT strives to provide high quality, effective, and efficient Information Technology (IT) services to those responsible for providing care to veterans on-site and throughout all the points of the veterans' health care in an effective, timely and compassionate manner. The VA depends on Information Management/Information Technology (IM/IT) systems to meet mission goals. Under Interagency Agreement (IA) VA 118-10-IA-0001, dated November 4, 2009 between the VA and the Department of the Navy, Space and Naval Warfare Systems Center, Atlantic (SPAWARSYSCEN Atlantic), the VA and OIT, requests SPAWARSYSCEN Atlantic to assist the VA in meeting its mission goals. This assistance will be in the areas of project management, configuration management, requirements management, integration with VA and non-VA systems, interfaces to VA and non-VA systems, Virtual VA conversion, web portal development, services development, testing, information assurance, architectural engineering support and in the integration of the various components/documents being delivered from various contractors and documentation to the VA.

The Veterans Benefits Administration (VBA) and OIT are seeking solutions for managing the backlog of paper claims, electronic documents, correspondence, and other content created and handled as part of their day-to-day business processes. As the volume of content grows exponentially, the VA faces increased pressure to operate more efficiently, while reducing costs and addressing veterans' needs for faster, simpler, and more efficient and effective services while adhering to industry best practices and compliance regulations.

To accomplish this, a comprehensive plan is in the process of being executed that will:

- Target deficiencies in workflows to maximize efficiencies and enable the veteran to receive benefits in a timely and efficient manner, as well as provide veterans with the ability to monitor the status of his/her benefits.
- Promote a solution that provides a unified content, process and compliant environment.
- Provide interoperability between databases, applications, operating systems, portals, security, servers, storage, systems management tools and web server environments.
- Deploy a Service Oriented Architecture (SOA) shared services platform to help reduce operations costs and simplify IT infrastructure.
- Provide operational efficiency across the organization by creating a common interface for building and deploying content and process applications.

Veterans Benefits Management System (VBMS) is a multi-year technology solutions project to transition the VBA from a paper-intensive claims processing environment to a paperless-based environment. The ultimate goal of this program is to provide the VBA with systems engineering support for the overall design of the VBMS solution (to include security components). This support also includes test and integration services for VBMS Phase 1 and the development of VBMS, the development of SOA objects, development of services for the user interface to communicate with the corporate database, development of the VBMS database, security to meet the requirements of VA Directive 6500 Information Security Handbook, and development of functionality to support end-to-end claims processing in an electronic environment. End-to-end claims processing provides functionality required for

establishment, development, rating, award, and appeal of a claim. It is the intent of this effort to deliver VBMS incrementally using an agile development methodology based upon the priorities determined by OIT.

1.1 Purpose

The purpose of this document is to create the specification for PDF document generation within VBMS. The guidance provided in the VBMS SOA Reference Architecture document outlined the patterns and practices for documenting a specification.

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this specification are to be interpreted in this specification as described in IETF RFC 2119. These keywords are thus capitalized when used to unambiguously specify requirements over protocol and application features and behavior that affect the interoperability and security of implementations. When these words are not capitalized, they are meant in their natural-language sense.

The following source provides additional information about IETF RFC 2119.

- http://www.ietf.org/rfc/rfc2119.txt

This specification is broken down into the following sections: The history and goals of the PDF specification are described, then the relevant ISO standards on which the VBMS PDF Document Specification is built are discussed in section two, followed by a background discussion on relevant PDF document metadata standards in section three, and finally the VBMS PDF Document Specification in section four.

1.2 History and Goals

The goal of PDF is to enable users to exchange and view electronic documents easily and reliably, independent of the environment in which they were created or the environment in which they are viewed or printed. At the core of PDF is an advanced imaging model derived from the PostScript® page description language. This PDF Imaging Model enables the description of text and graphics in a device-independent and resolution-independent manner.

A PDF document consists of a collection of objects that together describe the appearance of one or more pages, possibly accompanied by additional interactive elements and higher-level application data. A PDF file contains the objects making up a PDF document along with associated structural information, all represented as a single self-contained sequence of bytes.

A document's pages (and other visual elements) can contain any combination of text, graphics, and images. A page's appearance is described by a PDF content stream, which contains a sequence of graphics objects to be painted on the page. This appearance is fully specified; all layout and formatting decisions have already been made by the application generating the content stream.

In addition to describing the static appearance of pages, a PDF document can contain interactive elements that are possible only in an electronic representation. PDF supports annotations of many kinds for such things as text notes, hypertext links, markup, file attachments, sounds, and movies. A document can define its own user interface; keyboard and mouse input can trigger actions that are specified by PDF objects. The document can contain

interactive form fields to be filled in by the user, and can export the values of these fields to or import them from other applications.

Finally, a PDF document can contain higher-level information that is useful for interchange of content among applications. In addition to specifying appearance, a document's content can include identification and logical structure information that allows it to be searched, edited, or extracted for reuse elsewhere.

Figure 1 below depicts the goal of the VBMS PDF specification – to convert all relevant media types into a file format that is optimized for current/future usages, in expected scanning/rendering scenarios.

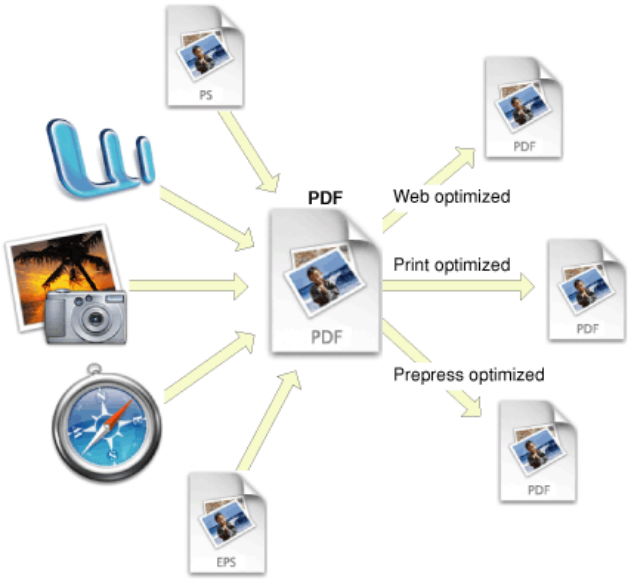

Figure 1: PDF Specification Goals

The following source provides additional information about PDF.

- PDF Reference, sixth edition, Adobe® Portable Document Format, Version 1.7, November 2006
  http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

2. ISO Standards

2.1 ISO 19005-1 (PDF/A)

The ISO 19005-1 standard for PDF/A-1 was published in 2005 and geared towards long-term preservation. It provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files and aims to support future access and future migration needs by providing frameworks for: 1) embedding metadata about electronic documents, and 2) defining the logical structure and semantic properties of electronic documents. The PDF/A-1 specifications are based on the specifications of PDF 1.4 and describe which aspects of a PDF are compulsory, optional or prohibited in a PDF/A-1 compliant file. PDF/A-1 defines two conformance levels:

- A ‘Level A’ conformant file (PDF/A-1a) SHALL adhere to all requirements of the ISO standard.
- A ‘Level B’ conformant file (PDF/A-1b) does not have to adhere to two requirements: the use of tagging to specify the document’s logical structure and reading order, and the use of Unicode character maps that map character codes to Unicode values.

As preserving the natural reading order and using Unicode are crucial for the preservation of the document as it was intended by the creator, the VBMS PDF specification will be a superset of PDF/A-1a conformance.

PDF/A-1 files MUST include:

- Embedded fonts
- Device-independent color
- XMP metadata

PDF/A-1 files MAY NOT include:

- Encryption
- LZW Compression
- Embedded files
- External content references
- PDF Transparency
- Multi-media
- JavaScript

The following source provides additional information about ISO 19005-1.

- PDF ISO 19005-1, December 2004 http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920

2.2 ISO 24517-1 (PDF/E)

ISO 24517-1 was published in 2008 and defines a format for the creation of documents used in engineering workflows and is based on the PDF version 1.6. It leverages U3D, another open standard, for the representation of 3D content.

2.3 PDF Healthcare (PDF/H)

PDF Healthcare is a “Best Practices Guide” (BPG) that is supplemented by an “Implementation Guide” (IG). The PDF Healthcare BPG and the IG are based on open, published specifications with direction specific to the health care industry. The BPG and IG describe attributes to facilitate the capture, exchange, preservation and protection of healthcare information. Such attributes include the ability for health care providers and consumers to develop a secure, electronic container that stores and transmits relevant health care information, important for maintaining and improving health. The health care information can include but is not limited to personal, handwritten documents, (structured or unstructured) clinical notes, (structured) laboratory test result reports, (unstructured) word-processed / text summary reports, electronic forms, scanned document images, digital diagnostic images, photographs, and signal tracings (e.g., electrocardiograms [ECGs]).

PDF Healthcare is NOT a proposed standard. The PDF Healthcare BPG and IG are intended to be used to guide the generation and consumption of secure and portable containers of personal

health information and electronic health record information rather than replacing existing standards or adding new standards for health care information interoperability.

The following source provides additional information about PDF/H.

- http://www.aiim.org/documents/standards/PDF-h_Implementation_Guide_2008.pdf

2.4 ISO 14289 (PDF/UA)

PDF/Universal Accessibility is the International Standard for accessible PDF documents. The mission of PDF/UA is to develop technical and other standards for the authoring, remediation and validation of PDF content to ensure accessibility for people that use assistive technology such as screen readers for users who are blind.

PDF documents must comply with the checkpoints specified in § 1194.22 (Web-based Intranet and Internet Information and Applications).

Constraints mandated by Section 508 are:

- All images MUST have alternate text
- Pages designed to convey information via color MUST be adequately tagged
- All low-contrast (shaded) text MUST be rendered as image (with tags) or removed
- Row and column headers MUST be included for data tables. Multi-level headings for row/columns or spanning cells SHOULD be avoided.

The following source provides additional information about PDF and 508 Compliance.

- http://www.access-board.gov/sec508/guide/1194.22.htm

2.5 ISO 32000-1

ISO 32000 specifies a digital form for representing PDF documents. PDF was developed and specified by Adobe Systems Incorporated beginning in 1993 and continuing until 2007 when the ISO standard was prepared. The Adobe Systems version PDF 1.7 is the basis for the ISO 32000 edition. The specifications for PDF are backward inclusive, meaning that PDF 1.7 includes all of the functionality previously documented in the Adobe PDF Specifications for versions 1.0 through 1.6.

PDF, together with software for creating, viewing, printing and processing PDF files in a variety of ways, fulfills a set of requirements for electronic documents including:

- Preservation of document fidelity independent 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,
- 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 for use with other file formats and applications, and
- Electronic forms to gather data and integrate it with business systems

This standard does not specify:

- Specific processes for converting paper or electronic documents to the PDF format;

- Specific technical design, user interface or implementation or operational details of rendering;
- Specific physical methods of storing these documents such as media and storage conditions;
- Methods for validating the conformance of PDF files or readers

The following source provides additional information about ISO 32000.

- PDF ISO 32000, January 2008, http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf

3. Metadata

3.1 Info Dictionary

The Info Dictionary has been included in PDF since version 1.0. The optional Info entry in the trailer of the file can hold a document information dictionary containing metadata for the document. Any entry whose value is not known should be omitted from the dictionary rather than included with an empty string as its value. It contains a set of document info entries, simple pairs of data that consist of a key and a matching value. Applications can add their own sets of data to the info dictionary.

3.2 Extensible Metadata Platform

XMP (Extensible Metadata Platform) is an Adobe technology for embedding metadata into files. It can be used with a wide variety of data files. With Acrobat 5 and PDF 1.4 (2001) this mechanism was also made available for PDF files. XMP is more powerful than the info dictionary, which is why it is used in a number of PDF-based metadata standards.

Extensible Metadata Platform (XMP) metadata travel with files, and can be embedded in many common file formats including PDF, TIFF, and JPEG. Metadata properties are grouped in XML schemas. Each schema is identified by a unique namespace URI and holds an arbitrary number of properties. The XMP specification includes more than a dozen predefined schemas with hundreds of properties for common document and image characteristics. The most widely used predefined XMP schema is called the Dublin Core, or dc. It includes general properties such as Title, Creator, Subject, and Description. In addition to predefined schemas, custom schemas can be defined to cover specific metadata requirements.

The following sources provide additional information about XMP and Dublin Core.

- XMP Specification Part 1, Data Model, Serialization, And Core Properties July, 2010
- XMP Specification Part 2, Additional Properties, July, 2010
- XMP Specification Part 3, Storage In Files, July 2010
- ISO 16684-1, Feb 2012
- http://dublincore.org/

Many libraries/tools exist to facilitate the setting/getting of XMP metadata content. The following source provides additional information about XMP toolkits.

- http://en.wikipedia.org/wiki/Extensible_Metadata_Platform

4. VBMS PDF Document Specification

This section describes the specifications each PDF placed in the VBMS system must adhere to in order to be in compliance. The source of the requirement is listed in parentheses after the requirement.

4.1 Requirements

To be in compliance with the VBMS PDF specification, each PDF MUST:

- Be PDF 1.3 or later in PDF Normal or PDF Searchable Image formats only. PDF Image Only formats are not acceptable. (ISO 32000)
- Be scanned at a resolution of 300 dots per inch (dpi) to ensure that the pages of the document are legible both on the computer screen and when printed, while at the same time minimizing the file size. (VBMS Constraint)
- Not be down-sampled, which is a process of decreasing the number of pixels in the image. Down-sampling is an option in PDF optimizing but can lead to poor quality images. (VBMS Constraint)
- Be despeckled, by removing isolated “dots” within the image which can cause recognition problems, making the result image cleaner. (VBMS Constraint)
- Be deskewed, improving OCR results by straightening crooked pages. (VBMS Constraint)
- Have all fonts embedded. Doing so improves text searching in the PDF. Some TrueType fonts have a setting added by the font designer that prevents the font from being embedded and should be avoided. (ISO 19005)
- Be a tagged PDF. Tags describe logical structural aspects (paragraphs, lists, tables, links, illustrations, etc.) and are used by rendering applications and screen-reader devices. (ISO 19005 and ISO 14289)
- Provide alternative text descriptions for all non-text document elements. (e.g. images that are part of the document's content) You This does not apply to images that provide decorative elements to the document. (ISO 14289)
- Contain no security restrictions (password, certificate, etc.). These restrictions could hamper future use or transformations of the documents, and can interfere with screen readers. (ISO 19005)
- Not contain any bookmarks or links (ISO 19005)
- Not contain crop marks, registration marks, date stamp, time stamp, or any other mark that does not appear in the original document. (ISO 19005)
- Specify color spaces in a device-independent manner. (ISO 19005)
- Be created from source documents using the “Optimize the PDF for fast web view” option to reduce file sizes and file opening times. (ISO 19005)
- Use JPEG 2000 for compression. The advantage of this compression algorithm is that it supports both lossy and lossless data compression. Lossless is required for text-based content where no loss is acceptable. Lossy is acceptable for image content. (ISO 19005)
- Use lower case characters and avoid using special characters except hyphens and underscores in file names. Special characters to avoid include punctuation, spaces, or other non-alphanumeric symbols (e.g., \ / : * ? < > | “ % # +) . (VBMS Constraint)

To be in compliance with the VBMS PDF specification, each PDF SHOULD:

- Use gray scale and color sparingly, as it significantly increases the file size. Gray scale and color should be used only when these features improve the reviewability of the material. (VBMS Constraint)

4.2 Metadata

Table 1 lists the metadata REQUIRED for all PDFs:

Element	Value	Comments
Contributor (dc)	Persons, Organizations, or Services	The entity responsible for making contributions to the content of the resource
Creator (dc)	Person, Organization, or Service	The entity primarily responsible for making the content of the resource
Date (dc)	String	The date(s) associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource
Description (dc)	String	Textual description of the content of the resource
Format (dc)	Number of documents/pages within the file; Size of the file; Dimensions of the resource	Designate the physical or digital manifestation of the resource
Identifier (dc)	String or number conforming to a formal identification system	Unambiguous reference to the resource within a given context
Publisher (dc)	Person, Organization, or Service	An entity responsible for making the resource available.
Subject (dc)	String	Describes the content of the resource
Title (dc)	String	A name given to the resource.
Type (dc)	Text, StillImage, or Collection	To describe the file format, physical medium, or dimensions of the resource, use the Format element.

Table 1: Required Metadata

Appendix A: Acronym List

Acronym	Definition
DCMI	Dublin Core Metadata Initiative
ISO	International Standards Organization
IT	Information Technology
IM/IT	Information Management/Information Technology
JPEG	Joint Photographic Experts Group
OIT	Office of Information & Technology
PD	Product Development
PDF	Portable Document Format
PDF/UA	Portable Document Format/User Accessibility
SOA	Service Oriented Architecture
SPAWARSYSCEN	Department of the Navy, Space and Naval Warfare Systems Center, Atlantic
TIFF	Tagged Image File Format
U3D	Universal 3 Dimensional
VA	Veterans Affairs
VBA	Veterans Benefits Administration
VBMS	Veterans Benefit Management System
XML	eXtensible Markup Language
XMP	eXtensible Metadata Platform

Appendix B: VBMS Architecture Attributes

The following architecture attributes were harvested from the VBMS technical requirements in the System Requirements Specification and were presented at the VBMS architecture planning offsite meeting. The VBMS architecture team has agreed that these attributes serve to create a solid foundation for VBMS and should be taken into account when making decisions or providing architectural guidance.

Available	Business goals driven
Coexist with Legacy	Discoverable
Documented	Allows for better, faster workflow
Encapsulated; do not negatively impact existing systems	Enterprise focused - 'Common Devices'
Expandable	Extensible
Flexible/Agile	Governance
Heterogeneous	Implementable
Interoperable	KIS - Keep it simple
Layered; not fragile	Loosely coupled
Low barrier to entry for users	Maintainable
Policy driven/compliant	Recoverable
Role based, access	Scalable
Secure; protect sensitive data	Service Oriented; Reusable
Standards based; compliance standards	Thin client/Web based
Useable	

Approval Signatures

This section is used to record the approval of the *PDF Specification Document* document during the Formal Review. Conduct the review face-to-face where signatures can be obtained 'live' during the review. If unable to conduct a face-to-face meeting, conduct the review via a teleconferencing medium and capture concurrence during the meeting. The Scribe should add names by each position cited.

The following members governing the document are required to sign. Please annotate signature blocks accordingly.

REVIEW DATE: <*date*>

SCRIBE: <*name*>

_____

Signed: Date:

James Barr

SPAWAR Project Manager

_____

Signed: Date:

Clayton Coleman

SPAWAR Lead System Engineer

_____

Signed: Date:

Tracie Loving

VA Program Manager