1. Introduction

1.1. PDF - Portable Document Format

PDF was developed in the early 1990s as a way to share computer documents, including text formatting and inline images. In those early years before the rise of the World Wide Web and HTML documents, PDF was popular mainly in desktop publishing workflows.

Why Use PDF?

A properly prepared PDF will maintain the original fonts, images, graphics as well as the exact layout of the file (think of it as an electronic snapshot). A PDF file can be shared, viewed, and printed by anyone using the free Adobe Reader software regardless of the operating system, original design application or fonts. Originally PDF was mostly used by graphic artists, designers and publishers for producing color page proofs. With its evolving technology, however, today PDF is used for virtually any data that needs to be exchanged among applications and users. It is an open file format specification and PDF is available to anyone who wants to develop tools to create, view or manipulate PDF documents.

Why is PDF important?

It is becoming increasingly easy to create PDF files as (from a user's stand-point) the process is almost as simple as printing. Essentially, anything that can be done with a sheet of paper can be done with a PDF. PDF technology is being used more frequently to produce offset printed documents (provided the designer properly embeds fonts and images). Adding to mainstream adoption, of course, is the fact that many applications allow users to save, import or export a document as a PDF (including popular publishing programs like QuarkXPress and CorelDraw), and you can also find a variety of third-party PDF conversion software tools available. With the capability to embed metadata (data about data) in a PDF file, along with the use of security options and electronic signatures PDF is also becoming a standard for data archiving. It may have taken a few years to perfect and years of dedication by the development team at Adobe, but today more and more people are turning to PDF as the solution for something not even thought of in 1993.

Technical details about PDF

A PDF file is basically a 7-bit ASCII file, except for certain elements that may have binary content. A PDF file starts with a header containing the magic number and the version of the format such as PDF-1.7. The format is a subset of a COS ("Carousel" Object Structure) format.[15] A COS tree file consists primarily of objects, of which there are eight types:

— Boolean values, representing true or false

— Numbers

— Strings, enclosed within parentheses ((...)), may contain 8-bit characters.

— Names, starting with a forward slash (/)

— Arrays, ordered collections of objects enclosed within square brackets ([...])

— Dictionaries, collections of objects indexed by Names enclosed within double pointy brackets

— Streams, usually containing large amounts of data, which can be compressed and binary

— The null object

1.2. Web Communication Methodologies

1.3. Current methods to convert data to PDF