PDF for Documents, XML, and Rich Content

Ed Chase

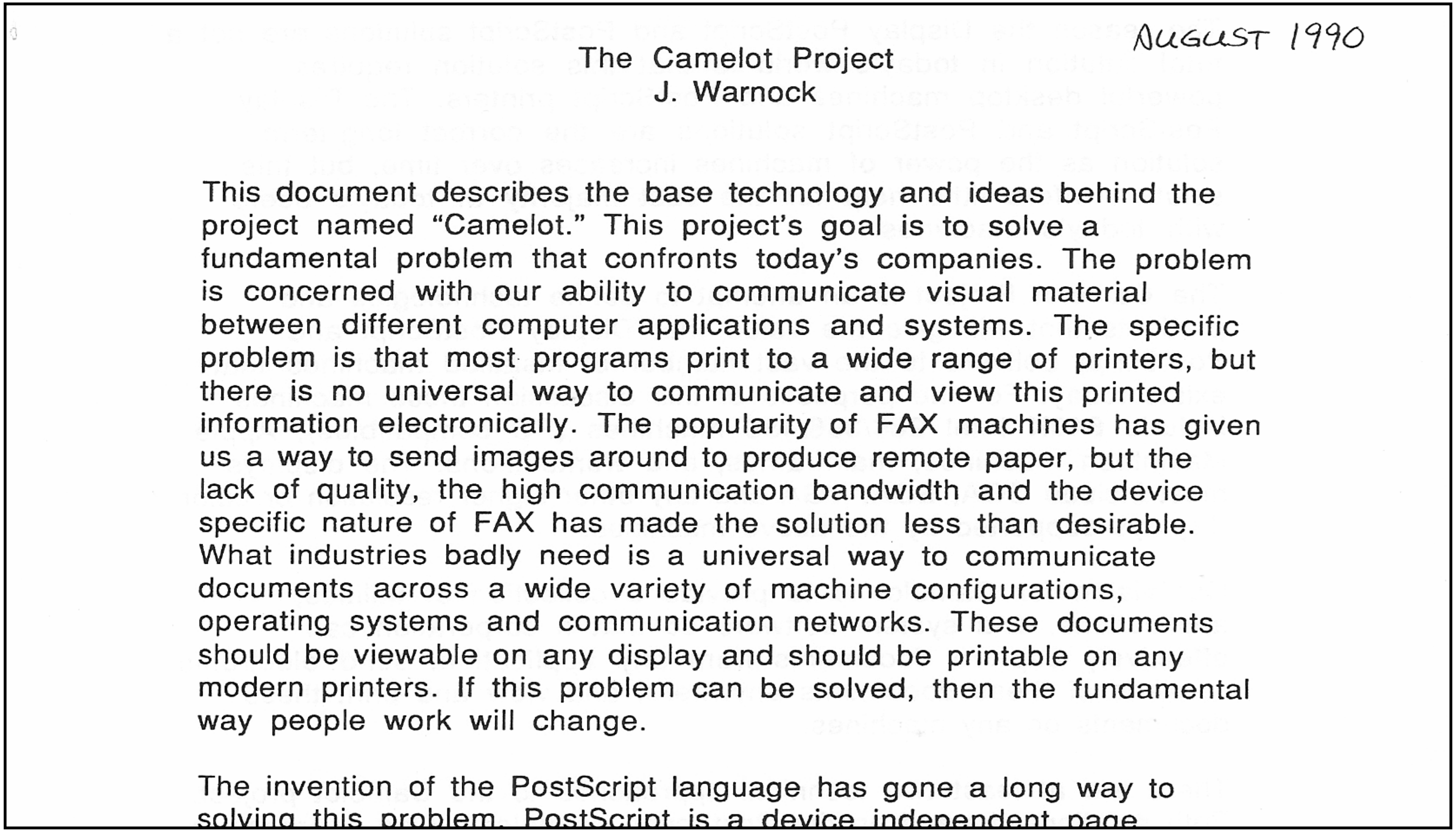
Worldwide Standards Engineer

Adobe Systems

April 2006

Part 1
- The Origins of PDF

The Camelot Paper



August 1990

The Camelot Project
J. Warnock

This document describes the base technology and ideas behind the project named "Camelot." This project's goal is to solve a fundamental problem that confronts today's companies. The problem is concerned with our ability to communicate visual material between different computer applications and systems. The specific problem is that most programs print to a wide range of printers, but there is no universal way to communicate and view this printed information electronically. The popularity of FAX machines has given us a way to send images around to produce remote paper, but the lack of quality, the high communication bandwidth and the device specific nature of FAX has made the solution less than desirable. What industries badly need is a universal way to communicate documents across a wide variety of machine configurations, operating systems and communication networks. These documents should be viewable on any display and should be printable on any modern printers. If this problem can be solved, then the fundamental way people work will change.

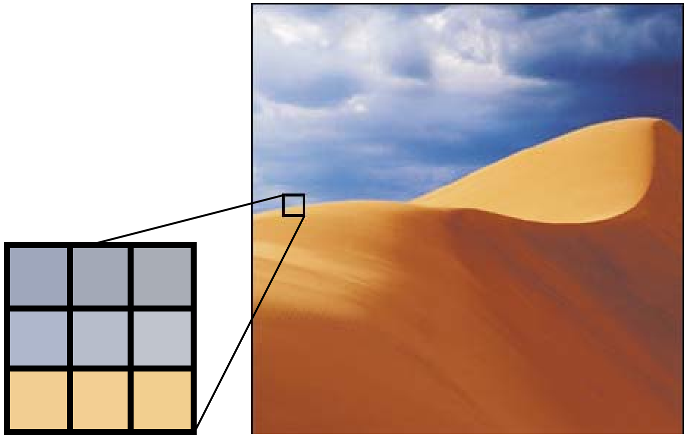
The invention of the PostScript language has gone a long way to solving this problem. PostScript is a device independent page

Adobe & PostScript

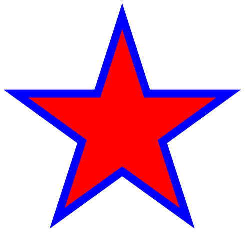
- **Original ideas date back to 1976**
  - John Warnock's concepts from a 3D graphics system invented to track ships in NY harbor
  - Adobe founded in 1982 by John Warnock & Chuck Geschke
- **PostScript (1985)**
  - Page description language
  - Replaced ASCII text with structured layout and precise font handling & *hinting*
  - Actually a programming language – running the program interprets the content
  - Includes flow commands & loops
  - Had to be processed by a RIP *(Raster Image Processor)* for display & print
  - Off-loaded processor intensive print jobs to printer
    - *The Apple LaserWriter was a more powerful computer than the Mac desktop PC*
  - Large files, external resources (fonts & images)
  - *EPS (Encapsulated PostScript)* includes a rendered bitmap for easier viewing

Imaging

- **Bitmap (Raster) Graphics**
  - Rows of pixels at different levels of color (or gray)
  - Photos & web images TIFF,GIF, JPEG, PNG
  - Resolution dependant
  - The larger the image, the larger the file



- **Vector Graphics**
  - X & Y coordinates, fixed colors & gradients, fonts, Bezier curves
  - PostScript, WMF, SWF, SVG
  - Resolution independent - decided by display/printer
  - Usually more efficient/smaller than bitmap

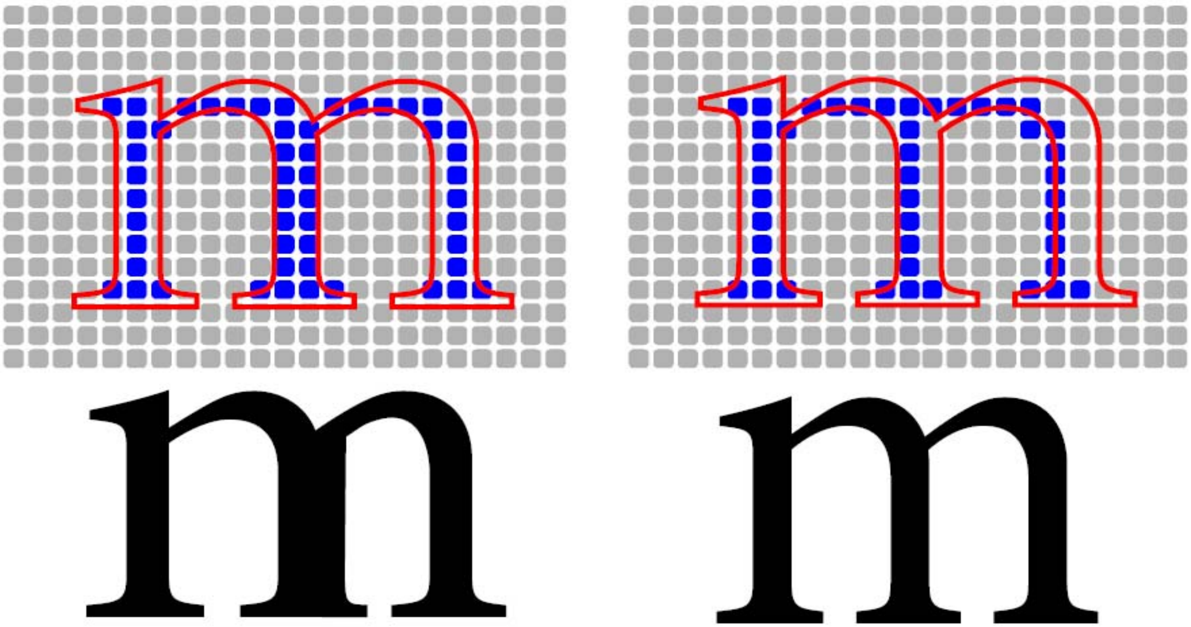


```
<polygon fill="red" stroke="blue"
stroke-width="10" points="350,75
379,161 469,161 397,215 423,301
350,250 277,301 303,215 231,161
321,161" />
```

Precise Typography

- Outline Fonts
  - Vector-based for efficiency at multiple sizes and consistent rendering
  - Scalable & rotate-able

- Anomalies
  - Uneven stem widths
  - Hints and/or anti-aliasing

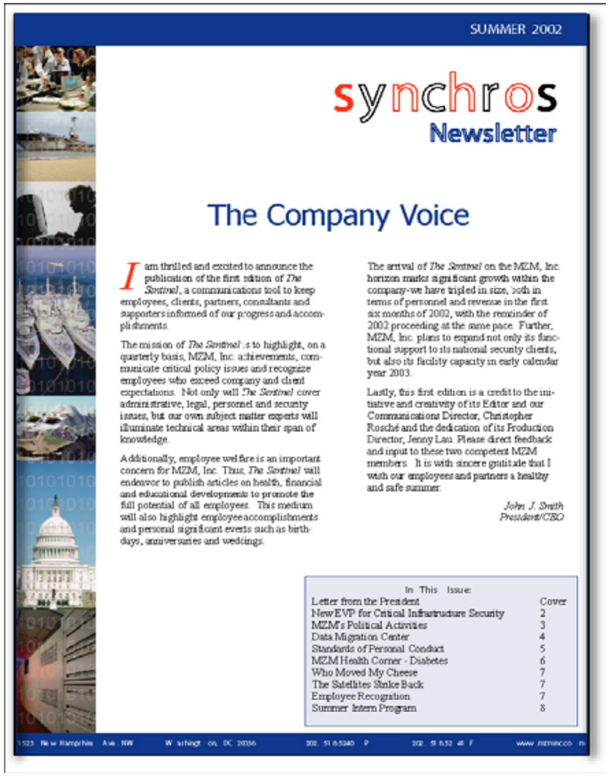


2005 Adobe Systems Incorporated. All Rights Reserved.

The Portable Document Format – PDF (1992)

- View & print rich content from any application across platforms & operating systems
- Simplified "object-oriented" PostScript for page layout and vector images
- Supports embedded bitmaps and image compression
- Rich font support with embedding and subsetting
- Single file packaging and structure format with compression
- Less computation-intensive, faster to open - interpreted results of PS code

Composite Documents



SUMMER 2002

synchros
Newsletter

The Company Voice

I am thrilled and excited to announce the publication of the first edition of *The Sentinel*, a communications tool to keep employees, clients, partners, consultants and supporters informed of our progress and accomplishments.

The mission of *The Sentinel* is to highlight, on a quarterly basis, MZM, Inc. achievements, communicate critical policy issues and recognize employees who exceed company and client expectations. Not only will *The Sentinel* cover administrative, legal, personnel and security issues, but our own subject matter experts will illuminate technical areas within their span of knowledge.

Additionally, employee welfare is an important concern for MZM, Inc. Thus, *The Sentinel* will endeavor to publish articles on health, financial and educational developments to promote the full potential of all employees. This medium will also highlight employee accomplishments and personal significant events such as birthdays, anniversaries and weddings.

The arrival of *The Sentinel* on the MZM, Inc. horizon marks significant growth within the company-we have tripled in size, both in terms of personnel and revenue in the first six months of 2002, with the remainder of 2002 proceeding at the same pace. Further, MZM, Inc. plans to expand not only its functional support to its national security clients, but also its facility capacity in early calendar year 2003.

Lastly, this first edition is a credit to the initiative and creativity of its Editor and our Communications Director, Christopher Rosché and the dedication of its Production Director, Jenny Lau. Please direct feedback and input to these two competent MZM members. It is with sincere gratitude that I wish our employees and partners a healthy and safe summer.

John J. Smith
President/CEO

In This Issue:

1523 New Hampshire Ave., NW Washington, DC 20036 202. 518.5340 P 202. 518.5240 F www.mzminc.com

Pages

Images

Graphics

Fonts

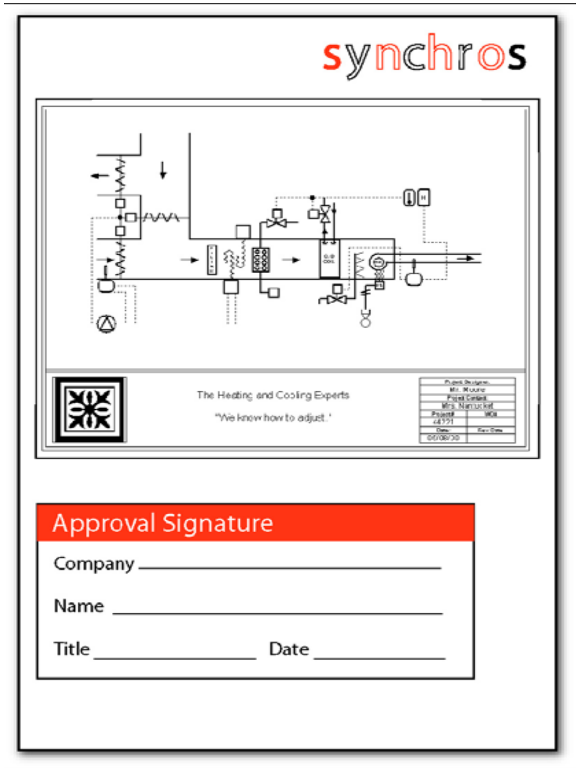
Colorspaces

Metadata

Annotations

Links

Digital signatures



synchros

The Heating and Cooling Experts

"We know how to adjust."

Project Designer
Mr. Moore
Project Contact
Mrs. Nantucket

Project#	WO#
44721	
Date	Rev Date
06/08/00	

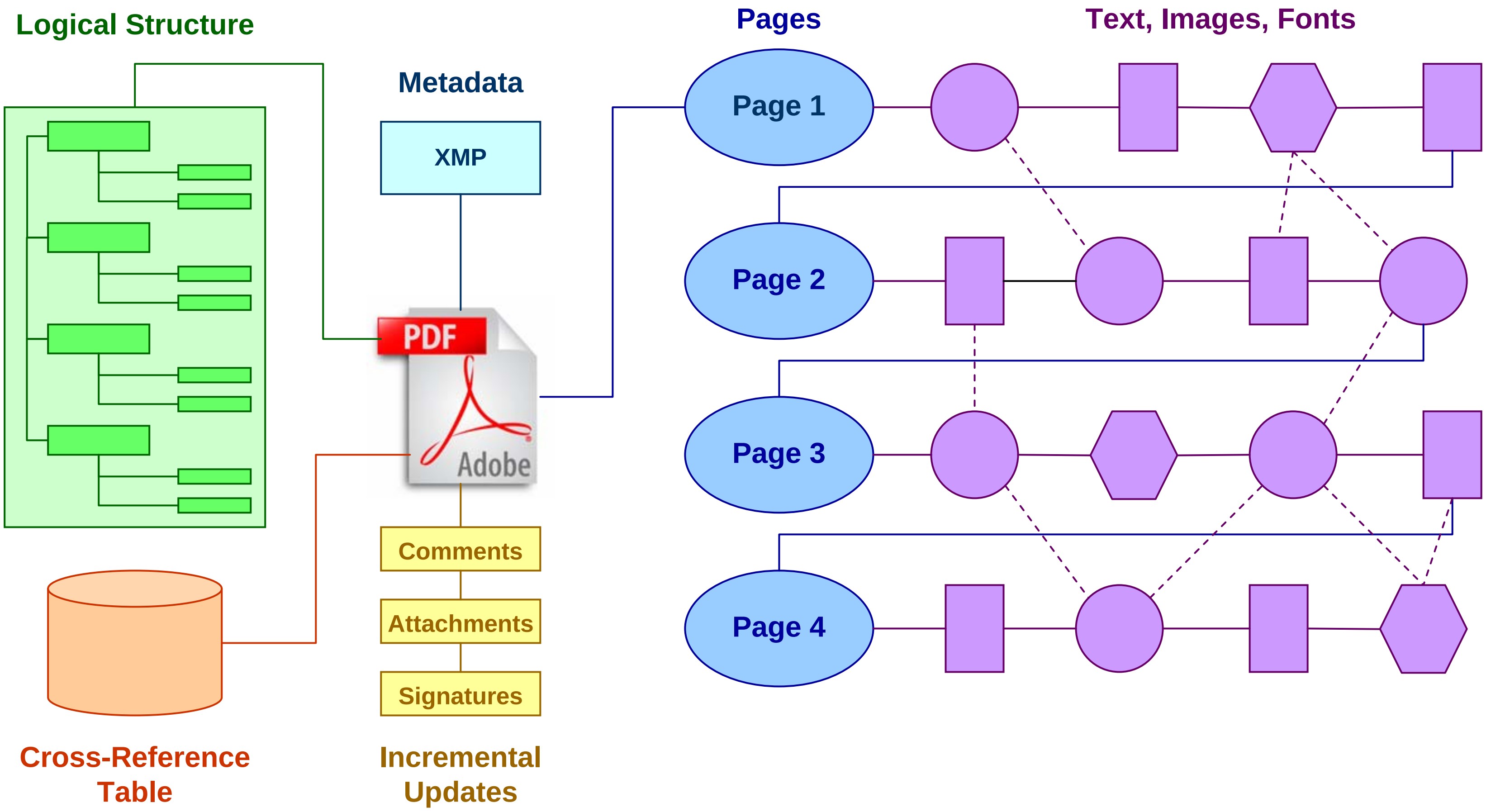
Approval Signature

Company ______________________________

Name ______________________________

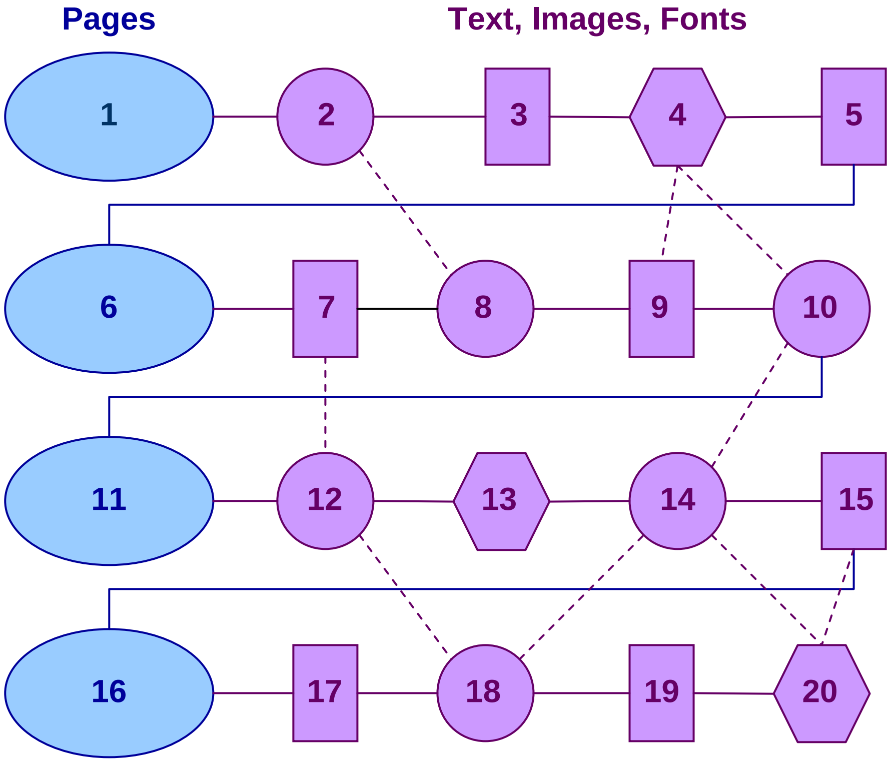
Title ______________ Date ______________

Inside of a PDF



Objects

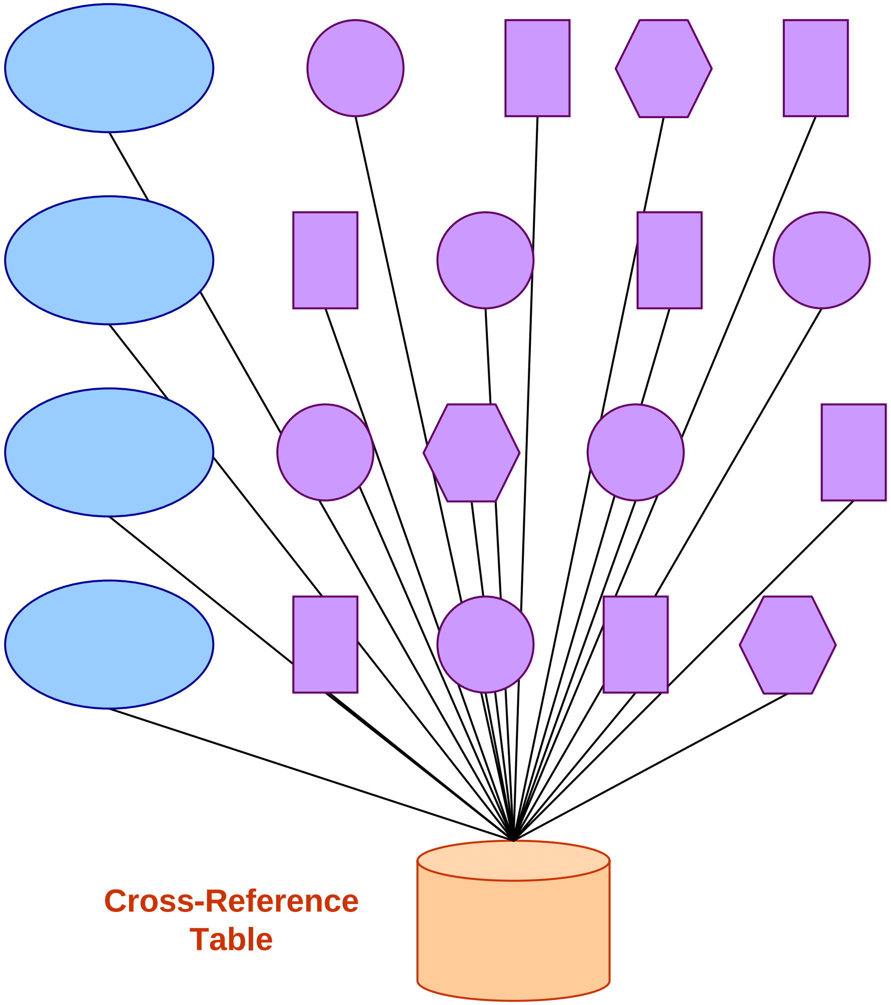
- PDF files are made from "objects"
- Objects are sequential, but can occur in any order in a file
  - Single-pass file generation
- Objects can refer to each other by number
- References can create a cross-linked set of objects (mathematical graph)
- Cross reference table maps object numbers to places within the file



Objects

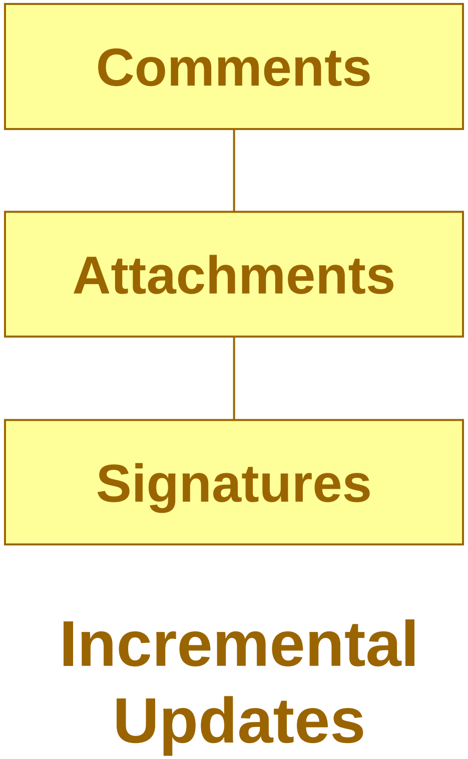
The Cross Reference Table

- No logical order to objects
  - Arbitrary references among objects – fonts, images, colorspaces
  - Processing follows references not sequence
- Cross-reference table provides exact locations of each object
- Allows fast and efficient *random access* to objects without having to parse entire file
- Located at end of file to allows single-pass generation of PDFs



Annotations & Incremental Updates

- Annotations
  - Comments, notes, highlights, sounds, movies, attachments, stamps, signatures
  - Interactive
  - May use external viewers & codecs
- Incremental Updates
  - Modifications are added to the end of the file
  - Cross-reference table is updated
  - Allows for versioning and detecting modifications between signatures



Metadata & Logical Structure

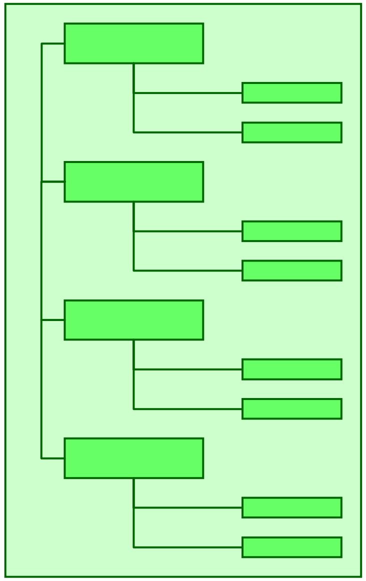
- Metadata
  - Descriptive information about a document
    - Subject, author, keywords
  - XMP – eXtensible Metadata Platform
  - Based on XML RDF
    - Dublin Core + Extensible
    - Published spec & Open Source tools

- Logical Structure
  - Tagged PDF
  - Presents contents in logical human reading order of a document
  - Re-flow (PDAs & Phones)
  - Accessibility

Metadata

XMP

Logical Structure

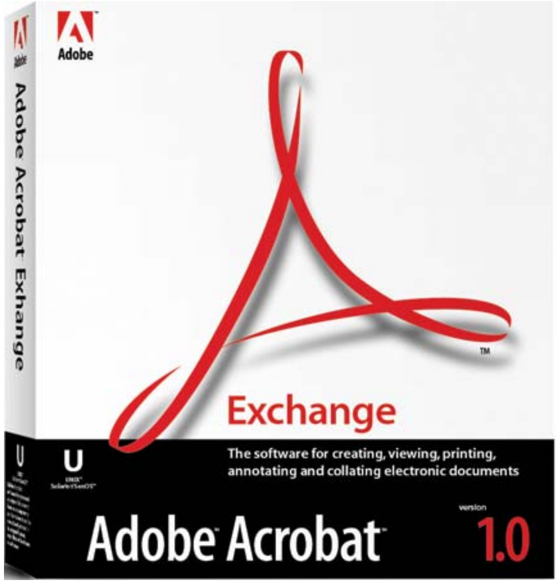


Part 2

- 15 Years of PDF

PDF 1.0 - 1993

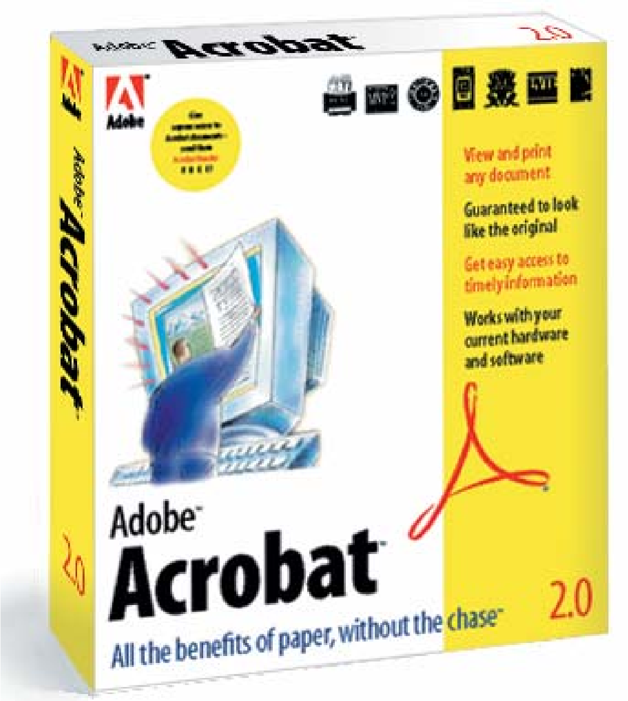
- PDF 1.0 Reference published
  - 214 pages
- View and print anywhere
- Simple find, link, annotations features
- Files were ASCII85 encoded
- Solved the font problem
  - Embedding & subsetting


Adobe
Adobe Acrobat Exchange
Exchange
The software for creating, viewing, printing, annotating and collating electronic documents
version
Adobe Acrobat 1.0

- Acrobat 1.0 – June, 1993
  - Reader was $50
  - Creation tools ~ $800

PDF 1.1 - 1994

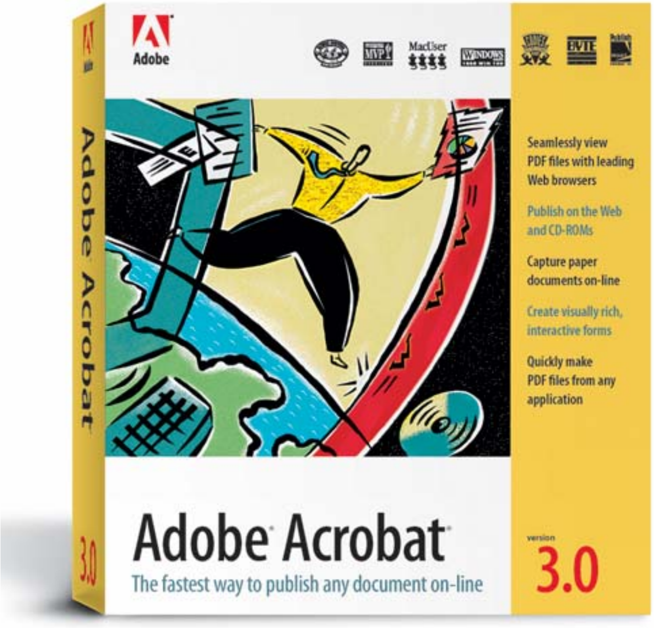
- Password Security
- Article threads
- Actions (inc. external links)
- No longer required ASCII85
- Binary bits at start of file
  - Transport issues with ASCII


Adobe Acrobat
View and print any document
Guaranteed to look like the original
Get easy access to timely information
Works with your current hardware and software
Adobe Acrobat
2.0
All the benefits of paper, without the chase

- Acrobat 2.0 – November, 1994
  - MS-DOS, UNIS, OS2 versions
  - Plug-ins for developers
  - Reader is free (still fits on a floppy disk)

PDF 1.2 - 1996

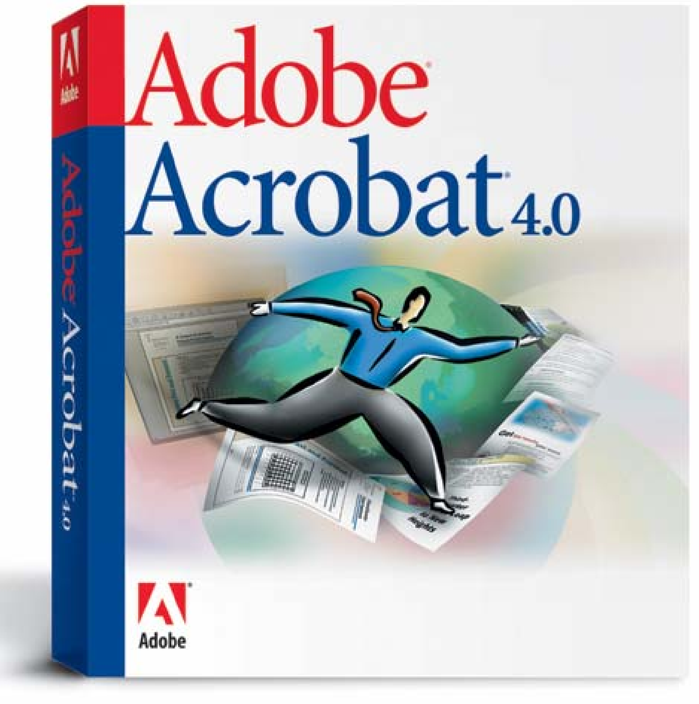
- The Acrobat “Internet” release
  - Browser window, byte serving, fast first page
- Support for CJK in March, 1997 (v. 3.5)
- Prof. publisher features (PS Level II)
  - Spot colors, halftones, OPI
  - Undercolor removal, overprint
- Simple fill-in forms
- Start of an event model
  - On open, buttons, mouse over, etc.
- Support for external movie files


Adobe
Seamlessly view PDF files with leading Web browsers
Publish on the Web and CD-ROMs
Capture paper documents on-line
Create visually rich, interactive forms
Quickly make PDF files from any application
Adobe Acrobat
The fastest way to publish any document on-line
version 3.0
Adobe Acrobat
3.0

- Acrobat 3.0 – November, 1996
  - First set of specialized features for the “Professional Publisher”

PDF 1.3 - 1999

- Collaboration
  - Sticky notes, highlights, round-trip comments
- File attachments
- PostScript III parity
- Digital signatures
- Programmable (JavaScript) forms
- Logical structure
- First eBooks published

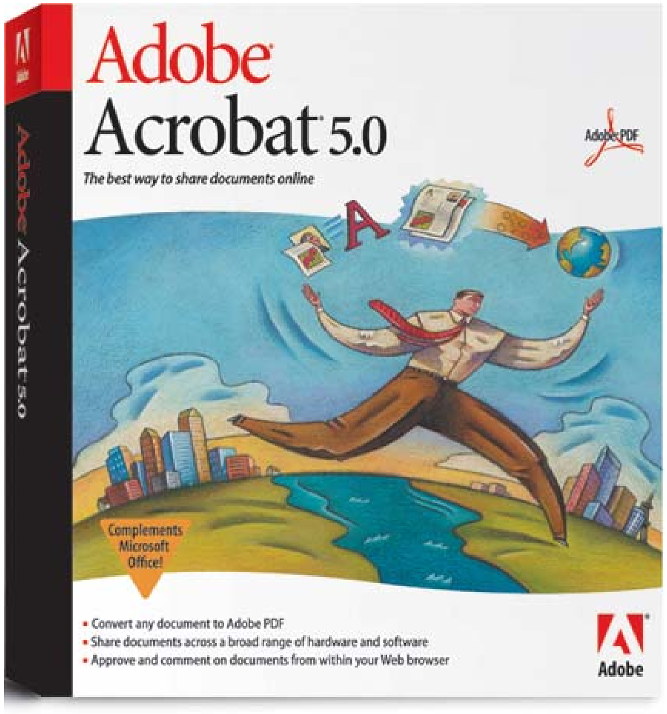


- Acrobat 4.0 – April, 1999
  - Dropped the name “Exchange”
  - Self-sign signatures

PDF 1.4 - 2001

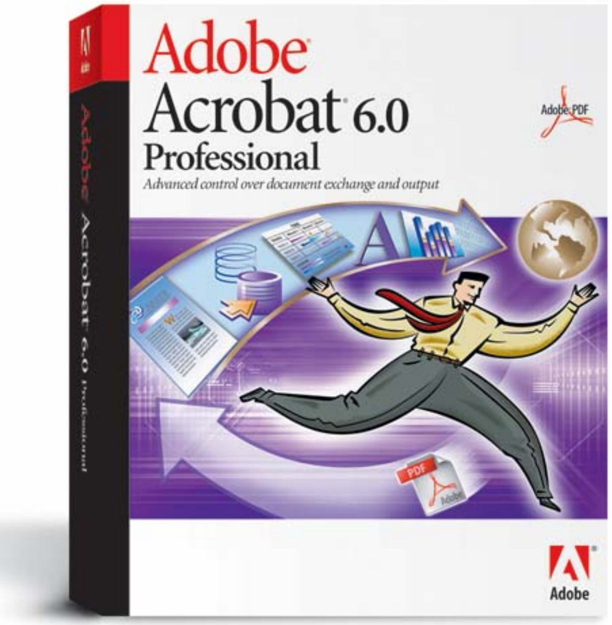
- Transparency
- Tagged PDF
- Accessibility
- JBIG2
- XMP metadata
- Reader Extensions
- Forms connected to backend databases
  - FDF
- *ISO PDF/X based on PDF 1.3 approved in 2001*

- Acrobat 5.0 – April, 2001
  - Easy PDF Creation with PDFMakers
  - Palm, PocketPC, and Nokia phone PDF Readers


Adobe Acrobat 5.0
The best way to share documents online
Adobe PDF
Complements Microsoft Office!
Convert any document to Adobe PDF
Share documents across a broad range of hardware and software
Approve and comment on documents from within your Web browser
Adobe
Adobe Acrobat 5.0

PDF 1.5 - 2003

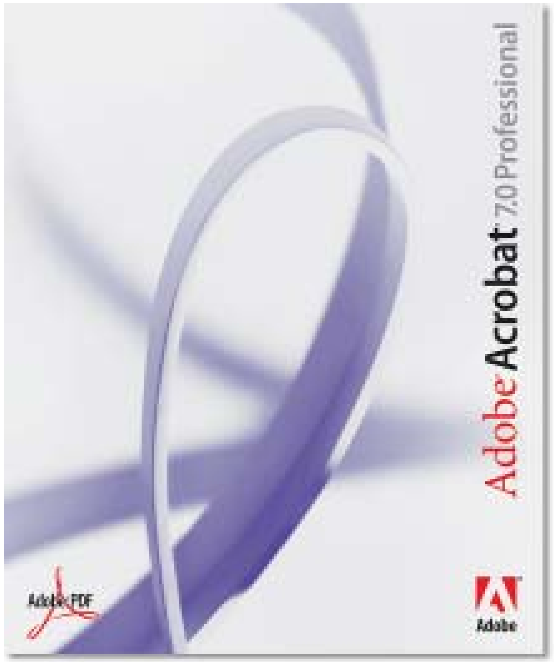
- XML (XFA) Forms
- Embedded multimedia
- Better file compression (object streams)
- Layers (optional content)
- JPEG2000
- Larger than 2 Gb files
- Certificate security (signatures & encryption)



- Acrobat 6.0 –May, 2003
  - Acrobat is split into “Standard” and “Professional”, and “Adobe Reader”
  - First set of features for AEC
  - eBook Reader

PDF 1.6 - 2004

- Embedded 3D (U3D)
- Larger than 200” pages
- OpenType font support
- Various features for AEC
  - Measure, Zoom, Object Data (CAD, Visio)
- PDF 1.6 Reference Manual
  - 1213 pages
- *ISO PDF/A based on PDF 1.4 approved in 2005*



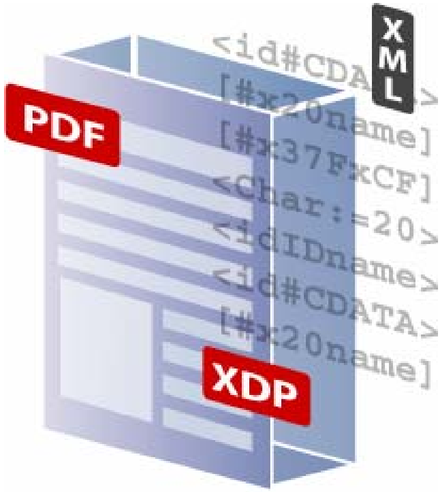
- Acrobat 7 – December, 2004
  - Launch time
  - XML Form Designer
  - Organizer
  - More AEC and Pro Publish tools
  - Acrobat 3D

Part 3
- PDF and XML

PDF & XML

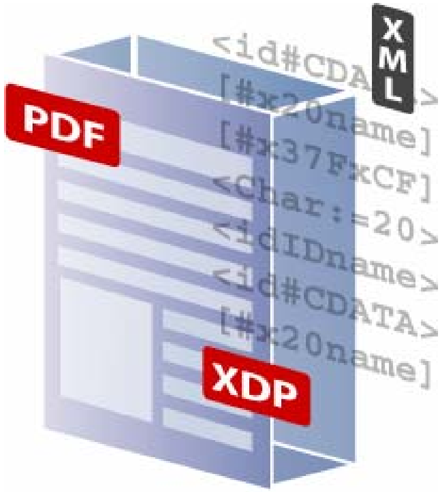
- XML Use with PDF
  - XML + XSL-FO for PDF generation
  - XMP: RDF XML
  - Export structure (tags) as XML
  - XFA & XDP – Adobe Intelligent Documents

- PDF or XML for Documents?
  - Difficult to produce the same flexibility and precision for compound documents with current generation of XML formats
  - Efficiency, file size & generation
  - Tagging, structure, metadata in PDF provide XML equivalents and export
  - Random access
  - Standards – PDF/X, PDF/A, PDF/E, PDF/UA
  - Ubiquity



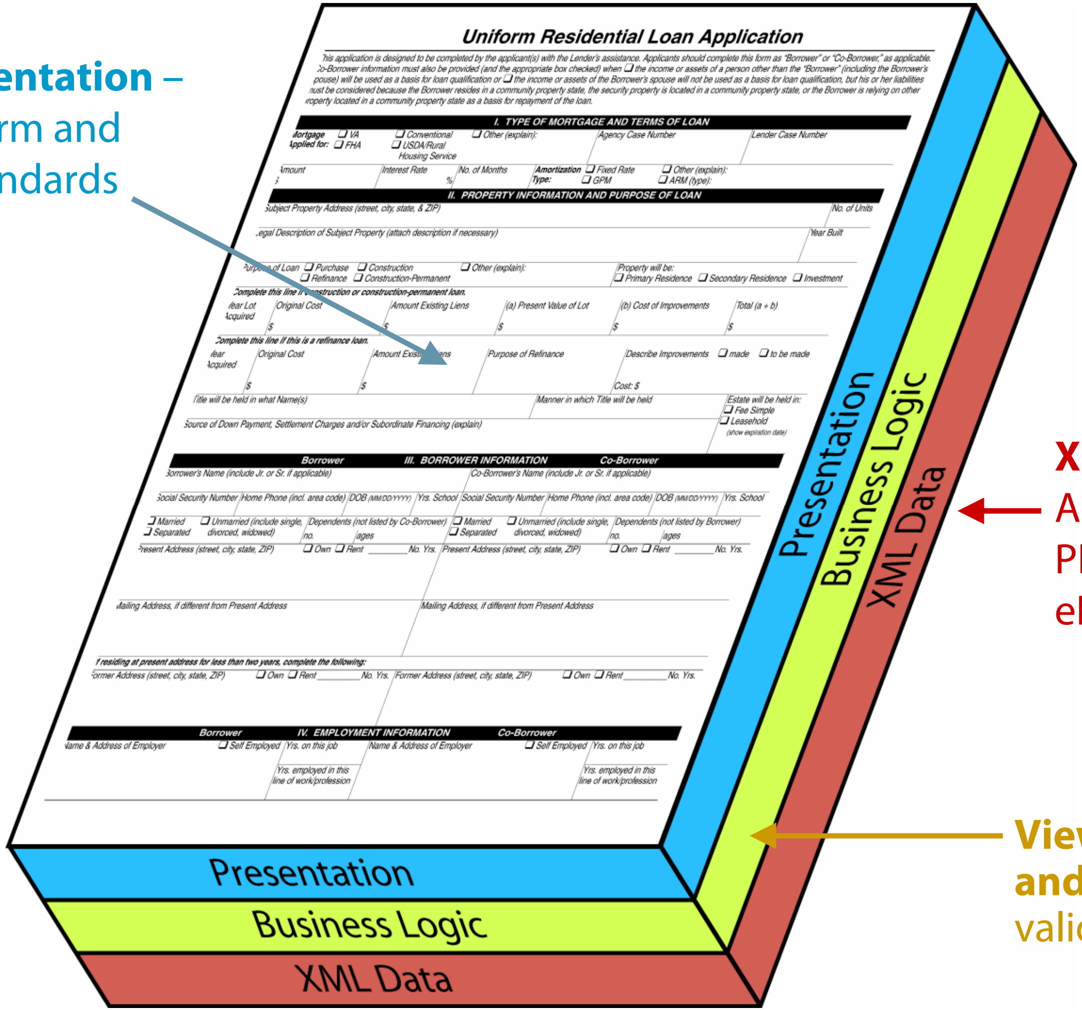
PDF & XML Part 2 – Intelligent Documents

- *Human-readable PDF + Machine-readable XML data*
- Combines PDF with XML industry standards for interoperability
- Simplifies development and deployment of standards-based applications
- PDF as a platform for industry XML standards
    - PDF can be the bridge across industry and LOB domains
- Secure container that can also leverage industry security and authenticity standards
- *Not an XML version of PDF, but uses XML as the template, data, & container*



PDF as a Platform for XML Standards

PDF Presentation – precise form and layout standards



Uniform Residential Loan Application

'his application is designed to be completed by the applicant(s) with the Lender's assistance. Applicants should complete this form as "Borrower" or "Co-Borrower," as applicable. Co-Borrower information must also be provided (and the appropriate box checked) when □ the income or assets of a person other than the "Borrower" (including the Borrower's pouse) will be used as a basis for loan qualification or □ the income or assets of the Borrower's spouse will not be used as a basis for loan qualification, but his or her liabilities nust be considered because the Borrower resides in a community property state, the security property is located in a community property state, or the Borrower is relying on other roperty located in a community property state as a basis for repayment of the loan.

I. TYPE OF MORTGAGE AND TERMS OF LOAN

Mortgage Applied for: □ VA □ FHA □ Conventional □ USDA/Rural Housing Service □ Other (explain):
Agency Case Number
Lender Case Number

Amount $
Interest Rate %
No. of Months
Amortization Type: □ Fixed Rate □ GPM □ Other (explain): □ ARM (type):

II. PROPERTY INFORMATION AND PURPOSE OF LOAN

Subject Property Address (street, city, state, & ZIP)
No. of Units

Legal Description of Subject Property (attach description if necessary)
Year Built

Purpose of Loan □ Purchase □ Refinance □ Construction □ Construction-Permanent □ Other (explain):
Property will be: □ Primary Residence □ Secondary Residence □ Investment

Complete this line if construction or construction-permanent loan.

Year Lot Acquired
Original Cost $
Amount Existing Liens $
(a) Present Value of Lot $
(b) Cost of Improvements $
Total (a + b) $

Complete this line if this is a refinance loan.

Year Acquired
Original Cost $
Amount Existing Liens $
Purpose of Refinance
Describe Improvements □ made □ to be made
Cost: $

Title will be held in what Name(s)
Manner in which Title will be held
Estate will be held in: □ Fee Simple □ Leasehold (show expiration date)

Source of Down Payment, Settlement Charges and/or Subordinate Financing (explain)

Borrower III. BORROWER INFORMATION Co-Borrower

Borrower's Name (include Jr. or Sr. if applicable)
Co-Borrower's Name (include Jr. or Sr. if applicable)

Social Security Number
Home Phone (incl. area code)
DOB (MM/DD/YYYY)
Yrs. School
Social Security Number
Home Phone (incl. area code)
DOB (MM/DD/YYYY)
Yrs. School

□ Married □ Separated □ Unmarried (include single, divorced, widowed)
Dependents (not listed by Co-Borrower) no. ages
□ Married □ Separated □ Unmarried (include single, divorced, widowed)
Dependents (not listed by Borrower) no. ages

Present Address (street, city, state, ZIP) □ Own □ Rent ________ No. Yrs.
Present Address (street, city, state, ZIP) □ Own □ Rent ________ No. Yrs.

Mailing Address, if different from Present Address
Mailing Address, if different from Present Address

If residing at present address for less than two years, complete the following:

Former Address (street, city, state, ZIP) □ Own □ Rent ________ No. Yrs.
Former Address (street, city, state, ZIP) □ Own □ Rent ________ No. Yrs.

Borrower IV. EMPLOYMENT INFORMATION Co-Borrower

Name & Address of Employer □ Self Employed
Yrs. on this job
Yrs. employed in this line of work/profession
Name & Address of Employer □ Self Employed
Yrs. on this job
Yrs. employed in this line of work/profession

XML Data Standards – ACORD, MISMO, OAGIS, PISCES, RosettaNET, UBL, eDocs, XBRL

View and Data Mapping and Validation – interactive validation and rendering

PDF as a Platform for XML Standards - Examples

- MISMO, PISCES, PRIA, REPI
  - Electronic mortgages & property standards
- ACORD
  - Insurance standards
- UBL, OAGIS, UN eDocs
  - ebXML CCTS-based b2b standards
- RosettaNet
  - Manufacturing
- XBRL
  - Financial reporting

- Paperless electronic mortgages - MISMO eMortgages
  - Diverse mortgage banking ecosystem with many different participants
  - Legally significant documents – notes, disclosures, credit reports, investors
  - Digital identities and signatures
  - MISMO XML data standards + PDF for mortgage documentation
- International trade documentation – UN eDocs
  - Cross-border trade documents – negotiable bills of lading, purchase agreements
  - Wide range of participants – from paper to full automation
  - PDF is recommended as a platform for UN eDocs XML standards

Part 4
- Demos

Thank You!

Ed Chase
Standards Engineer
Adobe Systems Worldwide Standards
chase@adobe.com