Archiving digital documents and E-Mails in PDF/A

***

Webinar
Wednesday, May 27, 2009

***

PDF Tools AG

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Introductory remarks

- The presentation will last around 45 minutes
- Afterwards there will be additional 15 minutes to answer your questions; please use the chat/question function to ask.
- We are not native English speakers, thanks in advance for your understanding ;-)

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

What is PDF/A?

- **ISO 19005 is a standard of the International Organisation for Standardization (ISO) and has been published on October 1, 2005, as …**
  - *ISO 19005-1: Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)*
- **The ISO norm defines the standard format PDF/A-1 for the long-term archiving of electronic documents. It is based on PDF version 1.4 of Adobe Systems.**
- **PDF/A (A stands for „Archiving") is a variant of PDF**
  - **It only contains elements that are suitable for long-term archiving** (no dynamic elements etc.)
  - **Elements that are necessary for a flawless reproduction are embedded into the document, such as fonts and color profiles**

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

PDF/A Competence Center – founded in 2006

The aim of the PDF/A Competence Center is to promote the exchange of information and experience in the area of long-term archiving in accordance with ISO 19005: PDF/A. This is achieved through these activities:

- **Promotion of the PDF/A Standard**
  - Classical and on-line marketing
- **Education about PDF/A**
  - Conferences, Seminars, Presentations
    Actually: 3rd International PDF/A Conference, June 16/17-18, 2009 in Berlin (www.pdfa.org)
- **Work on the ISO Standard**
  - National representatives in the ISO committee of USA, Japan, Germany, Austria and Switzerland
- **Technical Working Group**
  - Publications (TechNotes)
  - Coordination of the requests to the ISO committee
  - Testsuites for the certification of products

PDF-TOOLS.COM
Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

PDF/A Competence Center – ca. 100 members

Partner Members

callas software, compart, LuraTech, PDF-TOOLS.COM Premium PDF Technology, PDFlib, intarsys, SEALSYSTEMS THE DIGITAL PAPER FACTORY

Full Members

Adobe, ABBYY SOFTWARE HOUSE, NUANCE, KOFAX A RICOH GROUP COMPANY, RICOH, Docu Com Document Company, RSD, Open Limit, ACTINO software, atori, AuthentiDate, BARR, BRIDGE TECHNOLOGIES, coextant MAKING CONTENT FLOW, crawford Technologies, PETEC, docutec document understanding technologies, eDocuMAN, FileAffairs, forcont business technology gmbh, GMC SOFTWARE TECHNOLOGY, b.u.s micr filmdienst, Janich & Klass Computertechnik, OpenType corporation in Japan, OS OPTIMAL SYSTEMS, PAGETECH, PENTADOC, Samarcande, xenos, www.zoeller.de

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF Tools AG

- Founded as an independent spin-off company in 2002, in PDF market since 1993
- Server-based developer tools for creating, processing, converting, rendering and enhancing PDF and PDF/A documents
- International: Customers in over 60 countries, branch in Canada
- Swiss delegate in the ISO Working Group 171 (PDF/A, PDF 1.7) with voting rights
- Largest range of PDF/A compliant products worldwide

PDF-TOOLS.COM
Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Your hosts

- **Dr. Hans Bärfuss, Chief Executive Officer, PDF Tools AG**
  - Works on PDF technology since 1993
  - Active member of the ISO committee for PDF/A
  - Founder/vice president PDF/A Competence Center
- **Dr. Hans-Rudolf Aschmann, Chief Technology Officer of PDF Tools AG**
  - Also works for more than 15 years in the PDF world
  - Specialist for PDF/A from digital sources
  - Software architect of the Document Converter Service
- **Carlo Nessi, Head of Marketing of PDF Tools AG**
  - IT marketing since 1989 (3M, Canon, Swisscom)

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

Overview

You will learn …

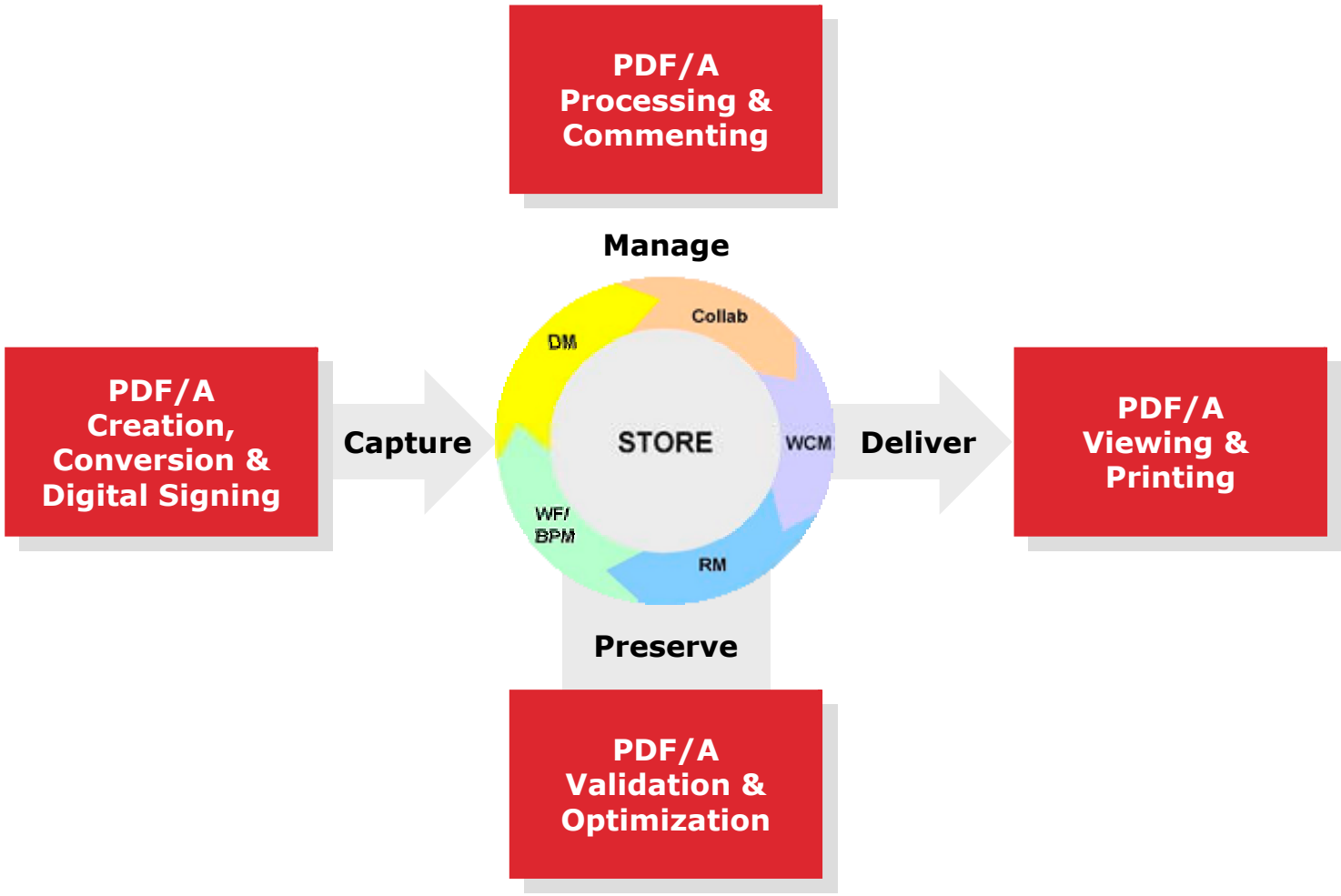
- How digital documents develop as archive material
- Which properties analog and digital source have
- Why it is worthwhile to convert digital sources to PDF/A for archiving
- How digital sources are converted to PDF/A (processes, challenges, special sources, font handling, digital signatures etc.)

Copyright © 2008 PDF/A Competence Center

PDF/A within the AIIM model for ECM

Copyright © 2008 PDF/A Competence Center

PDF/A within the AIIM model for ECM



Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Sources of digital documents

- Inbox
  - Scans with or without OCR (optical character recognition)
  - E-mails with or without attachments
- Office, graphics and construction
  - MS Word, Excel, Powerpoint, Visio, etc.
  - Illustrator, Indesign, Photoshop, etc.
  - CAD: Autocad, 3D Studio Max, etc.
- Elektronic data interchange
  - SWIFT, EDIFACT, etc.
- Outbox
  - Print data streams: PostScript, PCL, AFP, etc.
- Archive migrations
  - Masses of TIFF and other files, including source data (metadata, object relationships, etc.)

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Attributes of analog and digital sources

Attribute	Analog	Digital
Sources	Scanner, raster images	Standard and proprietary formats from applications and data streams, in file storage, mailboxes and attachments
Quality of the source	Good	Large differences
Complexity of the source	Low	Can be very high
Product differentiation	Compression rate, performance	Quality
Biggest challenge	OCR recognition rate	Loss of information during the conversion

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Testing of print pathes (1)

- The following samples are extracts from PDF/A compliant files
- The results show, that the conversion with low quality tools can be problematic

PDF-TOOLS.COM Premium PDF Technology

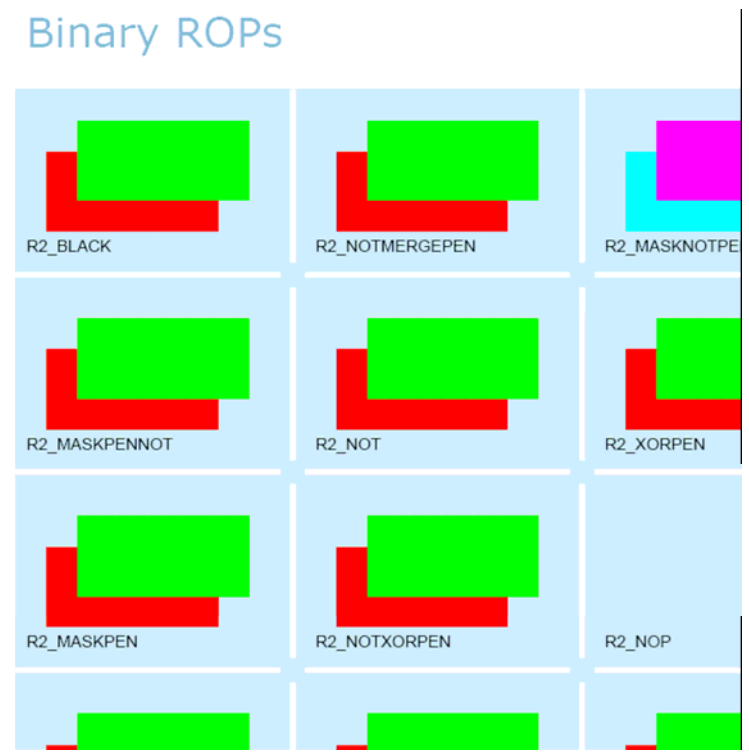
Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Testing of print pathes (2)

Original

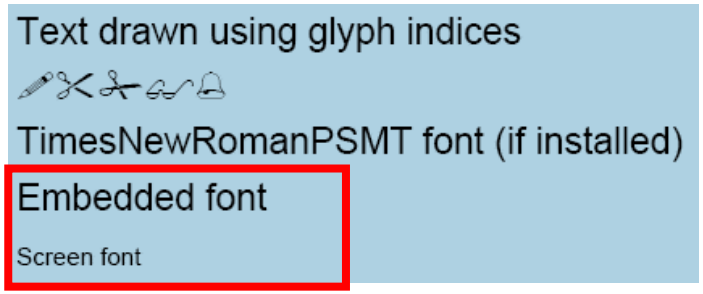


Incorrect Conversion

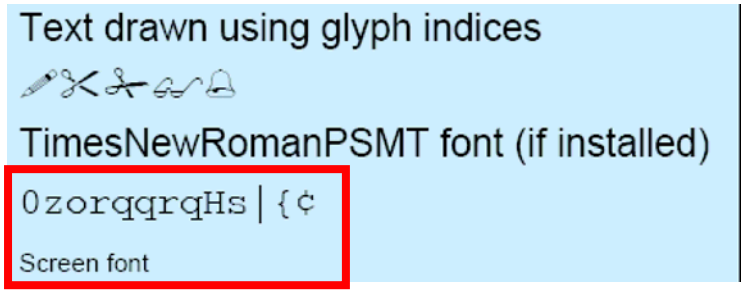
PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

Testing of print pathes (3): Fonts



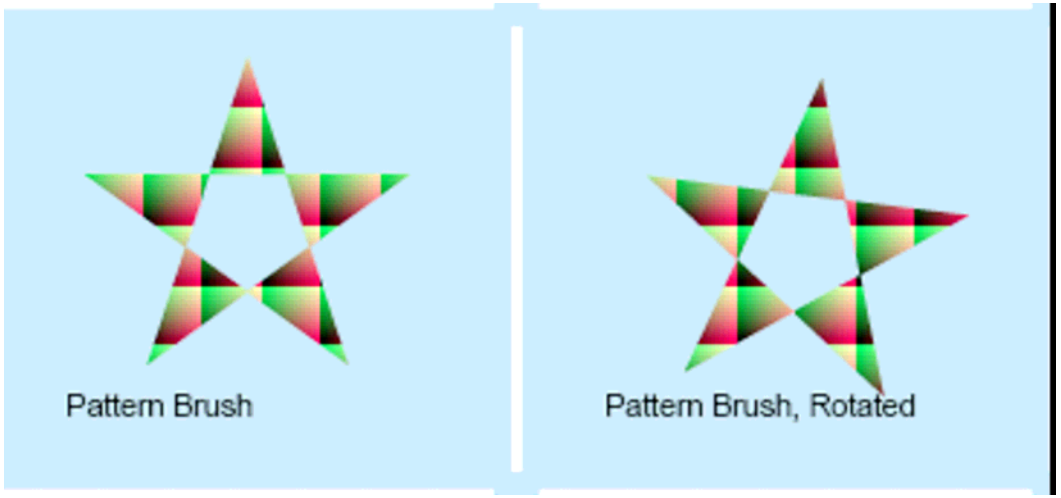
Original



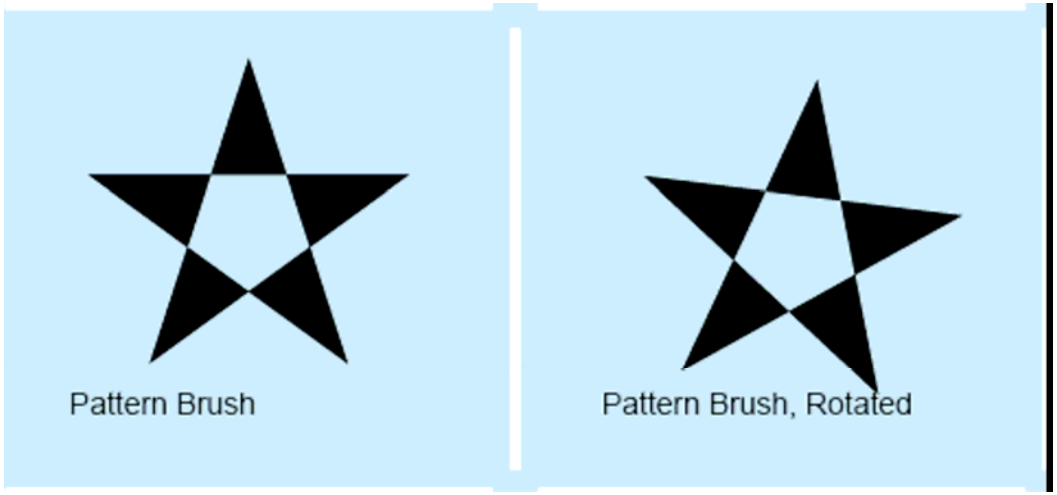
Incorrect Conversion

Copyright © 2008 PDF/A Competence Center

Testing of print pathes (4)



Original



Incorrect Conversion

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

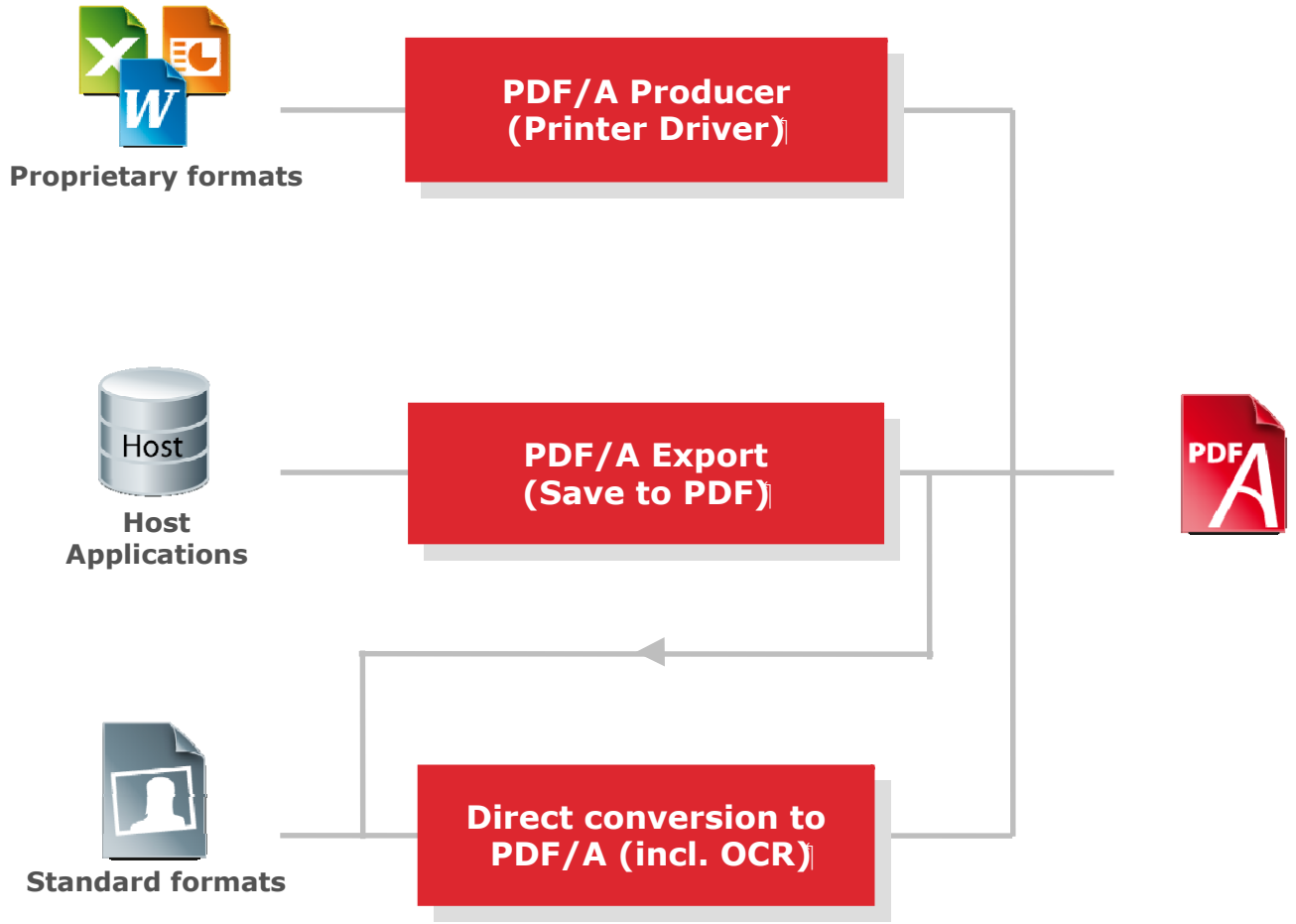
Why convert to PDF/A?

- The user does not have to maintain the original "native" applications and the platforms on which the applications operate, to view the documents
- Users depend less on software manufacturers because all of the relevant information is saved in one ISO-standardized format and this format is manufacturer-independent (PDF/A)
- Simplified processing due to the fact that the archived data is standardized into one format.
- Option to perform a full-text search in all of the stored data.
- These advantages involve an economic benefit that must not be underestimated.
- **Disadvantes:** loss of interactivity or the built-in "functionality" of the native format. **Solution:** Archiving as PDF/A and in the native format.

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

Conversion to PDF/A



Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Challenges of the conversion of digital documents to PDF/A

- **Colors:**
  - If the colour profiles from the sources are missing, assumptions have to be made about the color space
- **Fonts:**
  - If fonts (or glyphs) are missing, replacement fonts must be selected. To do this, the text must be a Unicode text
- **Transparency:**
  - The flattening of transparency is complex and may lead to the loss of information (fonts, vectors, etc.)
- **Levels, interactive and multimedia elements:**
  - Only the "Print Preview" is retained
- **Actions**
  - Functionality (JavaScripts etc.) is lost
- **Digitale Signatures**
  - Must be checked, documented and signed again

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Conversion of E-Mails to PDF/A

- E-Mails are digital-born documents
- The attachments of E-Mails can contain many different formats
  - Standard formats
  - Proprietary formats
  - Containers, which can also be nested
- E-Mails can be stored in different places:
  - Mailboxes of E-Mail servers
  - File system
- E-Mails contain different types of information:
  - Display as Text, HTML or RTF
  - Also contain header information
- Conversion of E-Mails to PDF/A
  - Body and attachments are converted separately
  - Merge to one single document
  - Handling of digital signatures

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Conversion of Websites to PDF/A

- Objective of the archiving of websites: To retain the contents of the (own) website in a way that is legally trustworthy, to be used as an evidence in legal procedures
- It is not useful to just „print" the website to PDF/A, as the layout is often changed in the printing function of a website; but it's important to keep the layout as it appears on the screen
- Solution:
  - Decide on one browser and browser-version
  - Define rules for archive-friendly webpage design
  - Decide which representation should be used (screen view or print view)
  - Capturing of the website as an image
  - Storage of other information such as texts, images, fonts, background, colors, flash previews etc.
  - Merge of the contents together with the link information to reproduce the website structure within the PDF document

PDF-TOOLS.COM Premium PDF Technology

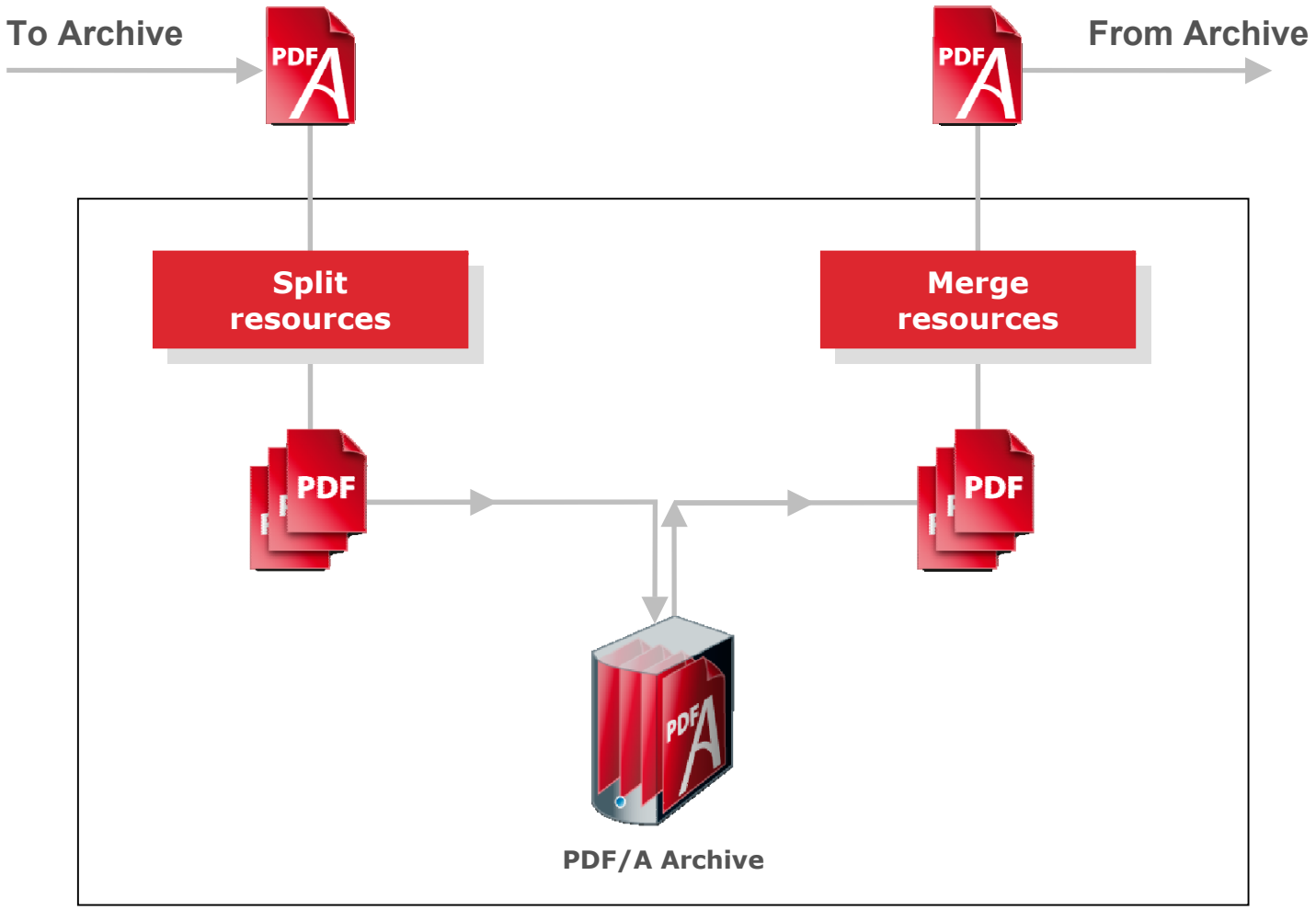
Copyright © 2008 PDF/A Competence Center

Conversion software: on client or server?

Attribute	Client	Server
Scaling workstations	Small amount	Large amount
Distribution	Complex	Simple
Robustness for the users	Depends on the creator-applications	Independent
Performance for the users	Restricted by the client	Scalable
Supported source formats	Restricted by the installation	Scalable
Application support	Local	Central

Copyright © 2008 PDF/A Competence Center

Font handling in mass archiving



Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Legal security with digital signatures

- A PDF/A compliant digital signature can be added to a PDF/A file
- Objective is the best possible legal security
- What can a digital signature really provide:
  - When (time) the digital signature has been applied
  - If the document has been manipulated since and if yes, what has been changed
  - Who/which process within a company has made the conversion
- A signature alone cannot guarantee:
  - Correctness of the content (analog to the source)
  - Proof of 100% visual similarity with the original
- **Possible solution:** Certification of the processes

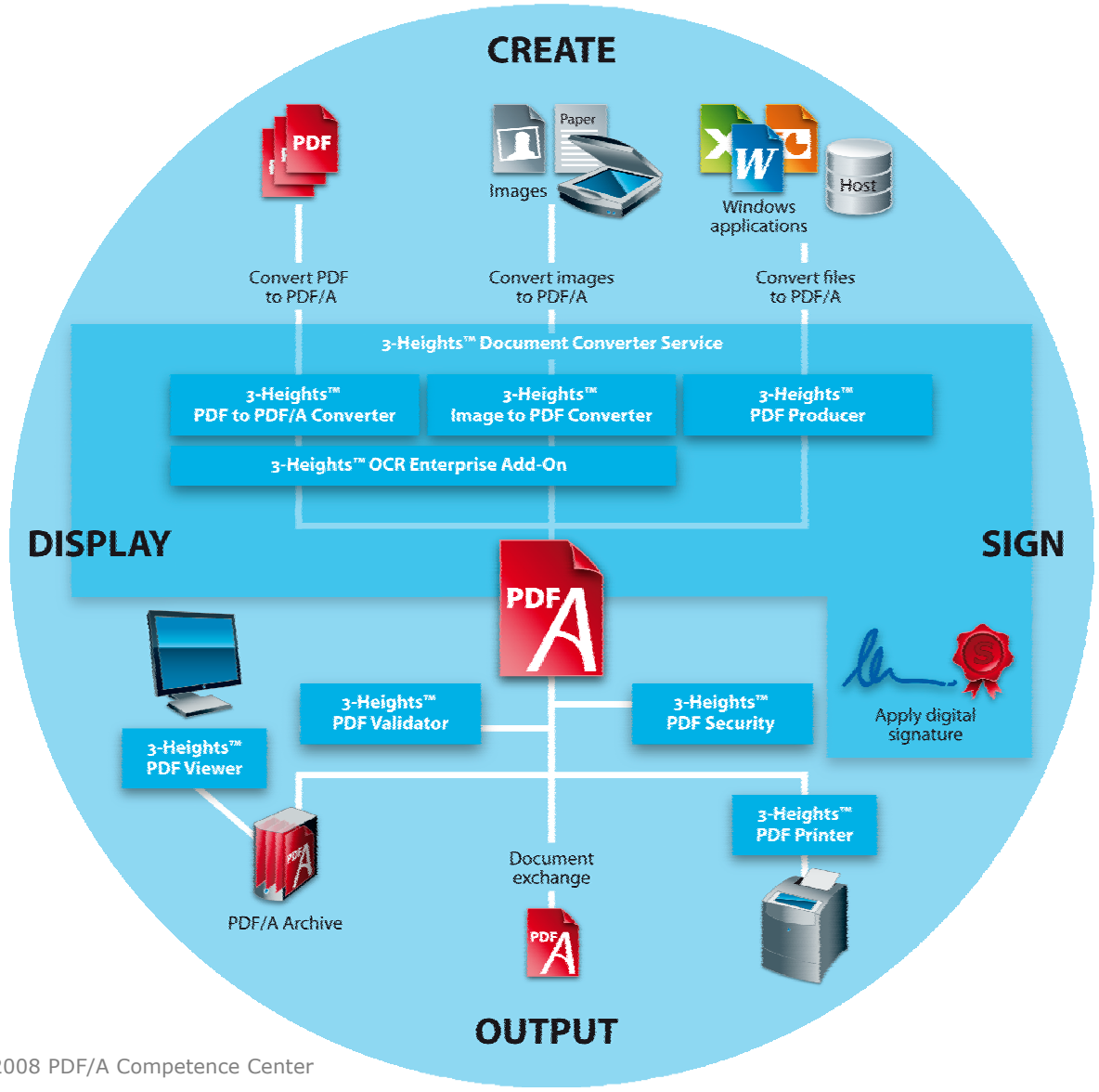
PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

PDF/A products of PDF Tools AG



PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

3-Heights™ Document Converter Service

- Converts images, Office documents, E-Mails incl. attachments, websites and existing PDF documents automatically to PDF/A
- Extensible service, for example for additional conversion functionalities (with plugins)
- Output formats are TIFF, PDF and PDF/A, incl. application of a digital signature
- Optional OCR Add-On
- Decentral use via many different interfaces: Windows Service with watched folders, Command Line, API, Explorer Plugin or direct in the mailbox (IMAP)
- This product is suitable for any volume and company size thanks to its scalability

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Thanks for attending this webinar!

Questions?

- … can now be asked using the chat/question function
- … or send us an e-mail to: pdfsales@pdf-tools.com
- … or call us on: Tel. +41 43 411 44 50

PDF Tools AG

www.pdf-tools.com

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A Competence Center

www.pdfa.org

Backup slides

- PDF/A - Features
- PDF/A - Advantages

PDF-TOOLS.COM Premium PDF Technology

Copyright © 2008 PDF/A Competence Center

PDF/A - Features

PDF/A: An ISO Standard

- ISO 19005 is an ISO (International Standards Organisation) Standard that was published on October 1, 2005:

  ISO 19005-1: Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)

- Defines a format (PDF/A) for the long term archiving of electronic documents and is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (implemented in Adobe Acrobat 5)

Two Levels of Compliance

- There are two levels of compliance for PDF/A:
  - PDF/A-1a: Level A compliance in Part 1
  - PDF/A-1b: Level B compliance in Part 1

- PDF/A-1a represents full compliance with all requirements of the ISO standard and guarantees both accessibility (e.g. full text search and support for devices for the disabled) and reproducibility

- PDF/A-1b is a slightly reduced set of requirements and the guarantee is limited to reproducibility

Copyright © 2008 PDF/A Competence Center

PDF/A - Advantages

Advantages

- Improved accessibility alone may substantiate the implementation of an electronic archive. Some advantages of a PDF/A archive over a TIFF or paper archive are:

Full-Text Search

- PDF/A stores text as objects, allowing for an efficient full-text search in an entire archive. TIFF must first be scanned.

File Size

- PDF/A files require only a fraction of the memory space of original or TIFF files, without loss of quality.

Optimization

- PDF/A format can be optimized. The optimization can be focused on images (e.g. scanned checks) or extracting structured data (e.g. voucher information).

Metadata

- Metadata like title, author, creation date, modification date, subject, keywords, etc. can be stored in a PDF/A file.

Copyright © 2008 PDF/A Competence Center