
XML, PDF and eForms

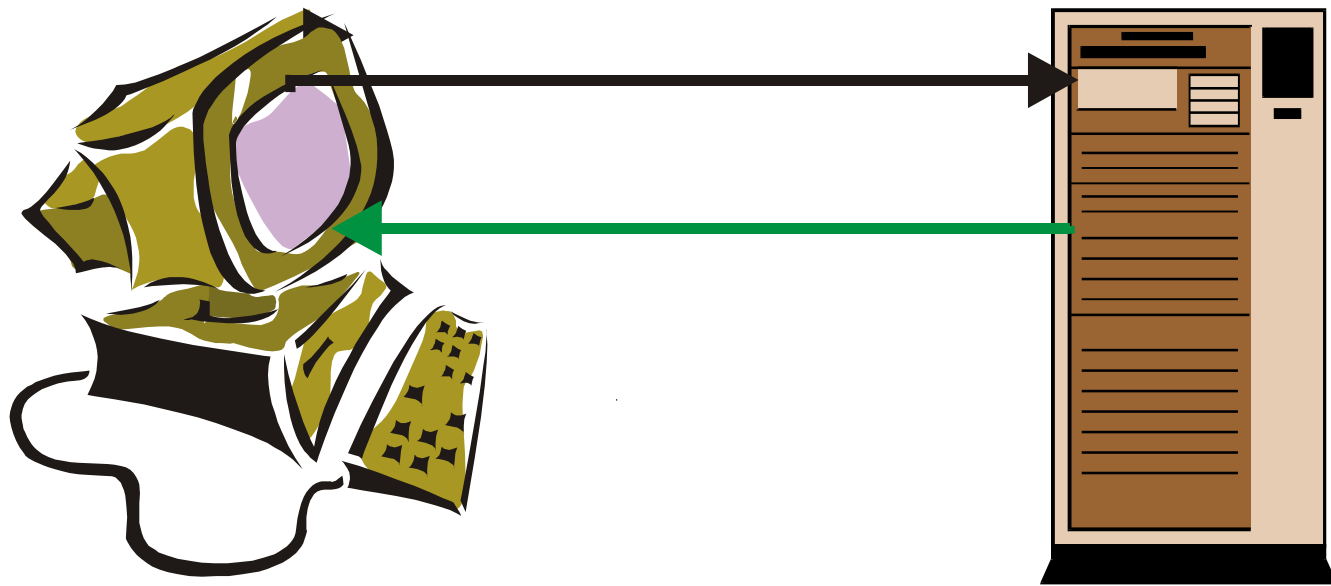
Subtitle, if any

Leonard Rosenthol
Chief Technology Officer
PDF Sages, Inc.

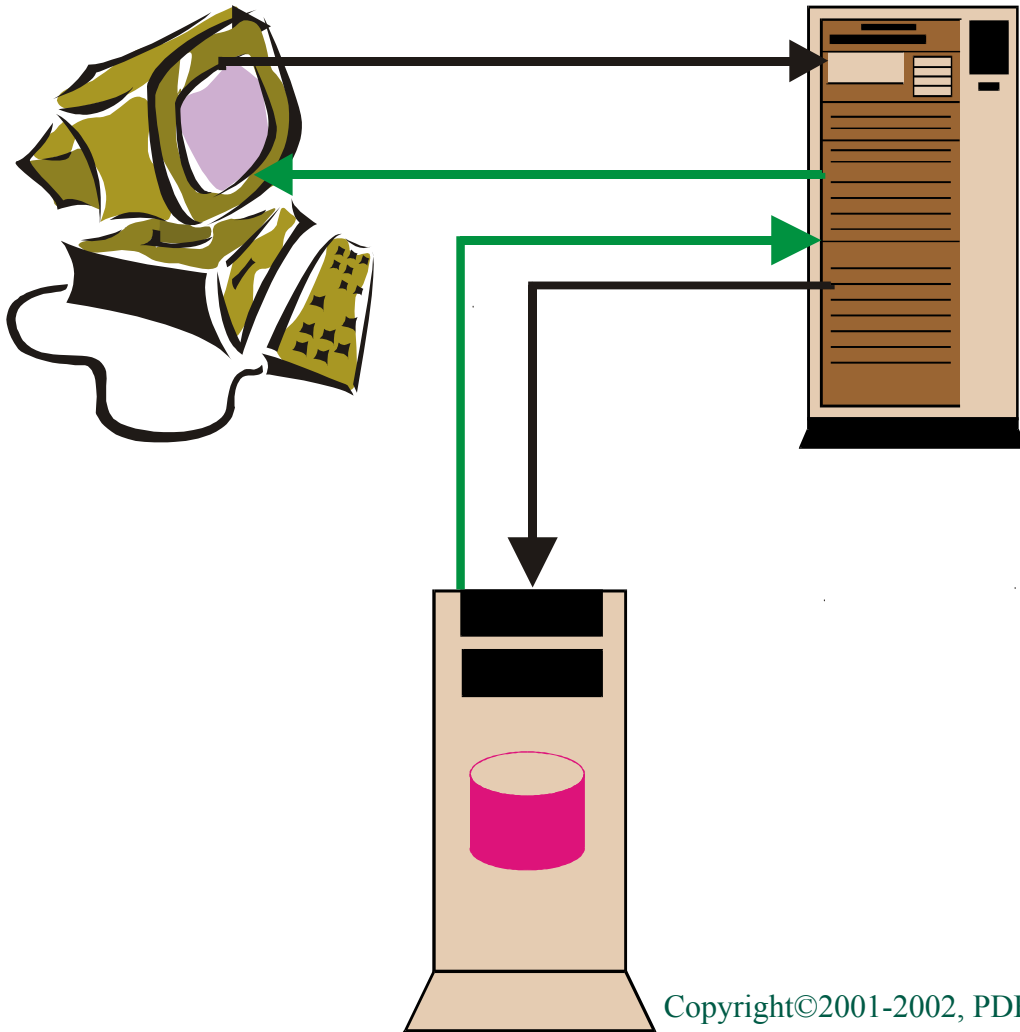
Overview

- Simple Data Exchange
- Server-Side Processing
 - Form Merging
 - Stamping/Overlaying
 - Dynamic Production

Simple Data Exchange aka. Client-Side Merging



Something missing?



A bit more in depth...

- Acrobat submits HTML, FDF, or XFDF via HTTP
 - Optionally you could use an HTML form & submit in std. HTML Format
- CGI/ASP on server receives the data
- It then processes it (potentially interacting with a database)
- It MAY then return a new set of data in FDF or XFDF format for display in the same (or a different) PDF document

What do I need to make it work?

- Web Browser
- Acrobat 3.5 or later
 - Acrobat is the only PDF viewer with forms support, which were introduced in 3.5
- Web server w/CGI or ASP support
 - Pick your favorite OS and server software along with integration language (Perl, VB, Python, PHP, etc.)
- Optionally you may have a database or other server-based tools to manipulate/massage the data

HTML vs. FDF vs . XFDF

- HTML form posting is a standard and supported by all CGI and ASP tools. It does not, however, provide support for all Acrobat features.
- FDF is the “old standard” for transferring PDF form data. Requires the use of the Adobe FDFToolkit (or similar tool) to parse the files (though generating them is easy enough with any good text processor)
- XFDF is an XML-based version of FDF. This means you get the benefits of FDF, with the ability to use your favorite XML tools.

Pros and Cons of Simple Exchange

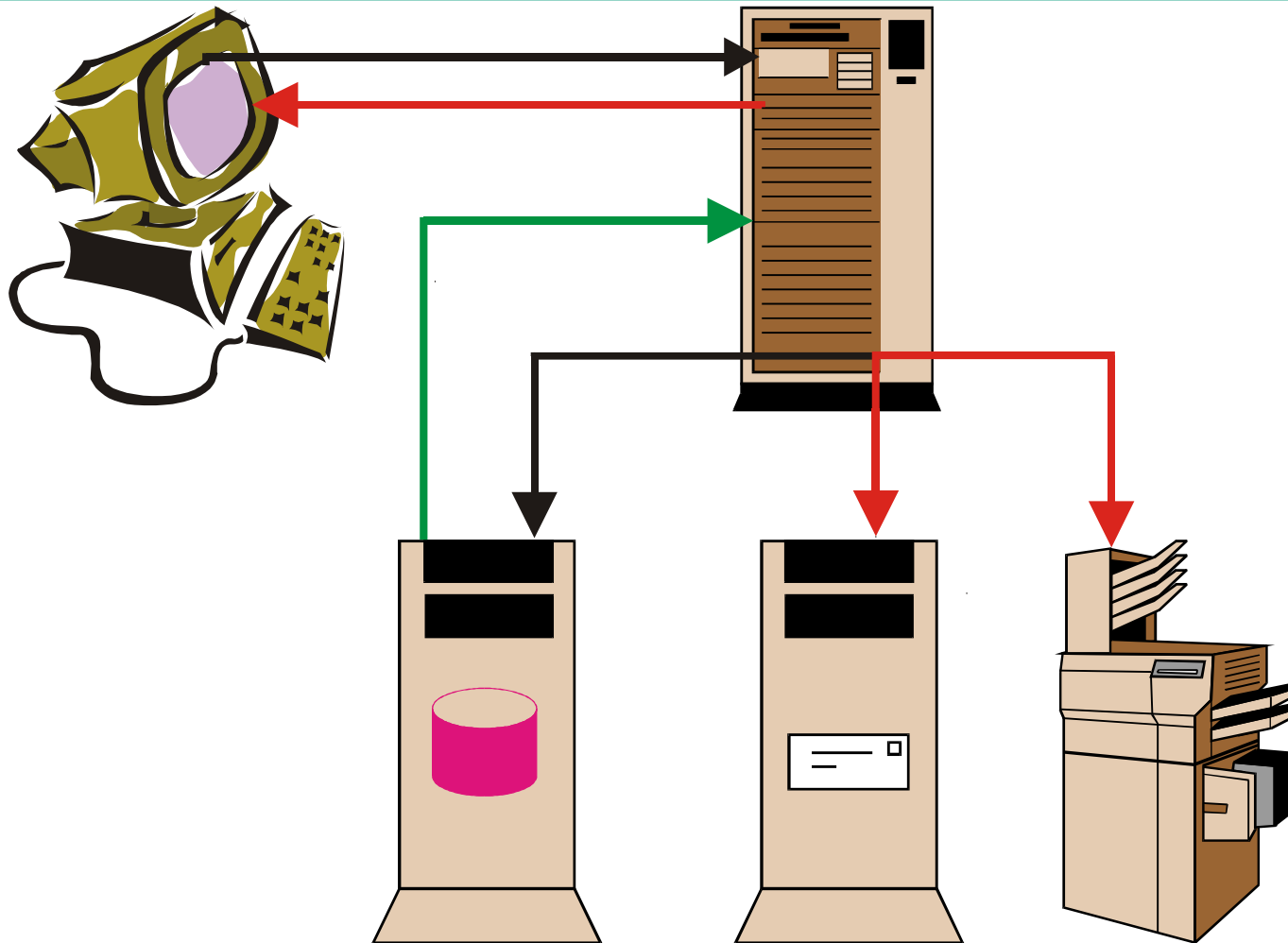
- Pros

- Doesn't require any 3rd party tools
 - Cheap
 - No need for admin access on server

- Cons

- Requires Acrobat
 - Mac OS & Windows
 - No Unix, Palm, PocketPC, etc.
- Reader uses can NOT save the filled in form
- Browser incompatibility issues
- Can't return data into separate window or frame

Server-Side Processing



Pros and Cons of Server-Side

- Pros

- Does NOT require Acrobat
 - So you can support Unix, Palm, PocketPC, etc.
- Reader uses CAN save the filled in form
- NO browser incompatibility issues
- CAN return data into separate window or frame

- Cons

- Requires 3rd party tools
 - \$\$ is involved
 - How much is determined by your specific needs
 - » OS platform, integration language, database support, etc.
 - NEED admin access on server to install/config

Server-Side Options

- Form Merging
- Stamping/Overlaying
- Dynamic Production

Form Merging

- Start with a standard PDF form
- Use a 3rd party tool to fill in the form fields
 - From FDF
 - From XML/XFDF
 - Direct from a Database
- Optionally “flatten” the fields
- Return the filled-in form to the user

Available Tools

- FDFMerge – Appligent
- PDFToolkit – ActivePDF

Pros & Cons of Form Merging

- Pros
 - Easy to setup and understand
 - No need to be a programmer
 - Some tools can automate the entire process
 - Not just text, but also checkboxes, buttons, etc.
- Cons
 - Limited formatting options for data
 - For non-flattened forms that will be compatible with Acrobat 4, you are limited to the Base14 fonts
 - Limited layout options
 - No reflow, extra white space, borders, etc.
 - Limited non-Roman text support

Stamping/Overlaying

- Start with any PDF – no form fields required!
- Use 3rd party tools to “stamp” data into PDF
 - May require creating new document and copying original first, then overlaying.
 - NOTE: these products will remove all interactive elements (bookmarks, hyperlinks, etc.) from the source PDF’s
 - Some of these tools use XML-based description languages to define where to place the data.
- Return new PDF to the user

Available Tools

- StampBatch – Appligent
- PDFToolkit – ActivePDF
- PDFlib – PDFlib, GMBH
- PDFever – Perl Studio
- ReportLab – ReportLab
- iText – Lowagie
- PDFLibrary – Big Faceless

Pros & Cons of Stamping

- Pros
 - Greater flexibility of formatting & layout options
 - May support non-Roman text
- Cons
 - Requires manual determination (trial & error?) of where elements are placed on the page.
 - Little to no support for non-textual elements
 - May require some programming

Dynamic Production

- Maintain layout and “template” content in a database or other content repository
 - eg. XML
- Combine the “template” with the data
 - eg. XSL-T
- Process the new content in conjunction with styling information
 - eg. XSL:FO
- Return the new PDF to the user

Available Tools

- PDFlib – PDFlib, GMBH
- PDFever – Perl Studio
- ReportLab – ReportLab
- iText – Lowagie
- Report Generator – Big Faceless
- FOP – Apache Project
- XEP – RenderX

Pros & Cons of Dynamic Production

- Pros
 - Greatest flexibility of formatting & layout
 - Pagination and document reflow
 - May include tabular formatting, charts, etc.
 - May support non-Roman text
- Cons
 - LOTS more setup/configuration work
 - Will require programming

Wrap Up

- There are lots of forms out there that deserve to be electronically processes
 - I know I WISH I could submit more this way!
- It's not as cut and dried issue – there are LOTS of options out there, and many things to consider.
- Make your decision based on your needs, then your desires, and NOT on your pocketbook.
 - Your pocketbook may point you in the wrong directions ☹

Q & A

