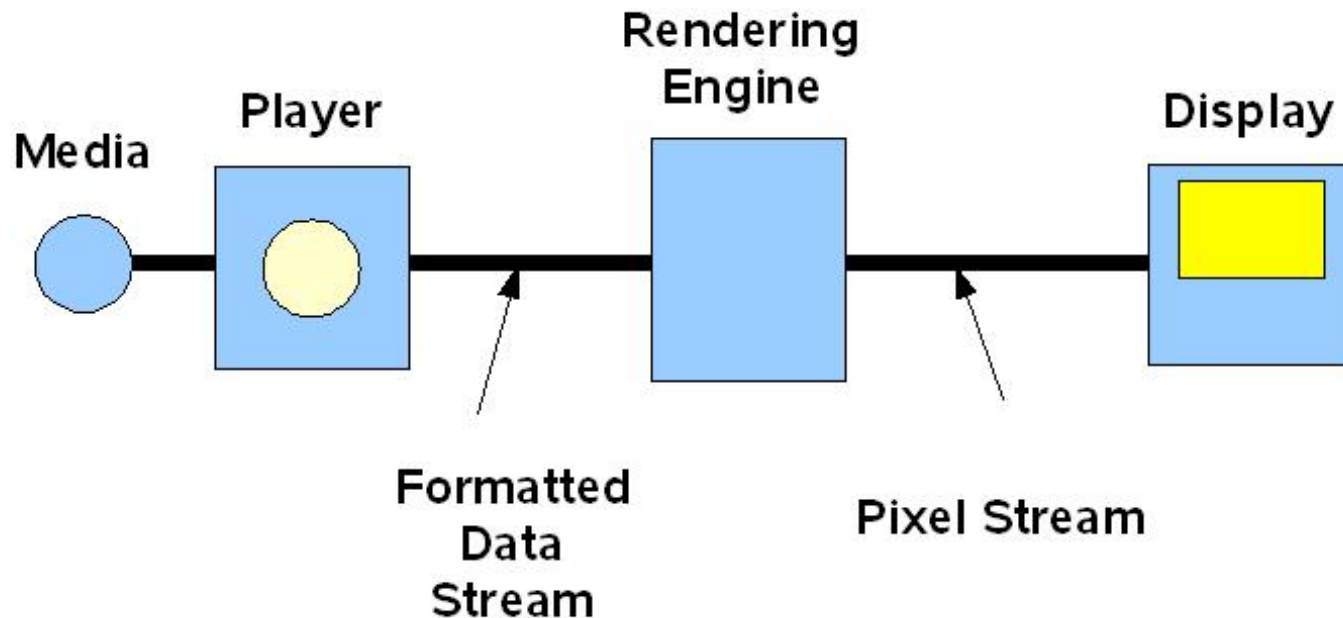


Collections

John R. Brinkema
Administrative Office of U.S. Courts
brinkema@teo.uscourts.gov
(202) 502-2747

The Archiving Problem

- You are not done yet ...
- Media, Player, Formatted Data Stream, Rendering Engine, Pixel Stream, Display



Why This Presentation

- Go beyond my parochial view
- To get you thinking
- Get requirements
 - I have been talking with Adobe about collections for several years

Collections

- Two or more bundles of information that are treated as one bundle
- I don't limit my bundles to just PDF, which is why I use the term 'bundle'.
- More on this later ...

Thinking About Paper Collections

- Scribblings and stamps on a page
- Documents and books
- Routing slips, notes, and yellow stickies
- Folders of documents, piles of books
- Cabinets of folders, shelves of books
- Room of cabinets, a library

I want a collection standard that can handle electronic versions of all these bundles

Basic Assumptions

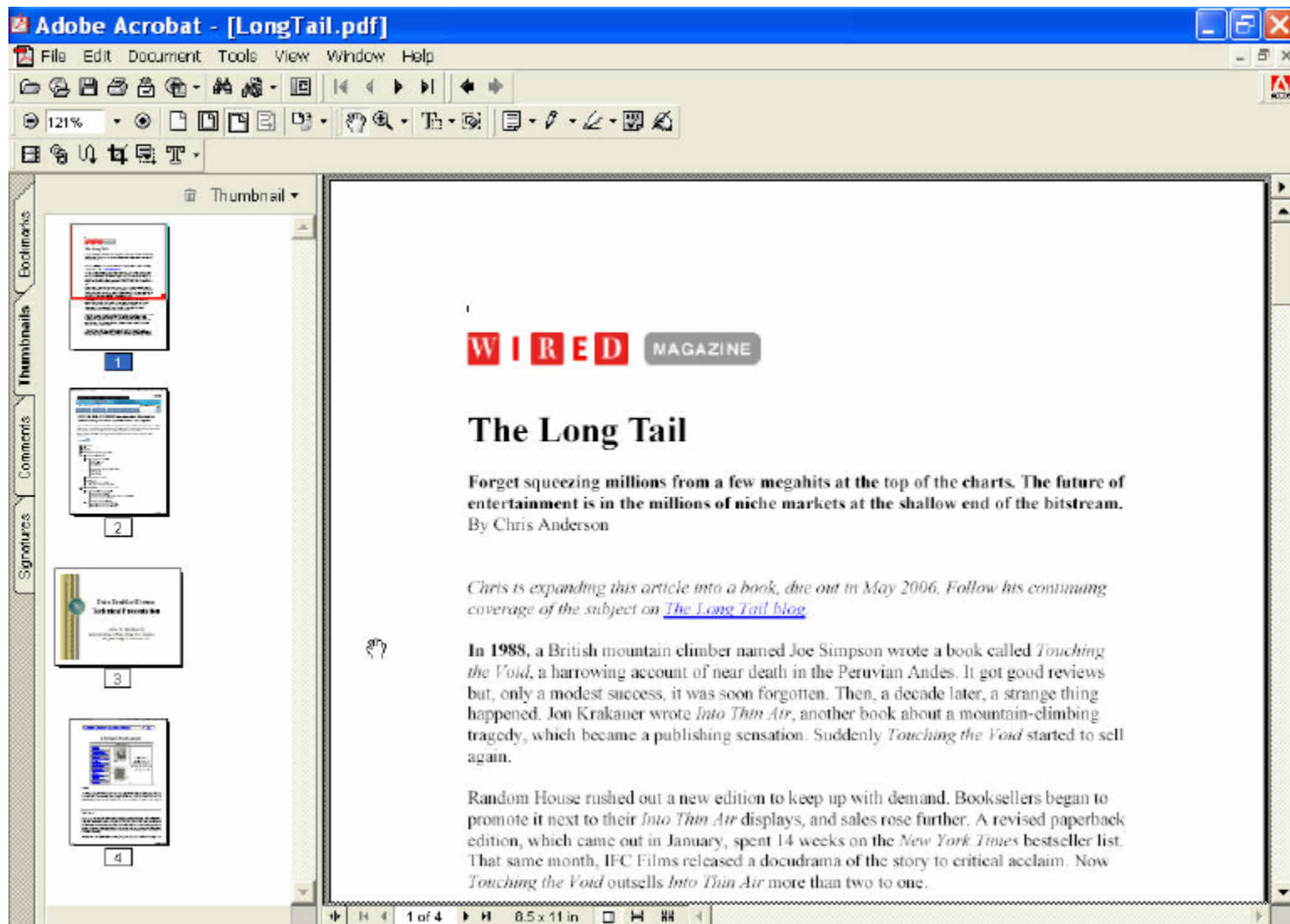
- Document Integrity
 - Collecting documents does not change the original document
- Documents can be added to the collection and removed from the collection without affecting other documents
- ‘Thumbnails’ of a collection are *frontispieces (thinking PDF reader)*

Frontispiece

From the OED:

- The first page of a book or pamphlet, or what is printed on it; the title-page including illustrations and table of contents; hence, an introduction or preface. Obs.
- *An illustration facing the title-page of a book or division of a book. (The current sense.)*

What a collection of PDFs might look like



More Basic Assumptions

- Collections are treated as a single document
- Collections of Collections are allowed
 - Collections have TOC as their frontispiece
- I expect the web (or its successors) is here to stay
- Most collections will not be physically on the readers computer

Little and Big Collections

- Collections in one file
- Collections in several files
- Collections distributed in several locations

Distributed Collections

- Extract only that part of a collection that is needed
- Document proxy
 - Proxy frontispiece
 - Stands in for the actual document
 - Describes where relevant parts of document / collection are located

More than just format

- Intelligent data slicing protocol
- Frontispiece for whole document
- Displayable view of portion of document

Links

- Hyperlinks are here to stay
- URL's and URI's
 - Locations and names
 - Things in the net move
- Handles
 - Point to service that tells the renderer where the document is currently located

Very Simple Collections

- Document + Meta data
 - Stamps, Routing Slips, Annotations
 - ‘Naturally’ XML
 - Require a renderable representation of every component in a collection
 - Like ‘appearance’ data in signatures
- Documents + factored-out common objects
 - Like font information

Overlays

- PDF *highlight* format
- Combine two or more elements in a collections and display them together
 - Stamps
 - Signatures
 - Annotations

Embedded Files

- In PDF: evil
- In collections: necessary evil
- Non-PDF (PDF/A) components
 - Originals of translations
 - Forensics
- Not necessarily PDF/anything
- Associated data
 - ‘Registered Standard’