

Internet Technologies for Digital Libraries

Larry Masinter
September 1998

THE DOCUMENT COMPANY
XEROX

For copies of slides...

- <http://www.parc.xerox.com/masinter>

- Send email to:

lm@acm.org

with your name, address, and telephone number,
and I will send you a copy.

Purpose of Tutorial

- Survey of Internet Technologies
 - important for digital libraries
- Introduce terms and set in context
- Standards and the Standards process

Outline of Tutorial

Part 1: Digital Libraries, Document Management, and the Web

- Overview, Motivation

Part 2: Internet Standards

- Organizations and processes

Part 3: Survey of Internet technology

- Document formats
- Naming and Metadata
- Network Protocols

About the speaker

Principal Scientist, Xerox Palo Alto Research Center

- Previous areas of work include programming languages and environments, document and image management systems
- Current project: Electronic commerce for document services
- Internet application standards:
 - World Wide Web Consortium
 - Advisory Board
 - Internet Engineering Task Force Applications Area
 - Chair, URI; Chair, HTTP
 - Editor or author of 18 Internet specifications
 - Web, Printing, Fax, document communication

Part I: Digital Libraries, Document Management, and the Web

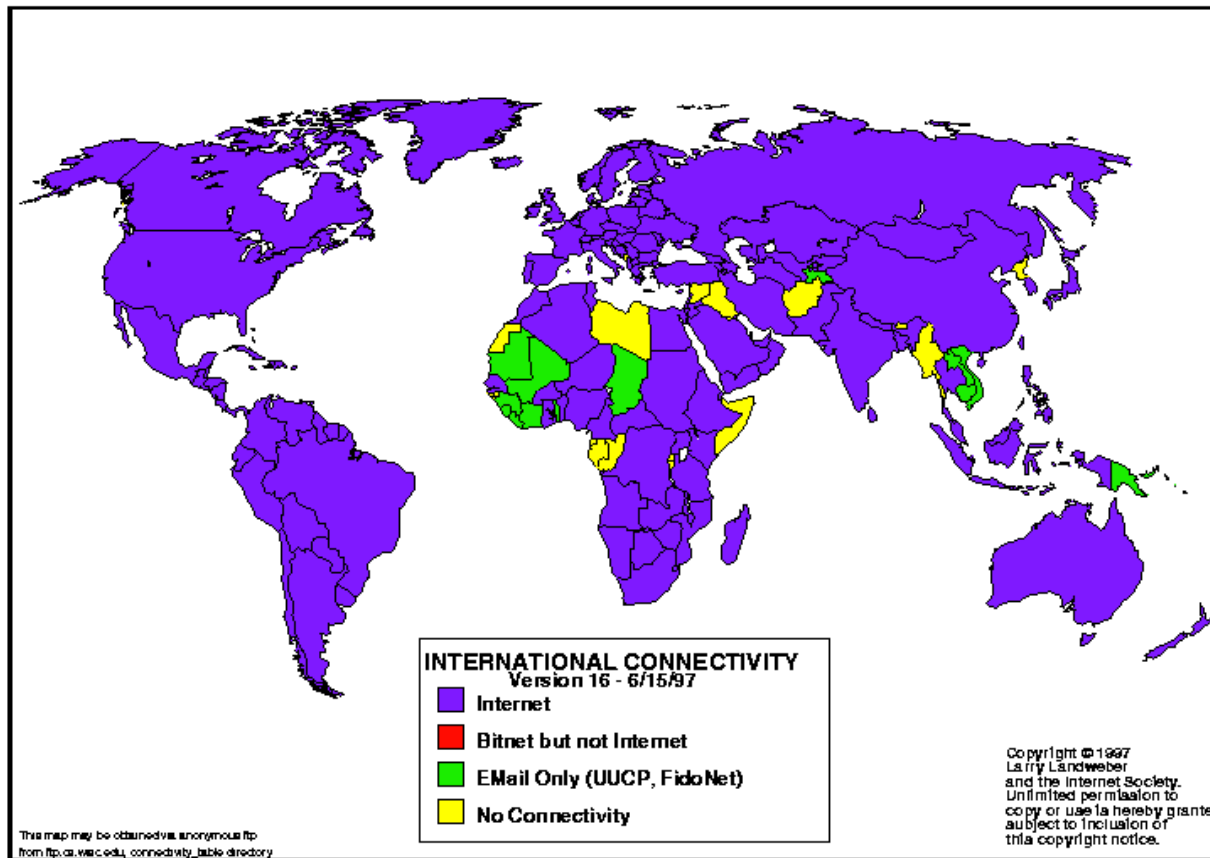
- Introduction and definitions
- Common technology elements
- The opportunity

Defining “Internet”: what’s important

- One network, everyone on it
- Multiple modes of communication
- Multiple media

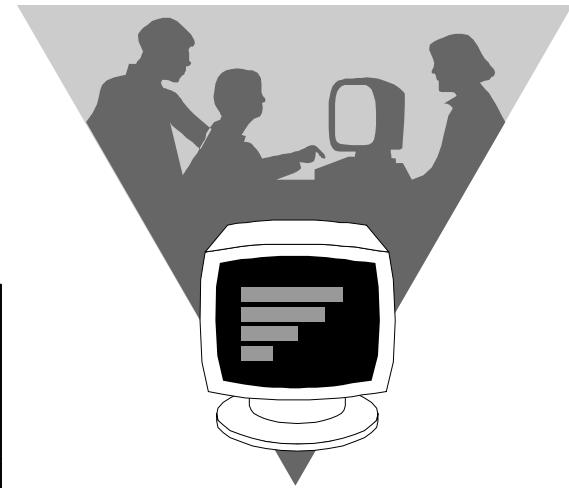
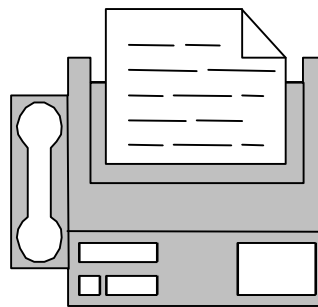
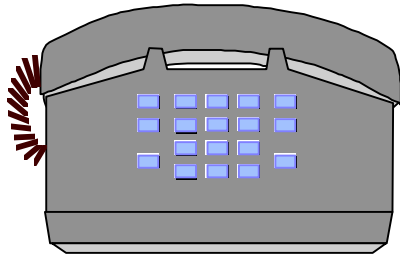
What's *important*?

One Network, Everyone On It



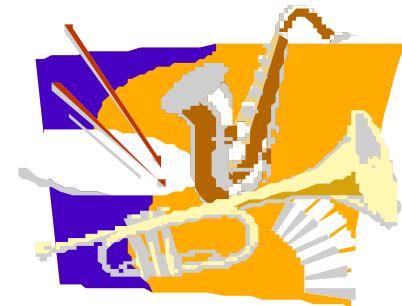
Many modes of communication

- publish/retrieve - web
- broadcast/subscribe - netnews, push
- interact - email



Many kinds of media

- documents
- video
- music
- interactivity



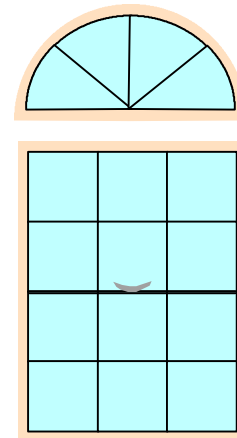
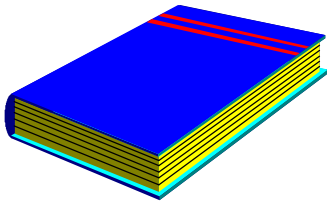
Internet: Global Interoperability

- Shared standards
- Working together
- International support
- Everyone connected, including
 - publishers, readers,
 - consumers, merchants



Digital Libraries: A Vision

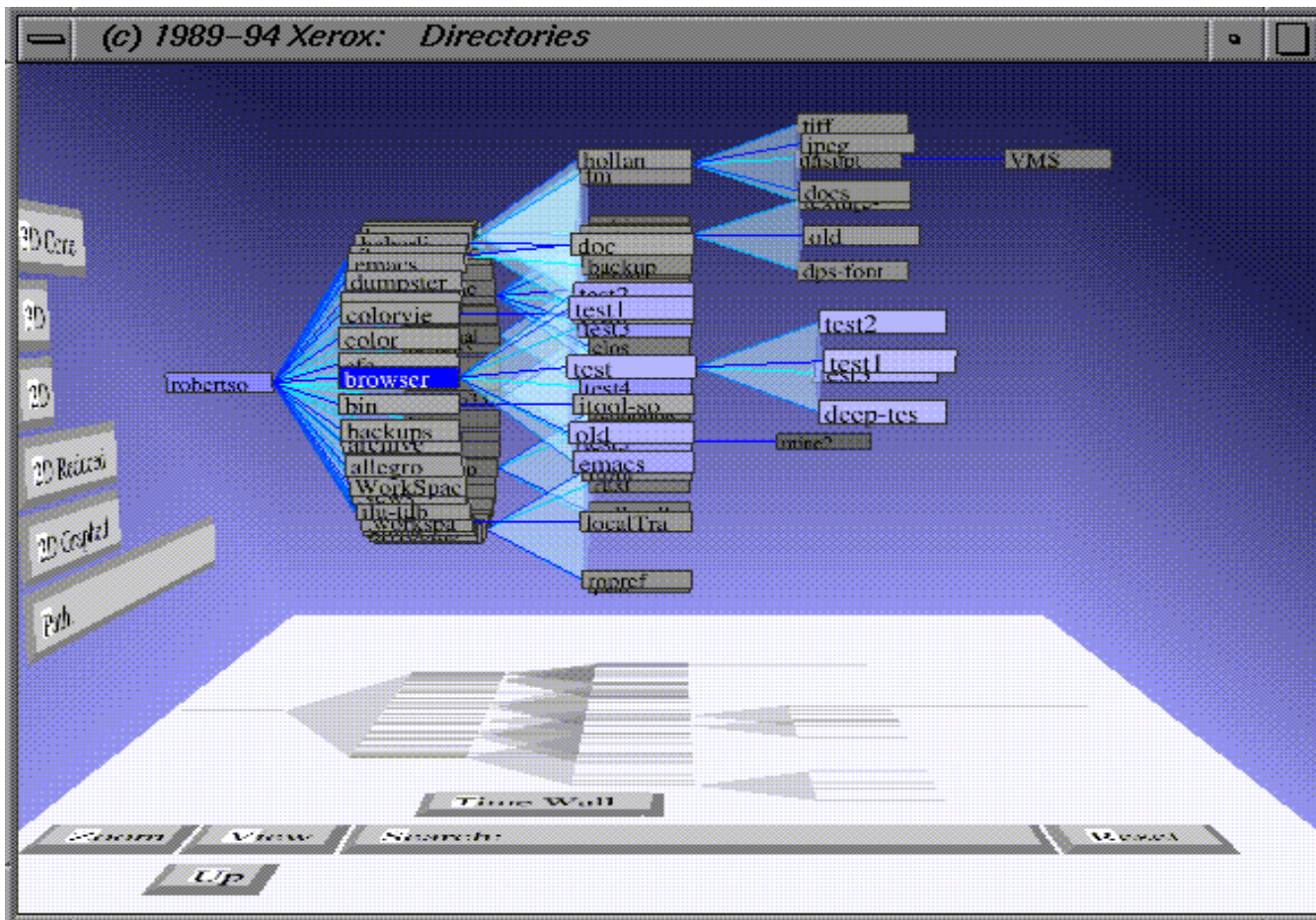
- Putting the world's knowledge online
- Preserve dual role of preservation & access
- Access to very large collections
- Document conversion and access



Putting the world's knowledge online

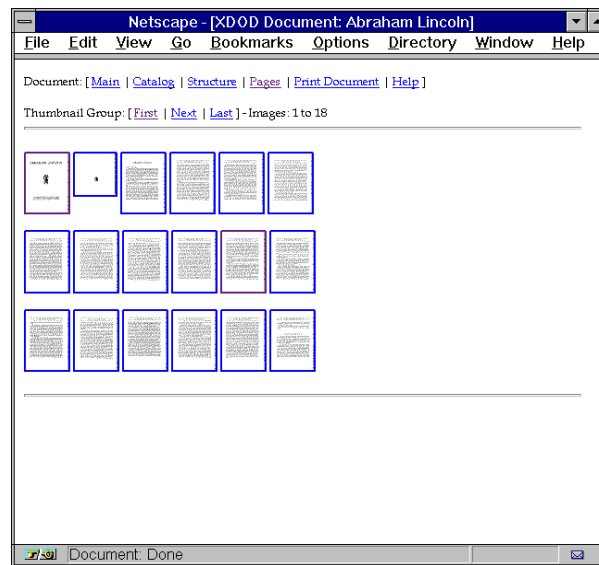


Access to very large collections



Document conversion and access

- Convert traditional library material
- Make available to local and remote users
- Focus on Indexing, finding, accessibility



Document Management: the Vision

- Helping groups and companies organize information and work



Document Management Industry

- Well-established set of products and services
- Broad marketplace, but primary applications:
 - Documents as Memory
 - Documents as Process
 - Documents for Products
 - Documents as Products

Documents as Memory:

Helping groups remember

- Users are “knowledge workers”
 - read information, analyze, produce more
- Law office
 - track correspondence
- Product organization
 - manage database of customer complaints, feature requests
- Marketing department
 - competitive profile

Document as Process: Helping managing work

Documents contain process information

- Insurance company
 - handling claims
- Financial services
 - processing loan applications
- Government offices
 - processing building permits
- Commercial organizations
 - customer billing, service

Document for Product: Describing something else

Examples from Manufacturing and Service

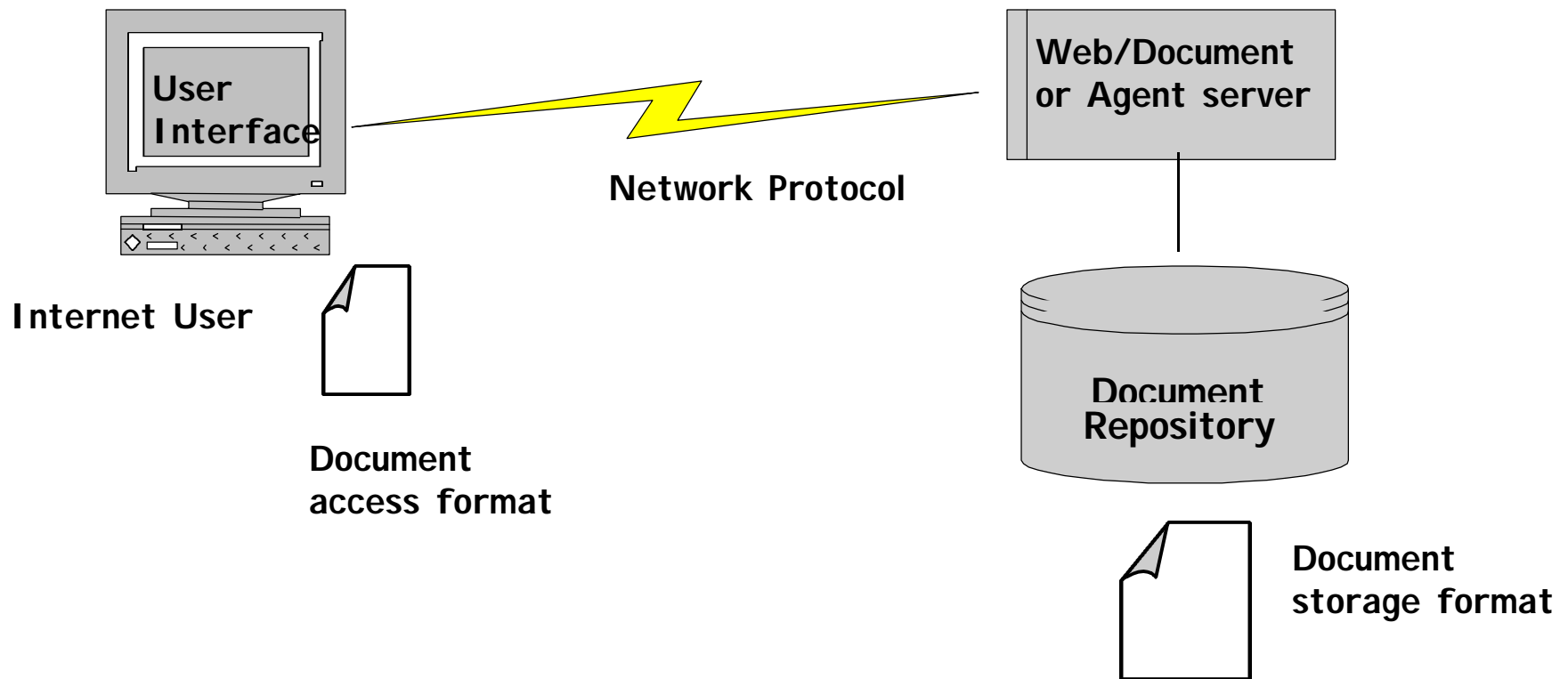
- Aerospace
 - Managing service, construction diagrams
- Pharmaceuticals
 - New Drug Application
- Financial services firm
 - Description of financial products
- Software
 - Manuals in many languages

Document as Product: “Intellectual Property”

Quality, timeliness, copyright management

- Entertainment
 - fiction, novels, trade
- Education
 - Course pack
 - Document assembly
- Production publishing
 - Just-in-time production
 - Books on demand

Common Technology Framework

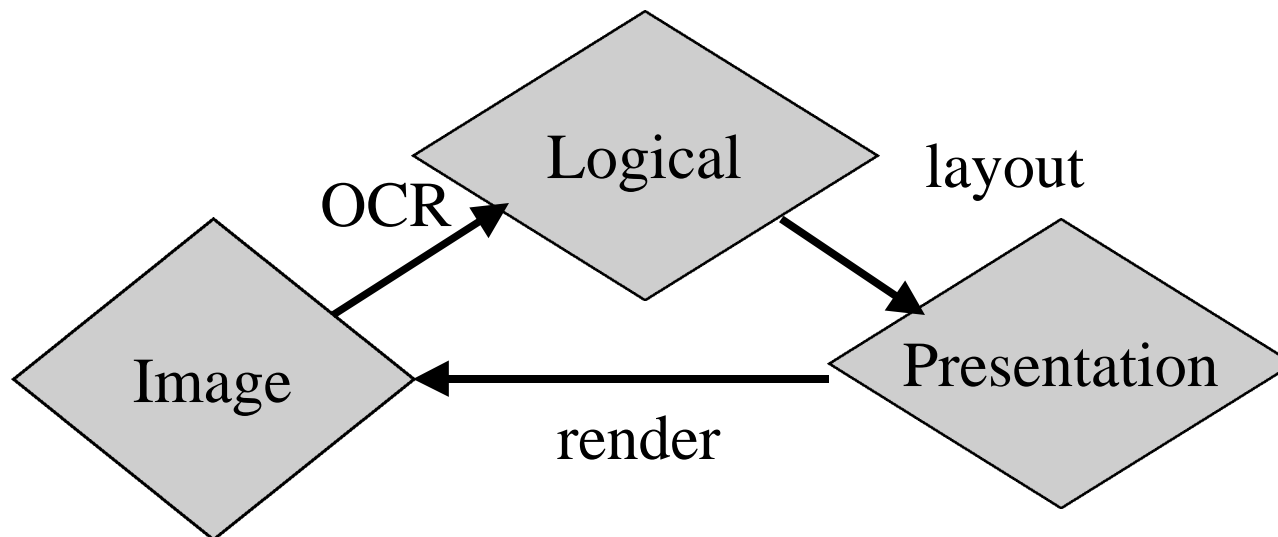


Common Technology Elements

- Documents
 - Formats, access methods, conversions
- Protocols
 - network and application interfaces
- References
 - naming, addressing, and description
- Security
 - authentication, authorization, accounting

Documents: Technology elements

- Electronic document formats:
 - logical, presentation, image
- Document conversion methods:



Protocols: Technology Elements

- Searching
 - Internet: Web forms, DASL
 - Libraries: Z39.50
- Document access and interaction
 - Internet: HTTP
- Metadata discovery and update
 - WebDAV
- User identification, authorization

References: Technology Elements

- Internet:
 - naming: URLs, URNs, URIs, Links
 - metadata: RDF
- Document management:
 - proprietary document handles
- Library:
 - ISBN, ISSN
 - MARC

Security: Technology Elements

- User Identity
 - login, smart card
- Key management
 - keeping passwords from travelling
- Property rights
 - recording and monitoring use
- Watermark
 - marking documents for later tracking

The Opportunity

- Document management and the Internet
 - Moving work beyond the enterprise
- Digital libraries and document management
 - long term preservation and community access
- The Internet and digital libraries
 - The world's knowledge available to the world

Summary

- Three applications
- Technology needs overlap
 - Different mechanisms
- Many opportunities
if we bring technology
together

