

WebDAV

Remote Collaborative Authoring and Electronic Records Management

Jim Whitehead

University of California, Santa Cruz
ejw@soe.ucsc.edu

WebDAV

- WebDAV is an:
 - Application layer network protocol
 - Extension to the Hypertext Transfer Protocol (HTTP 1.1)
 - Developed by the Internet Engineering Task Force (IETF)
- WebDAV
 - **Remote collaborative authoring** of Web resources
 - Overwrite prevention
 - Metadata management
 - Namespace operations

Major WebDAV Clients

- Application Software:
 - **Microsoft:** Office 2000/XP (Word, Excel, PowerPoint, Publisher)
 - **Adobe:** Photoshop 6, Acrobat 5
- XML editors
 - **Excsoft:** Documentor
 - **Altova:** XML Spy 3.5
- Web Site Authoring
 - **Adobe:** Go Live 5
 - **Macromedia:** Dreamweaver 4
- Remote File Access:
 - **Apple:** Mac OS X webdavfs & iDisk
 - OS X also ships with Apache and mod_dav (have to configure mod_dav to make it work)
 - **Microsoft:** Windows Web Folders
 - **Wind River Software:** WebDrive
 - Goliath (Mac, open source)
 - Nautilus (GNOME project, Eazel)
 - WebDAV Explorer (UC Irvine, Feise/Kanomata, open source)

Major WebDAV Servers

Microsoft: IIS 5/6, Exchange 2000, Sharepoint

Apache: mod_dav (over 10,000 sites)

Oracle: Internet File System

Adobe: InScope

Xythos: Web File Server

Novell: Netware 5.1, Net Publisher

W3C: Jigsaw

Endeavors: Magi-DAV

IBM: DAV4J (DeveloperWorks)

DataChannel: DataChannel Server (DCS 4.1)

Intraspect: 4i

OpenLink: Virtuoso

Hyperwave: Information Server 5.5

Using WebDAV

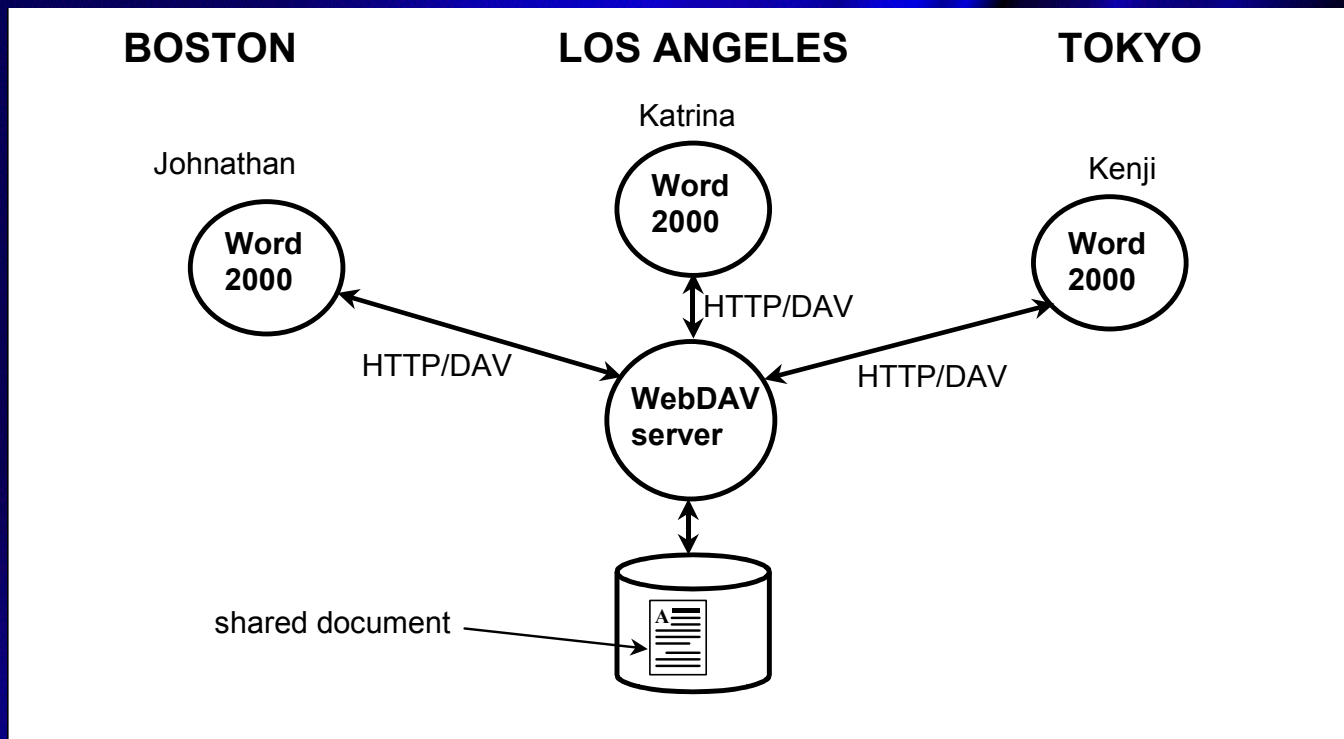
- You have to produce an XML document
 - Fire up your favorite XML processor & start editing
- You decide to bring on another author
 - Using the same XML processor, save to the Web
 - Give your collaborator the URL of the XML document
 - Start collaboration on the document by editing in-place on the Web
- A **seamless transition** from individual to collaborative work

Application Use of WebDAV



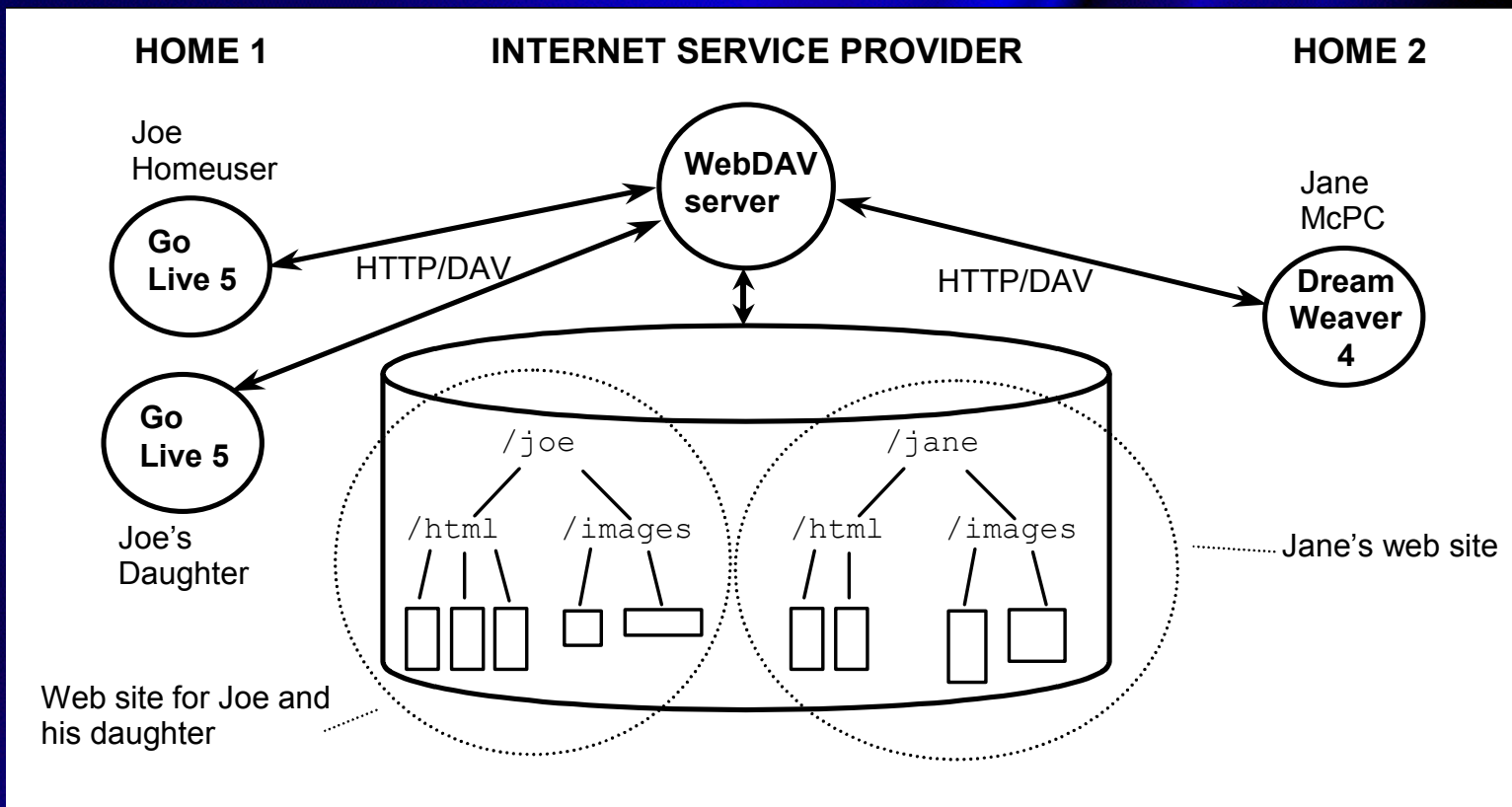
Collaborative Document Authoring

- Three collaborators, in different cities, use Word 2000 to collaborate on a report they are producing together.

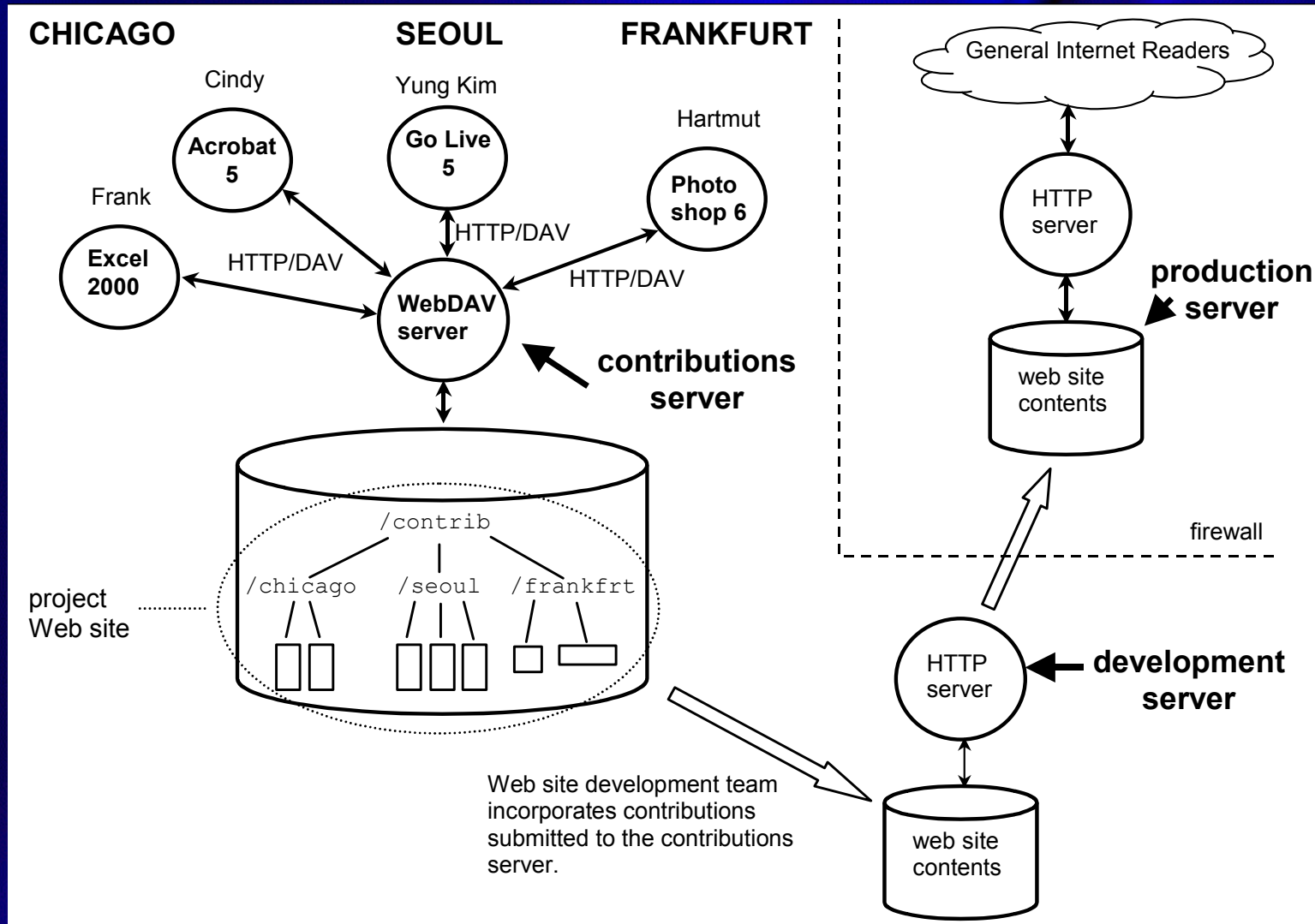


Collaborative Web Site Authoring

- Two homes each develop their family Web site using WebDAV to interact with their ISP.



Remote Authoring, Part of Staged Production

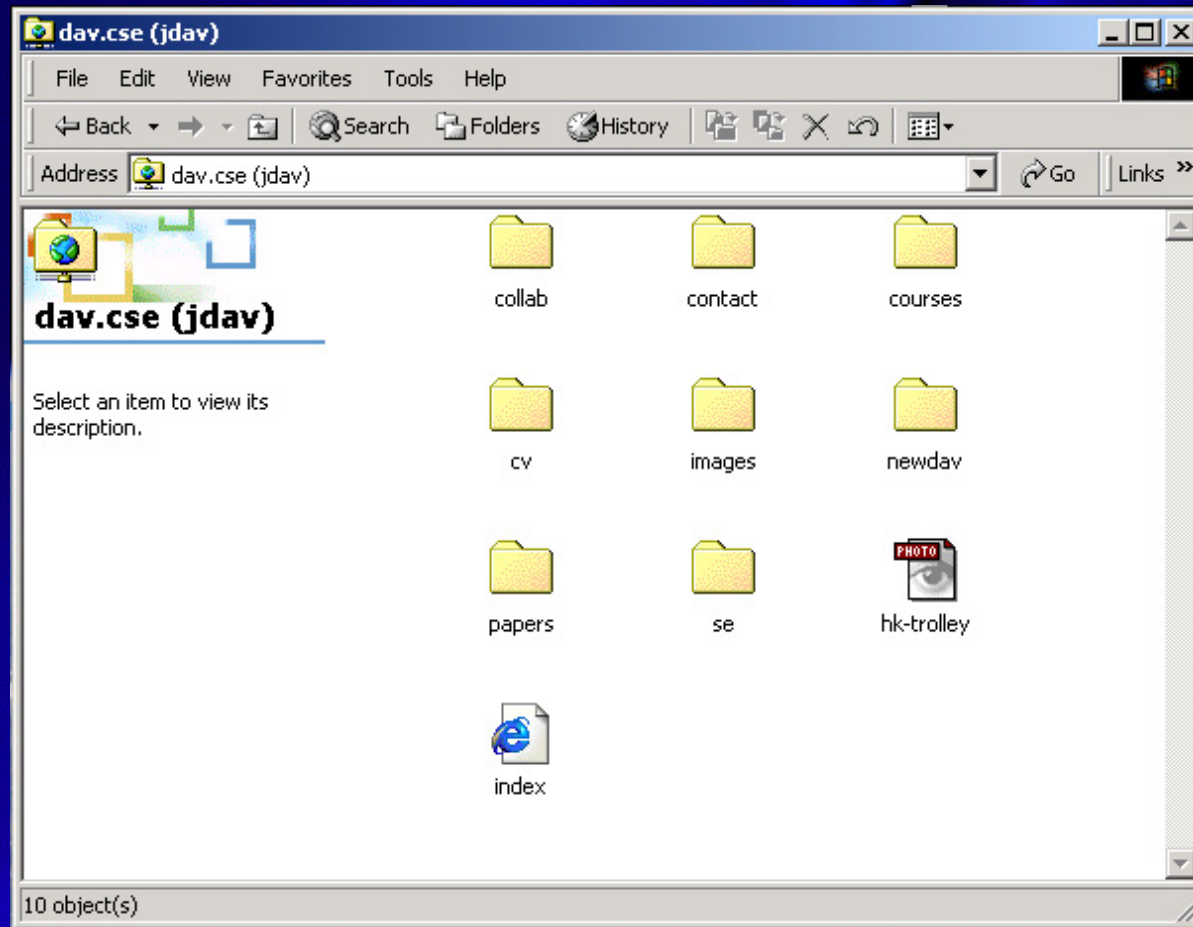


Visions for WebDAV

- Participants in WebDAV have many views on what it is:
 - A protocol for **collaborative authoring** of all document types
 - XML, HTML, word processing, spreadsheets,
 - A Web-based **network file system**, with nice high-latency behavior
 - Better performance than NFS and Samba over the Internet
 - A **data integration technology** for accessing a wide range of repositories
 - Document mgmt. systems, configuration mgmt. systems, filesystems, etc.
 - Remote **software engineering** infrastructure
 - A **replacement protocol** that can handle email, calendaring, directory lookup and more
 - Could replace: POP, IMAP, CAP, LDAP...
- *All views are correct!*

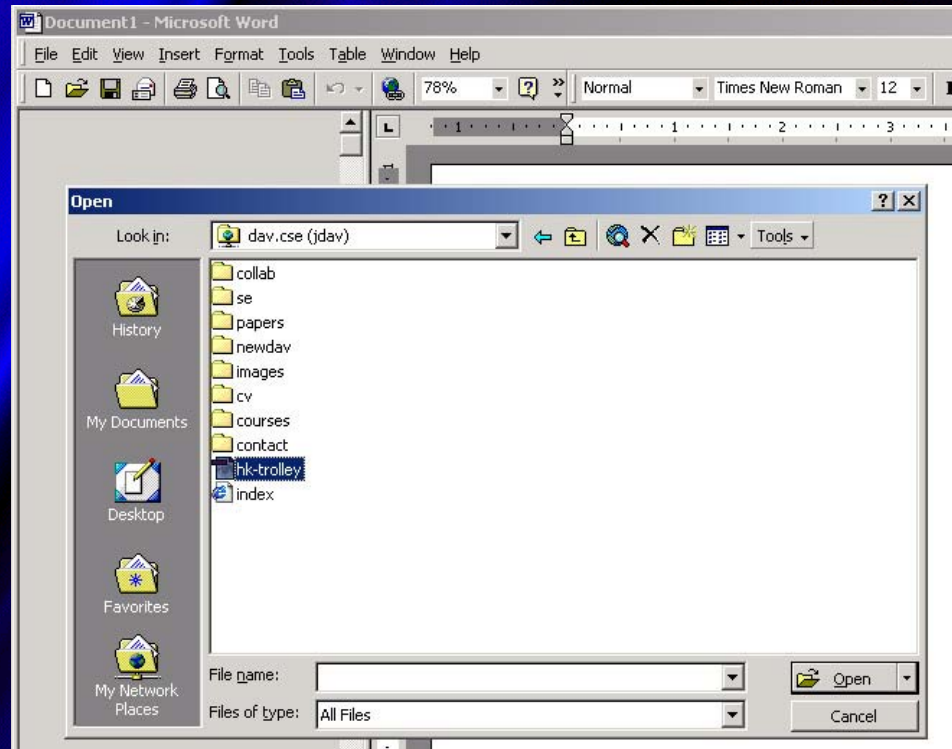
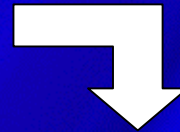
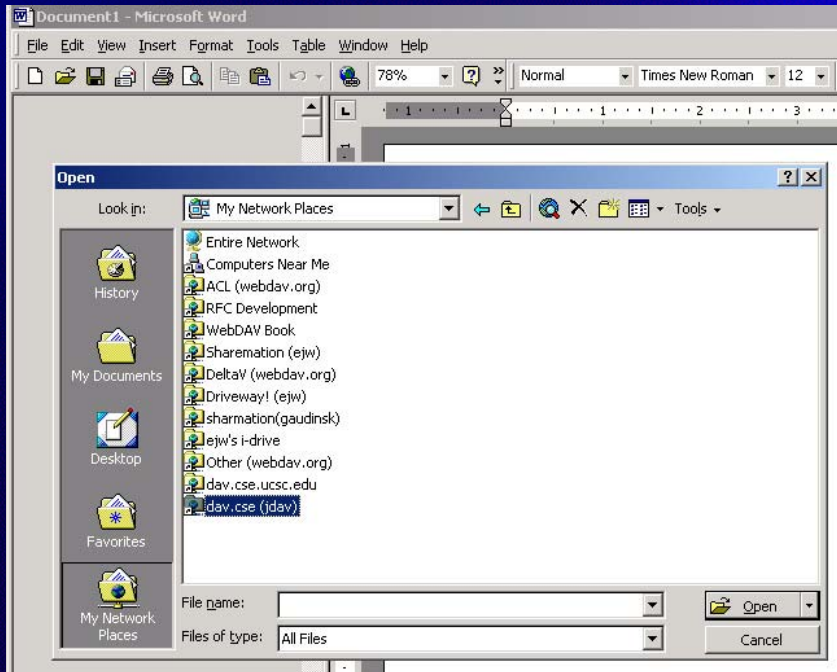
Filesystem View

- Exemplars: Web Folders, WebDrive, WebIFS, TeamDrive, Mac OS X



Document Authoring

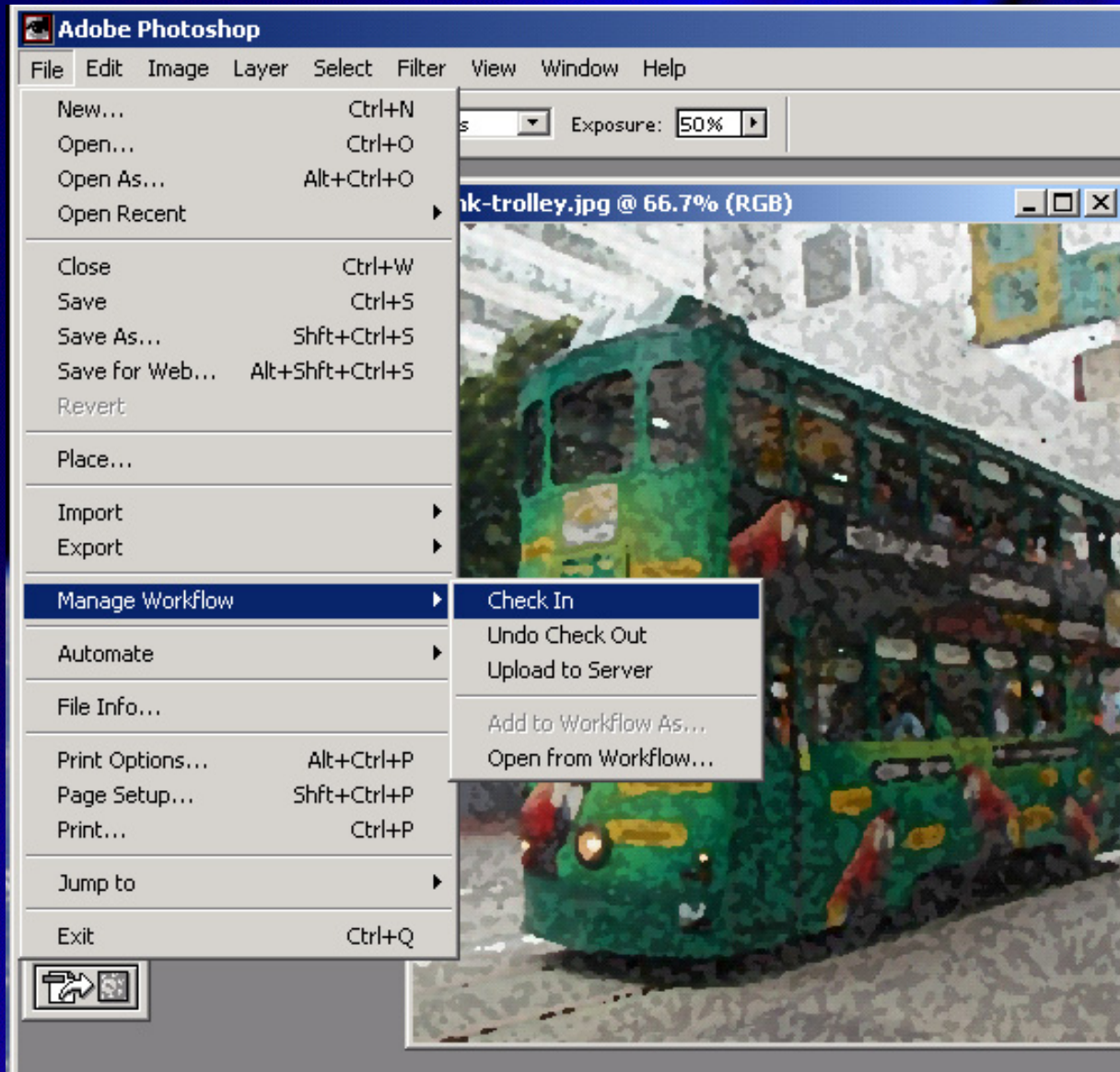
- Exemplars: Office 2000/XP: Word, Excel, PowerPoint, as well as Photoshop, Documentor, and XML Spy



Office: uses filesystem metaphor for WebDAV location

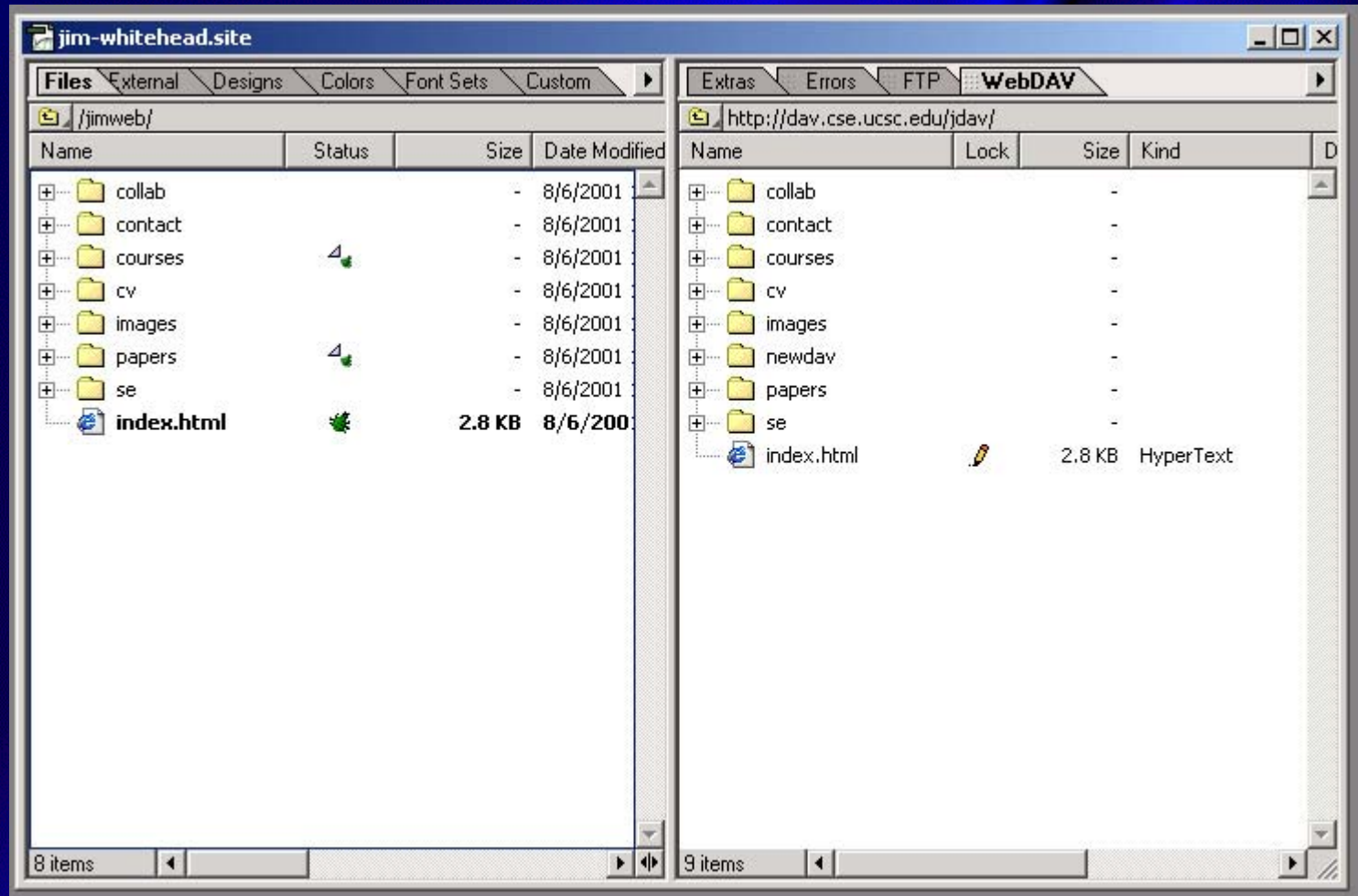
Photoshop

- **Workflow** metaphor for WebDAV location



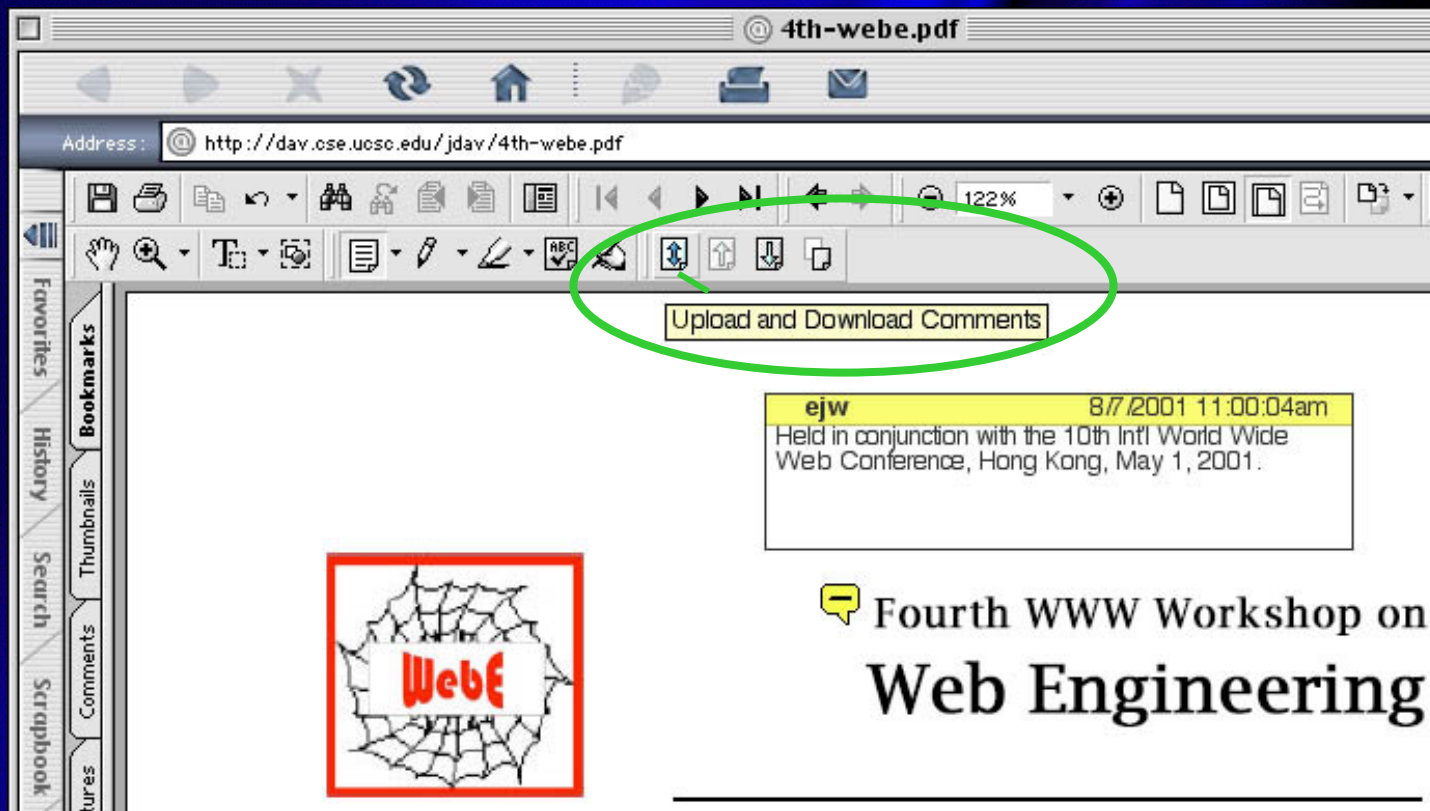
Web Site Authoring

- Exemplars: Go Live 5, Macromedia Director
- **Site** metaphor for WebDAV location



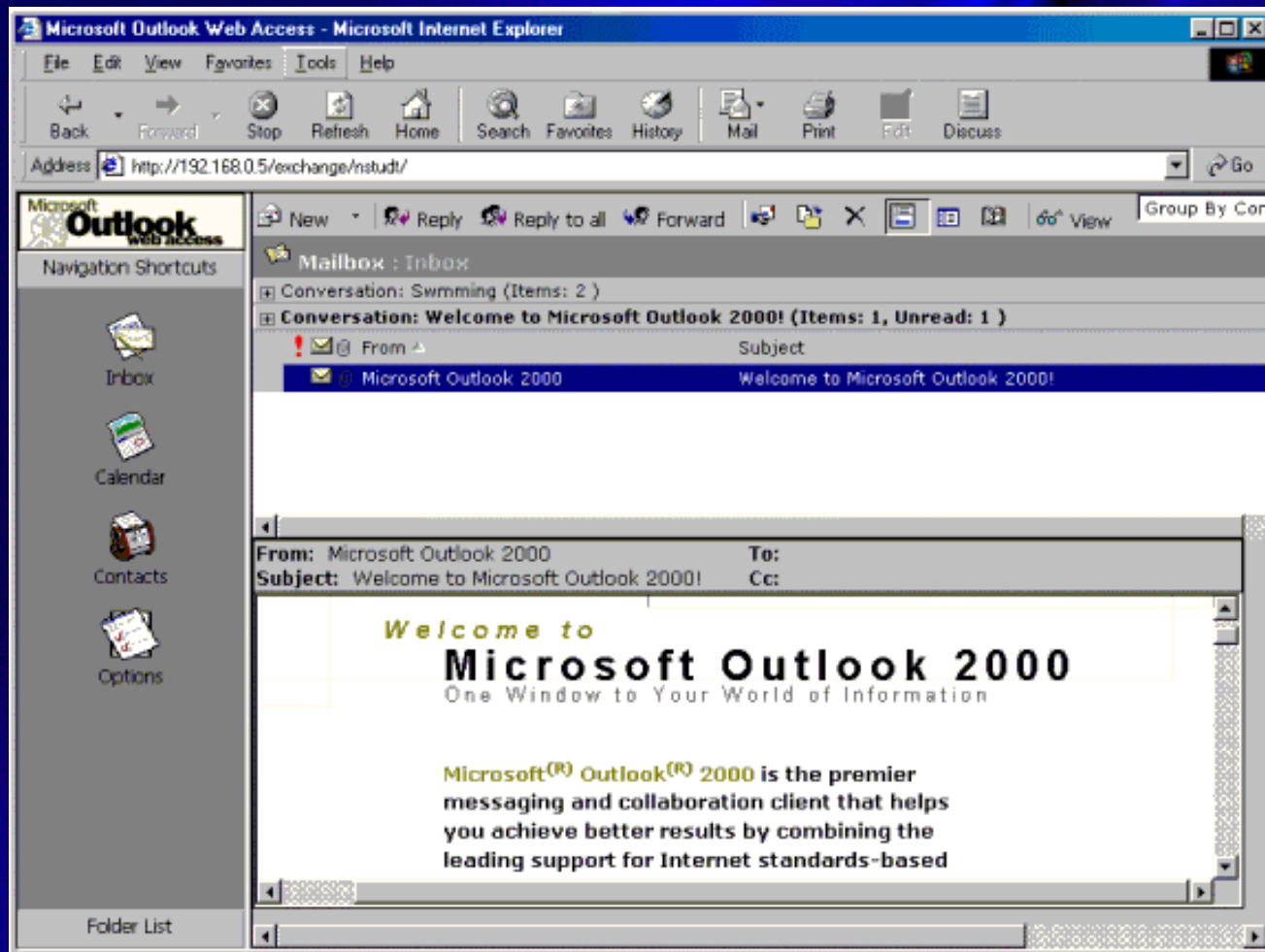
Remote Collaborative Annotation

- Acrobat 5 views a WebDAV location as a storage location for document annotations
 - Annotations are stored in resources separate from the PDF document
 - One collection per document
 - One annotation resource per user (in collection)



Email Access via WebDAV

- Outlook Web Access for Exchange 2000 uses WebDAV
 - plus some non-standard "batch" methods, for efficiency
- Client is DHTML running in the browser



Facets of WEBDAV

- There are many ways to view the DAV work:
 - Collaboration infrastructure
 - Metadata repository infrastructure
 - Namespace management infrastructure
 - Access control infrastructure
 - Searching infrastructure – DASL

Collaboration Infrastructure

- Whole resource locking supports:
 - remote collaborative authoring of any media type
 - Web pages, Word processing, Presentations, XML, ...
- Lock characteristics:
 - Long-duration locks
 - Not associated with network connection
 - Client receives a lock token identifying the lock
 - Can disconnect from network after receiving lock token
 - Locks automatically expire after a client specified timeout period
 - Lock single resources, or hierarchies of resources
- *Infrastructure* for asynchronous, widely distributed, hypertext-aware, collaborative editing tools.

Metadata Recording Infrastructure

- Metadata support
 - ***Properties*** are (name, value) pairs that can be created, modified, deleted, and read on Web resources.
 - Consistency of properties can be maintained by the server or the client
 - Property values are well-formed Extensible Markup Language (XML)
 - Can store RDF as well
 - Property name is a URI or URL
 - Extensible, global property namespace
- ***Infrastructure*** for recording information *about* Web data
 - A general purpose metadata repository

Namespace Management Infrastructure

- Remote name space management:
 - Copy and Move
 - individual resources
 - hierarchies of resources
 - to/from a locked hierarchy
 - Create and modify collections of resources
 - Retrieve listings of collection members
 - Useful for creating Save...As dialog boxes
 - Adds hierarchical navigation to the Web
 - Augments hypertext navigation and searching
- *Infrastructure* for remotely organizing and viewing collections of Web resources

Access Control Infrastructure

- Access Control:
 - The ability to remotely control who can read and write a resource
 - Key challenge:
 - Expose the access control capabilities of the repository...
 - ...while ensuring the client-side user interface can be simple (I.e., avoid lots of feature discovery)
 - Access control lists
 - Each entry grants/denies a privilege to a principal (or group)
 - Searching for principals
 - Can find people by name, title, organization (caseless string match)
- *Infrastructure* for remotely creating collaboration groups

Searching Infrastructure

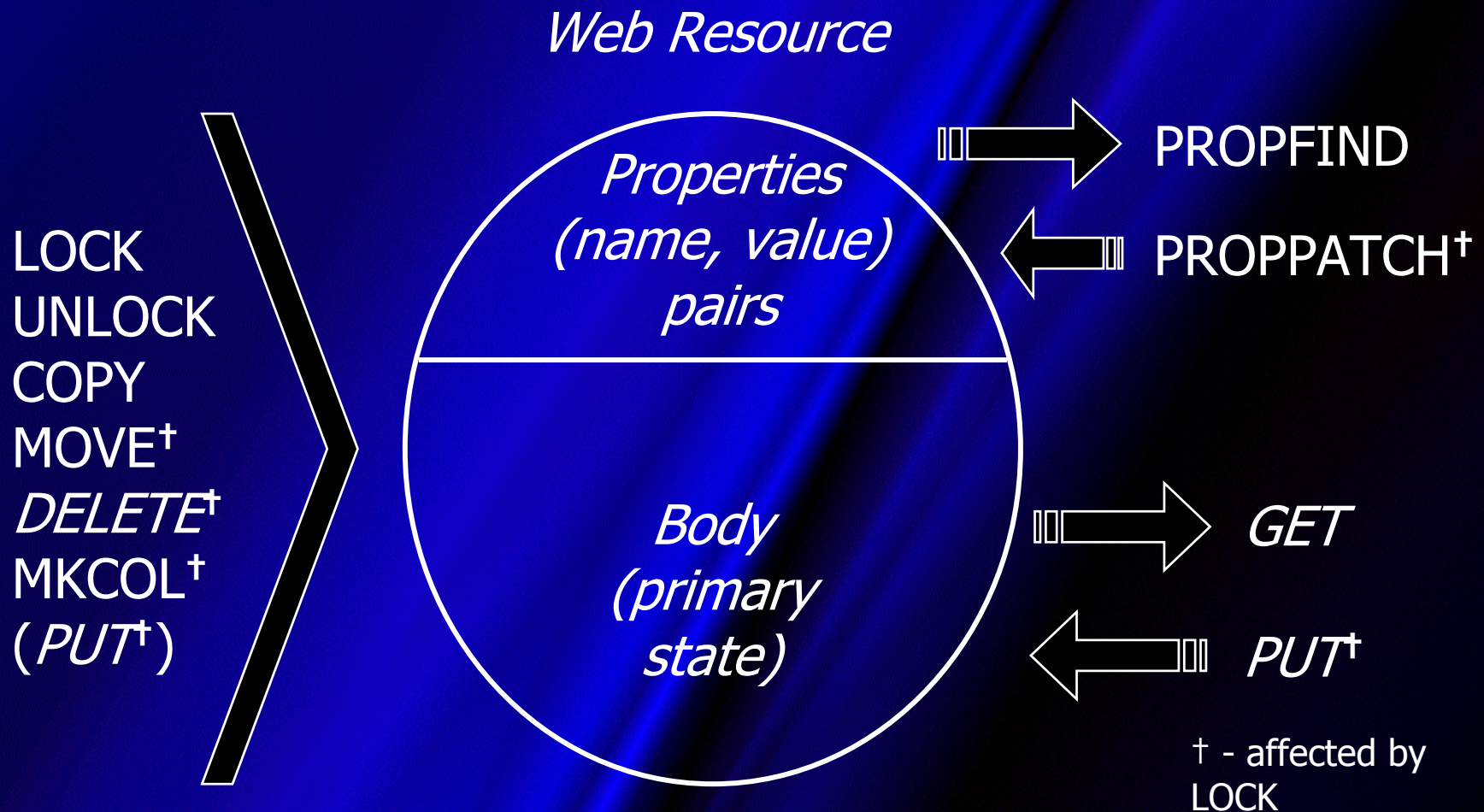
- Searching a WebDAV repository - DASL:
 - Search for resources with a given property, or a given property value
 - Search for a substring inside a resource body
 - Search scope can be one resource, a collection of resources, a hierarchy of resources, or a whole server
- Search syntax is extensible
 - Search specification specified as body submitted with SEARCH command
 - Could accommodate XML Query syntax
- *Infrastructure* for remote searching

WebDAV Methods and Data Model

WebDAV Methods

- **Overwrite Prevention:**
 - LOCK – prevents non-lock holders from writing to the resource
 - UNLOCK – removes a lock
- **Metadata Management:**
 - PROPFIND – read properties from a resource
 - Allprop – all property names and values
 - Propname – only return property names
 - Prop – just return specified properties
 - PROPPATCH – write properties on a resource
- **Namespace Management**
 - COPY – duplicate a resource
 - MOVE – move a resource (preserving identity)
 - MKCOL – create a new collection

Scope of WebDAV Methods



Electronic Records Management

ERM Standards Landscape

- **DoD 5015.2**

- “Design Criteria Standard for Electronic Records Management Software Applications”
- Detailed requirements that ERM systems must meet
- Joint Interoperability Test Command (JITC) performs compliance testing
- DoD policy is to only acquire 5015.2-compliant systems

- **ISO 15489**

- “Records Management”
- Rationale for records management
- High level requirements for ERM systems
- Compliance testing is difficult, if not impossible

ERM Standards Landscape (2)

- GILS
 - “Global Information Locator Service”
 - A profile of Z39.50 for government records
 - Z39.50 is a network protocol for searching bibliographic databases
 - Defines:
 - Standard metadata items
 - Search operators

ERM and Interoperability

- DoD 5015.2 is **not** an interoperability standard
 - It is a **requirements** standard
- Impact:
 - Tools **will not interoperate** with more than one 5015.2 compliant repository
 - **Metadata translation** is required when moving records across 5015.2 compliant repositories
 - No interoperability means **no network effects**
 - **No incentive to create tools** for 5015.2
 - Every tool must have a wrapper for **every** 5015.2 repository
 - **No possibility of open source support**
 - Data models of 5015.2 repositories are not generally available
- GILS **is** an interoperability standard for searching
 - But, 5015.2 lists support for GILS as “non-mandatory”

Goals for ERM Interoperability

- Two potential goals, in increasing order of difficulty:
- **Interchange:** the ability to replicate a file plan and its contents across records management systems such that the duplicate is perceived as identical by a human.
 - Includes file groups, files, disposition instructions, record categories, records, cutoffs & retention specifications
- **Functional interoperation:** Records management user agents should seamlessly work with a broad range of records management systems

Achieving Interchange

- A (UML?) data model based on 5015.2 entities showing relationships among:
 - File plan, file group, file, record category, disposition instruction, cutoff spec., retention spec., records
- Map metadata values into XML
 - Standard values for common metadata items
 - Media types, formats, vital records indicators
 - All future-proofed using XML
- A disposition instruction language
 - A language for specifying cutoff & retention policies suitable for cross-system interchange
- A language for linking related records (Xlink?)
- Packaging conventions for email & attachments
 - Just use MIME?
- Describing syntax of various code spaces
 - Record category code, file (group) code, etc.

Achieving Functional Interoperation: WebDAV and ERM

- Map data model onto DAV concepts
 - Resources, collections, version histories
- Use WebDAV properties to store XMLified metadata items
- Develop server support for
 - Interpreting disposition instructions
 - Automatic archiving/disposal of records
 - Searching of 5015.2 properties (via DASL)
- Develop client support for management of:
 - File plans, file groups, files, disposition instructions, record categories
 - Assignment of codes to records
 - Linking of records

Benefits of WebDAV for ERM

- Widespread deployment of interoperable, low-cost infrastructure for archiving electronic records
 - Extends who can benefit from ERM
 - Schools, small town governments, other under-resourced government functions
 - Better management of, and access to records
- Interoperability standard
 - Could be adopted by other 5015.2 compliant systems
 - Use of XML will help “future-proof” records metadata
 - Leverage existing base of DAV applications
 - Strong possibility of open source support
 - Could lead to a marketplace of 5015.2 supporting tools
 - Has utility far outside government realm
 - **Personal** records management

Benefits of WebDAV for ERM

- Integrated document authoring and records management
 - Manage records using the **same system** where the resource was initially authored
 - No artificial distinction between...
 - Where documents are created
 - Where documents become records
 - Where records are disposed of
 - Potentially even where records are archived for long-term
 - ERM is integrated into the entire document lifecycle from the moment a (future) record is created

The Road Ahead

1. Develop mapping of 5015.2 to WebDAV/DASL
 - Metadata representations in XML or RDF
 - Representation of 5015.2 abstractions (file plan, record category)
 - Mapping functions to WebDAV/DASL operations
2. Proof of concept
 - Server support (modified Apache mod_dav)
 - Metadata entry & search client
3. Refine initial mappings
4. Begin standards development process using research results as a starting point

WebDAV Resources

www.webdav.org

WebDAV

- **WebDAV Resources**

- <http://www.webdav.org/>
 - A central collection of pages and links to all things WebDAV.

- **WebDAV Working Group**

- <http://www.ics.uci.edu/pub/ietf/webdav/>
 - Contains links to active documents, and a complete list of WebDAV-supporting applications.

- **RFC 2518 – WebDAV Distributed Authoring Protocol**

- <http://www.ics.uci.edu/pub/ietf/webdav/protocol/rfc2518.pdf>
 - This is the WebDAV Distributed Authoring Protocol specification

- **WebDAV: A network protocol for remote collaborative authoring on the Web**

- *Proc. of the Sixth European Conference on Computer-Supported Cooperative Work*, Sept. 12-16, 1999, Copenhagen, Denmark, pp. 291-310.
- <http://www.ics.uci.edu/~ejw/papers/dav-ecscw.pdf>
 - An academic paper giving an overview of the WebDAV Distributed Authoring Protocol.

DeltaV

- **Delta-V Working Group web page**
 - <http://www.webdav.org/deltav/>
 - The home page for the IETF Delta-V Working Group, with links off to the most recent specifications.
- G. Clemm, J. Amsden, C. Kaler, J. Whitehead, "**Versioning Extensions to WebDAV**", Internet-Draft, work-in-progress, draft-ietf-deltav-versioning-20, October 23, 2001.
 - <http://www.webdav.org/deltav/protocol/draft-ietf-deltav-versioning-20.htm>
 - The most recent revision of the versioning and configuration management protocol specification.
- **The Future of Distributed Software Development on the Internet**
 - Web Techniques, Vol. 4, No. 10, October, 1999, pages 57-63
 - <http://www.webtechniques.com/archives/1999/10/whitehead/>
 - An introduction to WebDAV and DeltaV that describes the advantages of DeltaV over CVS for remote collaborative software development

DASL: DAV Searching & Locating

- DAV Searching and Locating page
 - <http://www.webdav.org/dasl/>
 - A web site containing links to the most recent WebDAV searching protocol specifications.
- A. Babich, J. Davis, R. Henderson, D. Lowry, Sa. Reddy, Su. Reddy, "DAV Searching and Locating", work-in-progress.
 - <http://www.webdav.org/dasl/protocol/draft-davis-dasl-protocol-00.html>
 - The most recently edited DASL protocol specification. This document was never submitted as an Internet-Draft.
 - In 1-2 years, will be submitted as a Proposed Standard
- GILS and DASL
 - <http://www.gils.net/dasl/workshop.html>
 - <http://lists.w3.org/Archives/Public/www-webdav-dasl/1999JanMar/0003.html>

Access Control Protocol

- Access Control page
 - <http://www.webdav.org/acl/>
 - A web site containing links to current access control protocol specifications.
- **WebDAV Access Control Protocol**
 - draft-ietf-webdav-acl-07, November 7, 2001.
 - <http://www.webdav.org/acl/protocol/draft-ietf-webdav-acl-07.htm>
 - The most recent revision of the access control protocol specification.

Advanced Collections

■ **WebDAV Bindings**

- draft-ietf-webdav-binding-protocol-02, December 17, 1999.
- <http://www.ics.uci.edu/pub/ietf/webdav/collection/draft-ietf-webdav-binding-protocol-02.txt>
 - The most recent revision of the WebDAV Bindings Protocol.

■ **WebDAV Redirect Reference Resources**

- draft-ietf-webdav-redirectref-protocol-02, December 17, 1999.
- <http://www.ics.uci.edu/pub/ietf/webdav/collection/draft-ietf-webdav-redirectref-protocol-02.txt>
 - The most recent revision of the Redirect Resources protocol.

■ **WebDAV Ordered Collections Protocol**

- draft-ietf-webdav-ordering-protocol-02, December 20, 1999.
- <http://www.ics.uci.edu/pub/ietf/webdav/collection/draft-ietf-webdav-ordering-protocol-02.txt>
 - The most recent revision of the Ordered Collections protocol.

Document Roadmap

WebDAV Working Group:

Distributed Authoring

Locking, Properties, Copy/Move
RFC 2518 complete

Access Control

Protocol for remote access control
Finish: January 2002

Advanced Collections

Requirements and protocol for
bindings, redirectors, ordered coll.
Finish: May 2002

DASL:

Searching

Requirements and protocol for
searching a WebDAV repository
Finish: late 2002

Delta-V Working Group:

Versioning & CM

RFC number imminent
Finish: December 2001

