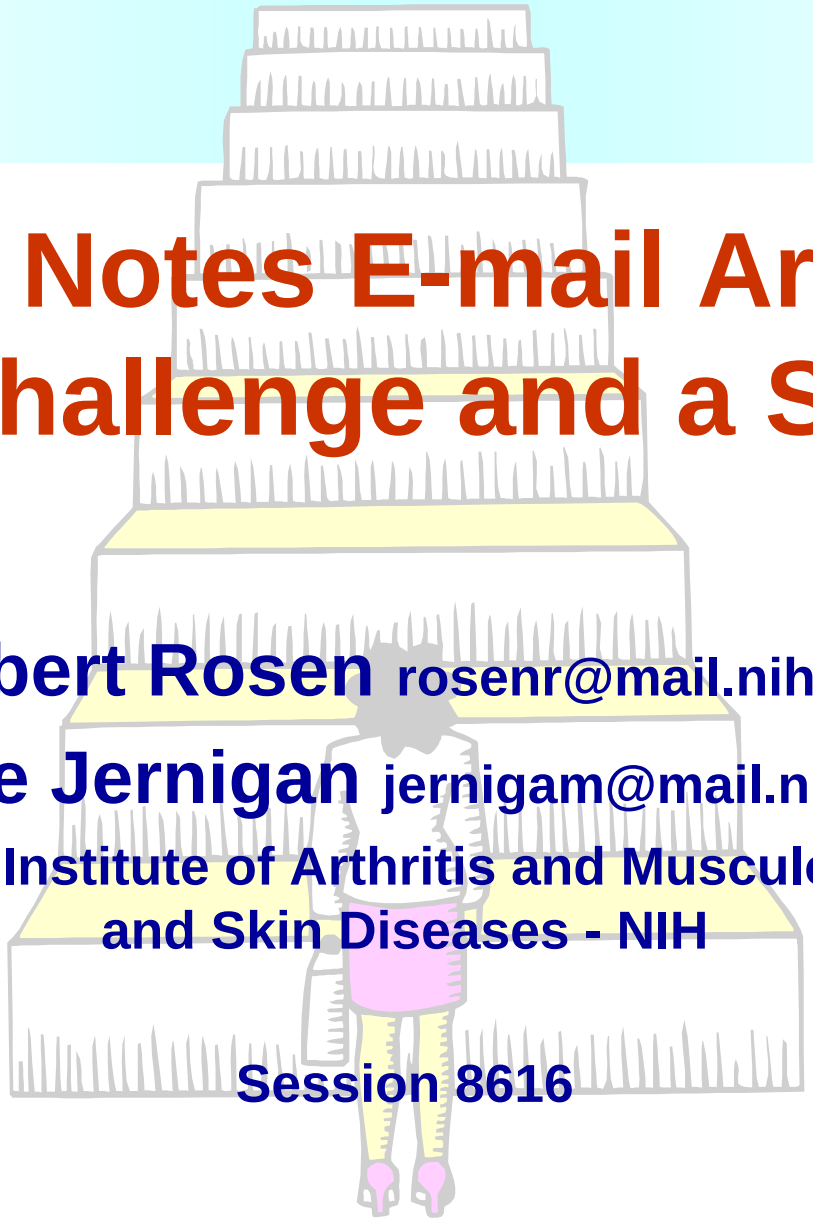
NIAMS

SHARE
Technology • Connections • Results

Lotus Notes E-mail Archiving
The Challenge and a Solution



Robert Rosen rosenr@mail.nih.gov

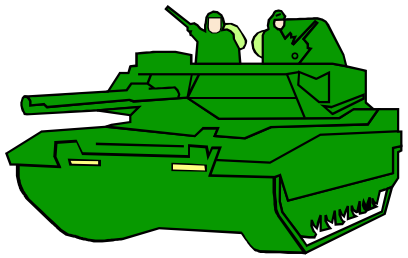
Millie Jernigan jernigam@mail.nih.gov

National Institute of Arthritis and Musculoskeletal and Skin Diseases - NIH

Session 8616

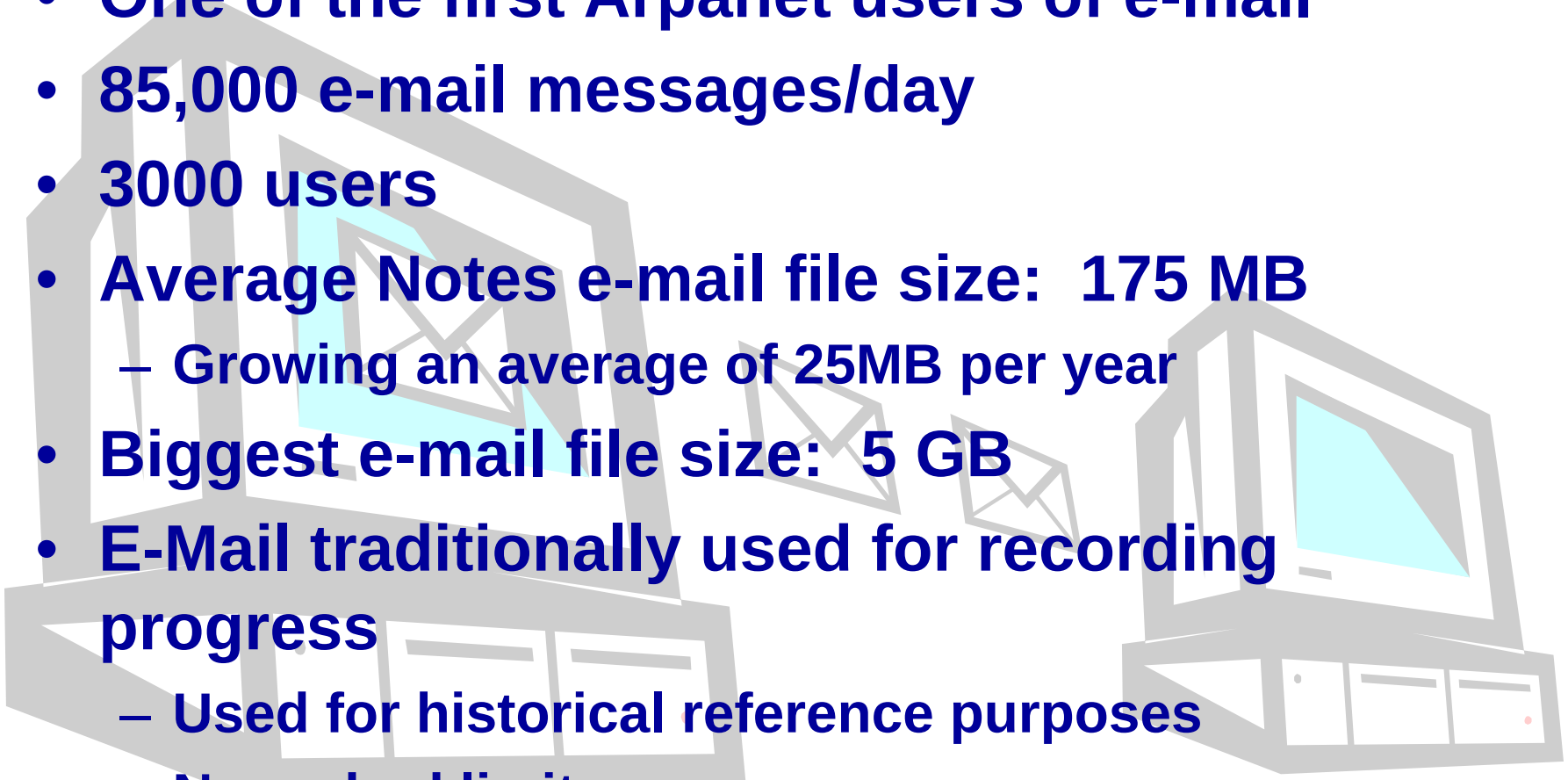
The Lab

- **The lab addresses the key military technologies for Army and Department of Defense programs.**
- **E-Mail is a critical tool used in the operation of the lab and the conduct of research and development**



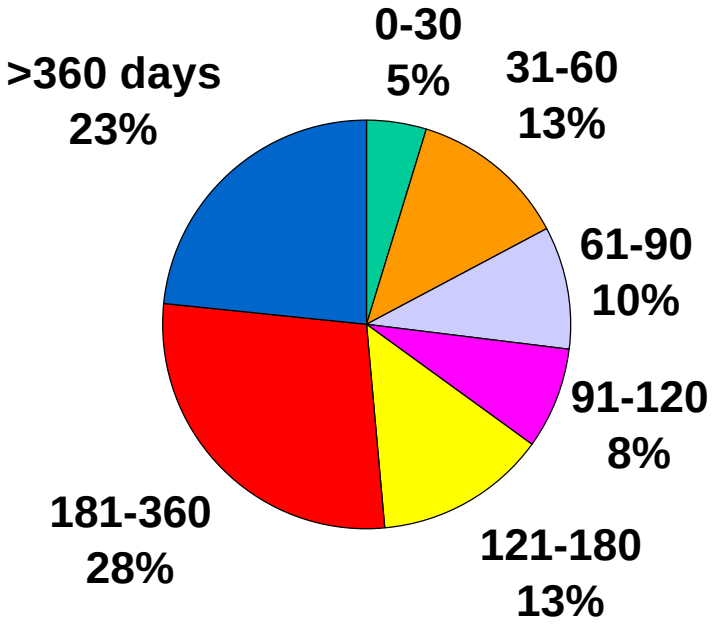
The E-Mail Challenge

- One of the first Arpanet users of e-mail
- 85,000 e-mail messages/day
- 3000 users
- Average Notes e-mail file size: 175 MB
  - Growing an average of 25MB per year
- Biggest e-mail file size: 5 GB
- E-Mail traditionally used for recording progress
  - Used for historical reference purposes
  - Never had limits



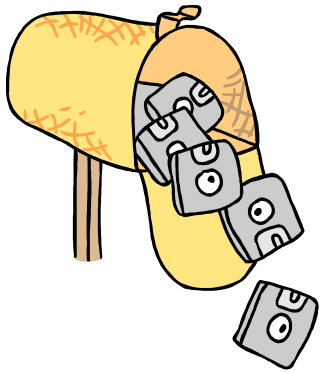
The Environment

- 14 e-mail servers (and cluster backups)
- Servers at 98% storage capacity
- Requests to delete unneeded e-mail ineffective
- ½ mail >6 months old



Proposed “Policy”

- E-mail 0-3 months on server
- E-mail 3-15 months on MSS
  - E-mail stub left behind
  - Click on stub retrieves e-mail
- E-mail 15-60 months on LTO tape
  - E-mail stub left behind
  - Click on stub retrieves e-mail
- E-mail > 5 years deleted



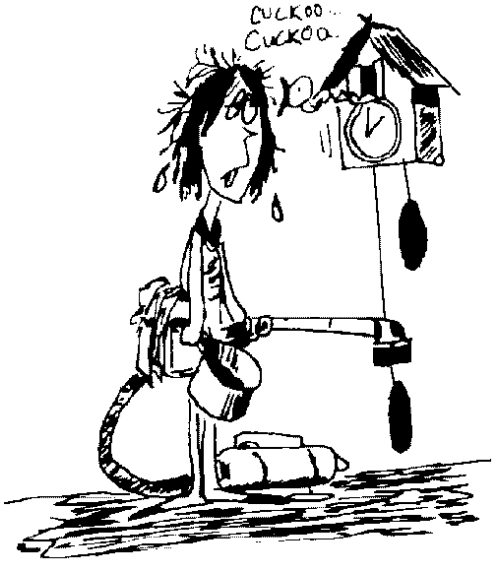
How to do it?

Common Store for Lotus Domino (CSLD)

- **Multipart Product**
  - **Common Store identifies and serves up documents from Domino into Content Manager**
  - **Content Manager stores the archived document**
  - **Content Manager maintains the database index of archived documents in DB2**
  - **Tivoli software manages DB2/docs on tape**

Implementation

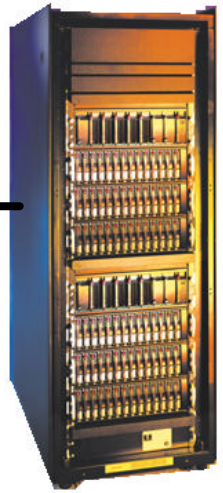
- Turn Key?
  - NO!!
- What do you need
  - Tailoring
    - Notes developer skills
  - Optimizing
    - CSLD, Content Manager, DB2, TSM and hardware skills



- Specifics
  - Installing MSS and LTO along with CSLD
  - Installing Content Manager server (Netfinity)

Our Hardware

Fibre Channel



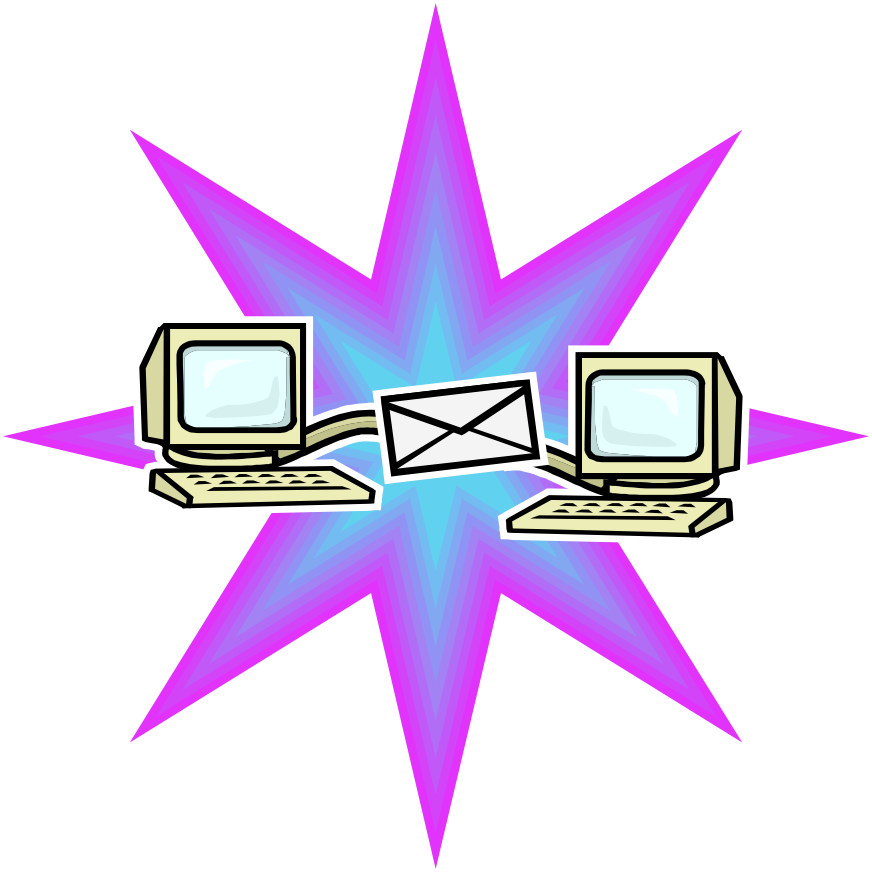
SCSI

Netfinity 7100
1 GB RAM, 145GB Storage
Combined
Content Manager/
CSLD Server

LTO 3583
36 Slots
2 Drives

MSS 2106
1.4TB Storage

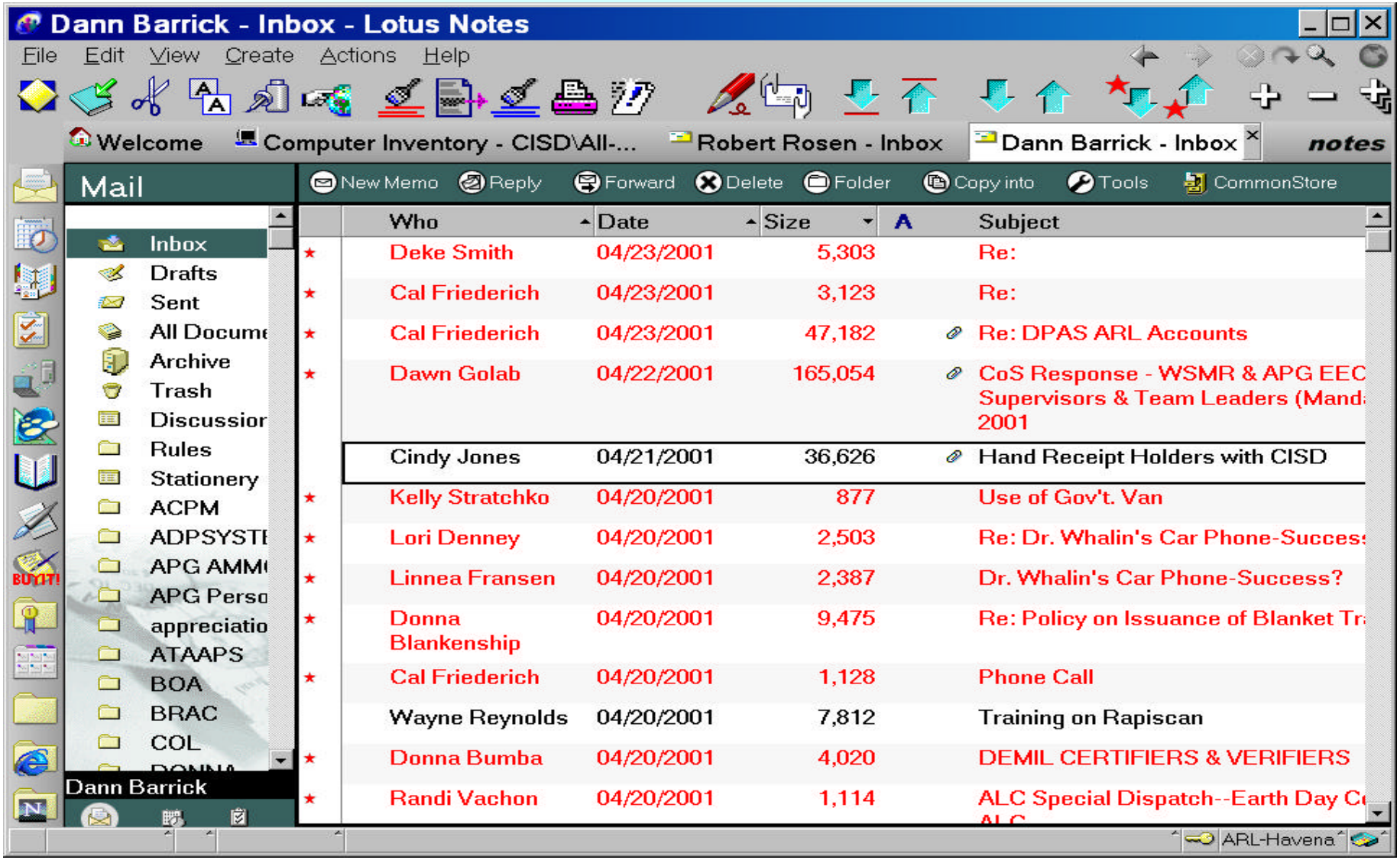
Let’s Look at it In Operation



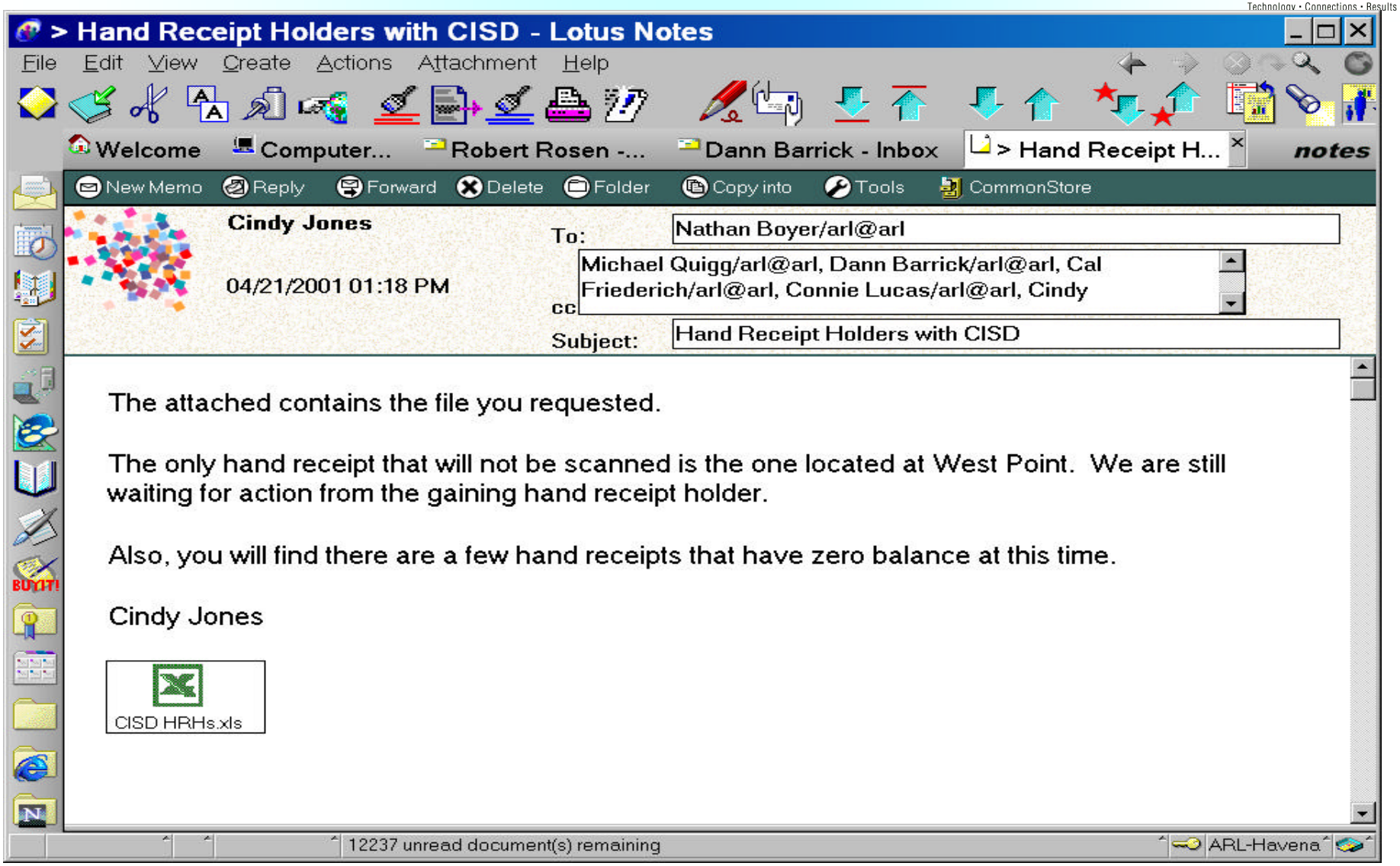
IBM **Common Store**

For Lotus Domino

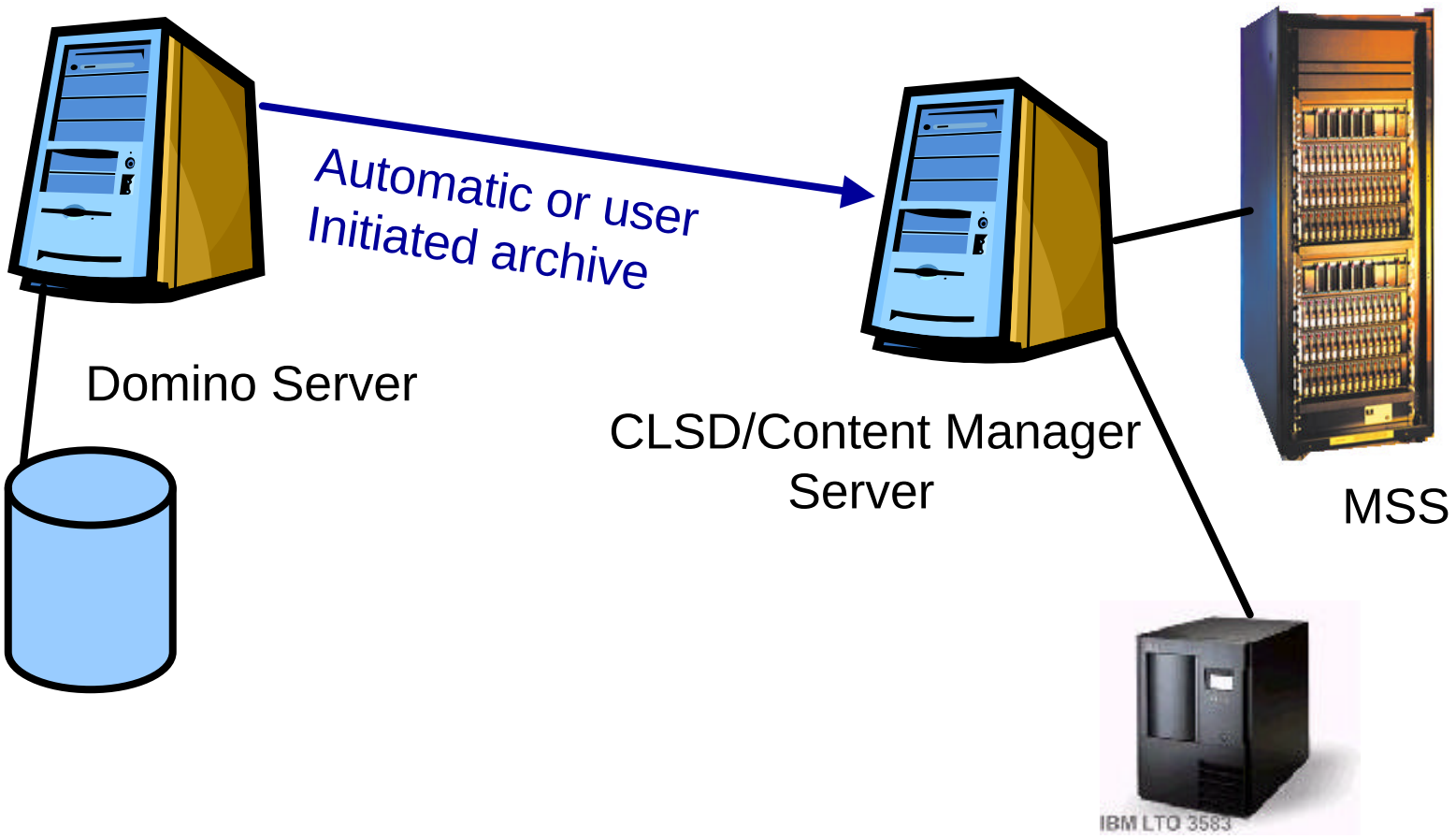
User Sees


Dann Barrick - Inbox - Lotus Notes
File Edit View Create Actions Help
Welcome
Computer Inventory - CISD\All-...
Robert Rosen - Inbox
Dann Barrick - Inbox
notes
Mail
New Memo
Reply
Forward
Delete
Folder
Copy into
Tools
CommonStore
Inbox
Drafts
Sent
All Docume
Archive
Trash
Discussior
Rules
Stationery
ACPM
ADPSYSTE
APG AMM
APG Perso
appreciatio
ATAAPS
BOA
BRAC
COL
Dann Barrick
Who
Date
Size
A
Subject
Deke Smith 04/23/2001 5,303 Re:
Cal Friederich 04/23/2001 3,123 Re:
Cal Friederich 04/23/2001 47,182 Re: DPAS ARL Accounts
Dawn Golab 04/22/2001 165,054 CoS Response - WSMR & APG EEC Supervisors & Team Leaders (Mand 2001
Cindy Jones 04/21/2001 36,626 Hand Receipt Holders with CISD
Kelly Stratchko 04/20/2001 877 Use of Gov't. Van
Lori Denney 04/20/2001 2,503 Re: Dr. Whalin's Car Phone-Succes
Linnea Fransen 04/20/2001 2,387 Dr. Whalin's Car Phone-Success?
Donna Blankenship 04/20/2001 9,475 Re: Policy on Issuance of Blanket Tr
Cal Friederich 04/20/2001 1,128 Phone Call
Wayne Reynolds 04/20/2001 7,812 Training on Rapiscan
Donna Bumba 04/20/2001 4,020 DEMIL CERTIFIERS & VERIFIERS
Randi Vachon 04/20/2001 1,114 ALC Special Dispatch--Earth Day C
ARL-Havena

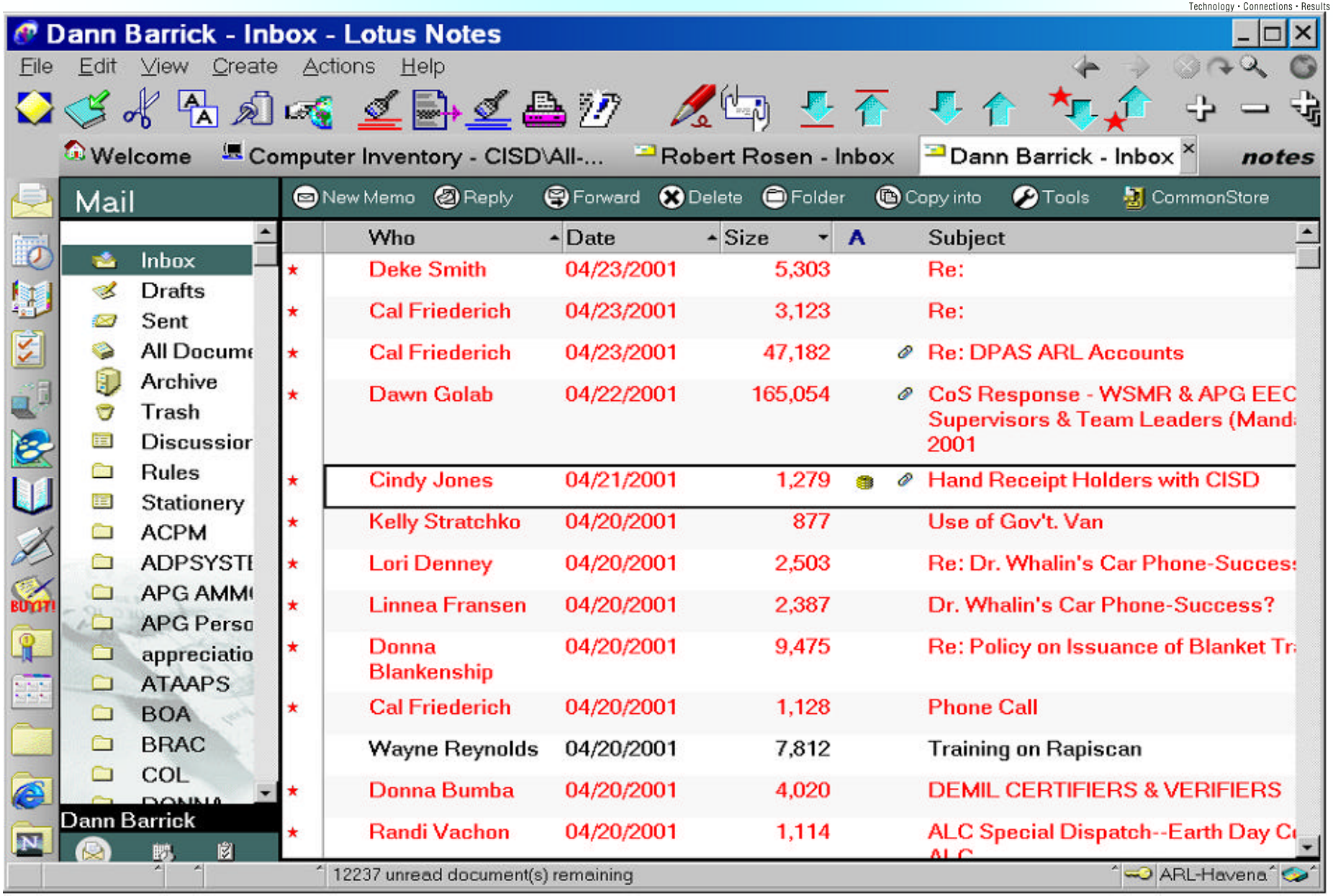
User's E-Mail Message



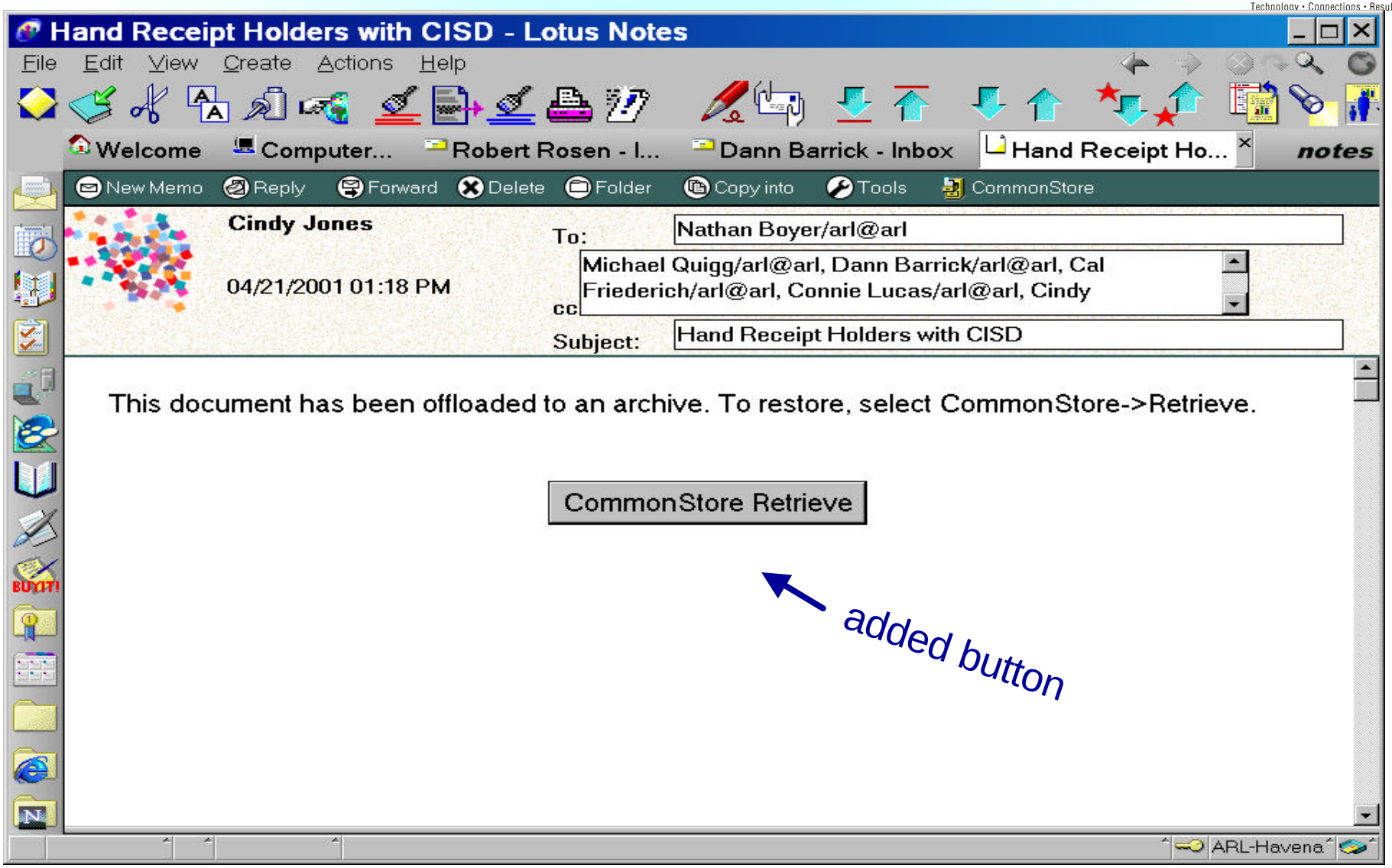
CSLD - Archiving



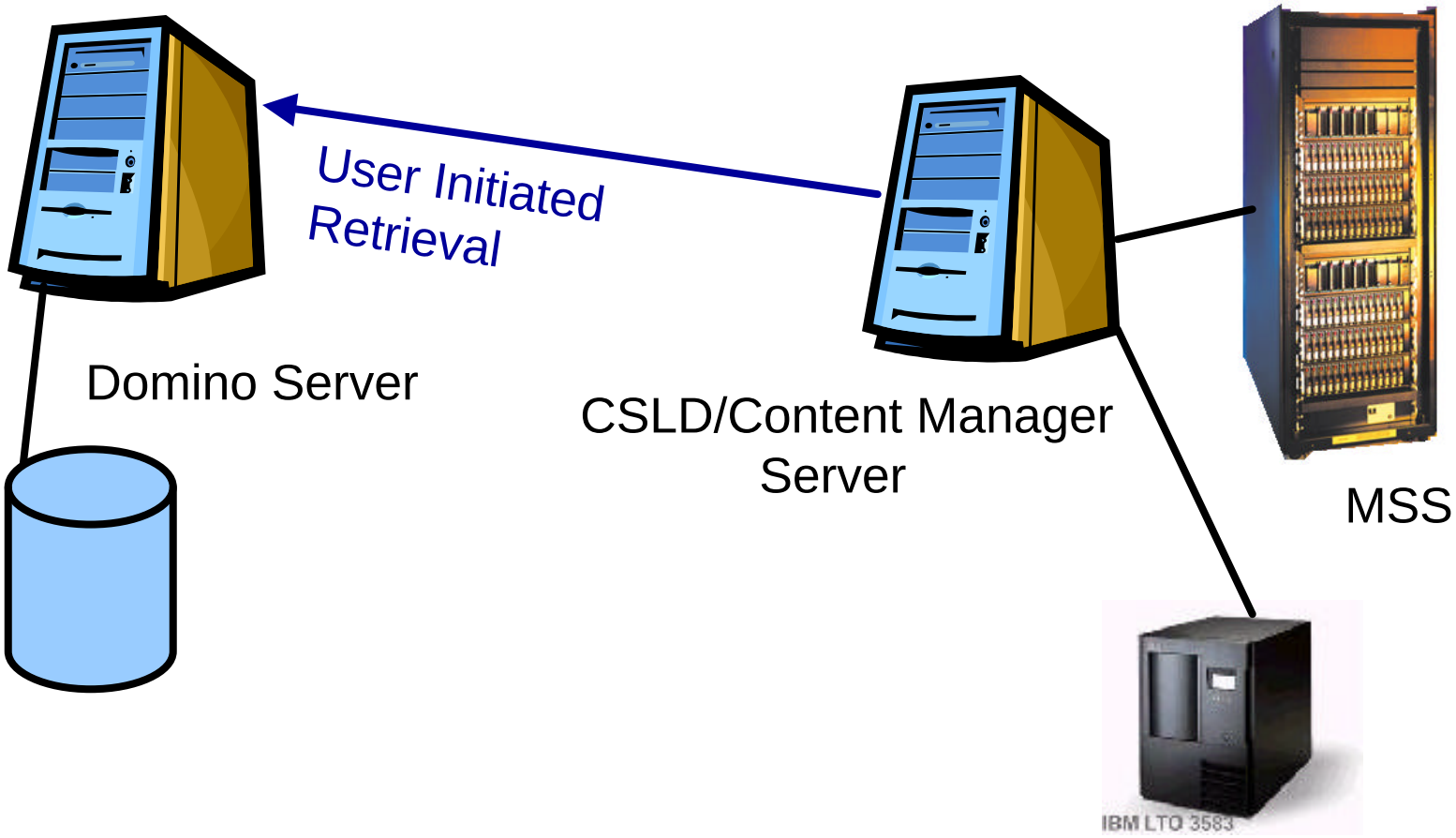
User Now Sees



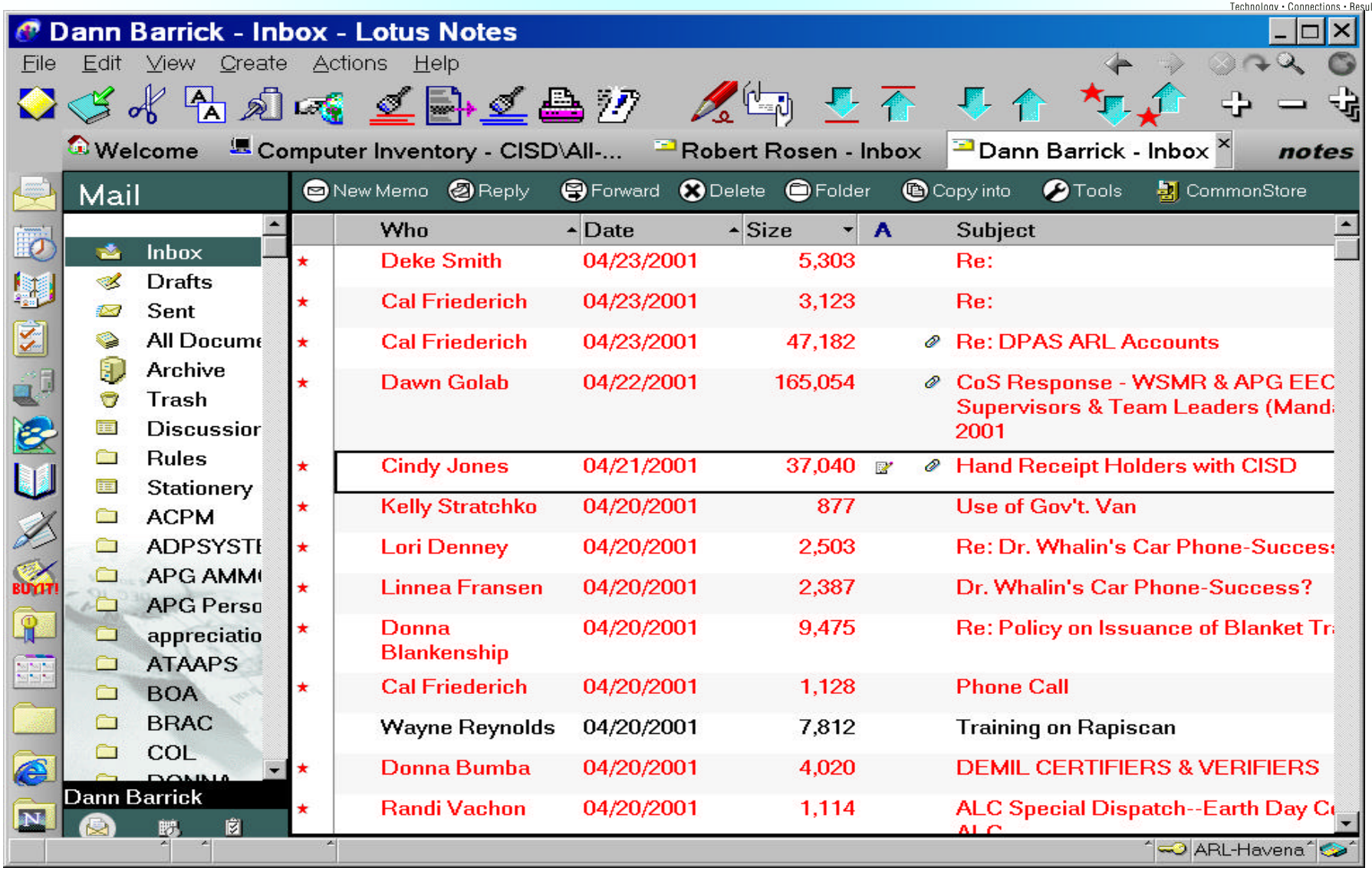
And the E-Mail Looks Like



CSLD - Retrieval

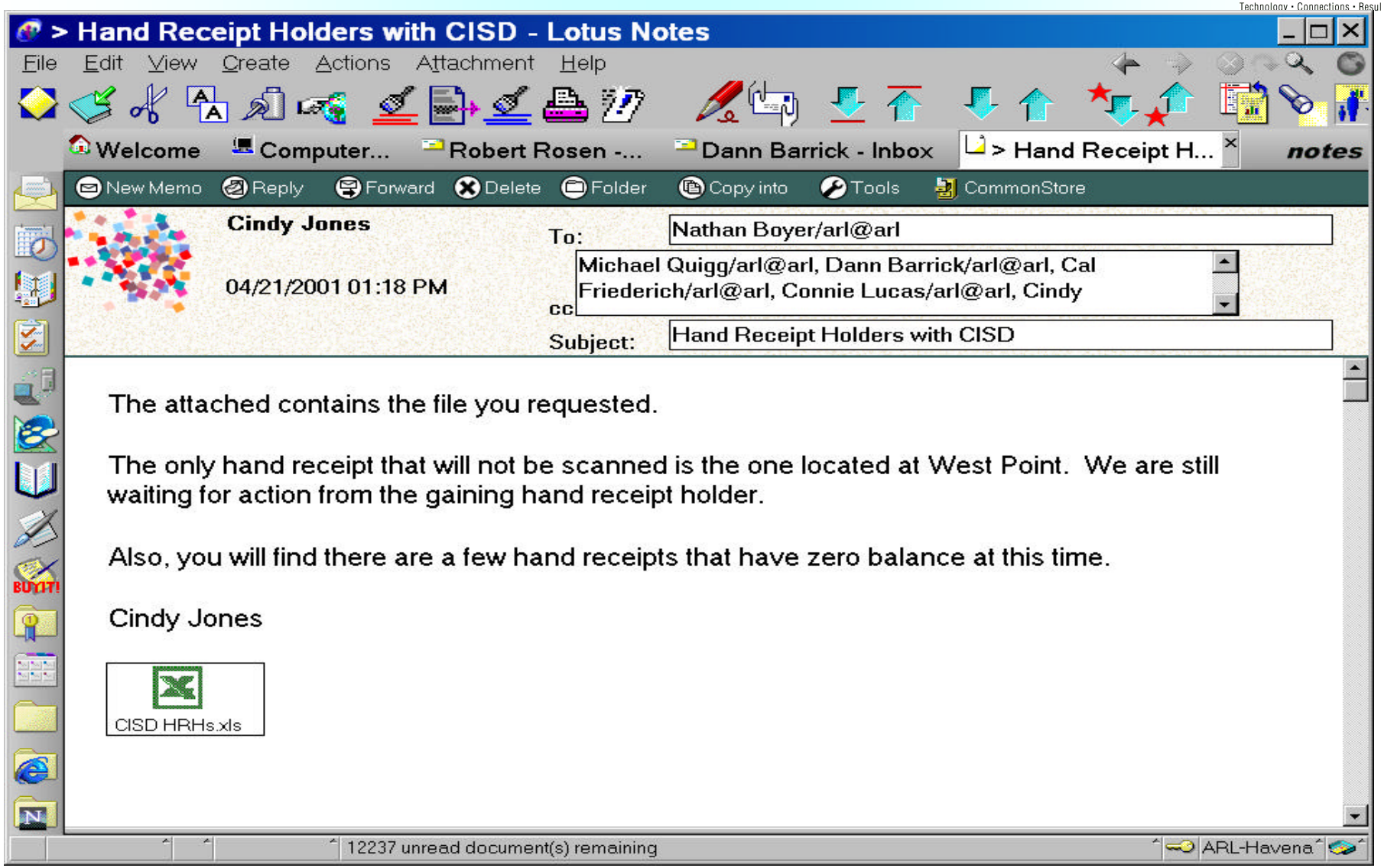


User Inbox Now Looks Like



User's E-Mail Restored

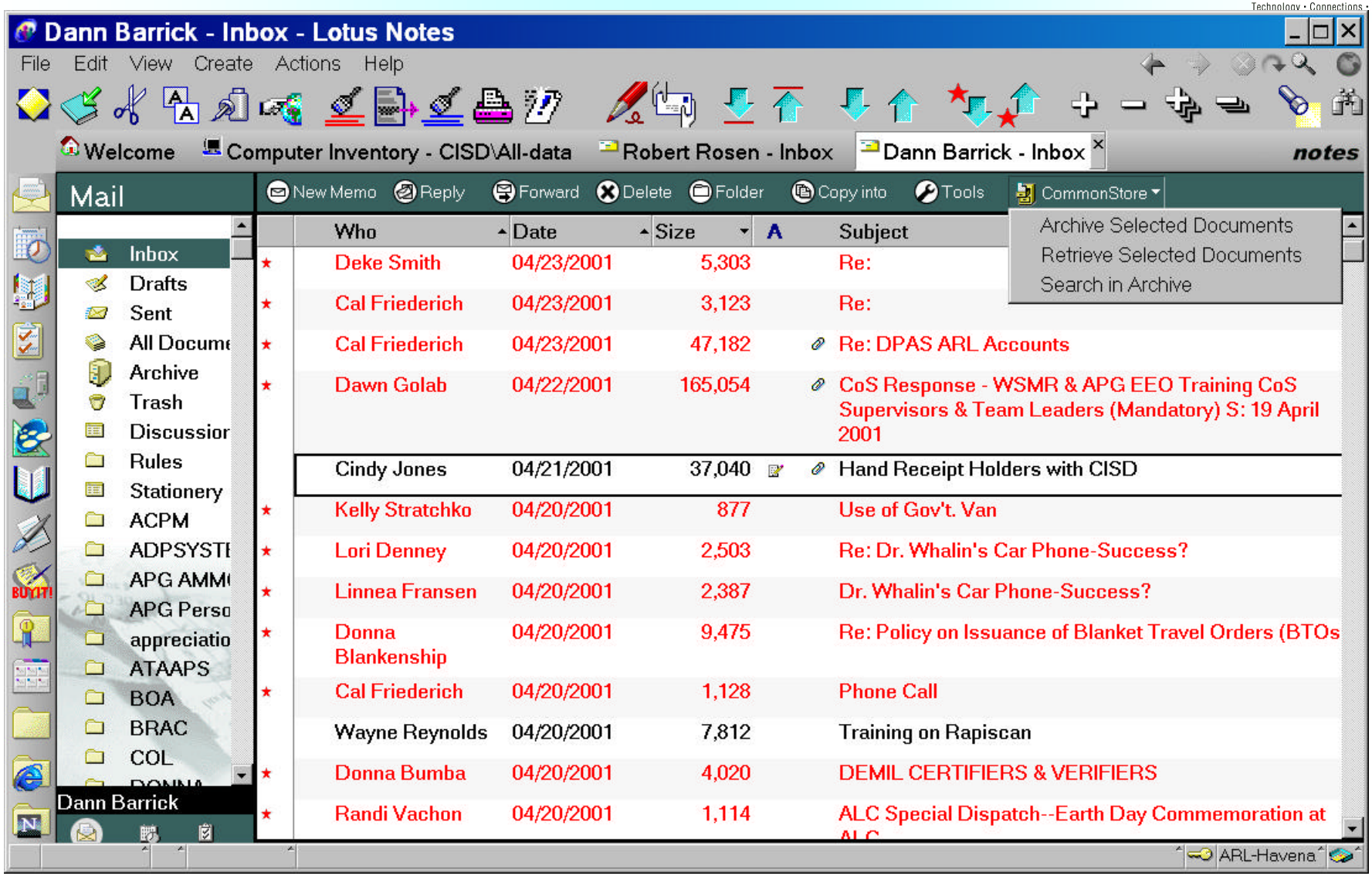
(for two weeks)



Common Store Options

NIAMS

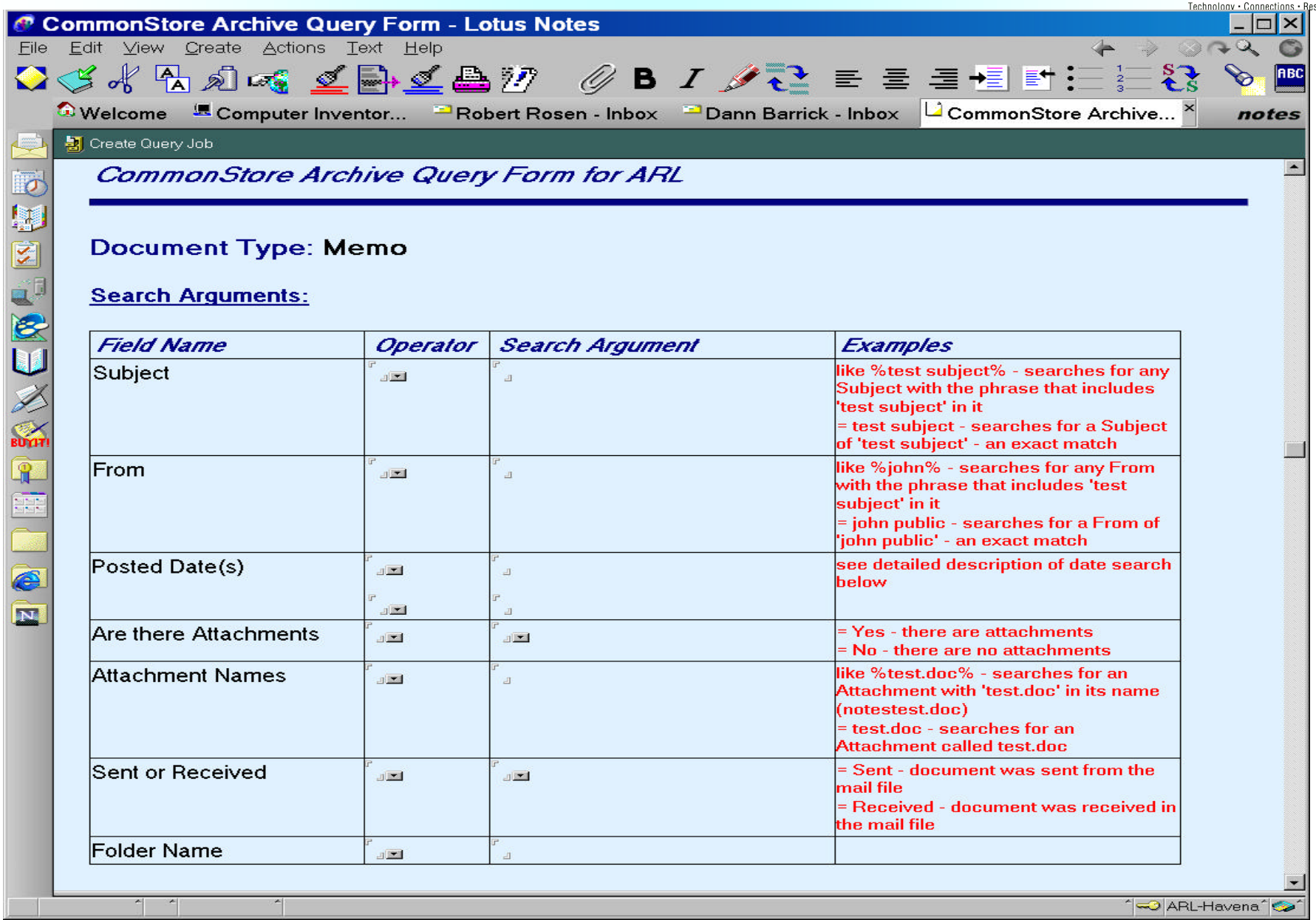
SHARE
Technology · Connections · Results



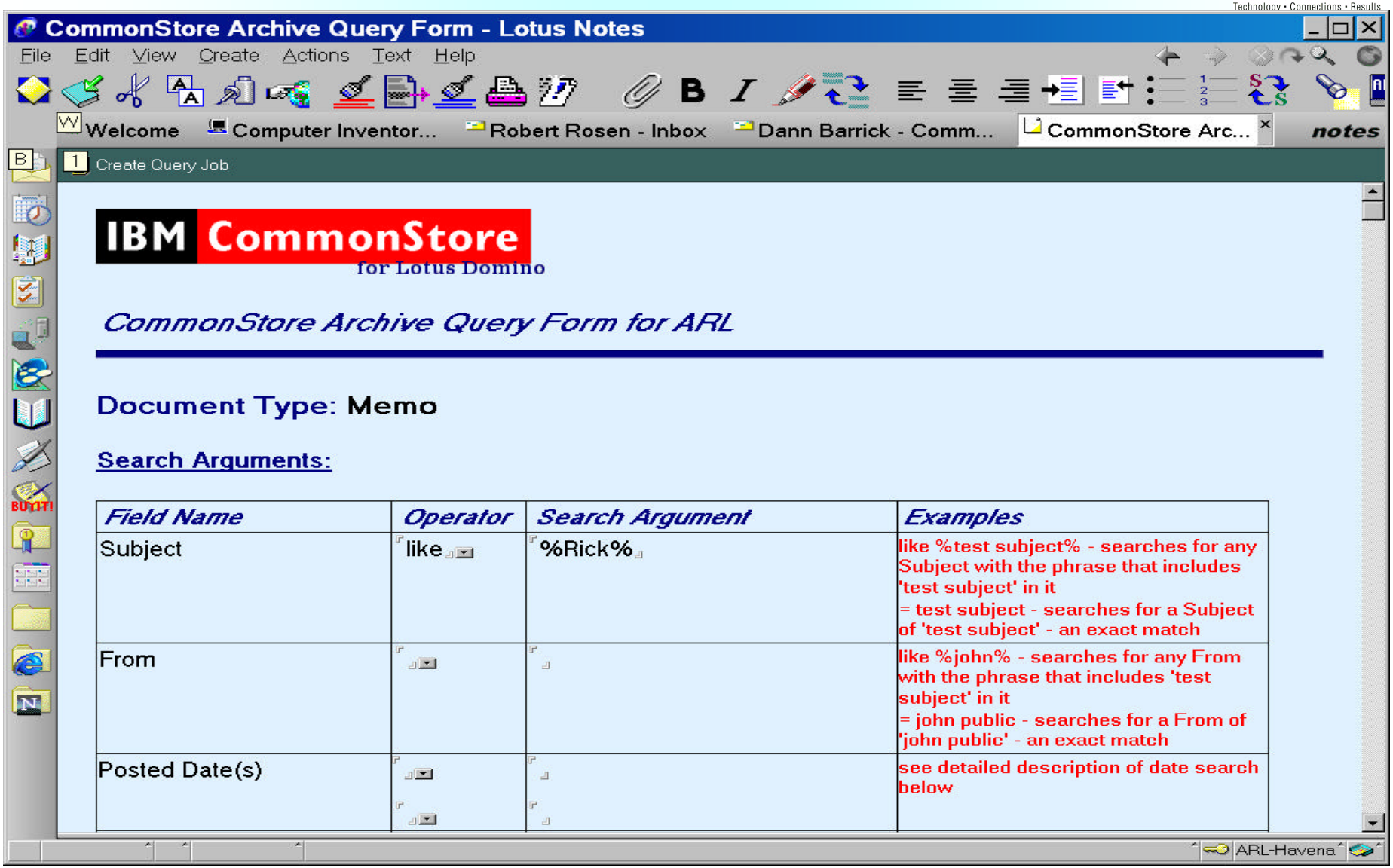
Archive Searches

- Oops, I deleted my stub but need the document.
- I have too much stuff in my in-box to find it.
- etc.

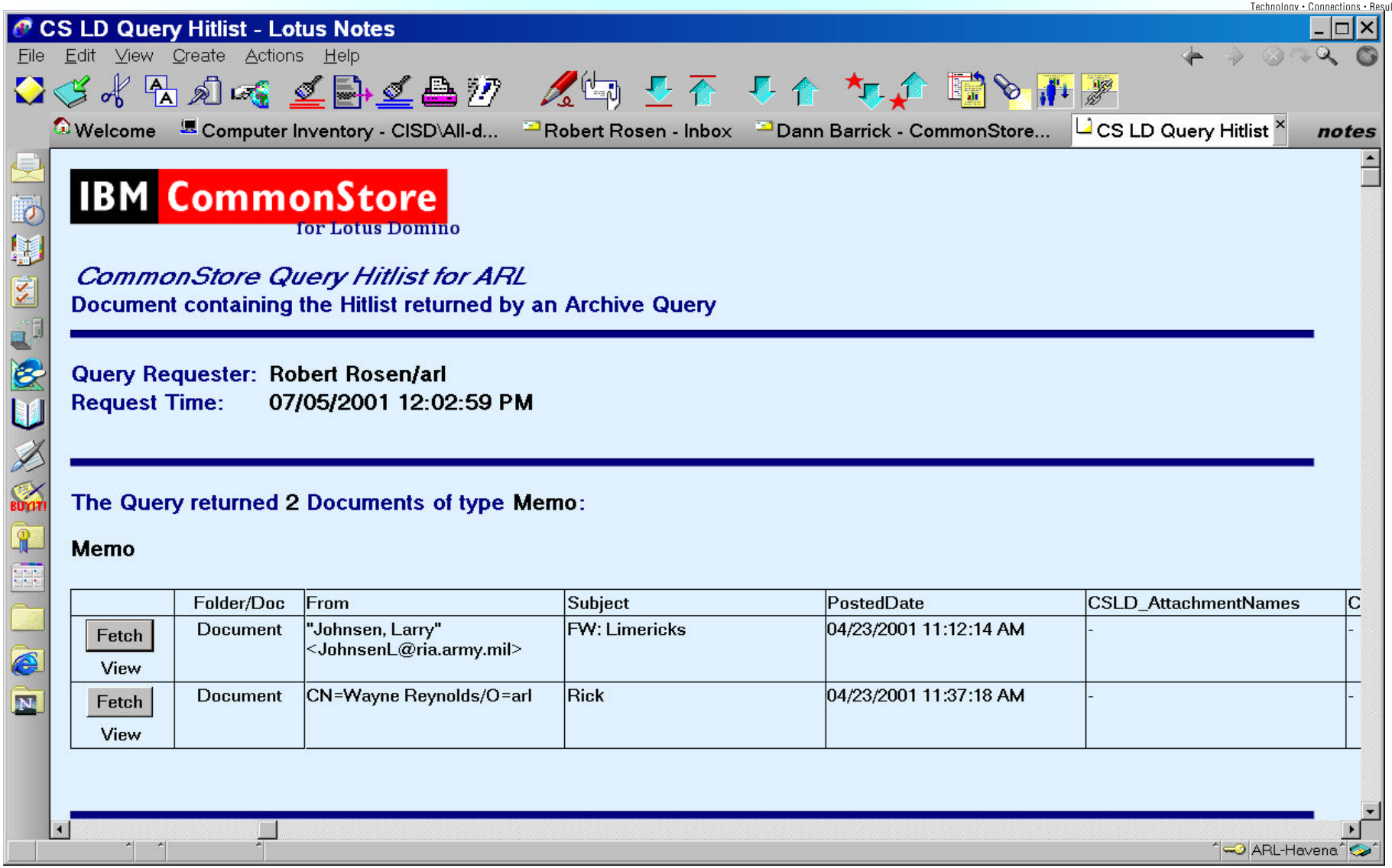
Query Form


CommonStore Archive Query Form - Lotus Notes
File Edit View Create Actions Text Help
Welcome
Computer Inventor...
Robert Rosen - Inbox
Dann Barrick - Inbox
CommonStore Archive...
notes
Create Query Job
CommonStore Archive Query Form for ARL
Document Type: Memo
Search Arguments:
Field Name
Operator
Search Argument
Examples
Subject
like %test subject% - searches for any Subject with the phrase that includes 'test subject' in it
= test subject - searches for a Subject of 'test subject' - an exact match
From
like %john% - searches for any From with the phrase that includes 'test subject' in it
= john public - searches for a From of 'john public' - an exact match
Posted Date(s)
see detailed description of date search below
Are there Attachments
= Yes - there are attachments
= No - there are no attachments
Attachment Names
like %test.doc% - searches for an Attachment with 'test.doc' in its name (notestest.doc)
= test.doc - searches for an Attachment called test.doc
Sent or Received
= Sent - document was sent from the mail file
= Received - document was received in the mail file
Folder Name
ARL-Havena

Query Form


CommonStore Archive Query Form - Lotus Notes
File Edit View Create Actions Text Help
Welcome
Computer Inventor...
Robert Rosen - Inbox
Dann Barrick - Comm...
CommonStore Arc...
notes
Create Query Job
IBM CommonStore
for Lotus Domino
CommonStore Archive Query Form for ARL
Document Type: Memo
Search Arguments:
Field Name
Operator
Search Argument
Examples
Subject
like
%Rick%
like %test subject% - searches for any Subject with the phrase that includes 'test subject' in it
= test subject - searches for a Subject of 'test subject' - an exact match
From
like %john% - searches for any From with the phrase that includes 'test subject' in it
= john public - searches for a From of 'john public' - an exact match
Posted Date(s)
see detailed description of date search below
ARL-Havena

Search Return



What's Done

- **Mail template extensively modified**
- **Nightly archive and stub agents**
  - **Run manually until active mail contains only 90 days**
  - **Agents will then revert to nightly schedule**
- **Hid the folder archive function**
  - **Conflict on restores if folder contained previously archived documents**
  - **Not available in R4**

Performance

- Good Question
  - Archive
    - Single document, <5 seconds
    - 163,000 documents in 9.5 hours without restub agent (when DB2 table is optimized)
  - Retrieve
- Single document, <10 seconds
    - From tape, ~90 seconds

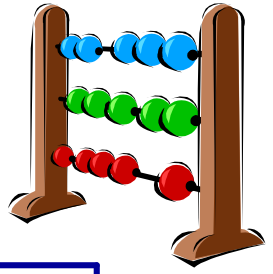
Just gone live, no extensive data

Performance
(continued)

NIAMS

SHARE
Technology • Connections • Results

- **Statistics to date**



Job Run	Docs older than	Number of Databases	Docs Archived	Archive Time
1	5 yrs	222	31,771	~ 2 hrs
2	4 yrs	397	77,580	~ 4 hrs
3	3 yrs	927	276,793	~ 12 hrs
4*	2 ½ yrs	968	350,915	~ 14 hrs
5	2 yrs	1,547	957,339	~ 336 hrs**
TOTAL			1,976,398	

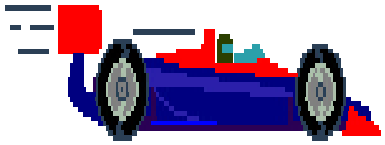
* Job did not complete; but had to be stopped for disk maintenance

** Excessive time caused by errors encountered in archiving docs

Performance

- Function of network speed
- Domino server speed
- Content Manager server speed

✓ Know a site that archived >200,000 documents in 6 hours (Content Manager was on a mainframe)



Issues We Encountered

- Second archive request submitted on same document before first request completes causes stub to be archived, erasing link to original document
- Search hit lists are not deleted from user mail files automatically
- CSLD executable can be held by a corrupt process which stops processing and requires an orderly shutdown and reboot
- CSLD tasks seem to drop or hang when they encounter network interruption
- DB2 table needs to be optimized frequently
- DB2 drops one of its NT Services once or twice a month

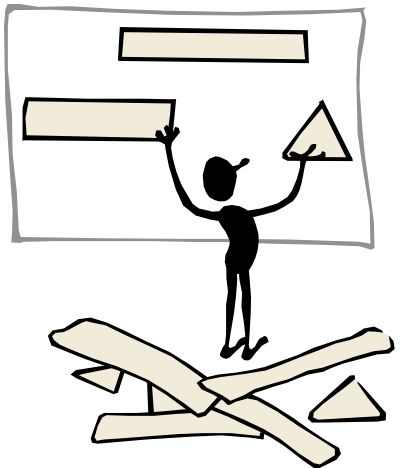
Things We Did Right

- **Extensive testing using a copy of live data**
  - **This flushed out problems such as:**
    - **A bug in the compiler which affected the CSLD code, caused a huge memory leak, and slowed processing down to a crawl.**
    - **A bug in the Notes API toolkit which caused us to be unable to generate a hit list on a search function.**
    - **Application errors caused by carriage returns in fields not set for rich text and forwarded documents that contained the original form code.**
    - **Initial field lengths to be indexed were too small which caused many documents to not be archived (e.g., the whole message was in the subject field).**

- **Separated the stubbing agent from the archiving agent**
  - **Doubled the document archive speed**

Things We Might Do Differently

- Place CSLD and Content Manager on separate servers
- Ensure contract language provides firm deadlines
- Ensure latest versions of software are provided
- NOT be the first one to do it!

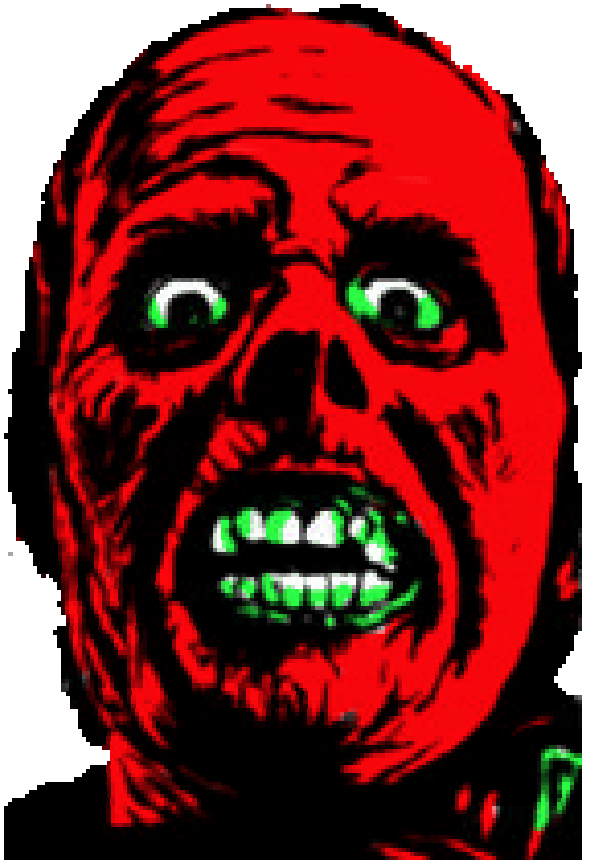


NIAMS

The Dark Side

SHARE
Technology · Connections · Results

- Product knowledge in IBM is rare
- Performance data is non-existent
- Experts are non-US citizens (security concerns)



http://www.redbooks.ibm.com/pubs/pdfs/redbooks/sg246405.pdf

Are We Satisfied

- Too soon to tell
- Concept is great
- Testing shows it does what we need
- Customer acceptance ?
  - Change is unpopular
  - Loss of “infinite” e-mail storage
  - Check back at next SHARE

NIAMS

Questions

SHARE
Technology • Connections • Results