



Capabilities Briefing

CMWG
Herndon, Virginia
March 11, 2009

PTFS has Evolved Significantly Since 1995



1995

- Founded company
- 4 employees
- Focus on desktop imaging, Integrated Library Systems, and custom search systems



2000

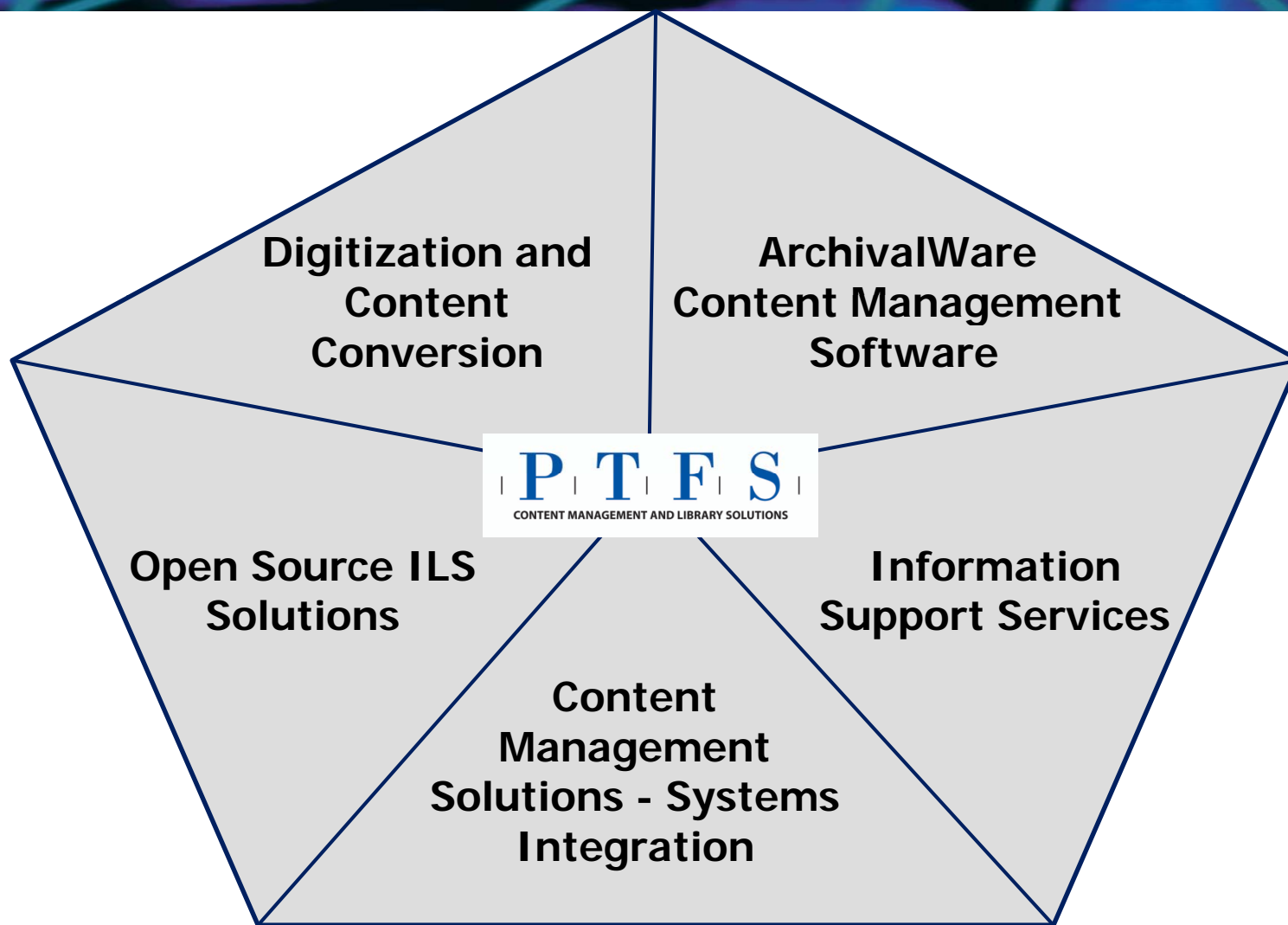
- Large Commercial project: Chicago Tribune
- First release of ArchivalWare
- Started commercial digitization service bureau



2009

- TS Facility Clearance, SCI staff, accredited IS conversion, safeguarding at Secret level
- ArchivalWare 4.2 released
- 70th ArchivalWare installation
- Launched ArchivalWare online discussion forum
- 120 employees

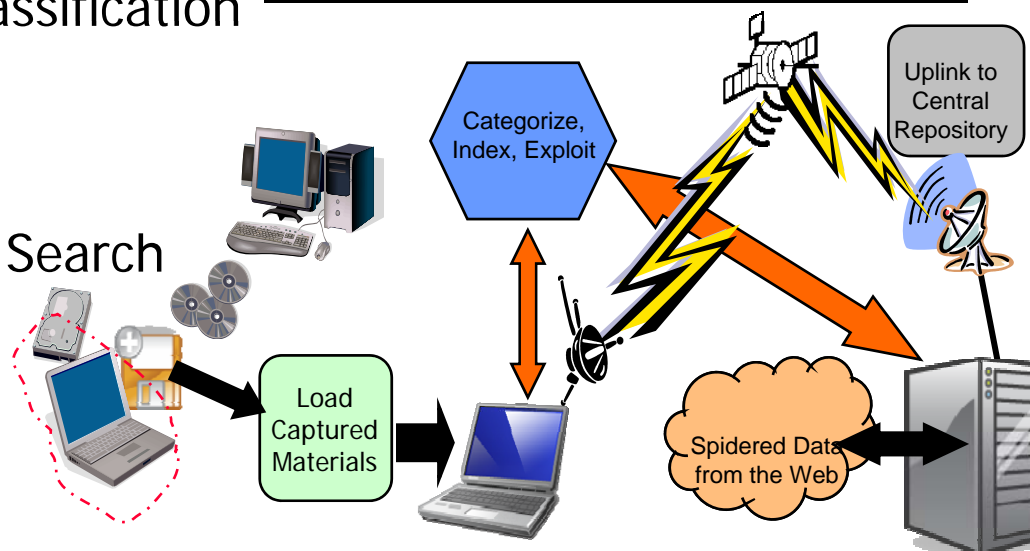
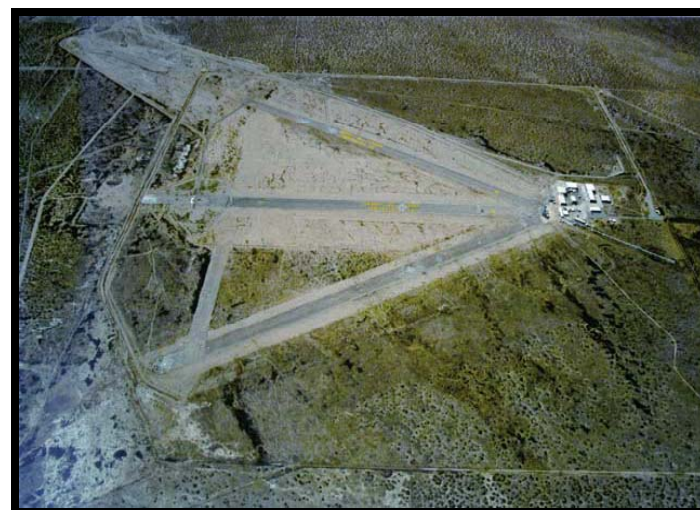
PTFS Focuses on Core Competencies



Partial Client List

- Libraries – Federal Government
 - Marines MWR
 - National Library of Medicine
 - Naval Research Laboratory
 - Department of Justice
 - U.S. Marine Corps (16 locations World Wide)
 - U.S. Attorney's Office
 - Department of State
 - Library of Congress
 - International Trade Commission Library
 - U.S. Army, Army Heritage Education Center
 - Department of Labor
 - Manitoba Legislative Library
 - National Library of Education
- Libraries – State Government
 - Maryland State Law Library
 - New Hampshire State Library
 - Wyoming State Library
 - Utah State Library
 - Library of Virginia
- Libraries – Public
 - Whittier Public Library
 - Independence Township Library
 - City of San Diego
 - City of Orange Public Library
- Libraries - Academic
 - Southern Oregon University
 - Embry Riddle Aeronautical University
 - National Defense University
 - Suffolk County Community College
 - Bryn Athyn College
 - Eastern Virginia Medical School / American College of Surgeons
 - Carnegie Mellon University
 - U.S. Air Force Academy
- Government Institutions – non-library
 - National Archives and Records Administration
 - Department of Interior (Labatt)
 - National Labor Relations Board (NLRB)
 - Department of Defense (Joint Chiefs)
 - National Security Agency
 - National Institutes of Health
 - Missile Defense Agency/Northrop Grumman
 - Commodity Futures Trading Commission
 - U.S. Army Medical Research Infectious Disease
 - Office Secretary of Defense (OSD)
 - National Geospatial Intelligence Agency (NGA)
 - Defense Threat Reduction Agency (DTRA)/L-3

- ArchivalWare Geospatial Functionality
 - Combine all data types; Imagery, digital video, documents
 - DIB integration
 - Integration with Google Earth
- Redaction Work-flow
 - Identity Protection, FOIA, Declassification
- Arabic Language Synthetic Intelligence Network
 - English to Arabic Cross-Lingual Search
- Open Source Data Collection
 - Spidering for doc collection
- DOMEX



ArchivalWare™ Offers a Robust Digital Archiving Solution

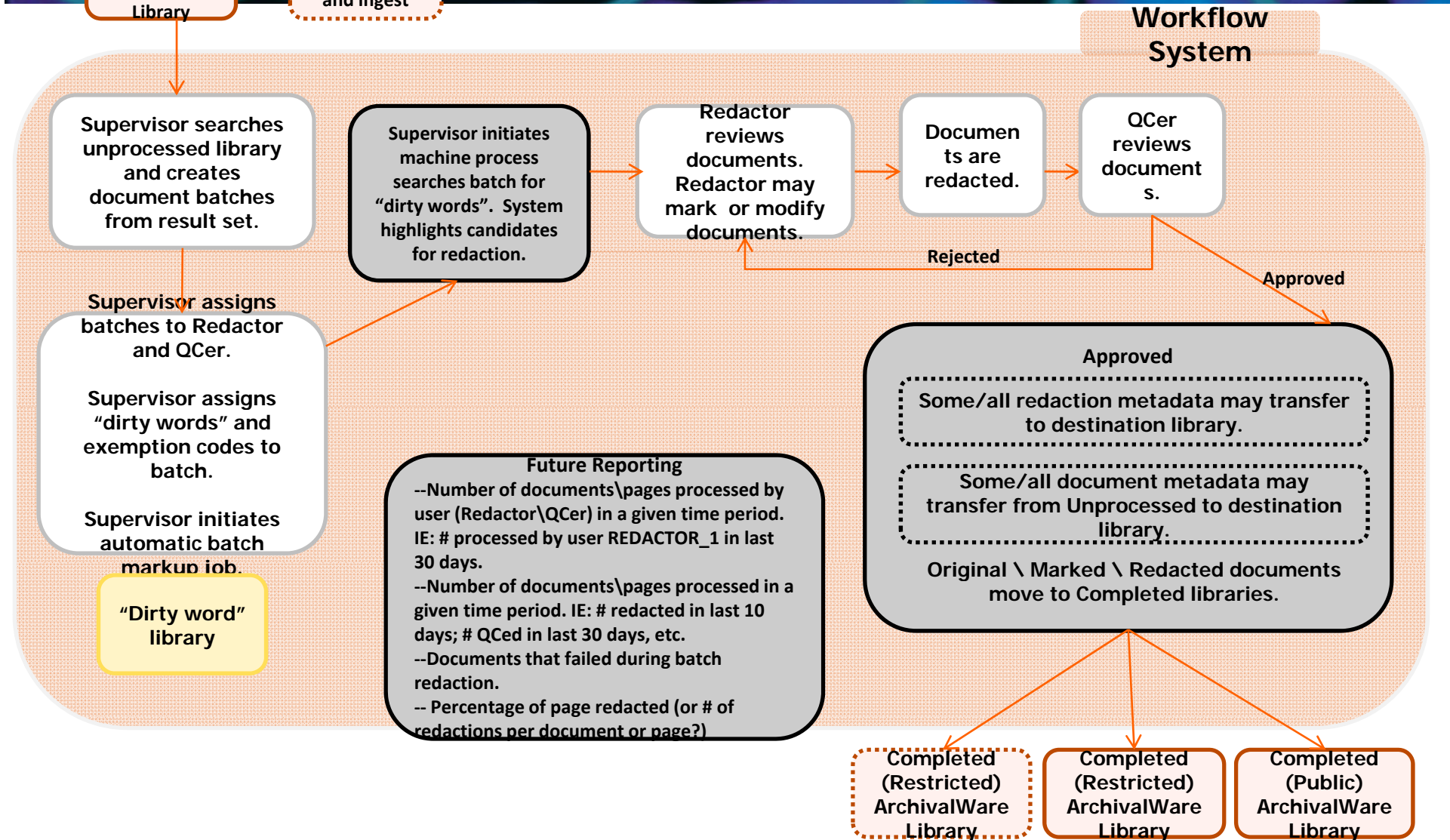
- Store, search, retrieve, manage: collections, users
 - Web-base
 - Full text search and retrieval system
 - Tools for building, managing & integrating digital archives
- User friendly, powerful search functionality: supports diverse users
 - Searchers with no searching experience
 - Power Searchers
- Supports three types of searching
 - Boolean, Concept, Pattern
- Sophisticated language tools

PTFS has Flexible Approach to Projects

- Complete entire project at PTFS Bethesda facility
 - PTFS develops complete digital object: archive quality image, derivative image (s) and metadata record
- Setup digitization facilities at customer's site, mobile facilities
 - PTFS selects hardware, software and processes.
 - PTFS provides training
 - PTFS or customer supply staff
- Hybrid
 - Duties are split between PTFS and customer: Document Preparation, Scanning, OCR, Creating Derivatives, Metadata input

Declassification

ArchivalWare Declassification Workflow



Declassification: Step 1 View and Assign Batches

- Group documents into batches
- Assign batches to analyst
- Maintain “dirty word” file

The screenshot shows the ArchivalWare web application running in a Windows Internet Explorer browser. The browser's address bar displays the URL: `http://10.1.1.17:8180/workflow/welcomeUser.seam?cid=123`. The page features the PTFS Digital Archiving Solutions logo and a navigation menu on the left. The main content area includes a description of ArchivalWare, a section for new features in version 4.0, and a list of capabilities.

ArchivalWare - Windows Internet Explorer provided by PTFS Inc.

http://10.1.1.17:8180/workflow/welcomeUser.seam?cid=123

ArchivalWare

P T F S
DIGITAL ARCHIVING SOLUTIONS

ArchivalWare is a content and digital asset management system developed specifically to allow users to store, access, and manage all digital archive collections within one system. ArchivalWare, in its latest release 3.0, combines a suite of high volume digital capture and production tools with a sophisticated web-based retrieval engine and advanced administrative features.

Check out what's new in ArchivalWare 4.0!

- Browse functionality allows you to create a rich hierarchical structure to organize your document archive.
- An administrative GUI enables simple system administration of management components including fields, search, browse, import, export and indexing, which customizes the user interface.
- Application Programming with J2EE components provide a complete API that facilitates easy integration of enterprise Interface (API) applications.

ArchivalWare also has these capabilities:

- Allows researchers to search the metadata associated with digital objects as well as the full text of a large quantity of full-text document formats.
- Provides a powerful relevance ranking capability that makes sure that information you are most interested in is shown first at the top of the "hit list".
- Highlights search hits documents that have been saved in PDF format.
- Utilizes a sophisticated pattern (fuzzy text), natural language (concept), and Boolean search engines that find search results with pinpoint accuracy.
- Is a web-based solution that includes a "Web Metadata Editor" and a digital object import utility.
- Supports both high volume digitization and OCR solution as well as light desktop scanning applications.
- Source formats can include digital documents, books, news clippings, photographs, video/audio collections, and electronically published materials.

Supervisor Functions

- View Batch
- View Process Definitions
- Synchronize Documents
- View Dirty Words
- View Reports
- Help
- Logout

Search For:

SEARCH

Created Date

Creator

Folder

Series

Source

Title

Resource Type

Declassification: Step 2 Recommend Redactions, Revise Redactions

- Select Document
- Assign Redactor & QCer
- Select markup candidates
- Assign Codes
- Proceed to redact mark-up words

The screenshot displays the ArchivalWare Redactor software interface. On the left is a sidebar with the ArchivalWare logo and a menu titled 'Redactor Functions' containing options: View Batch, View Redacted Documents, View Reports, Help, and Logout. The main window shows a document titled 'Haig Tells Of Theories On Erasure' by George Lardner Jr., a Washington Post Staff Writer. The document text includes 'White House' and 'erased', both of which are highlighted with red rectangular boxes. A redaction code '(b)(1)' is visible in the top right corner of the document area. The interface also features a toolbar with various icons for editing and navigation.

Declassification: Step 3 Finalize Redactions, QC

- Redact Document
- Assign to QCer
- QCer has read only privileges
- QCer approves document if satisfied
- Reassigns if unacceptable

The screenshot displays the PTF S Content Management and Library Solutions interface. On the left, a sidebar menu titled 'Redactor Functions' includes options: 'View Batch', 'View Redacted Documents', 'View Reports', 'Help', and 'Logout'. The main content area shows a document titled 'Haig Tells Of Theories' by George Lardner Jr., a Washington Post Staff Writer. The document text is partially redacted with black boxes labeled '(b)(1)'. The visible text includes: 'Haig Tells Of Theories On (b)(1)', 'By George Lardner Jr. Washington Post Staff Writer', '(b)(1) (b)(1) chief of staff Alexander M. Haig Jr. said yesterday he and (b)(1) (b)(1) lawyers had discussed fears that "some sinister force" (b)(1) one of President Nixon's subpoenaed Watergate tapes. Testifying in federal court here. Haig told of what he

Open Source Intelligence

■ Challenge

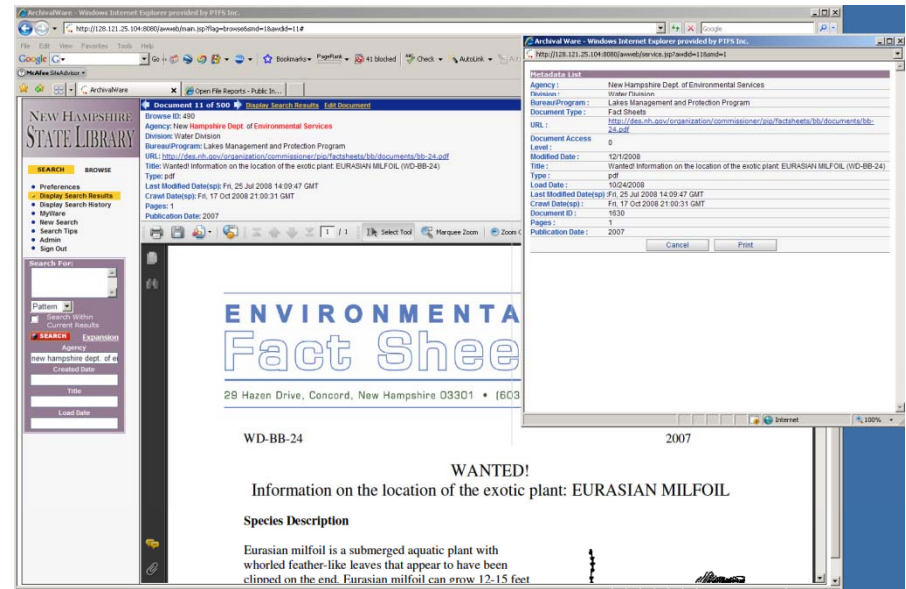
- Establish repository for all state publications

■ Scope

- Configure ArchivalWare in ASP environment
- Develop Spider capabilities to routinely scrape all state agency websites
- Load existing documents, routinely refresh repository

■ Status

- Site went live in August 2008



NHSL Step 1: Add URL's

Spider Administration

File Spider Service Crawl Job Help

localhost:2112

NHSL_CrawlJob

Configuration Management

General Traversal Filters Forms Authentication Connection Store Schedule Logging Environment

Seed URLs

http://des.nh.gov/organization/commissioner/pip

Crawling Information

Host Distance: Same Host

URL Distance: Unlimited

URL Redirection: 10

Maximum Documents: Unlimited

☒ Auto Detect Source Character Encoding

Default Source Character Encoding: ISO-8859-1 "Latin-1"

☒ Use Cookies

☐ Use Depth First Crawling

☐ Interpret JavaScript

☐ Normalize Host

☐ Case Sensitive URLs

☐ Prepend WWW

☒ Honor Robot Exclusion Protocol

☐ Honor NOINDEX Tag

☒ Filter for Domino Duplicates

Apply Revert Help

Spider Administration

FileSpider ServiceCrawl JobHelp

localhost:2112

NHSL_CrawlJob

ConfigurationManagement

GeneralTraversalFiltersFormsAuthenticationConnectionStoreScheduleLoggingEnvironment

Crawl FiltersDocument Store Filters

Allow Filters

Name	Type	Regular Expression
NHSL COMMISSIONER	URL	http://des\.\nh\.gov/organization/commissioner/pip

UpDownNewRemove

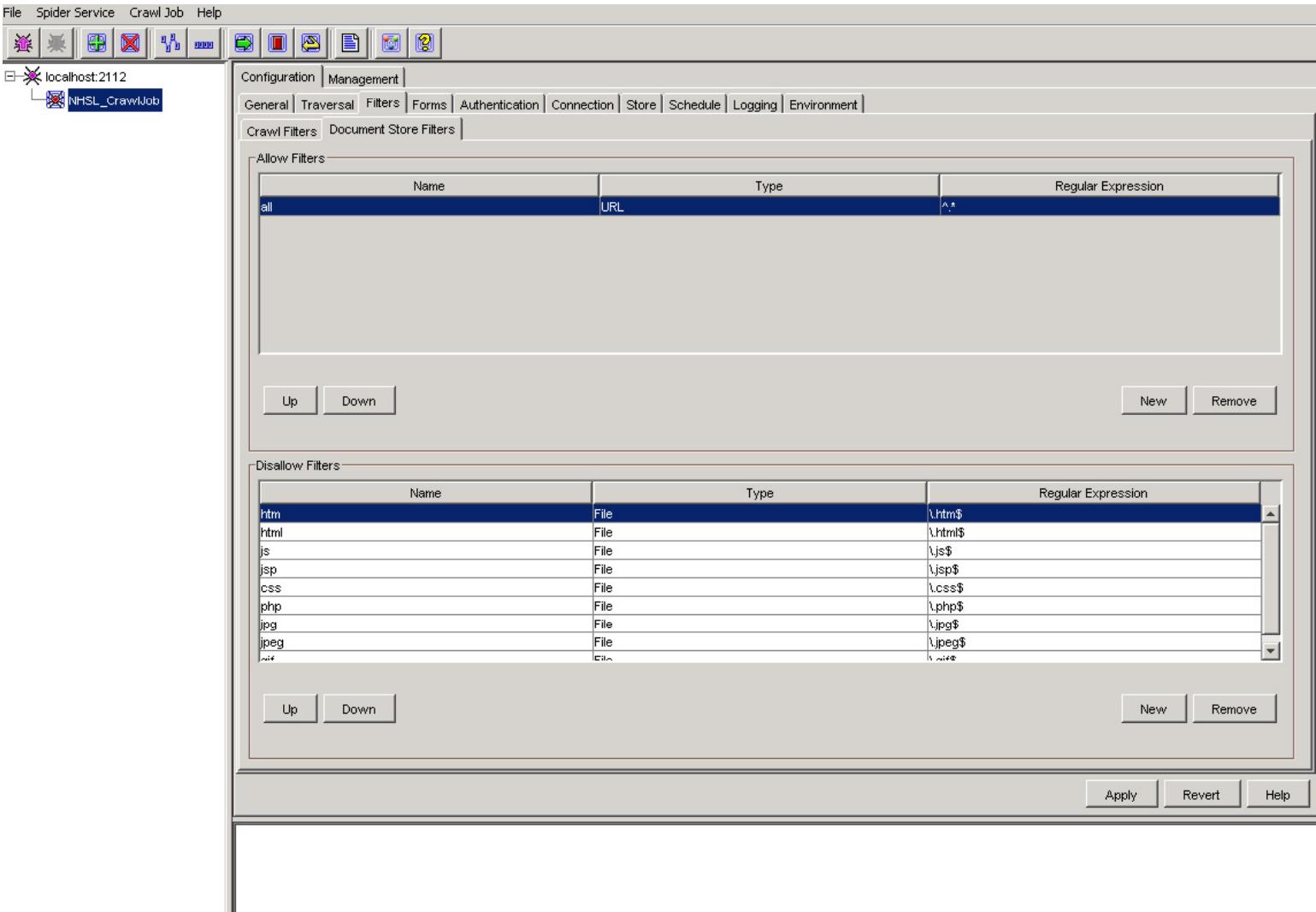
Disallow Filters

Name	Type	Regular Expression
NHSL COMMISSIONER FORMS	URL	http://des\.\nh\.gov/organization/commissioner/pip/forms

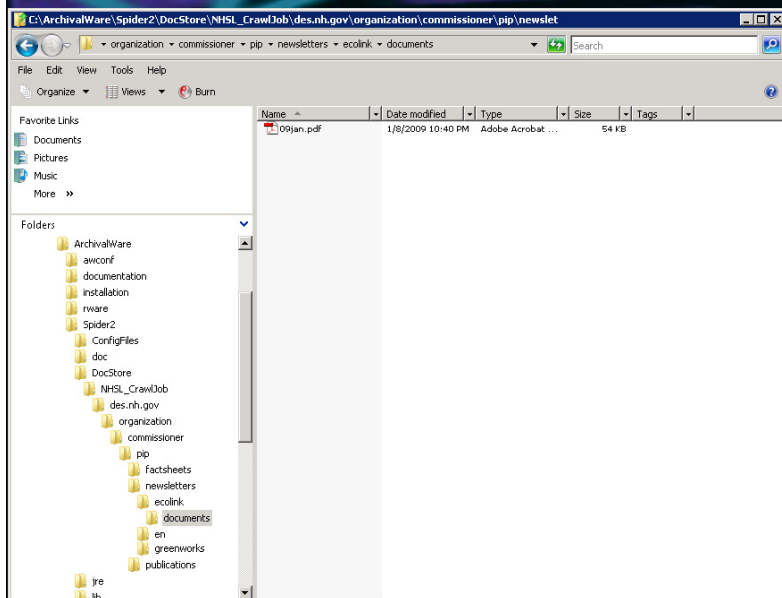
UpDownNewRemove

ApplyRevertHelp

NHSL Step 3: Apply file filters



NHSL Step 4: Spidered content, stats returned



	url_id	first_visited_date	last_visited_date	last_mod_date_header	crawl_id	response_code	status_id	depth	parent_referer_url_id	redirect_url_id
1	10003	2008-10-17 08:01:04.470	2009-01-09 03:24:03.007	Mon, 10 Nov 2008 07:00:00 GMT	14	200	2	0	NULL	10002
2	10427	2008-10-17 08:18:10.203	2009-01-09 03:40:04.663	Tue, 06 Jan 2009 18:30:37 GMT	14	200	2	2	10057	NULL
3	10799	2008-10-17 08:50:46.323	2009-01-09 04:12:25.943	Tue, 30 Dec 2008 17:10:25 GMT	14	200	2	3	10395	NULL
4	14006	2009-01-09 03:40:21.340	2009-01-09 03:40:21.340	Fri, 02 Jan 2009 16:35:24 GMT	14	200	1	2	10057	NULL
5	14007	2009-01-09 03:40:36.610	2009-01-09 03:40:36.610	Thu, 08 Jan 2009 20:54:00 GMT	14	200	1	3	10113	NULL
6	14008	2009-01-09 03:44:39.387	2009-01-09 03:44:39.387	Wed, 07 Jan 2009 16:48:58 GMT	14	200	1	3	10120	NULL
7	14009	2009-01-09 03:46:06.730	2009-01-09 03:46:06.730	Mon, 22 Dec 2008 20:59:46 GMT	14	200	1	3	10121	NULL
8	14010	2009-01-09 04:09:03.583	2009-01-09 04:09:03.583	Mon, 10 Nov 2008 07:00:00 GMT	14	200	1	3	10394	NULL
9	14011	2009-01-09 04:09:59.243	2009-01-09 04:09:59.243	Fri, 02 Jan 2009 14:18:35 GMT	14	200	1	3	10395	NULL
10	14012	2009-01-09 04:11:50.440	2009-01-09 04:11:50.440	Fri, 02 Jan 2009 14:18:35 GMT	14	200	1	3	10395	NULL

	url_id	field_id	field_value
1	10427	3	Fri, 09 Jan 2009 15:44:21 GMT
2	10427	4	Tue, 06 Jan 2009 18:30:37 GMT
3	10427	5	Is it safe to eat the fish we catch? Mercury & Other Poll...
4	10799	3	Fri, 09 Jan 2009 16:16:42 GMT
5	10799	4	Tue, 30 Dec 2008 17:10:25 GMT
6	10799	5	ARD-21 Controlling VOC Emissions from Industrial Sourc...
7	14006	3	Fri, 09 Jan 2009 15:44:36 GMT
8	14006	4	Fri, 02 Jan 2009 16:35:24 GMT
9	14006	5	A Summary of 2008 Shoreland Protection Act
10	14007	3	Fri, 09 Jan 2009 15:44:53 GMT
11	14007	4	Thu, 08 Jan 2009 20:54:00 GMT
12	14007	5	January 2009
13	14008	3	Fri, 09 Jan 2009 15:48:55 GMT
14	14008	4	Wed, 07 Jan 2009 16:48:58 GMT
15	14008	5	January/February 2009
16	14009	3	Fri, 09 Jan 2009 15:50:23 GMT
17	14009	4	Mon, 22 Dec 2008 20:59:46 GMT
18	14009	5	Recycle This
19	14010	3	Fri, 09 Jan 2009 16:13:20 GMT
20	14010	4	Mon, 10 Nov 2008 07:00:00 GMT
21	14010	5	CO-GE0-9 Automated Streamflow Frequency and Rech...

	url_id	doc_file_name	doc_size	content_type	path_id
1	10427	ard-ehp-25.pdf	50195	application/pdf	3212
2	10799	ard-21.pdf	100255	application/pdf	3212
3	14006	cspa_brochure.pdf	501036	application/pdf	3203
4	14007	09jan.pdf	55287	application/pdf	3213
5	14008	09jan.pdf	571127	application/pdf	3218
6	14009	08dec.pdf	150102	application/pdf	27
7	14010	geo-9.pdf	252103	application/pdf	3228
8	14011	ard-ehp-28.pdf	49305	application/pdf	3212
9	14012	ard-ehp-29.pdf	51072	application/pdf	3212

NHSL Step 5: Process with Spider Synchronizer

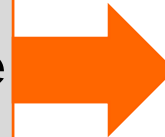
Spider

- Spider Data
 - New/modified
 - File
 - Metadata
 - URL



Spider
Synchronizer

- File
- Metadata
- Browse Structure



ArchivalWare

- Digital Objects
 - File
 - Metadata
 - Browse
 - Descriptive

NHSL Step 6: Review content in ArchivalWare

The screenshot displays the ArchivalWare web interface within a Windows Internet Explorer browser. The main window shows a document titled "ENVIRONMENTAL Fact Sheet" with the following details:

- Document ID:** 490
- Agency:** New Hampshire Dept. of Environmental Services
- Division:** Water Division
- Bureau/Program:** Lakes Management and Protection Program
- URL:** <http://des.nh.gov/organization/commissioner/pip/factsheets/bb/documents/bb-24.pdf>
- Title:** Wanted! Information on the location of the exotic plant: EURASIAN MILFOIL (WD-BB-24)
- Type:** pdf
- Last Modified Date(sp):** Fri, 25 Jul 2008 14:09:47 GMT
- Crawl Date(sp):** Fri, 17 Oct 2008 21:00:31 GMT
- Pages:** 1
- Publication Date:** 2007

The document content is displayed below the header, showing the title "ENVIRONMENTAL Fact Sheet" and the address "29 Hazen Drive, Concord, New Hampshire 03301 • (603) 271-2200". The document is dated "WD-BB-24" and "2007". The main text reads:

WANTED!
Information on the location of the exotic plant: EURASIAN MILFOIL

Species Description
Eurasian milfoil is a submerged aquatic plant with whorled feather-like leaves that appear to have been clipped on the end. Eurasian milfoil can grow 12-15 feet

An inset window on the right side of the browser shows the "Metadata List" for the document, providing additional details such as the document type (Fact Sheets), document ID (1630), and publication date (2007).

Geospatial Capabilities



ArchivalWare Geospatial Capabilities

- Situation

- Imagery (motion and stills) and related documents are stored in many different databases; and viewed with different tools

- Challenge

- Aggregate all source intelligence (imagery and documents) on an area of interest in a single screen; a single view, with a single tool

- The Solution

- Automatically integrate all imagery sources; link metadata, extract sample, display source and common format of imagery; and link geospatial with multi-term search tools

Single click, on a single tool; to retrieve all image, and links to all source intelligence on a geo-time defined area of interest.

Keywords

Boolean

Title

Description

Location

Spatial Criteria

Longitude(X1) 35.22229766845703

Latitude(Y1) 62.731075286865234

Longitude(X2)

Latitude(Y2)

Radius 0

Reset

Peckham

village

U.S.S.R.

237 N.

NAMED ON SON

khino 4

rhine)

1990

(Rev. 3-11-66)

1. [Perkhino, RS \(UFI:-2980114\) -00000037.pdf/Perkhino \(Open Document\)](#) 35.55,59.683333
2. [Perkhino, RS \(UFI:-2980116\) -00000041.pdf/Perkhino \(Open Document\)](#) 35.05,62.55
3. [Per'kho, RS \(UFI:-2980123\) -00000063.pdf/Per'kho \(Open Document\)](#) 37.233333,58.333333
4. [Chkalov Stairs, Nizhniy Novgorod \(Open Document\)](#) 44.009445,56.330059
5. [Perkhloyda, RS \(UFI:-2980120\) -00000053.pdf/Perkhloyda \(Open Document\)](#) 36.85,59.8
6. [Perkhovo, RS \(UFI:-2980125\) -00000071.pdf/Perkhovo \(Open Document\)](#) 33.920833,56.746389
7. [Moscou, RS \(UFI:-2960561\) -00000861.pdf/Moscou \(Open Document\)](#) 37.615556,55.752222
8. [Dmitrievskaya tower, Nizhniy Novgorod \(Open Document\)](#) 44.00566,56.32709
9. [Perkhinskaya, RS \(UFI:-2980118\) -00000045.pdf/Perkhinskaya \(Open Document\)](#) 34.316667,60.666667
10. [Perkhino, RS \(UFI:-2980117\) -00000043.pdf/Perkhino \(Open Document\)](#) 40.117778,60.874167
11. [Perkhovo, RS \(UFI:-2980124\) -00000069.pdf/Perkhovo \(Open Document\)](#) 31.316667,54.616667
12. [Perkhachevo, RS \(UFI:-2980113\) -00000027.pdf/Perkhachevo \(Open Document\)](#) 40.458333,64.476667
13. [Ostrova Perkhlyudy, RS \(UFI:-2980121\) -00000055.pdf/Perkh Islands \(Open Document\)](#) 36.460556,64.327778
14. [Per'kho, RS \(UFI:-2980123\) -00000059.pdf/Per'kho \(Open Document\)](#) 37.233333,58.333333
15. [Perkherevo, RS \(UFI:-2980119\) -00000049.pdf/Perkherevo \(Open Document\)](#) 33.214444,56.143611

State Department Global Real Property Operations

- **Challenge**
 - Manage 38,000 real estate properties in 170 countries and 380 cities
 - Records are paper & digital; in every legal language in the world
 - Records are scattered between Headquarters, Embassies and Consulates
- **Goals**
 - Digitize records from Headquarters, Embassies, Consulates
 - Establish rich metadata architecture, apply to all past & future documents
 - Establish multi-term -cross language search, retrieval, and work flow tools
- **Scope**
 - Scan, OCR, index, & 600K pages of back files; dating back to 18th century
 - Create Standard PDF/A format and XMP Metadata for paper & digital documents
 - Convert any document in any of 250+ legacy formats to PDF/A and XMP metadata

State Department Global Real Property Operations



State Department – OBO/RE Archive & Search System

Example – Concept & Fixed Term Search Capability – English => “Assessment” => Appraisals

Query:
Fixed Term
“La Cumbra”
and Concept
“Assessment”

Search
Concept
“Assessment”
Found As:
“Appraisals”

The screenshot displays a web-based search interface. On the left is a sidebar with the State Department seal and a navigation menu. The main content area shows a search result for a cable text. The search criteria entered are "La Cumbra" and "Assessment". The result is a cable text dated February 27, 2002, titled "UNCLASSIFIED TELEGRAM". The subject line is "REQUEST FOR FUNDS FOR APPRAISALS". The body of the cable text describes a request for funds for appraisals of two government properties, "LA CUMBRA" and "LA PIEDRA".

Page 1 of 1
Cable Text:
UNCLASSIFIED
TELEGRAM
February 27, 2002

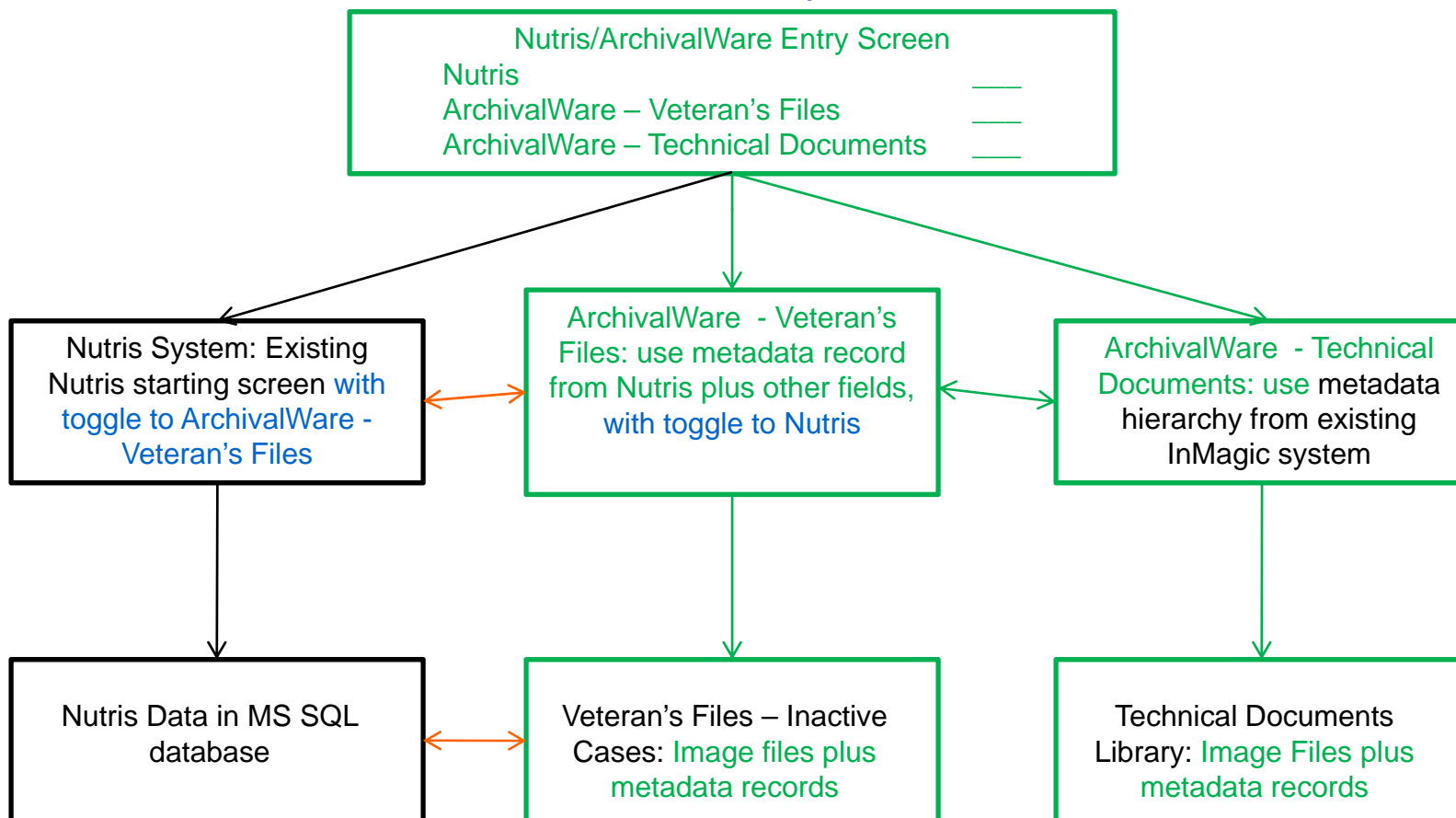
To: SECSTATE WASHDC - PRIORITY
Action: FGO
From: AMEMBASSY BOGOTA (BOGOTA 1913 PRIORITY)
TAGE: AELD, AMST, ASBC, KSEP
Captions: None
Subject: REQUEST FOR FUNDS FOR APPRAISALS
Ref: 01 SECSTATE 157807

PRE OBO/RE REQUEST, AMEMBASSY BOGOTA CONTRACTED TO HAVE 2
APPRAISALS DONE FOR TWO GOVERNMENT PROPERTIES IDENTIFIED AS:
LA CUMBRA (OBU #401221) AND LA PIEDRA (OBU #401222). THE
APPRAISALS WERE COMPLETED ON JANUARY 4TH 2002. OBO RECEIVED
THE WRITTEN REPORTS IN ENGLISH ON JANUARY 21ST 2002. THE
TOTAL AMOUNT FOR THE SERVICES IS \$500,000. PLEASE PROVIDE
FUND SITE.

PATTERSON

Other Case Studies

Veteran's System



Black: In Production today

Green: Phase I

Blue: Phase II

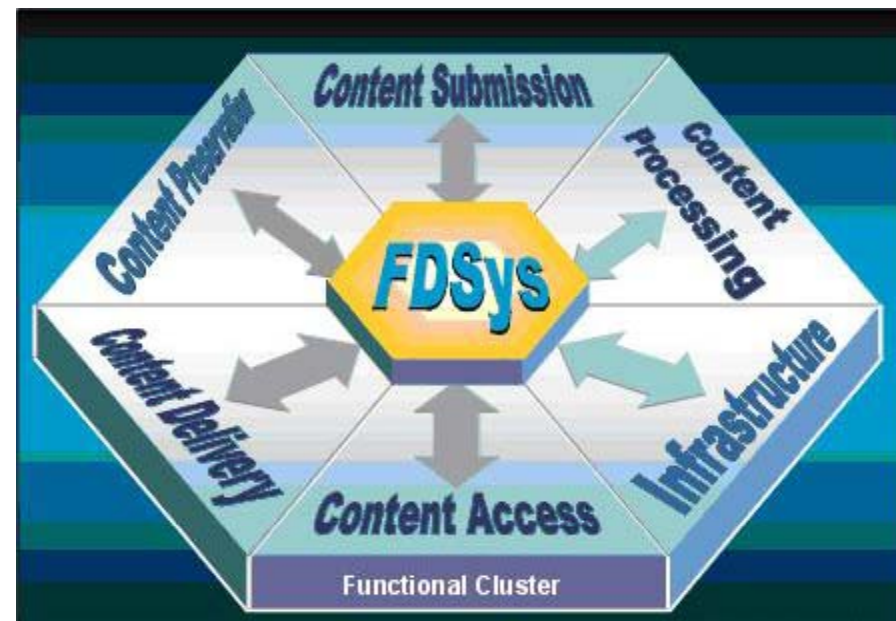
- ExLib – Explosive Library Network
 - The Situation: Data on conventional explosives, ordnance & mines - and IEDs is scattered throughout DoD and IC.
 - The Challenge: USG enlisted, contractors and coalition need quick, complete and accurate identification, analysis and operations information.
 - The Solution: ExLib Network Library aggregates digital data; in multiple text, image, and technical formats; integrates metadata; and provides network based notifications for help with identification, analysis, and de-arming and removal.

Lengthy IED report response time for can be reduced to seconds; and full content can be made available, complete and accurate.

The local, theater, and command can be notified immediately of the situation, analysis, response and disposition.

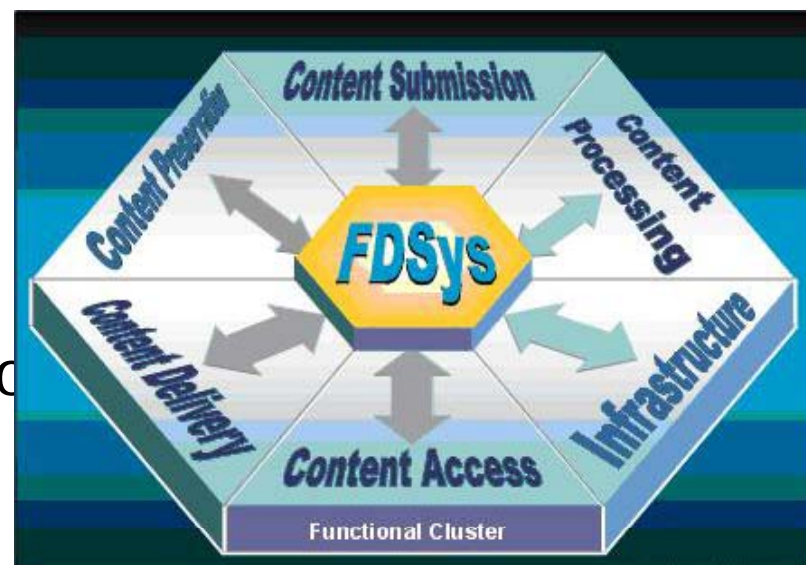
GPO FDSys Project

- One of the largest Government Content Management Systems ever built
- GPO will become repository of record for all published Federal documents
 - Over 120 million digitized government publications, terabytes of born-digital documents
- \$29MM, 3-year project
- PTFS is responsible for web development, infrastructure support (100+ servers) and system testing



GPO FDsys Project

- FDsys beta deployed 1/2/09
- 8 document collections from GPO Access migrated from existing repositories. 09 plans include migrating remaining 45 collections to include the Congressional Record.
- Ongoing work to include support of congressional and other government agency direct document submission functionality
- PTFS is responsible for web development, infrastructure support (100+ servers) and system testing



National Geospatial Intelligence Agency (NGA)

Background

- Office of Inspector General (OIG)
- Audit and Investigation divisions files
- 540,000 pages.
- Classified & unclassified

Requirement

- Turn-key solution; hardware, software, digital conversion, training, support.
- Convert hardcopy to a digital searchable format.
- Develop metadata records for search and browse
- Deliver the digital collection in an archiving application that allows personnel to store, search, retrieve and manage the collection.

Solution

- Shipped material to PTFS' secure facility
- Digitized the classified documents following processes certified for classified documents.
- Built server and installed ArchivalWare, PTFS' application to store, search, retrieve, browse and manage a collection
- Configured the software, loaded data and tested
- Train NGA personnel at PTFS' facility in Bethesda.

