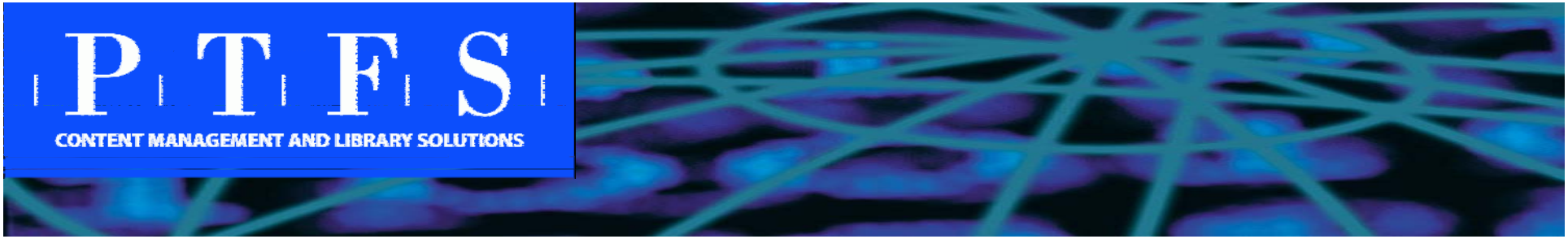


P T F S

CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Capabilities Briefing

CMWG
Herndon, Virginia
March 11, 2009

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

PTFS has Evolved Significantly Since 1995

1995

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

2000

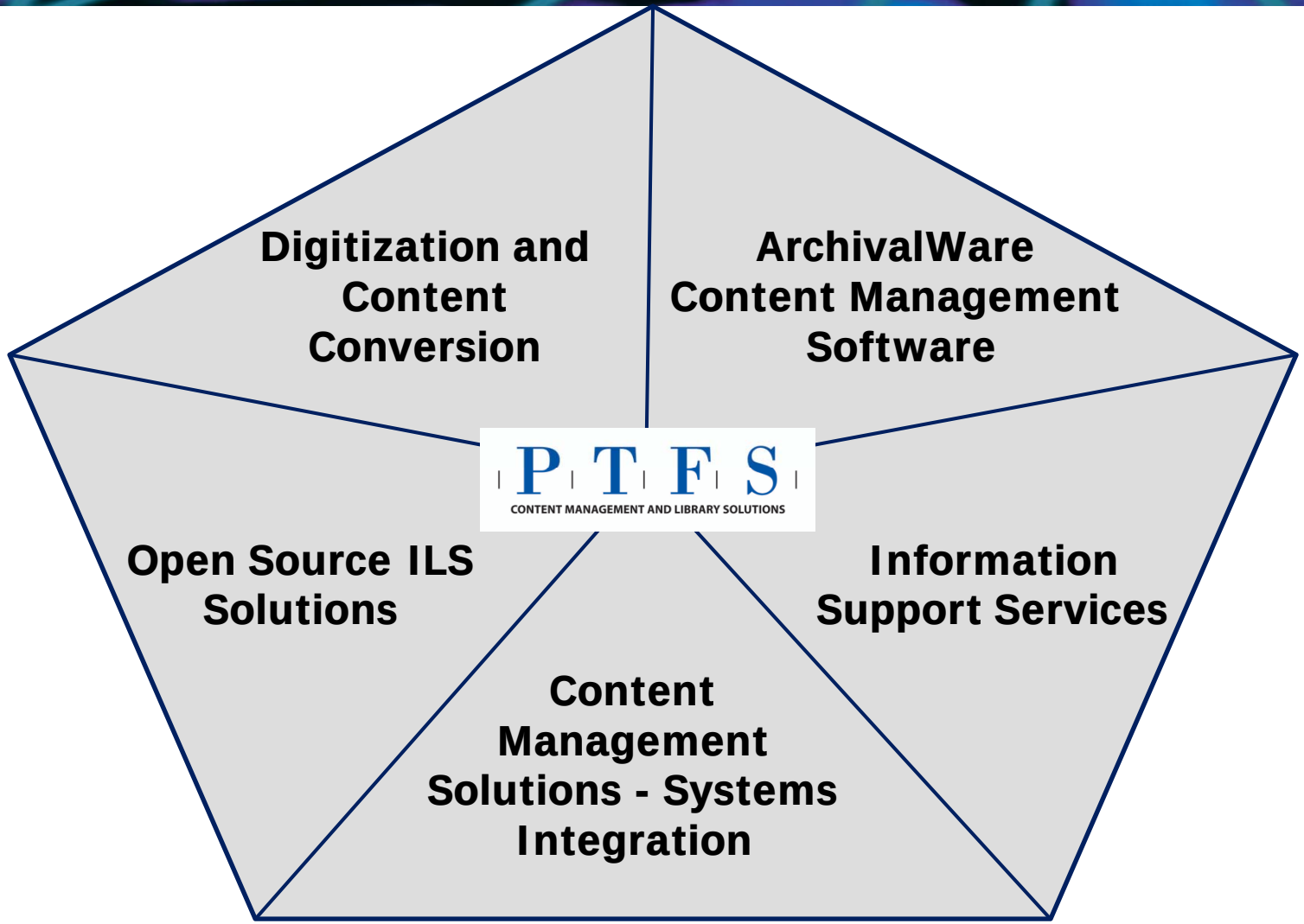
- Large Commercial project: Chicago Tribune
- First release of ArchivalWare
- Started commercial digitization service bureau

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
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

PTFS Focuses on Core Competencies



PTFS
CONTENT MANAGEMENT SOLUTIONS

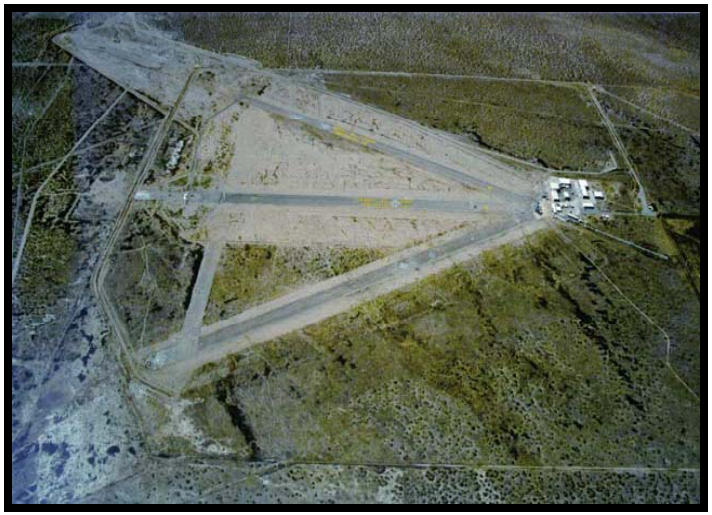
Partial Client List

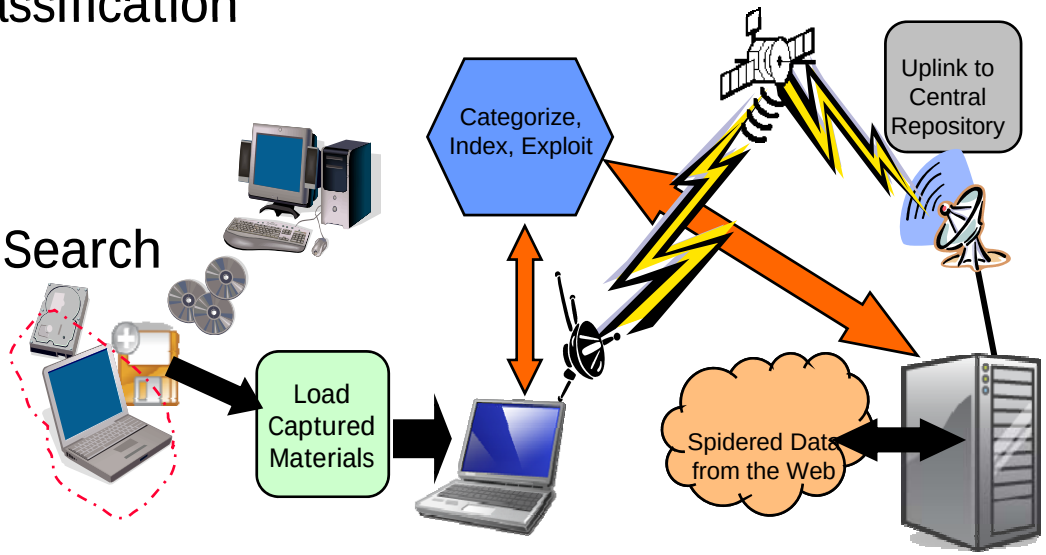
- Libraries – Federal Government
  - Marines MWR
  - National Library of Medicine
  - Naval Research Laboratory
  - Department of Justice
  - U.S. Marine Corps (16 locations World Wide)
  - U.S. Attorney's Office
  - Department of State
  - Library of Congress
  - International Trade Commission Library
  - 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

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Current R&D Development

- 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

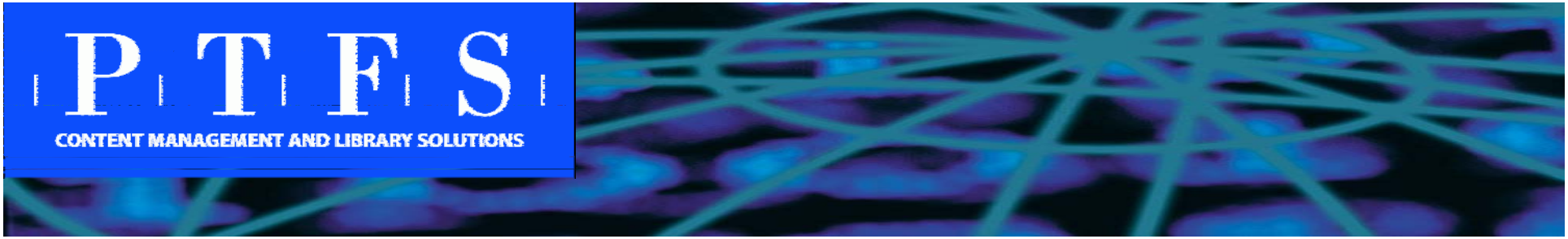
- Store, search, retrieve, manage: collections, users
  - Web-base
  - Full text search and retrieval system
  - Tools for building, managing & integrating digital archives
- User friendly, powerful search functionality: supports diverse users
  - Searchers with no searching experience
  - Power Searchers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

PTFS Confidential

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

PTFS has Flexible Approach to Projects

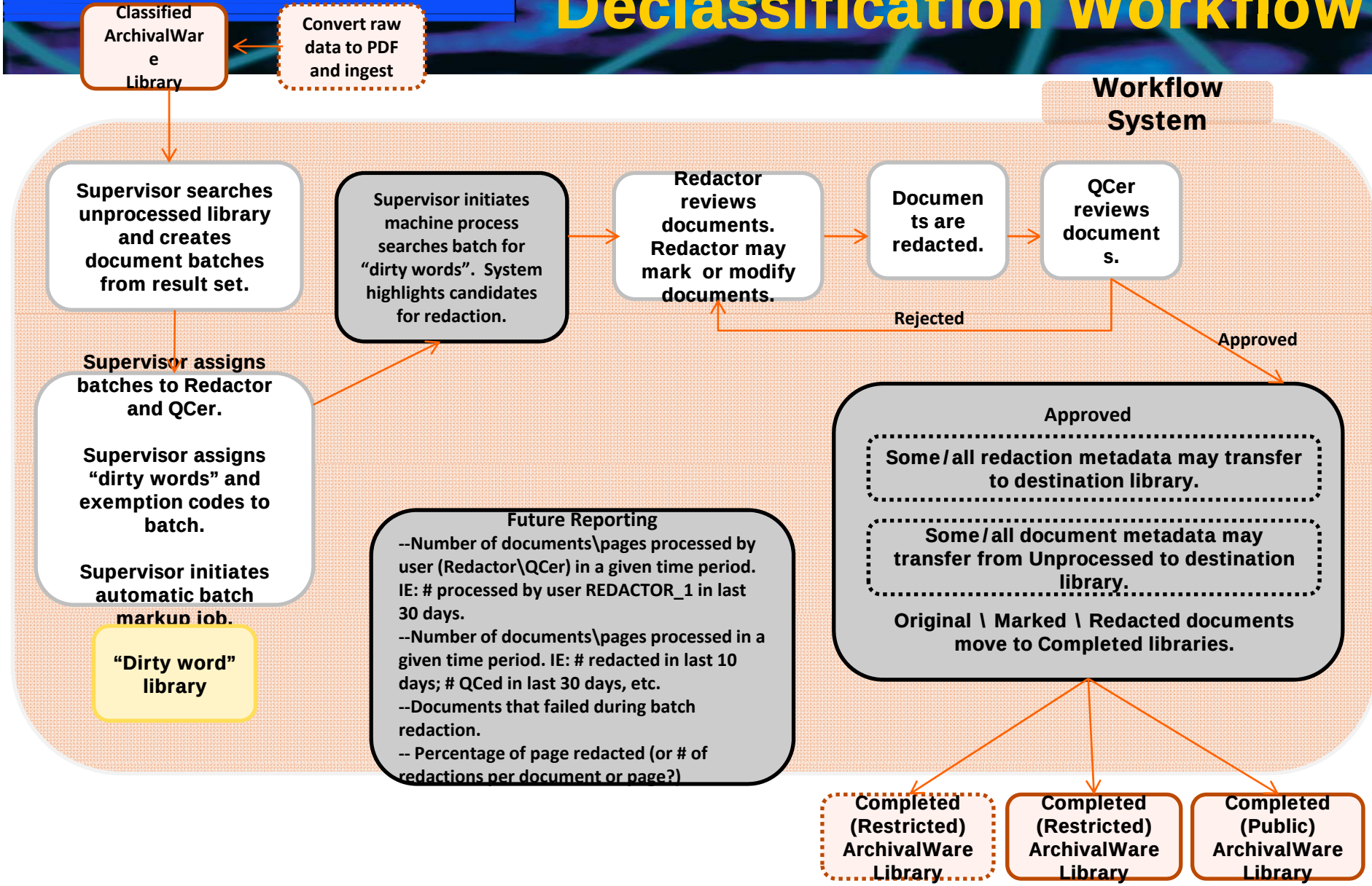
- Complete entire project at PTFS Bethesda facility
  - PTFS develops complete digital object: archive quality image, derivative image (s) and metadata record
- Setup digitization facilities at customer's site, mobile facilities
  - PTFS selects hardware, software and processes.
  - PTFS provides training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



Declassification

P T F S
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

ArchivalWare Declassification Workflow

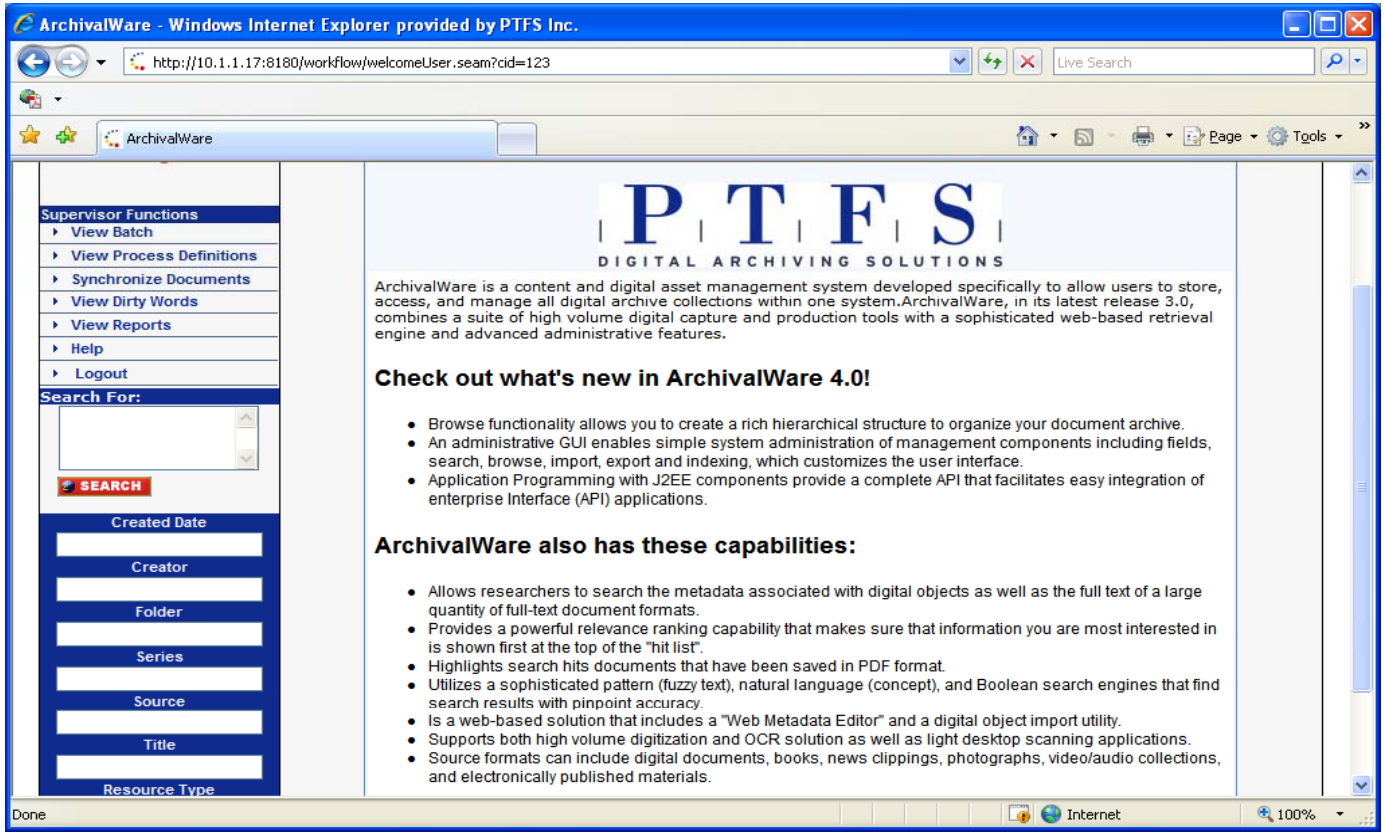


USG Unclassified / Commercial Confidential

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Declassification: Step 1 View and Assign Batches

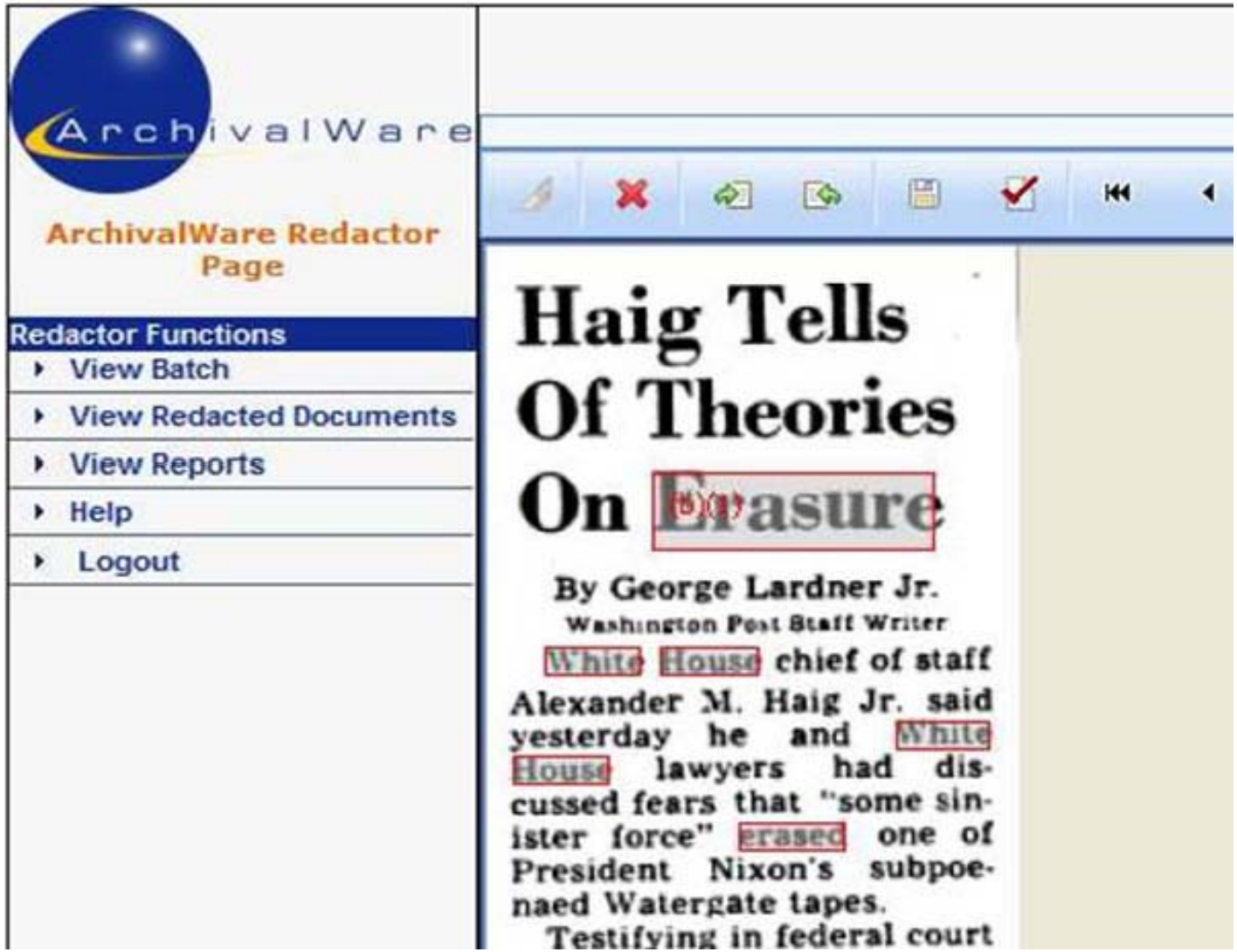
- Group documents into batches
- Assign batches to analyst
- Maintain “dirty word” file



PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Declassification: Step 2 Recommend Redactions, Revise Redactions

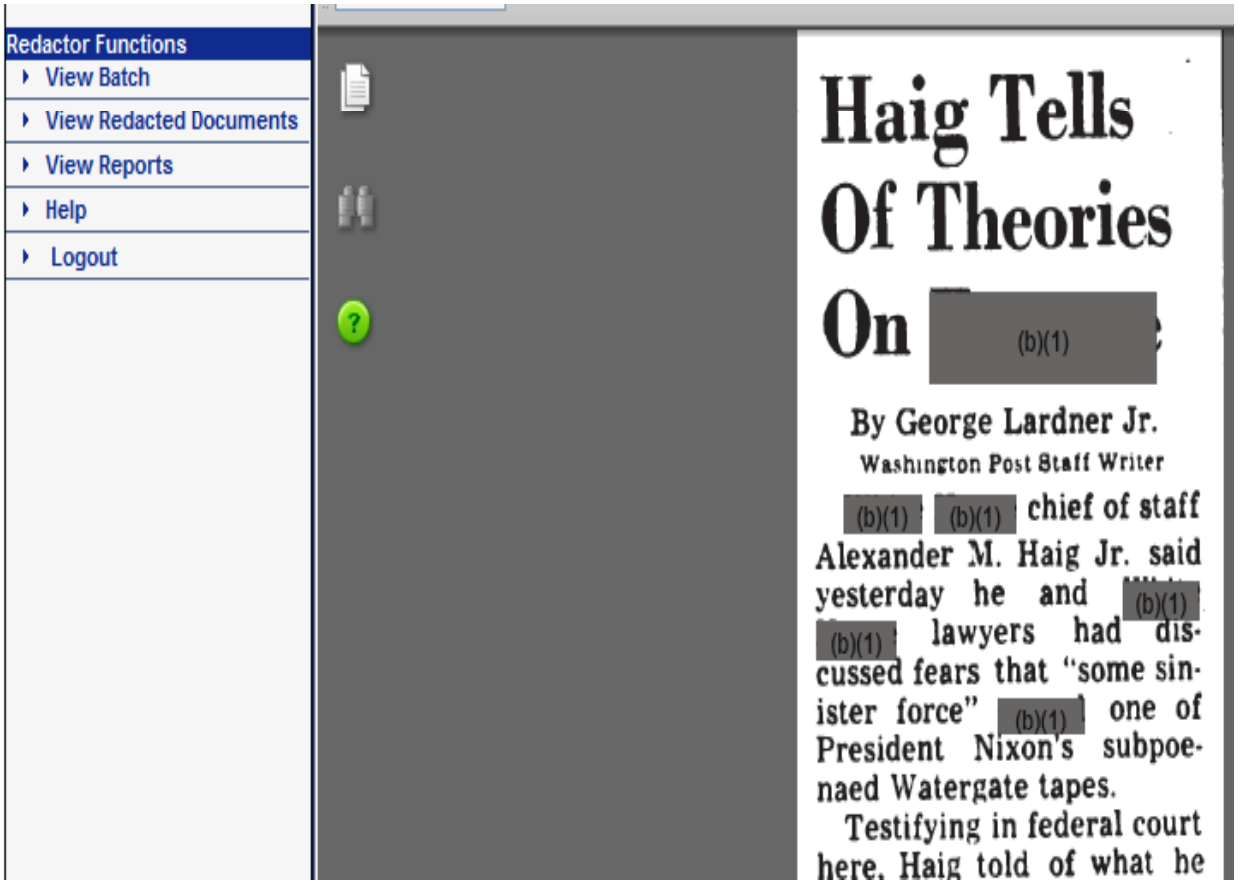
- Select Document
- Assign Redactor & QCer
- Select markup candidates
- Assign Codes
- Proceed to redact mark-up words

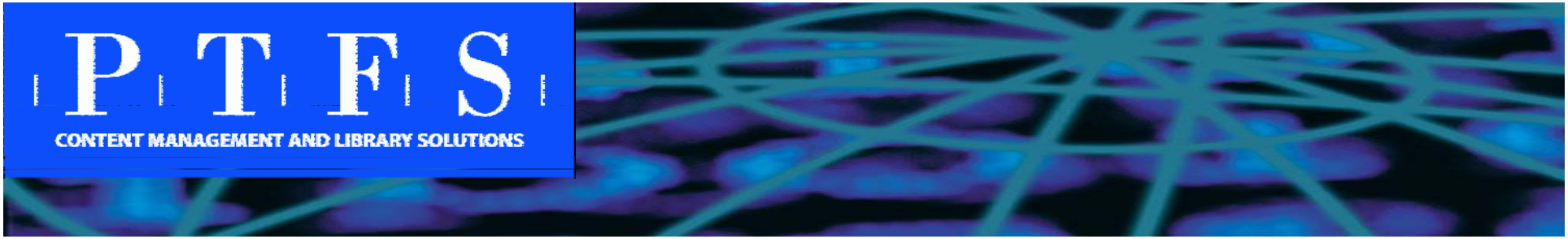


PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Declassification: Step 3 Finalize Redactions, QC

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

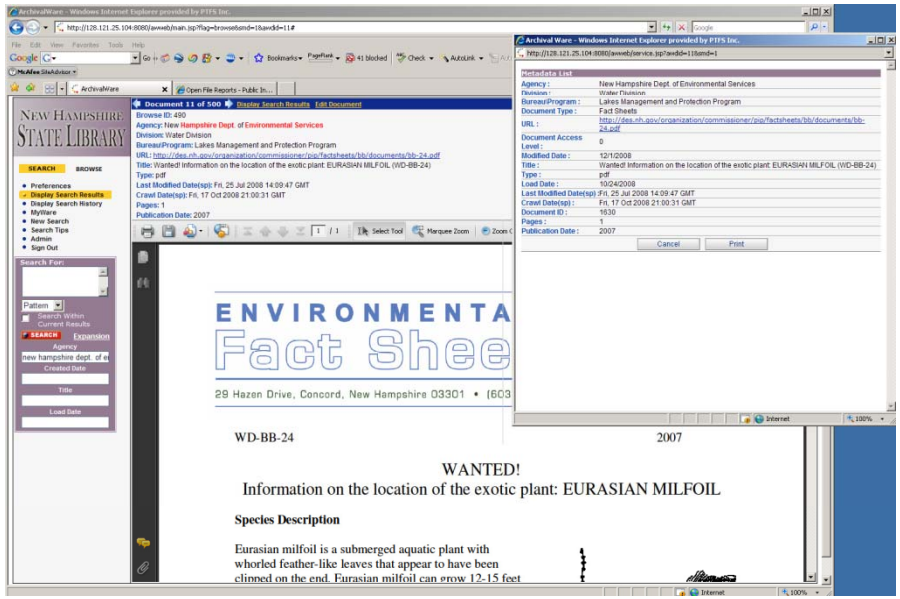




Open Source Intelligence

New Hampshire State Library

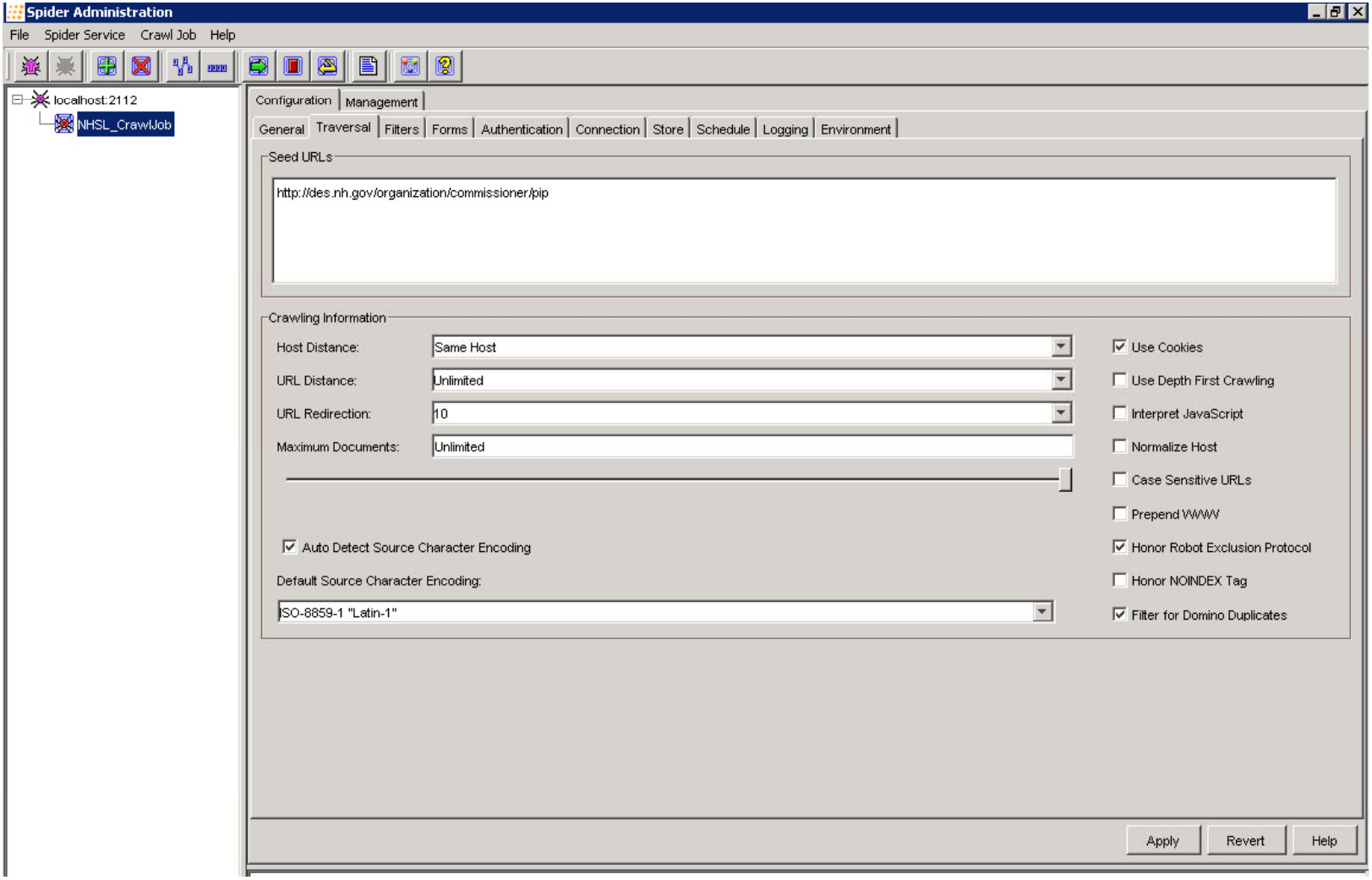
- Challenge
  - Establish repository for all state publications
- Scope
  - Configure ArchivalWare in ASP environment
  - Develop Spider capabilities to routinely scrape all state agency websites
  - Load existing documents, routinely refresh repository
- Status
  - Site went live in August 2008



PTFS Confidential

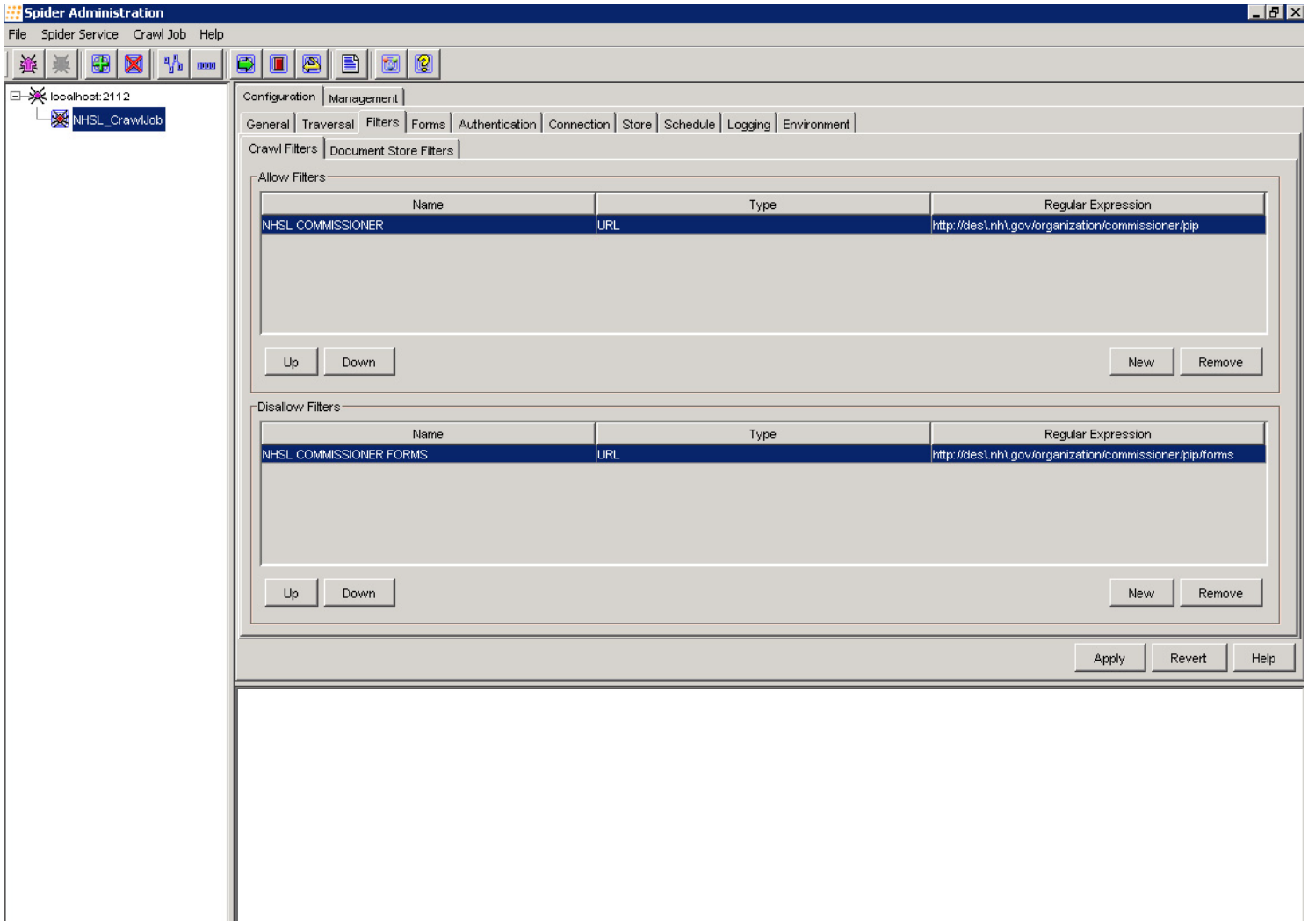
P T F S
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

NHSL Step 1: Add URL's



PTFS Confidential

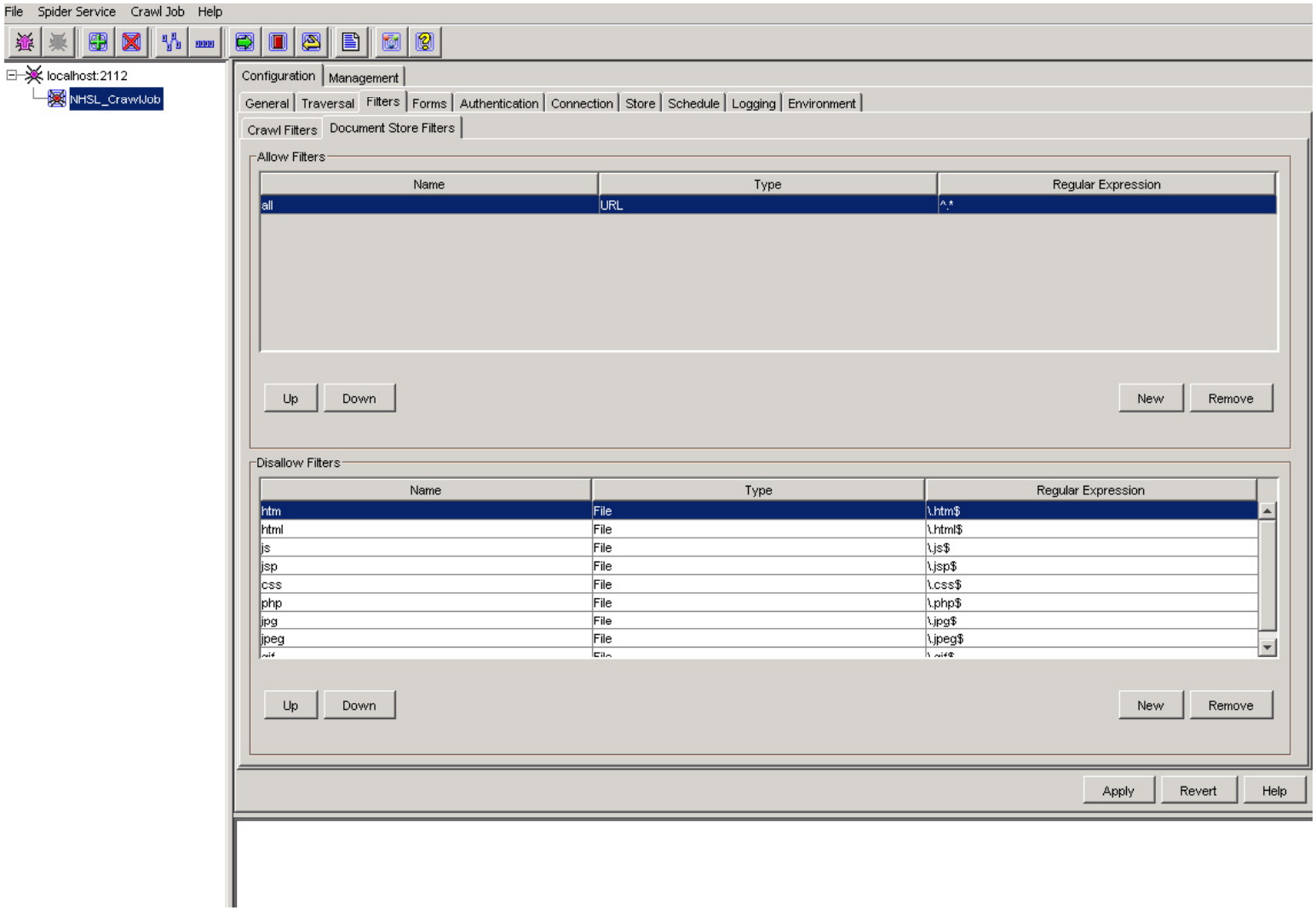
NHSL Step 2: Apply URL page filters



PTFS Confidential

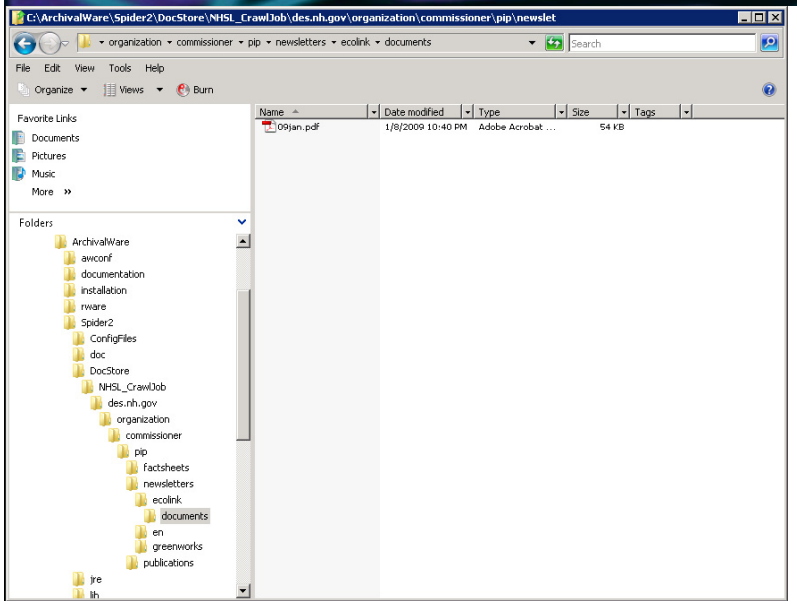
P T F S
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

NHSL Step 3: Apply file filters



PTFS Confidential

NHSL Step 4: Spidered content, stats returned



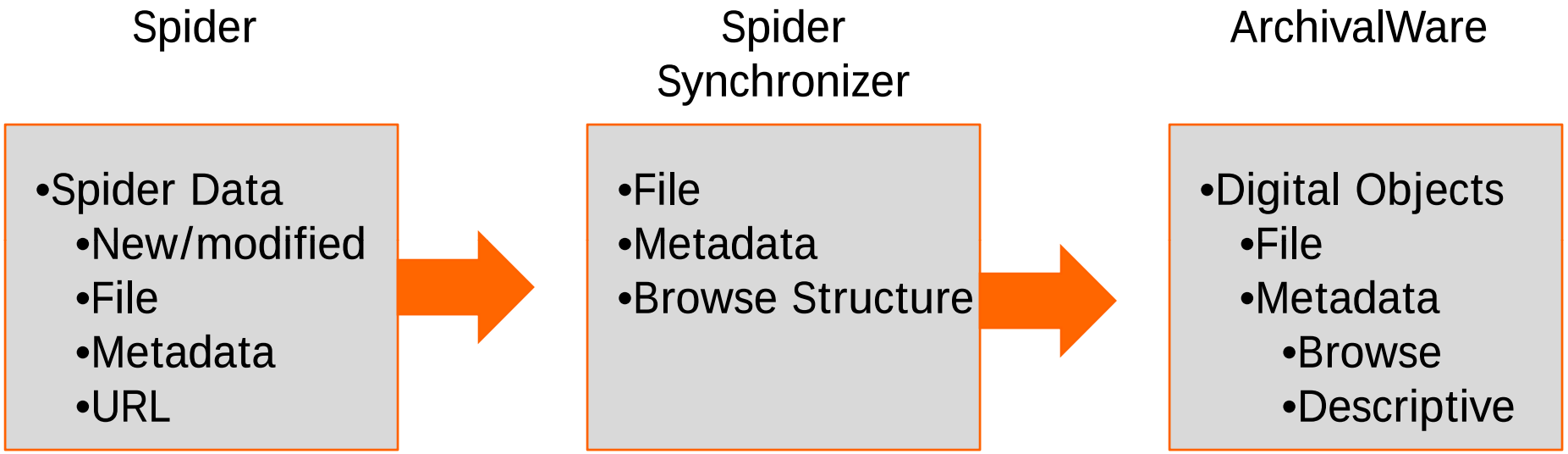
	url_id	first_visited_date	last_visited_date	last_mod_date_header	crawl_id	response_code	status_id	depth	parent_referer_url_id	redirect_url_id
1	10003	2008-10-17 08:01:04.470	2009-01-09 03:24:03.007	Mon, 10 Nov 2008 07:00:00 GMT	14	200	2	0	NULL	10002
2	10427	2008-10-17 08:18:10.203	2009-01-09 03:40:04.663	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-GEO-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

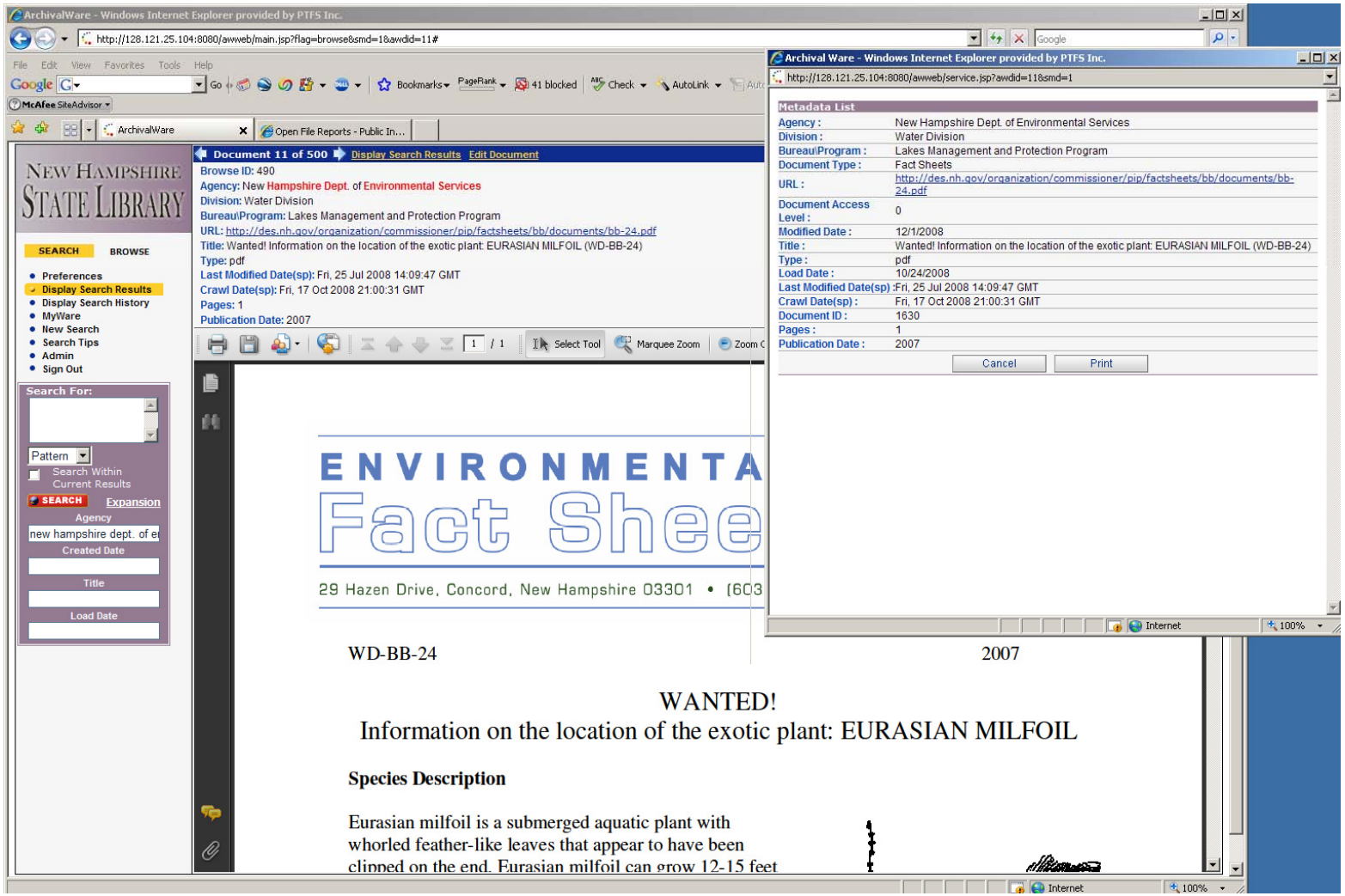
PTFS Confidential

NHSL Step 5: Process with Spider Synchronizer



PTFS Confidential

NHSL Step 6: Review content in ArchivalWare



PTFS Confidential

PTFS

CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Geospatial Capabilities

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

ArchivalWare Geospatial Capabilities

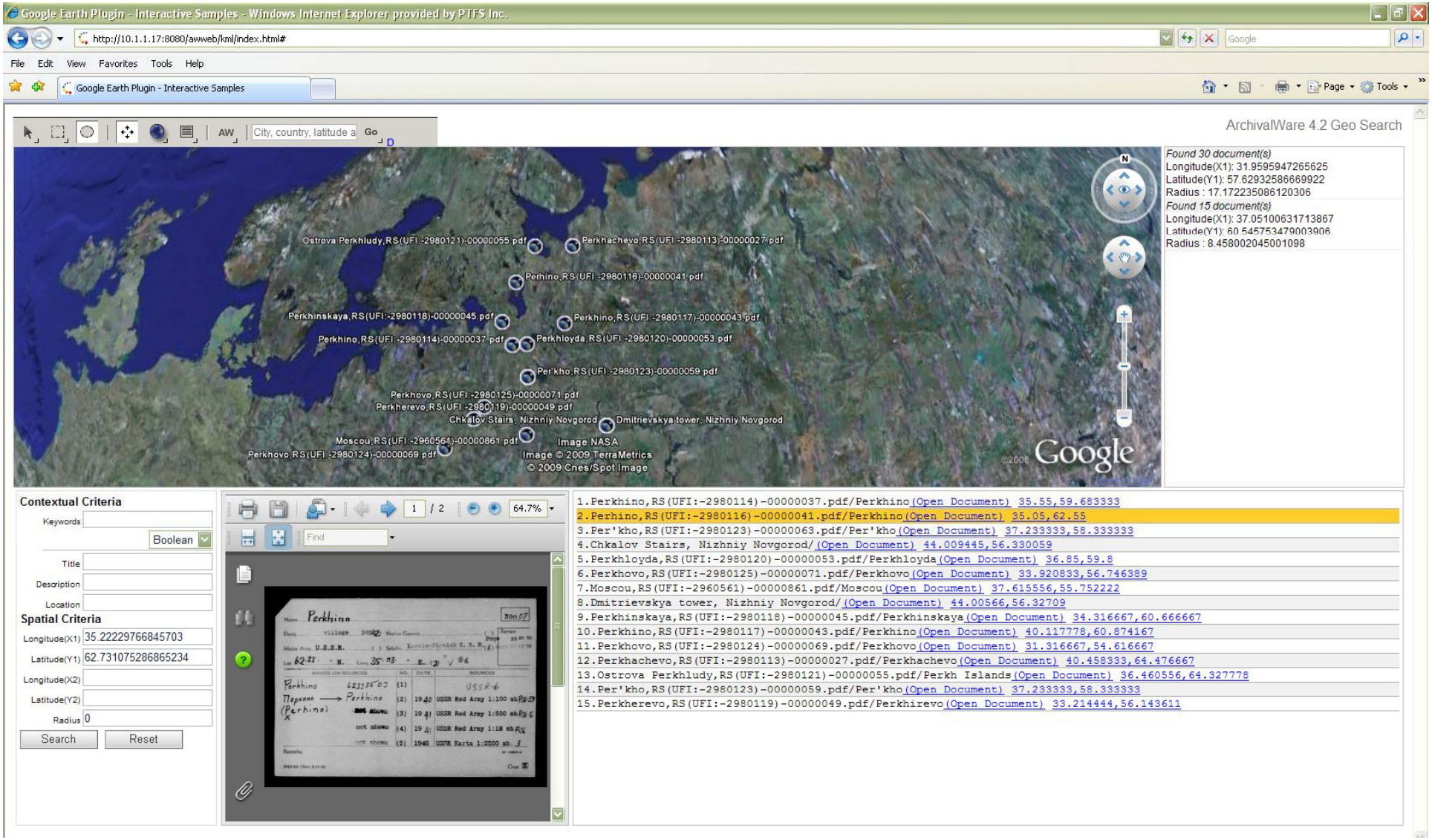
- Situation
  - Imagery (motion and stills) and related documents are stored in many different databases; and viewed with different tools
- Challenge
  - Aggregate all source intelligence (imagery and documents) on an area of interest in a single screen; a single view, with a single tool
- The Solution
  - 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.

P T F S

CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

ArchivalWare Geospatial Google Earth Integration



PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

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

P.T.F.S.
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

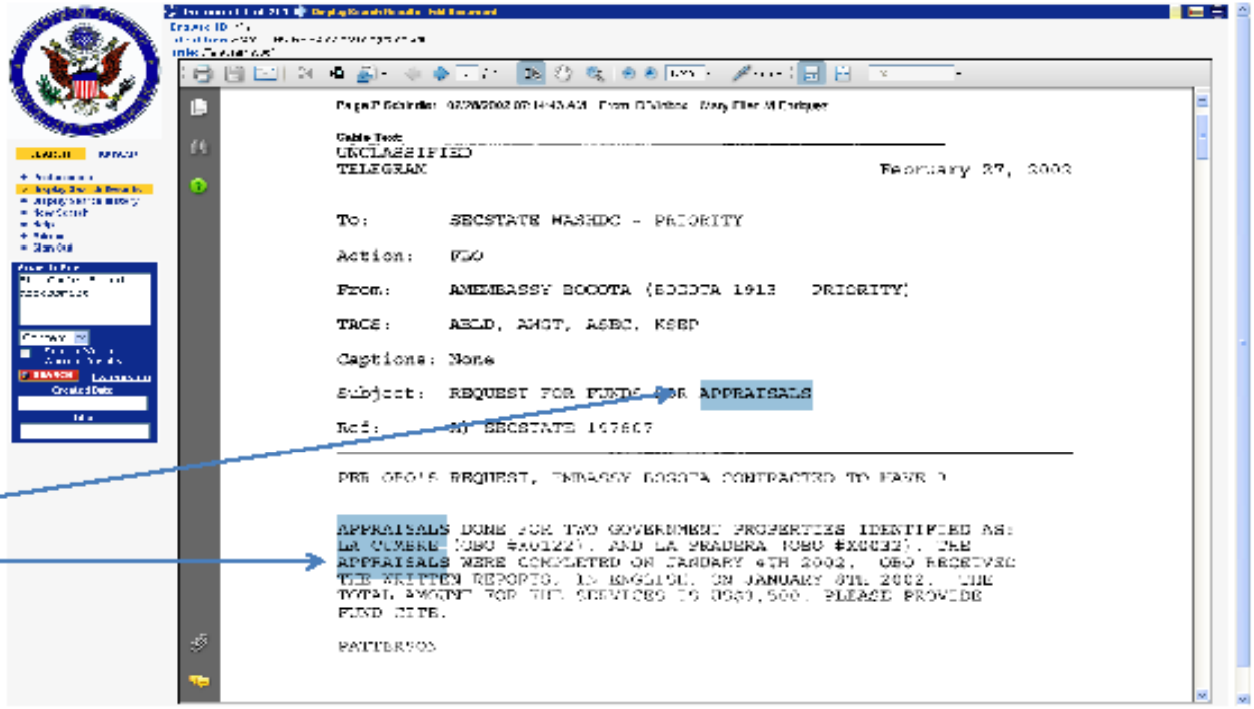
State Department Global Real Property Operations

State Department – OBO/RE Archive & Search System

Example – Concept & Fixed Term Search Capability – *English => "Assessment" => Appraisals*

Query:
Fixed Term
"La Cumbra"
and *Concept*
"Assessment"

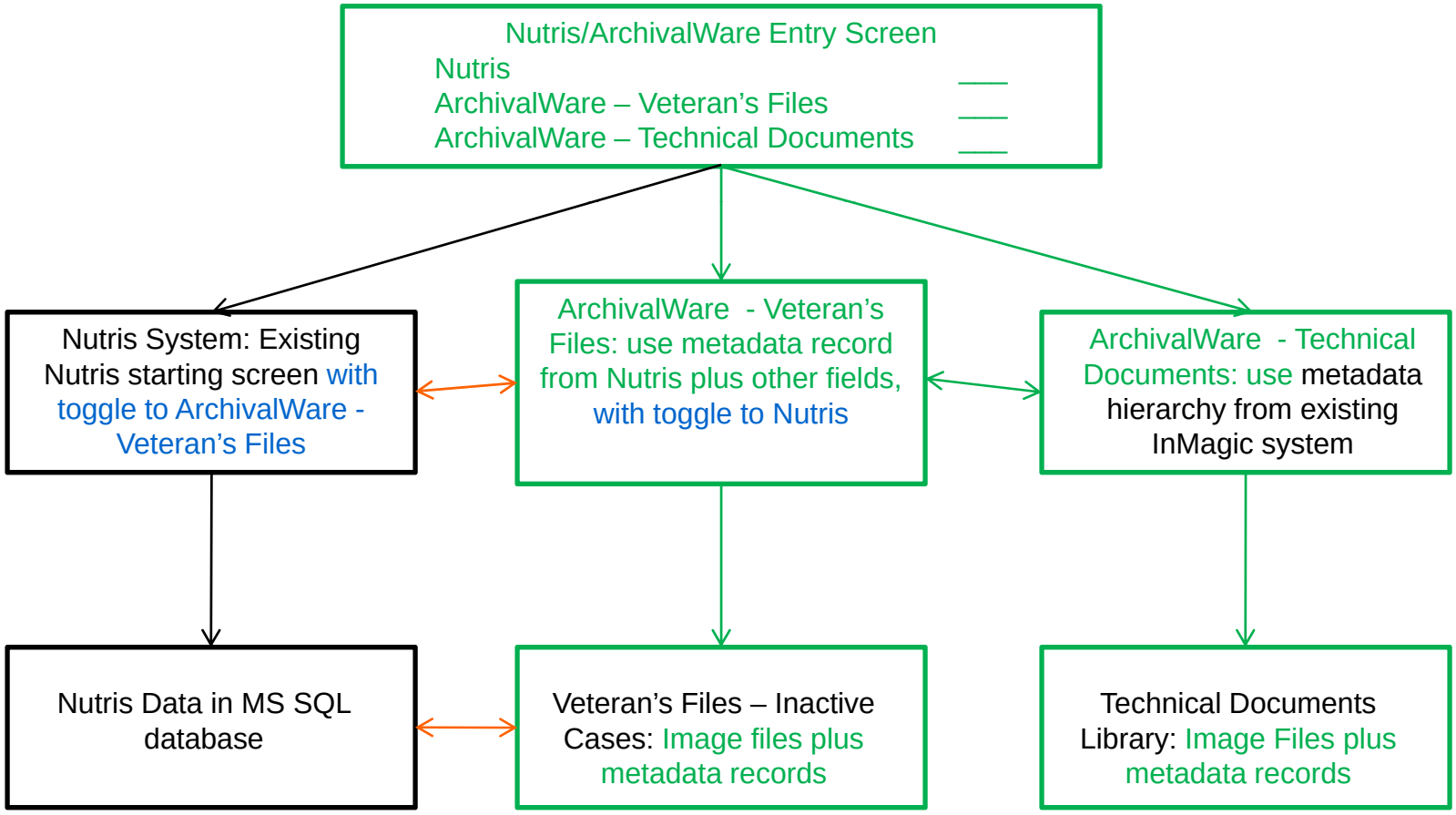
Search Concept
"Assessment"
Found As:
"Appraisals"



State Department – OBO/RE Sample Screen Shot

Other Case Studies

Veteran's System



Black: In Production today
Green: Phase I
Blue: Phase II

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

Explosive Library Network

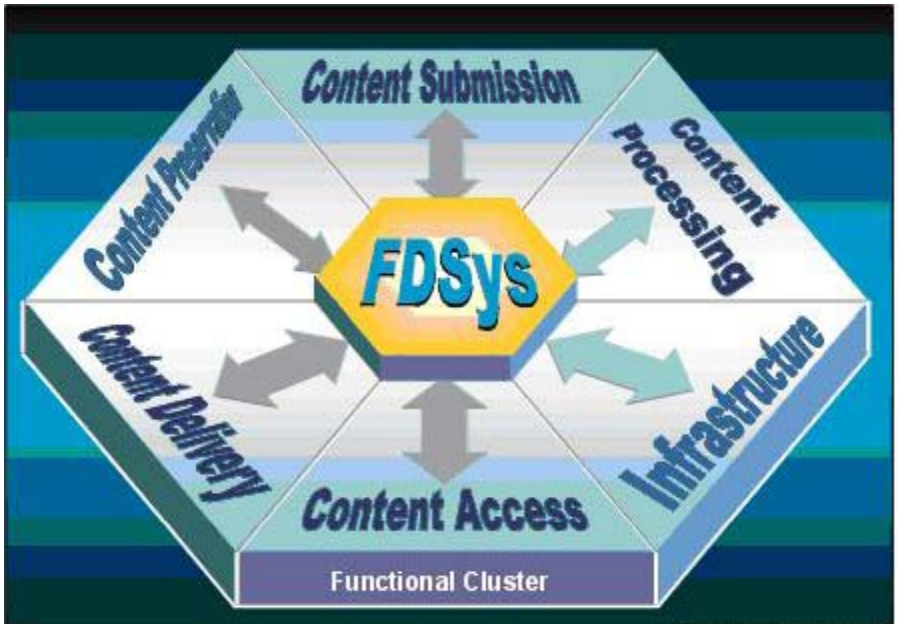
- ExLib – Explosive Library Network
  - The Situation: Data on conventional explosives, ordnance & mines - and IEDs is scattered throughout DoD and IC.
  - The Challenge: USG enlisted, contractors and coalition need quick, complete and accurate identification, analysis and operations information.
  - 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

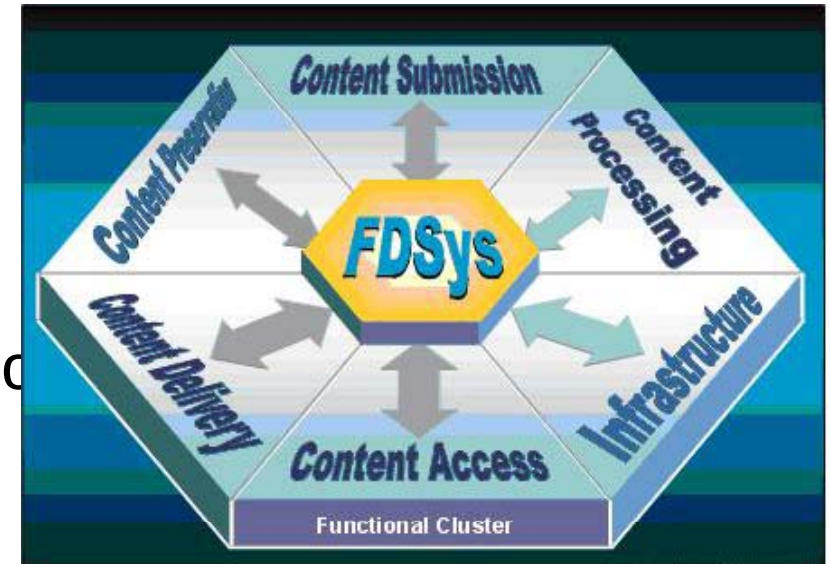
- One of the largest Government Content Management Systems ever built
- GPO will become repository of record for all published Federal documents
  - Over 120 million digitized government publications, terabytes of born-digital documents
- $29MM, 3-year project
- PTFS is responsible for web development,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



PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

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



PTFS Confidential

PTFS
CONTENT MANAGEMENT AND LIBRARY SOLUTIONS

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.