



Data Cleanup: The Never-Ending Story

Mark Martin
Kathy Waldrop

Scientific and Technical Information Program Working Meeting
Atlanta, Georgia
April 23, 2009



hidden documents stuck PII patent pending
failed urls unable-to-process deleted



Data Cleanup Statistics

DOE Sites

E-Link Records Needing DOE Site Reconciliation

■ Hidden Documents	9,935
– (most were hidden in 2007 for potential PII)	
■ US Dissemination Only Documents	0
■ Unable-to-Process Records	4,467
■ Patent Pending Documents	24
■ Failed URLs	392
■ Potential Protected PII Incidents	845



Hidden Documents Awaiting Reconciliation

Total Hidden Documents	9,935
– Hidden Documents for PII	8,362

- Have unclassified unlimited access limitations
- Hidden because there were concerns that they may contain sensitive information
- The documents need to be reviewed by the originating research organization/site to determine if sensitivities exist
- Unhide the documents that are accurately marked unclassified unlimited
- Edit E-Link records to reflect appropriate access limitation and review confirmation for documents that should be protected
- Replace full-text document as appropriate



Unable-to-Process Records

Unable-to-Process Records 4,467

- These records have not processed because of data errors
- The description of the error is listed in the 'Error Description' column
- Some records have been in queue for several years.
- **Majority of Errors:**
 - **Patent data wrong product type**
 - **Media contains unchecked possible PII**
 - **Missing required media**
 - **Empty research org**



Patent Pending Documents

Patent Pending Documents

24

- These documents have Patent Pending access limitations yet the publication dates have passed by more than 5 years
- These E-Link records should be revised to reflect the appropriate access limitations



Failed URLs

Failed URLs

392

- These documents failed the URL check upon submission to OSTI
- The URLs should be revised to reflect the corrected links directly linking to the documents or the document should be uploaded.
 - Errors include:**
 - **404 - Not Found – location changed**
 - **901 - Host Name Not Found – Server is off, gone, or newly named**
 - **403 - Forbidden – secure area; requires authentication**
 - **302 - Moved Temporarily – basically same as 404, system administrator has posted so we'll know it will return to this location**
- Regardless of error, OSTI can't get to STI
- You should contact your system administrator



Potential Protected PII Incidents

Potential Protected PII Incidents

845

- These documents include unclassified unlimited and OpenNet STI submitted to OSTI since February 15, 2008.
 - The site indicated that a review for PPII was not done before the document was uploaded
 - An algorithm search was performed on these documents and they were identified as potentially containing Protected PII.
 - These documents are on hold (aka stuck) until resolved. They have not completed processing by OSTI.
- To complete processing of these documents, review the information that caused the document to be flagged and either mark the radio button that the document does not contain Protected PII or replace the document if it does contain Protected PII.

Energy Link - Reports - Windows Internet Explorer

http://test.osti.gov:8080/elink/reports.jsp

Edit View Favorites Tools Help

Energy Link - Reports

Home | About E-Link | Help | User Update | Log Off waldropk

ENERGY LINK
Office of Scientific and Technical Information
STI EXCHANGE STIP | OSTI | DOE

HOME | FORMS: 241.1 - 241.3 | SEARCH/EDIT: SAVED FORMS - PENDING RELEASE | UPLOAD:HISTORY | REPORTS | ADMIN | AUTHORITIES

E-Link Reports

Reconciliation Reports:

The reports below list E-Link records that need to be reconciled by the respective site(s). See descriptions at the individual reconciliation report level.

- [Hidden Documents](#)
- [US Dissemination Only Documents](#)
- [Unable-to-Process Records](#)
- [Patent Pending Documents](#)
- [Failed URLs](#)
- [Potential Protected PII Incidents](#)
- [Documents Hidden for PII - 2007 Or Earlier](#)

Status Reports:

See descriptions about the various status reports at the individual report level.

- [Unhidden Documents](#)
- [Deleted Documents](#)
- [OSTI-to-Site TAP Records*\(formerly "Harvest Records"\)](#)
- [R&D Project Documents](#)
- [Access Limitation Changes](#)
- [Full Text Changes](#)

Note: Until August, 2008, the OSTI-to-Site TAP process was referred to as "Harvesting".

Deleted Records - Issue Identified in FY 2009



- Sites deleted records for which OSTI has paper/hard copy full-text. The ability to delete these records has been removed for items input by OSTI. What can be done about those already deleted?
 - Identify the FY the item was deleted
 - Identify deleted items for which OSTI has paper/hard copy



New Elink Reports

Access Limitation Change & Full Text Change

- Facilitation tool for STI management and digitization efforts
- Also beneficial for Sites – identify consistency issues between OSTI's bibliographic database and site maintained bibliographic database
- Additionally notification of these types of changes, i.e. Access Limitation and Full Text, is forthcoming
- Examples – although not written in stone, i.e. comments are welcome

Search for Access Limitation Change

Start Date: (MM/DD/YYYY)
End Date: (MM/DD/YYYY)
Site Code:
OSTI ID:

Access Limitation Change Report

This report shows reports in E-Link where the access limitation has changed. Also, included for each record is the date the change to the access limitation was made, the current access limitation, the previous access limitation and the name of the individual who made the change.

Total results: 9

OSTI ID	Report Number	Site Code	Input Code	Title	Change Date	Current Access Lim	Previous Access Lim	Access Lim Other	Changed By
1387	DE-FC21-93MC30097--55	NETL	NETL	Task 1.14 - Benzene Removal From Gasoline Using Subcritical Water	12/11/2008	UNL	PAT	Previously Access Limitation 'PAT'; confirmation of public accessibility required prior to posting digitized text	Susan Olson
1688	DE-FC21-93MC30097--30	NETL	NETL	Subtask 1.17 - Subcritical Water Extraction of Mercury From Soils and Sediments. Final report	12/11/2008	UNL	PAT	Previously Access Limitation 'PAT'; confirmation of public accessibility required prior to posting digitized text	Susan Olson
2284	SAND98-2655	SNL	SNL	Covert, Distributed Bio-Sensors for UXO/CVW Based on Amplified Immunoassays Conducted in Porous Inorganic Media	12/11/2008	UNL	PAT	Previously Access Limitation 'PAT'; confirmation of public accessibility required prior to posting digitized text	Teresa Gilbert
705264	SAND--91-0559	SNL-A	TIC	Cavity-coupled optical link	03/10/2009	OUC; ECI	PAT; ECI		Teresa Gilbert
751140	SAND98-8666	SNL	SNL	A Lightweight, Link-layer, Source-based routing Protocol for LEO Satellite Networks	03/10/2009	UNL	PAT	Previously Access Limitation 'PAT'; confirmation of public accessibility required prior to posting digitized text	Teresa Gilbert
10103846	DUH--10969	PNNL	PNNL	Project 9336: Hanford-retention basin and effluent dilution	11/12/2007	UNL	USO		David Long
10103846	HW--82940	PNNL	PNNL	Irradiation Processing Department monthly report, July 1964	11/12/2007	UNL	USO		David Long
10104492	HW--82944	PNNL	TIC	Irradiation Processing Department monthly report, November 1964	11/12/2007	UNL	USO		David Long
10104723	HW--82942	PNNL	TIC	Irradiation Processing Department monthly report, September 1964	11/12/2007	UNL	USO		David Long



[Website Policies/Important Links](#)



Last Updated: 04/16/2009

Search for Full Text Change

Start Date: (MM/DD/YYYY)
End Date: (MM/DD/YYYY)
Site Code:
OSTI ID:

Full Text Change Report

This report shows reports in E-Link where the full-text document has changed in one way or another. Whether the full-text was replaced, the URL changed (edited or added), or an electronic document attached for the first time will be reflected in this report. The type of change will be indicated in the Change Description column. Also, included for each record is the date the change to the full-text document was made, the access limitation, the media status, and the name of the individual who made the change.

Total results: 13

OSTI ID	Report Number	Site Code	Media	Input Code	Title	Change Date	Access Lim	Changed By	Change Description	Media Status
750027	SAND--99-3117	SNL	P	SNL	Geometric simplification of analysis models	01/06/2000	UNL	Teresa Gilbert	Full Text Replaced	Latest
750029	SAND--99-3149/1	SNL	P	SNL	Final report for the high performance commodity interconnects for clustered scientific and engineering computing. Part 1 of 3	01/06/2000	UNL	Teresa Gilbert	URL Added	Latest
750030	SAND--99-3149/2	SNL	P	SNL	Final report for the high performance commodity interconnects for clustered scientific and engineering computing. Part 2 of 3	01/06/2000	UNL	Teresa Gilbert	Full Text Replaced	Latest
750031	SAND--99-3149/3	SNL	P	SNL	Final report for the high performance commodity interconnects for clustered scientific and engineering computing. Part 1 of 3	01/06/2000	UNL	Teresa Gilbert	Full Text Replaced	Latest
750112	WSRC-TR-99-00342	SRS	P	SRS	Effect of sludge solids to mono-sodium titanate (MST) ratio on MST-treated sludge	01/06/2000	OUO; SSI	Wileva Dunbar	Full Text Replaced	Latest
750083	SRT-PUM--97-0017	SRS	P	SRS	Results of an inter-laboratory study of glass formulation for the immobilization of excess plutonium	01/06/2000	UNL	Wileva Dunbar	Full Text Replaced	Latest
750083	SRT-PUM--97-0017	SRS	P	SRS	Results of an inter-laboratory study of glass formulation for the immobilization of excess plutonium	01/06/2000	UNL	Wileva Dunbar	Full Text Replaced	Historical
750086	WSRC-TR--99-00277	SRS	P	SRS	Hydrogen generation during melter feed preparation of Tank 42 sludge and salt washed loaded CST in the Defense Waste Processing Facility	01/06/2000	UNL	Wileva Dunbar	Full Text Replaced	Latest
750087	WSRC-TR--99-00302	SRS	P	SRS	Hydrogen generation and foaming during tests in the GFPS simulating DWPF operations with Tank 42 sludge and CST	01/06/2000	UNL	Wileva Dunbar	Full Text Attached	Latest
750090	WSRC-RP--99-00836	SRS	P	SRS	The effect of temperature and radiation on the cesium adsorption ability of IONSV/256 IE-910 and IONSV/256 IE-911	01/06/2000	UNL	Wileva Dunbar	Full Text Attached	Latest
750091	WSRC-RP-99-00849	SRS	P	SRS	Idealized mixing impacts	01/06/2000	UNL	Wileva Dunbar	URL Added	Latest
750092	WSRC-TR--99-00286	SRS	P	SRS	Phase IV testing of monosodium titanate adsorption with radioactive waste	01/06/2000	UNL	Wileva Dunbar	Full Text Attached	Latest
750094	WSRC-TR-99-00265	SRS	P	SRS	M-Area and Metallurgical Laboratory Hazardous Waste Management Facilities Groundwater Monitoring and Corrective-Action Report, First and Second Quarters 1999, Volumes I and II	01/06/2000	UNL	Wileva Dunbar	URL Added	Latest

Discussion



Thank You

Kathy Waldrop
E-Link Product Mgr.
241user@osti.gov
waldropk@osti.gov

Mark Martin
OSTI Support Services
Contract Mgr.
martainm@osti.gov