

OSTI's STI Announcement Web Service For 241.6 Data September 2012

CONTENTS

- 1.0 About
 - 1.1 [Announcement Notice 241.6](#)
 - 1.2 [OSTI's Web Service for AN 241.6](#)
- 2.0 Using the STI Announcement Web Service
 - 2.1 [Authentication](#)
 - 2.2 [POST Command: Creating and Modifying Records](#)
 - 2.3 [GET Command: Requesting Records](#)
- 3.0 AN 241.6 Web Service Metadata
 - 3.1 [AN 241.6 Required Fields](#)
 - 3.2 [AN 241.6 Optional Fields](#)
 - 3.3 [Examples of XML 241.6 Metadata Records](#)

1.0 About

1.1 Announcement Notice 241.6

Announcement Notice (AN) 241.6 provides to the U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) metadata needed to identify/announce publicly available datasets resulting from work funded by the DOE or performed in DOE facilities. The submitted information also allows OSTI to assign Digital Object Identifiers (DOIs) to datasets and register them with [DataCite](#) as a service to researchers. This value-added step facilitates visibility, helps ensure long-term preservation, and supports better linkage between DOE's published research results and the underlying data. The DOI assigned to each announced dataset is included in the XML response to each successful submittal. The primary contact identified in the metadata will also be notified by an automated email of successfully submitted records and the DOIs assigned.

The datasets themselves must be publicly available and maintained at a DOE site, DOE Data Center, or other publicly accessible location, such as an open repository. Submitting sites agree to ensure data persistence, which includes storing and managing data such that access and usability are provided indefinitely. The URL associated with each DOI should point to an HTML "landing page" that provides context for the dataset or to a notice page ("tombstone page") for data that has had to be retired. Datasets are not uploaded or stored at OSTI.

.The AN 241.6 web service is the easiest way to submit multiple records on a regular basis. However, an [interface for manual entry](#) of one AN 241.6 record at a time is available on E-Link, if desired.

1.2 OSTI's Web Service for AN 241.6

OSTI's Web Service for AN 241.6 provides an easy-to-use mechanism to submit new Announcement Notices, edit existing Announcement Notices, and retrieve metadata for Scientific and Technical Information (STI) Announcement Notices. A user manipulates a record by performing HTTP operations on the web service URL and providing XML metadata. The POST command/request allows a submitter to create or modify AN 241.6 records. The GET command/request allows a submitter to view records.

Submitters must coordinate with OSTI and submit records to a test environment before POSTing their first file. The test URL is <https://www.osti.gov/elinktest/2416api>. OSTI's production web service for announcing datasets is available at <https://www.osti.gov/elink/241.6api>.

All transactions with the web service require authentication through an active E-Link user account. For details or to request access to E-Link and obtain an active account, please visit <https://www.osti.gov/elink/register.jsp>.

2.0 Using the STI Announcement Web Service

2.1 Authentication

Each request requires authentication through an active E-Link user account. OSTI's STI Announcement Web Services supports HTTP Basic authentication over SSL. With this method, the client connects an HTTPS URL (e.g., <https://www.osti.gov/elink/2416api>). The POST and GET verb commands will pass along the standard Authentication HTTP header (base64 encoding).

If authentication is successful, the input body of the request is read as an XML document, parsed, and submitted appropriately. HTML status codes (200=OK, 401=Unauthorized, 500=System error, etc.) and an XML response, which includes certain metadata fields, are returned at the end of processing each request. Additional authentication steps may be taken prior to editing a metadata record.

2.2 POST Command: Creating and Modifying Records

When submitting data via the 2416 web service, a secure connection is created, and the login/authentication information and the metadata records are streamed through the connection. The OSTI web service will generate a response indicating success or failure for each records. An error message will be included for each failed record, as well. Multiple records can be submitted in a single submission.

The following sample of Java code illustrates a submission with the POST command. The sample name “testPost” can, of course, be changed to any name desired.

```
protected static boolean testPost() throws IOException {
    HttpURLConnection c = null;
    DataOutputStream out = null;
    InputStreamReader in = null;
    BufferedReader br = null;

    try {

        /** build the URL to connect to */
        StringBuilder url = new StringBuilder();
        // urlPart is defined elsewhere - it is the "https://www.osti....."
        url.append(urlPart);
        URL u = new URL(url.toString());
        c = (HttpURLConnection) u.openConnection();
        c.setRequestMethod("POST");
        c.setRequestProperty("Accept", "application/xml");
        c.setRequestProperty("Content-Type", "application/xml;charset=UTF-8");

        /** add login authentication
            usr_passwd is a string defined elsewhere. It is USERNAME + ":" +
PASSWORD for the elink user account
        */
        String auth =
org.apache.commons.codec.binary.Base64.encodeBase64URLSafeString(
(usr_passwd).getBytes() );
        c.setRequestProperty("Authorization", "Basic " + auth);

        /** write out the metadata stream */
        c.setDoOutput(true);
        out = new DataOutputStream(c.getOutputStream());

        /** make a string that contains the record(s) in XML format
            YOUR SITE DATA GOES IN THIS STRING
        */
        String rec = [SITE DATA];

        /** add the rec string to the output stream */
    }
}
```

```

byte[] buf = (null==rec) ? "".getBytes("UTF-8") : rec.getBytes("UTF-8");
out.write(buf, 0, buf.length);
out.close(); out = null;

/** open the connection */
c.connect();

/** get the response and appropriate response information */
int responseCode = c.getResponseCode();
if (responseCode>=400) in = new InputStreamReader(c.getErrorStream());
else in = new InputStreamReader(c.getInputStream());

br = new BufferedReader(in);
StringWriter writer = new StringWriter();
char[] buffer = new char[1024];
int n=0;
while ((n=br.read(buffer))!=-1) {
    writer.write(buffer, 0, n);
}
writer.close();

// see what the test results are from the API servlet
StringBuffer outBuf = writer.getBuffer();
System.out.print("Response returned " + outBuf.toString());

/** log any errors */
if (responseCode!=200) {
    log.error("OSTI ID: " + m.getOstiId() + " failed to post new DOI, error code="
+ responseCode);
    log.error("Message: " + writer.toString());
}
/** finished */
return (200==responseCode);
} finally {
    try {
        if (br!=null) br.close(); br = null;
        if (in!=null) in.close(); in = null;
        if (out!=null) out.close(); out = null;
    } catch ( Exception e ) {
        log.error("URL Close Error: " + e.getMessage());
    }
}
}
}

```

2.3 GET Command: Requesting Records

Metadata can be retrieved for records previously submitted from your site by using a GET request and supplying the osti_id argument on the command line. Authentication is required and is handled in the same fashion as a Create/Modify command. Metadata is returned as XML.

The following sample of Java code illustrates a GET request. The sample name “testGet” can, of course, be changed to any name desired.

```
protected static boolean testGet( ) throws IOException {

    HttpURLConnection c = null;
    DataOutputStream out = null;
    InputStreamReader in = null;
    BufferedReader br = null;

    try {
        /** build the URL to connect to */
        StringBuilder url = new StringBuilder();
        // urlPart is defined elsewhere - it is the "https://www.osti...../2416API"
        url.append(urlPart);
        /** append the osti id parameter**/
        url.append("?osti_id=1001628");
        URL u = new URL(url.toString());
        c = (HttpURLConnection) u.openConnection();
        c.setRequestMethod("GET");
        c.setRequestProperty("Accept", "application/xml");
        c.setRequestProperty("Content-Type", "application/xml;charset=UTF-8");

        /** add login authentication
            usr_passwd is a string defined elsewhere. It is USERNAME + ":" +
PASSWORD for the elink user account
        */
        String auth =
org.apache.commons.codec.binary.Base64.encodeBase64URLSafeString(
(usr_passwd).getBytes() );
        c.setRequestProperty("Authorization", "Basic " + auth);
        c.connect();

        /** get the response and appropriate response information */
        int responseCode = c.getResponseCode();

        if (responseCode>=400) in = new InputStreamReader(c.getErrorStream());
        else                in = new InputStreamReader(c.getInputStream());

        br = new BufferedReader(in);
        StringWriter writer = new StringWriter();
        char[] buffer = new char[1024];
        int n=0;
```

```

while ((n=br.read(buffer))!=-1) {
    writer.write(buffer, 0, n);
}
writer.close();

/** log any errors */
if (responseCode!=200) {
    log.error("OSTI ID: 10011628; error code=" + responseCode);
    log.error("Message: " + writer.toString());
}
/** finished */

} finally {
    try {
        if (br!=null) br.close(); br = null;
        if (in!=null) in.close(); in = null;
        if (out!=null) out.close(); out = null;
    } catch ( Exception e ) {
        log.error("URL Close Error: " + e.getMessage());
    }
}
}
}

```

3.0 AN 241.6 Web Service Metadata

3.1 AN 241.6 Required Fields

The following is a list of the required fields for the AN 241.6 STI Announcement Web Service. Required fields are designated by an asterisk (*). Records without required fields will fail to load into E-Link for processing and DOI registration. Only 15 metadata and/or administrative fields are required. The other fields available for your use are optional, though some, such as the Abstract/Description, are highly encouraged.

Note that the OSTI ID is a required field for all POST requests where the intent is to edit or update records. The GET request must also include the OSTI ID and will allow retrieval of a record previously submitted by your site.

REQUIRED

	Field Name	XML Tag Name	Additional Information
1	OSTI ID+	<osti_id>	Note that the OSTI ID is required in all requests intended to edit or update records. When POSTing new records to OSTI, no <osti_id> tag is needed in the XML. E-Link automatically assigns an OSTI ID to each record successfully submitted; you will

			receive it in the XML response returned to your site by the OSTI webservice.												
2	Site Code*	N/A	Automatically determined by the authenticated E-Link User account												
3	Dataset Type*	<dataset_type/>	Dataset Type refers to the main content of the dataset. Only one value is allowed. Use the two-letter code shown below:												
			<table border="0"> <thead> <tr> <th style="text-align: left;">Code</th> <th style="text-align: left;">Definition</th> </tr> </thead> <tbody> <tr> <td>AS</td> <td>Animations/Simulations</td> </tr> <tr> <td>GD</td> <td>Genome Data - Information that is numeric or alpha-numeric in nature (such as gene sequences) or that is a specialized mix of text and non-text information conveying results of genetics/genome research</td> </tr> <tr> <td>IM</td> <td>Interactive Data Map(s) – A non-static interface and the GIS data and/or shape files that generate it.</td> </tr> <tr> <td>ND</td> <td>Data primarily expressed with numbers; other content is secondary and supporting.</td> </tr> <tr> <td>IP</td> <td>Still Images or Photos - A collection of images or photographs produced by a scientific instrument or that convey scientific results of experiments. Scientific images that might constitute a data set could be images of cells or molecules that are typically taken with electron microscopes, 3-D structures of proteins or nanomaterials, images captured during an accelerator run, images from astronomy, etc.</td> </tr> </tbody> </table>	Code	Definition	AS	Animations/Simulations	GD	Genome Data - Information that is numeric or alpha-numeric in nature (such as gene sequences) or that is a specialized mix of text and non-text information conveying results of genetics/genome research	IM	Interactive Data Map(s) – A non-static interface and the GIS data and/or shape files that generate it.	ND	Data primarily expressed with numbers; other content is secondary and supporting.	IP	Still Images or Photos - A collection of images or photographs produced by a scientific instrument or that convey scientific results of experiments. Scientific images that might constitute a data set could be images of cells or molecules that are typically taken with electron microscopes, 3-D structures of proteins or nanomaterials, images captured during an accelerator run, images from astronomy, etc.
Code	Definition														
AS	Animations/Simulations														
GD	Genome Data - Information that is numeric or alpha-numeric in nature (such as gene sequences) or that is a specialized mix of text and non-text information conveying results of genetics/genome research														
IM	Interactive Data Map(s) – A non-static interface and the GIS data and/or shape files that generate it.														
ND	Data primarily expressed with numbers; other content is secondary and supporting.														
IP	Still Images or Photos - A collection of images or photographs produced by a scientific instrument or that convey scientific results of experiments. Scientific images that might constitute a data set could be images of cells or molecules that are typically taken with electron microscopes, 3-D structures of proteins or nanomaterials, images captured during an accelerator run, images from astronomy, etc.														
4	Dataset Title*	<title/>													
5	Creator(s)/ Principal Investigator(s)*	<creators/>	Format is Last Name, First Name, MI Separate multiple authors with a semi-colon followed by a space.												
6	Dataset Product Number(s)*	<product_nos/>	The most important identifying numbers given to the dataset by the host or originating organization. Separate multiple values with a semi-colon followed by a space. ‘None’ is an acceptable value when necessary.												
7	DOE Contract Number(s)*	<contract_nos/>	Use the format of the contract “as is,” but please leave off any preceding “DE”. If multiple contract and/or grant numbers apply, separate with a semi-colon followed by a space.												
8	Originating Research Organization*	<originating_research_org/>	Use the spelled-out text exactly as shown in the Originating Research Organization Authority at https://www.osti.gov/elink/authorities.jsp If work for this product was done at more than one research organization, multiple values may be listed; they should be												

			separated by a semicolon and a space. The primary DOE organization should be listed first, followed by any others. If non-DOE orgs are included, input the spelled-out, full name of the organization.
9	Publication/Issue Date*	<publication_date/>	Use one of these three Publication Date formats: <ul style="list-style-type: none"> • mm/dd/yyyy • yyyy • yyyy Month
10	Language*	<language/>	Up to 75 characters; use format from OSTI's Language Authority; e.g. English Authority values are available at https://www.osti.gov/mlink/authorities.jsp
11	Country of Origin/Publication*	<country/>	Use two character code from OSTI's Country Code Authority; e.g. US Authority values are available at https://www.osti.gov/mlink/authorities.jsp
17	Sponsoring Organization(s)*	<sponsor_org/>	Use the spelled-out text as shown in the Sponsoring Organization Authority at https://www.osti.gov/mlink/authorities.jsp If funding for this product was provided from more than one organization, multiple values may be listed; they should be separated by a semicolon and a space. The primary DOE sponsor should be listed first, followed by any others. If any of the others are not included in the Sponsor Organization Authority (non-DOE organizations, for example), please include the spelled-out, full name of the other sponsoring organization.
12	Site URL*	<site_url/>	OSTI can not accept, store, or post datasets. Datasets must be publicly available, and the submitted metadata must include a valid URL. The URL should link to an html "landing page" for the dataset.
13	Contact Name and Position*	<contact_name/>	Admin info only; it will not be displayed in public databases.
14	Contact Organization*	<contact_org/>	Admin info only; it will not be displayed in public databases.
15	Contact E-mail*	<contact_email/>	Admin info only; it will not be displayed in public databases.

3.2 AN 241.6 Optional Fields

The following is a list of optional fields for the AN 241.6 STI Announcement Web Service. Inclusion of some of these fields, such as the Abstract/Description, is highly encouraged, however.

OPTIONAL

1	Creator(s)/PI Email Address(es)	<creators_emails/>	Admin info only; it will not be displayed in public databases.
2	Related Resource	<related_resource/>	This is a place to provide the bibliographic info on the key paper(s) that the dataset supports.
3	Availability	<availability/>	Normally used to provide the name of an organization, a division within a lab, a specific employee's title, etc. to which a request for further information may be made.
4	Contributor Organizations	<contributor_organizations/>	Provide the names of any organizations that have significantly contributed to the gathering, formatting, analysis, etc. of the dataset. These are organizations that would not otherwise be credited because they will not be listed in the Originating Research/Submitting Organization field, or in the Sponsoring Organization field. Separate multiple entries with a semicolon and a space.
5	Other Identifying Numbers(s)	<other_identifying_nos/>	Any other numbers that users might wish to retrieve on or need to recognize. If there are multiple values in this field, separate them with a semicolon followed by a space.
6	Subject Categories	<subject_categories_code/>	Use the complete value (numerical code and spelled-out category title) as shown in the Subject Category Authority at https://www.osti.gov/elink/authorities.jsp . As many multiples as needed are allowed in this tag set; separate them with a semicolon and a space. List the primary subject category first.
7	Keywords	<keywords/>	
8	Description/ Abstract	<description/> 4000 character limit	Provide a clear, concise summary of the content of the dataset, as well as specialized parameters that describe the data. Specialized parameters may include a date range during which information was taken (such as May, 01 2002 - December 31, 2002), geographic information (such as a specific state, region, country, latitude and longitude, etc.), information such as well depth ranges, temperature ranges, etc.
9	DOI	<doi/>	Provide the DOI if one has been assigned prior to the dataset being announced to OSTI.

10	Dataset's File Extension	<file_extension/>	Some common file extensions are .txt, .csv, .ps, etc.
11	Software needed to utilize dataset	<software_needed/>	Specialized software tools are often developed to allow a user to manipulate data in various ways. If these tools are available for the user but do not have to be used with the data, they do not need to be listed. However, if there is a piece of software without which a user cannot open, see, or use the dataset, that software should be noted in this field
12	Dataset Size	<dataset_size/>	Optional. Indicate approximate size in number of files, in megabytes, or in other ways appropriate for the dataset's content.
13	Contact Phone	<contact_phone/>	Admin info only; it will not be displayed in public databases.

3.3 Examples of XML 241.6 Metadata Records

Here is an example of a 241.6 dataset record as it would come to OSTI's 241.6 web service. Immediately following it is what the xml response from web service back to the submitting service would be in case of a successful submission and what you would see if the submission failed. POST SUBMISSION SAMPLE:

```
<?xml version="1.0" encoding="UTF-8"?>
<records>
<record>
<osti_id></osti_id>
<dataset_type>ND</dataset_type>
<title>ARM Climate Modeling Best Estimate Lamont, OK (ARMBE-CLDRAD SGPC1)</title>
<creators>Renata McCoy; Shaocheng Xie;</creators>
<creators_emails></creators_emails>
<related_resource></related_resource>
<product_nos>none</product_nos>
<contract_nos>AC05-00OR22725</contract_nos>
<other_identifying_numbers>sgpC1amrbe-cldrd-v3</other_identifying_numbers>
<availability></availability>
<contributor_organizations>Pacific Northwest National Laboratory (PNNL); Brookhaven National Laboratory (BNL); Argonne National Laboratory (ANL); Oak Ridge National Laboratory (ORNL)</contributor_organizations>
<publication_date>05/14/2012</publication_date>
<language>English</language>
<country>US</country>
<sponsor_org>USDOE Office of Science (SC), Biological and Environmental Research (BER)</sponsor_org>
<subject_categories_code>54 Environmental Sciences</subject_categories_code>
<keywords>Cloud fraction profiles; Total, high, middle, and low clouds; Liquid water path and precipitable water vapor; Surface radiative fluxes; TOA radiative fluxes</keywords>
<description>The ARM CMBE-ATM [Xie, McCoy, Klein et al.] data file contains a best estimate of several selected atmospheric quantities from ACRF observations and NWP analysis data.</description>
<site_url>http://iop.archive.arm.gov/arm-iop/0showcase-data/cmbe/cmbe/sgpC1/cmbe-cldrad/</site_url>
```

```
<doi></doi>
<file_extension>cdf</file_extension>
<software_needed></software_needed>
<dataset_size>12544 KB</dataset_size>
<contact_name> ARM Archive User Services</contact_name>
<contact_org> ORNL</contact_org>
<contact_email> armarchive@ornl.gov</contact_email>
<contact_phone> 888-276-3282</contact_phone>
</record>
</records>
```

POST SUCCESSFUL – SAMPLE RETURN MESSAGE

```
<?xml version="1.0" encoding="UTF-8"?>
<records>
<record>
<osti_id>1035366</osti_id>
<product_nos>none</product_nos>
<title>ARM Climate Modeling Best Estimate Lamont, OK (ARMBE-CLDRAD SGPC1)</title>
<contract_nos>AC05-00OR22725</contract_nos>
<doi>http://dx.doi.org/10.5439/1035366</doi>
<status>SUCCESS</status>
<status_message></status_message>
</record>
</records>
```

POST FAILURE – SAMPLE RETURN MESSAGE

```
<?xml version="1.0" encoding="UTF-8"?>
<records>
<record>
<osti_id>0</osti_id>
<product_nos>none</product_nos>
<title>ARM Climate Modeling Best Estimate Lamont, OK (ARMBE-CLDRAD SGPC1)</title>
<contract_nos>AC05-00OR22725</contract_nos>
<doi></doi>
<status>FAILURE</status>
<status_message>Data too long, maximum number of characters for dataset type is 2</status_message>
</record>
</records>
```

The “FAILURE” tells the submitting organization that no record was loaded into E-Link in this instance; that’s why the OSTI ID number is 0. There’s no DOI number assigned to the dataset because DOI assignment happens during processing...which never took place. The status message identifies the error that made the POST submission fail. In this case, if the submitted record had had a three letter code in the dataset_type field instead of the correct 2 character ND, this would have been the error.

It is the submitting site’s responsibility to review the returned messages, correct any errors, and resubmit the failed records.

