

# OSTI's Data ID Service

## A Web Service/API for Announcement Notice (AN) 241.6

Version 4.0, July 2014

### CONTENTS

- 1.0 [About AN 241.6 and the Data ID Service](#)
  - 1.1 [General Requirements](#)
  - 1.2 [Understanding Key Components](#)
- 2.0 [Using the Automated Data ID Service](#)
  - 2.1 [Authentication](#)
  - 2.2 [POST \(Create/Update\) and GET \(Retrieve\) \[Client to OSTI\]](#)
  - 2.3 [Response and Notifications \[OSTI to Client\]](#)
- 3.0 [Metadata for the AN 241.6/Data ID Service](#)
  - 3.1 [AN 241.6 Required Fields](#)
  - 3.2 [AN 241.6 Optional Fields](#)
  - 3.3 [Example of XML 241.6 Metadata Record](#)

### Appendix I

[Landing Page Examples](#)

## About AN 241.6 and the Data ID Service

---

Announcement Notice (AN) 241.6 provides to the U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) the metadata needed to identify/announce datasets resulting from work funded by the DOE or performed in DOE facilities. The submitted information also allows OSTI, via its Data ID Service, to assign Digital Object Identifiers (DOIs) to datasets and register them with [DataCite](#). 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 Data ID Service is the official name covering both the submission process for metadata about datasets and the DOI registration process for those datasets. The Data ID Service allows you to submit, edit, or retrieve dataset records. Users can manipulate records by performing HTTP operations on the web service URL and providing XML metadata to create/update records (POST) or to retrieve records for review (GET). The web service allows automated submission of 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.

## 1.1 General Requirements

These general requirements have been established for data registration. The submitter must:

- Provide, at minimum, the required metadata to enable basic, bibliographic citation
- Have authority (though not necessarily ownership) to make public the assets for which they wish to assign the DOIs.
- Guarantee the persistence of registered data
  - By ensuring that data will be stored and managed such that access and usability are provided indefinitely.
  - By maintaining and updating as necessary all URLs associated with the DOI
- Establish and maintain appropriate landing pages for registered DOIs

Note that the datasets themselves are not uploaded or stored at OSTI.

## 1.2 Understanding Key Components

### Defining Datasets: Content and Granularity

Determining how DOIs should be applied to your data, i.e. at what level of granularity, is an important step in planning data registration. Some datasets are similar to collections in that they include multiple data files. The DOI can be assigned at the collection level, with the included data files listed individually on the dataset's landing page. Another option is to break the collection down in some logical way by topic, by geographic location, by time period, etc. and assign a DOI to each of these smaller "sets" of multiple files. On the other extreme, a dataset may also be as simple as a spreadsheet.

Datasets are not always numeric in nature. A computer model and sets of "canned" input could be a dataset, while each of the simulations or output files could be assigned their own DOIs. A collection of photographs or a graphic of a workflow can be registered as a dataset. Interactive resources or audiovisual items may be assigned DOIs. Data "studies" can be registered also; these may include a variety of materials associated together as a "package."

Subject expertise and knowledge of how your audiences normally look for your data will help you determine the appropriate content and boundaries of the datasets that you intend to register. OSTI can provide advice and share examples of best practices, but granularity decisions are ultimately the responsibility of the people who know the data best, i.e. the creators, the analysts, or the data center experts at your submitting organization.

## The Digital Object Identifier (DOI)

Each time a new submitting entity becomes a new “data client,” OSTI requests from DataCite a numeric prefix which will belong only to that one, specific entity. A data client may be a data center, a lab division, or a facility. Often, the data client is defined as one specific project. DataCite and OSTI’s Data ID Service ensure that no two data clients, regardless of how they are defined, will ever have the same DOI prefix.

A DOI consists of the number 10, followed by a period, followed by the data client’s prefix, a forward slash, and a numeric suffix. The suffix assigned by OSTI is the OSTI ID, a unique identifier that every record processing through any of OSTI’s systems must receive. If needed, data submitters may discuss with OSTI, prior to beginning production, the option of including their own unique character string to be inserted into the center of OSTI’s DOI format.

What makes the DOI not only a unique identifier, but a “persistent” identifier is the commitment behind it. Data clients commit to update the URLs for their DOIs whenever a posted landing page and its associated dataset must be moved to a new online address. The DOI will never change, but the URL, which is invisibly “bonded” to a DOI in the DataCite Registry, *can* change. Changing the URL in DataCite’s Registry is accomplished through a simple POST operation to update the affected record.

DataCite recommends that DOIs in a citation or database be displayed as live links, i.e. with <http://dx.doi.org/> preceding the beginning of the actual DOI numbers or characters. This is not a requirement, but it is a “best practice.”

Example: <http://dx.doi.org/10.5439/1021460>

## The Landing Page

A “landing page” or introductory page is recommended by DataCite for all registered data and actually required in two cases (See next paragraph). The DOI points to the landing page, which in turn, links out to the dataset itself. The landing page, which must always be publicly accessible, provides context for the dataset. The landing page may contain one or more of the following:

- Information on how a full, formal citation of the data should appear,
- Access information, including a notice of temporary restriction, if a dataset is not yet ready for public access.
- Links to software or toolkits needed to open, download, or analyze the data,
- A listing of the individual data files that may be included in the registered dataset (with links to them),
- Update and version information,
- Contact information.

DataCite notes two cases when a landing page is mandatory rather than recommended. A landing page is mandatory for any data that cannot be viewed using standard desktop software. A landing page is also mandatory when the data temporarily has restricted access. Because DOIs are not deleted from the DataCite Registry, a landing page can also serve as a “tombstone page” when a dataset must become unavailable. The tombstone page should tell a user who has followed the original DOI to that point what has happened to the data, why, and provide a contact for further questions.

In cases where neither a landing page nor a dataset can be accessed yet, but a DOI is needed to share among members of a project or collaboration, submitters may discuss with OSTI how to “reserve” a DOI for future registration.

[See examples of good landing pages at the end of this document.](#)

## 2.0 Using the AN 241.6 Web Service

---

### 2.1 Authentication

Access to the OSTI Data ID Service requires the submitting site to have an active, approved account through an approved web service account for [E-Link](#). Once an account is established, the account holder then submits records to a test environment and does a coordinated review of the results with OSTI before POSTing the first “live” file in production mode. The test URL is <https://www.osti.gov/mlinktest/2416api>. OSTI’s production web service for announcing datasets is available at <https://www.osti.gov/mlink/2416api>.

OSTI’s AN 241.6 Web Service supports HTTP Basic authentication over Secure Socket Layer (SSL), and each POST or GET request requires authentication. With this method, the client connects to one of the HTTPS URLs; the POST and GET verb commands will pass along the standard Authentication HTTP header (base64 encoding).

### 2.2 POST (Submit/Update) and GET (Retrieve) [Client to OSTI]

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. An error message will be included for each failed record, as well.

The POST request is used to submit metadata for a new record that will receive a new DOI. It is also used to edit or update an existing record that has previously been submitted. When POSTING to an existing record in order to update it, the OSTI ID must be included in the XML record. The updated information will then overwrite the

previously submitted information. **NOTE: If the OSTI ID is not included, the Data ID System will attempt to re-register and obtain a new DOI. This situation requires significant effort to correct.**

Metadata can be retrieved for any record previously submitted by your organization 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/Update command. The requested OSTI ID returns that specific XML record. A GET operation will not update a record, however.

### 2.3 Response and Notifications from the AN 241.6 Web Service [OSTI to Client]

OSTI's server makes an immediate response when the client-side server sends a POST or a GET request. The primary goal of the response is two-fold: let the client server "know" if it connected successfully and provide a status message about the record that was POSTed or retrieved. The status says SUCCESS if the record correctly transmitted and the metadata loaded into the processing system. The status says FAILURE if the record was unable to load into the processing system for any reason. The FAILURE response includes a status message indicating the error or problem that kept the record from loading. Errors leading to a FAILURE response occur when a required metadata field is missing, when a site URL cannot be validated as correct by the automated validation process, when a value is too long for the field it must load into, etc.

An example of each response type appears below. Note that the successful record returns its newly assigned OSTI ID and its newly assigned DOI. The response for the failed record has neither, since OSTI IDs and DOIs cannot be given to records that never actually make it into the processing system.

POST SUCCESSFUL – SAMPLE RESPONSE FROM AN 241.6 WEB SERVER

```
<?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 RESPONSE FROM AN 241.6 WEB SERVER

```
<?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 characters for dataset type is
2</status_message>
</record>
</records>
```

Once testing is complete and the submitting site is working in the production environment, OSTI also sends an automated email to those at the site that wish to be on the distribution list. The email reports the same information as the server response message but can reach a wider audience, if necessary.

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

## 3.0 AN 241.6 Web Service Metadata

---

The following two tables list the required and optional metadata fields, their XML tags, and the related business rules governing submission to the AN 241.6 Web Service. Records without required fields will fail to load will not receive DOIs. The other fields available for use are optional, though several, such as the Description, are highly encouraged. Third party indexing services often integrate with the DataCite Registry to reflect the DOIs and associated information in their own products. The more metadata these services can pick up from DataCite, the more retrievable your datasets will be.

### 3.1 AN 241.6 Required Fields

The following table lists the required fields for the AN 241.6 Web Service. Records without required fields will fail to load into E-Link for processing. They will not receive OSTI IDs or DOIs. The OSTI ID is required 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.

Please note that there are two ways to submit Creator/PI/Author names. These two methods cannot be combined in one record. A record must use either the single tag

<creators> and combine all names into one character string **OR** if any of the names have an accompanying ORCID, a record must use the <creatorsblock> which includes several tags for each creator name.

**REQUIRED**

	<b>Field Name</b>	<b>XML Tag Name</b>	<b>Additional Information</b>
<b>1</b>	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 web service.
<b>2</b>	Site Code	N/A	Automatically determined by the authenticated E-Link User account
<b>3</b>	Dataset Type*	<dataset_type/>	Dataset Type refers to the <b>main</b> content of the dataset. Only one value is allowed. Use the two-letter code shown below:
		<p><b>Two Digit Code</b></p> <p><b>AS</b> Animations/Simulations</p> <p><b>GD</b> Genome/Genetic 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</p> <p><b>IM</b> Interactive Data Map(s) – A non-static interface and the GIS data and/or shape files that generate it.</p> <p><b>ND</b> Data primarily expressed with numbers; other content is secondary and supporting.</p> <p><b>IP</b> Still Images or Photos – Images or photos 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.</p> <p><b>FP</b> Figures/Plots – These datasets are composed mainly of data diagrams, charts, drawings, and data plots.</p> <p><b>SM</b> The key factor for this dataset type is that it’s more than just a mixture of “stuff” on a Web page. It has structure, organization, search and retrieval capability, and the way the information is put together is what gives it meaning. The information does not exist elsewhere except in pieces. There should be some non-text content in the mix as well.</p>	
<b>4</b>	Dataset Title	<title/>	Full title of the dataset. Include version numbers and specific date ranges when applicable.
<b>5.1</b>	Creator(s)/ Principal	<creators/>	You may use this field <b>if you are NOT submitting ORCID IDs with any of the</b>

	Investigator(s)*/ Authors		<b>creator/PI (or author) names</b> in the record. Format for names is last name, first name, middle initial. Multiple creator names, separated with a semi-colon and a space are allowed.
5.2	Creator(s)/ Principal Investigator(s)*/ Authors	<creatorsblock> <creators_detail> <first_name></first_name> <middle_name></middle_name> <last_name></last_name> <affiliation></affiliation> <private_email></private_email> <orcid_id></orcid_id> </creators_detail> </creatorsblock>	Use the <creatorsblock> tags <b>if you ARE including ORCID number with any of the creators/Pis (authors)</b> in the submitted record.  When the submitted record has the <creatorsblock> tags, <b>ALL</b> creator/PI/author names <b>MUST</b> be broken out into this block format ( <b>even if only one of the names has an accompanying ORCID</b> ).
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 <a href="https://www.osti.gov/mlink/authorities.jsp">https://www.osti.gov/mlink/authorities.jsp</a> 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"> <li>• mm/dd/yyyy</li> <li>• yyyy</li> <li>• yyyy Month</li> </ul>
10	Language*	<language/>	E-Link will default to English. Authority values are available for other languages at <a href="https://www.osti.gov/mlink/authorities.jsp">https://www.osti.gov/mlink/authorities.jsp</a>
11	Country of Origin/ Publication*	<country/>	E-Link will default to U.S. If other country names are needed, Authority values are available at <a href="https://www.osti.gov/mlink/authorities.jsp">https://www.osti.gov/mlink/authorities.jsp</a>

12	Sponsoring Organization(s)*	<sponsor_org/>	Use the spelled-out text as shown in the Sponsoring Organization Authority at <a href="https://www.osti.gov/mlink/authorities.jsp">https://www.osti.gov/mlink/authorities.jsp</a> 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.
13	Site URL*	<site_url/>	OSTI cannot accept, store, or post datasets. The URL should link to a valid html "landing page" for the dataset.
14	Contact Name and Position*	<contact_name/>	Admin info only; it will not be displayed in public databases.
15	Contact Organization*	<contact_org/>	Admin info only; it will not be displayed in public databases.
16	Contact E-mail*	<contact_email/>	Admin info only; it will not be displayed in public databases.
17	Contact Phone	<contact_phone/>	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 Description, is highly encouraged, however.

#### OPTIONAL

	Field Name	XML Tag Name	Additional Information
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 <a href="https://www.osti.gov/mlink/authorities.jsp">https://www.osti.gov/mlink/authorities.jsp</a> . 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/>	Words or phrases or both may be included. Separate multiple entries with a semi-colon and a space.
8	Description	<description/>	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 <b>if one has already been registered prior to the dataset being submitted to OSTI.</b>
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.

### 3.3 Example of an XML 241.6 Metadata Record

Below is an example of a 241.6 dataset record in XML format as it would be POSTED to OSTI's 241.6 web service. This would be treated by the Data ID Service as a new record and would receive an OSTI ID and be assigned a DOI. These two identifiers would be returned to the submitting organization in both the server response and the results email.

```

<?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_nos>sgpC1amrbe-cldrd-v3</other_identifying_nos>
<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>

```

## Appendix I

### Landing Page Examples from CXIDB and ARM

(See images on the next page)

CXIDB ID 5

Citation Details

<b>Title:</b>	High-resolution x-ray diffraction microscopy of specifically labeled yeast cells
<b>Authors:</b>	Johanna Nelson et al.
<b>Journal:</b>	PNAS
<b>Year:</b>	2010
<b>DOI:</b>	<a href="https://doi.org/10.1073/pnas.0910874107">doi:10.1073/pnas.0910874107</a>

Description

This is the second of five exposures of the same sample at different tilts. This one is at +15 degrees tilt.

Check CXI IDs 4 to 8 for the complete set.

Experimental Conditions

<b>Method:</b>	Single Particle X-ray Diffraction Imaging
<b>Sample:</b>	Gold labeled frozen dried <i>Saccharomyces cerevisiae</i> yeast cells.
<b>Wavelength:</b>	1.653 nm
<b>Lightsource:</b>	ALS
<b>Beamline:</b>	9.0.1

Deposition Summary

<b>Depositor:</b>	Johanna Nelson
<b>Contact:</b>	<a href="mailto:jjnelson@slac.stanford.edu">jjnelson@slac.stanford.edu</a>
<b>Deposition date:</b>	2011-07-22
<b>Last modified:</b>	2011-07-22

Data Files

Diffraction Pattern: [cxidb-5.cxi](#)

Link to actual dataset

© 2014 CXIDB. All rights reserved.

Designed by Filipe Maia based on Zenlike.

The screenshot shows the ARM Climate Research Facility website. The main navigation bar includes: About, Science, Campaigns, Sites, Instruments, Measurements, Data, News, Publications, Education. The current page is titled "Datastream : AOSPC" and provides the following information:

- AOS:** condensation particle counter
- Active Dates:** 2011.03.09 - 2013.07.14
- Measurement Categories:** Aerosols
- Originating Instrument:** Aerosol Observing System (AOS)
- Measurements:** Only measurements considered **scientifically relevant** are shown below *by default*.
 

Measurement	Units	Variable
Aerosol concentration	parts/cc	concentration ( time )
Particle concentration		
- Locations:**
  - North Slope Alaska (NSA)
  - Southern Great Plains (SGP)
  - ARM Mobile Facility (PGH)
  - ARM Mobile Facility (PVC)

On the right side of the page, there are several utility sections:

- Documentation:** Data Quality Plots
- Citation:** DOI: 10.5439/1025152. Includes a "GENERATE CITATION" button.
- Order Data:** BUILD AN ORDER button.
- Comments?:** A form for providing feedback.

Note DOI and a link to text explaining how DOIs are assigned to ARM data

To order specific data files included in this datastream