

Science Data Catalog API Ingest Service (SDC-CMS)

Overview for USGS Data Stewards

Overview

1. Acronyms
2. USGS CMS
3. Data as a Content Type
4. SDC-CMS Overview
5. How Can Data Stewards Help

Acronyms

- CMS: USGS content management system
- SDC: USGS Science Data Catalog
 - API: application programming interface
 - JSON: JavaScript object notation
- SDC-CMS: Science Data Catalog-Content Management System API Ingest Service

USGS Content Management System (CMS)

- Combines and centralizes web content and systems management and enables integration of other USGS information systems (e.g., Science Data Catalog)
- Starting in 2016 all public USGS web sites (n=300+) were required to migrate to the CMS.
- In FY2018 the ELT and the Deputy Director began advising all remaining sites they are required to start, if not complete, migrations in FY18.
- The Director is preparing to set a hard deadline for migration completions.

USGS CMS (cont.)

There are currently ~180 sites still migrating and two of the struggles we hear about are:

- Migrating existing datasets that pre-date current data release policies,
- Managing data release content

Data Stewards are, therefore, critical to the immediate migration as well as the long-term maintenance of USGS web sites.

USGS CMS (cont.)

If you take away only 2 things from this presentation:

- Migrating existing, previously public datasets requires nothing more than verifying/re-hosting the dataset(s) so the content manager to proceed.
- Managing data release content creation and management is now automated via the SDC public API and the SDC-CMS service minimizing work for content managers.

Data as a Content Type

DT-Section	DT-Field
Content	Intro
Content	Non-USGS Funding Source
Content	Release Date
Content	SDC Catalog ID
Content	Thumbnail
Content	Title
Content	Type
Content	URL
Overview	Owned By
Records Management	Is a record
Records Management	Is part of litigation hold
Taxonomy	Keywords
Taxonomy	Mission Areas
Taxonomy	Programs
Taxonomy	Science Centers
Taxonomy	Science Support
Taxonomy	Special Topics

The Data and Tools content type:

- Is simple. Content items do not have a public view page, only a link to the DOI-based landing page.
- Has 17 fields total, only 6 of which are managed by SDC-CMS.

SDC-CMS: Service Overview

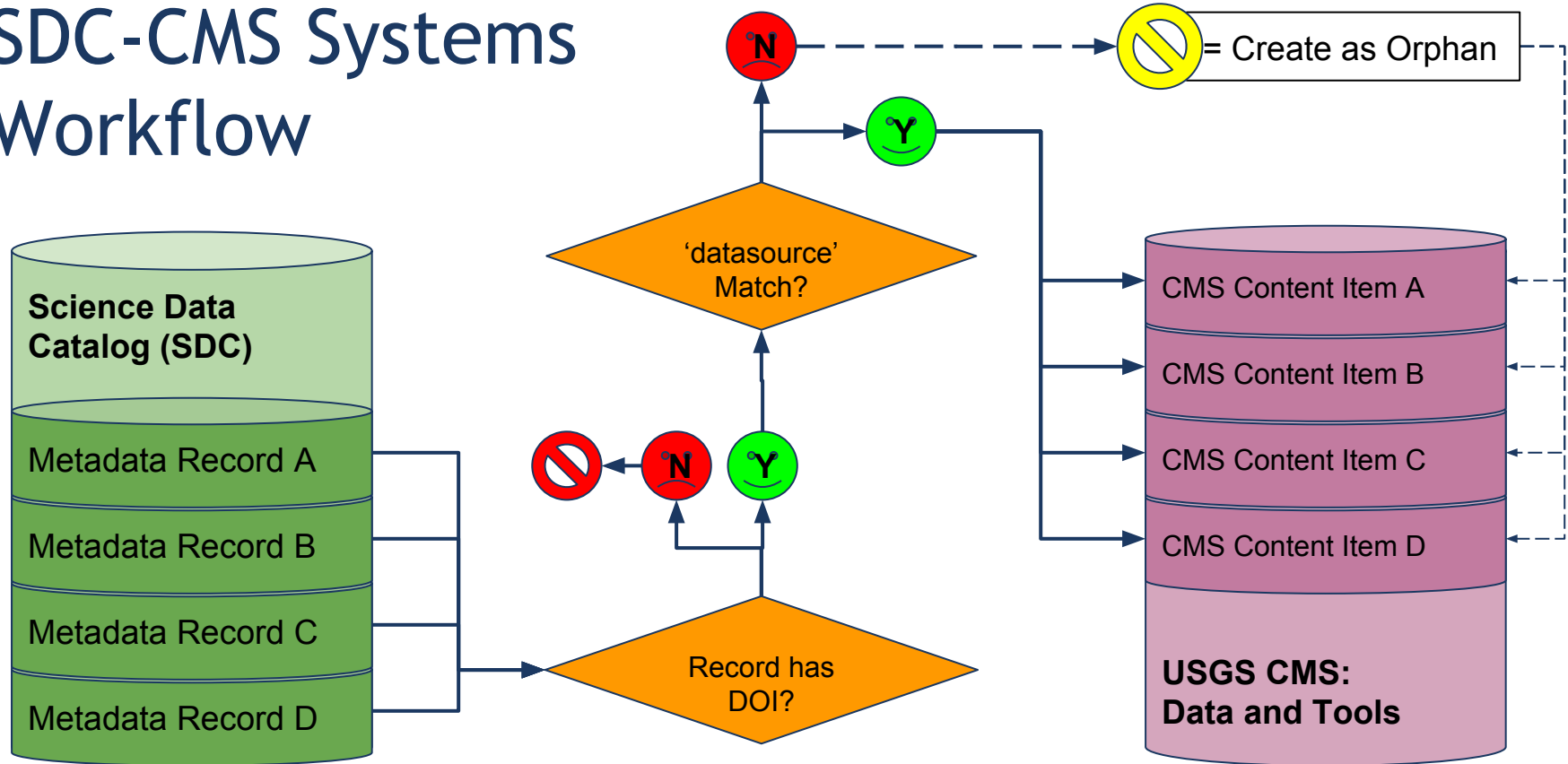
3 Steps to create and maintain CMS dataset records:

1. CSDGM metadata file added to the SDC catalog.
2. SDC-CMS identifies new/updated records and creates/updates CMS items.
3. Content manager simply identifies the dataset content in the CMS, verifies and tags, and publishes.

SDC-CMS: Field Mapping

SDC-JSON-Field	Processing-Action	DT-Section	DT-Field
abstract	TRIM to 400 characters and WRITE to 'Intro'	Content	Intro
--	--	Content	Non-USGS Funding Source
pubDate	WRITE to 'Release Date'	Content	Release Date
--	CONCATENATE "[data_url_format]&[Title]" and WRITE to 'SDC Catalog ID'	Content	SDC Catalog ID
--	INSERT USGS Logo/Placeholder	Content	Thumbnail
title	WRITE to 'Title'	Content	Title
--	SET to "Science Datasets"	Content	Type
web_url	WRITE to 'URL'	Content	URL
ORGdatasource	IF 'ORGdatasource' matches a CMS group/microsite name TAG 'Owner' to that science center Group.	Overview	Owned By
--	--	Records Management	Is a record
--	--	Records Management	Is part of litigation hold
keywords	WRITE keywords as a comma delimited string to 'Keywords'	Taxonomy	Keywords
--	--	Taxonomy	Mission Areas
--	--	Taxonomy	Programs
ORGdatasource	IF 'ORGdatasource' matches a CMS group/microsite name TAG 'Science Center' to that science center Group.	Taxonomy	Science Centers
--	--	Taxonomy	Science Support
--	--	Taxonomy	Special Topics

SDC-CMS Systems Workflow



How Can Data Stewards Help?

1. Keep doing what you're doing! High quality CSDGM files are ultimately the source for the CMS. Two common recommendations:
 - a. Only use the DOI-based URL
 - b. Case sensitive keywords of all types
2. If you haven't already, reach out to your content manager(s) and let them know you are available to help.
3. Help content managers find new hosts for data from an old USGS web site.

Questions + Answers = Discussion

Lance Everette

everettel@usgs.gov
970-226-9225