



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

CSW Best Practices

ESIP Summer Meeting 2016

July 19-22 2016

Doug Newman (NASA Earthdata – Raytheon)

Andrew Mitchell (NASA Earthdata)

The material is based upon work supported by the National Aeronautics and
Space Administration under Contract Number **NNG15HZ39C**

Agenda

- What is CSW?
- Why CSW and Best Practices?
- GetCapabilities
- GetRecords
- GetRecordById
- DescribeRecord
- GetDomain

WHAT IS CSW?

Catalog Service for the Web

From OGC website:

‘Catalog services support the ability ... search collections of descriptive information (metadata) for data, services, and related information objects. ... Catalogue services are required to support the discovery and binding to registered information resources within an information community.

OGC Catalog interface standards specify the interfaces, bindings, and a framework for defining application profiles required to publish and access digital catalogs of metadata for geospatial data, services, and related resource information...’

WHY CSW & BEST PRACTICES?

Why?

- It is a widely-used earth data discovery standard
- CWIC maintains a CSW framework for interoperability
- OGC specs are verbose and, in our opinion, difficult to interpret. Best practices may help

GETCAPABILITIES

Take this seriously

- Many implementations do a poor job at describing their service properly
- This method is key to getting clients to use your API
- This is the gateway to your API. If you do this poorly then confidence will be low with respect to the rest of your API

GETRECORDS

Support GET

- GET is more user-friendly
- But complex queries are difficult with GET
so...
- Support a subset of your queries with GET
 - Free text
 - Bounding box
 - Temporal range

Is summary necessary?

- Getting a list of ids seems useful
- Getting some intermediary representation of the metadata seems a throw back to the days when 'bandwidth was tight'

GETRECORDBYID

Is this necessary?

- It is mandatory but...
- GetRecords allows you to get the full metadata associated with a result set.
- Might be useful from a 'canonical metadata reference' context but that should really be the domain of other APIs or landing pages

DESCRIBEDRECORD

Of little use

- The ability to return schemas that are publicly available and defined in the GetCapabilities response seems unnecessary
- Being able to subset that schema seems a throw back to times when bandwidth was tight.

GETDOMAIN

Of some use

- For controlled vocabularies this is useful functionality.
- Elsewhere, not so much

IN GENERAL

Support GET wherever you can

- GET is much more user-friendly than POST.
- Reserve POST support for complex queries like GetRecords

Support formats beyond minimum

- OGC mandates the GMD ISO profile. This is somewhat out of date

Showing it's age

- A lot of the architectural considerations of CSW have made based on the way the world was some years ago.
- Verbose metadata results are no longer a problem
- One would hope 3.0 improves this but...

Comparison with OpenSearch

- It's not nearly as much fun
- It's not RESTful – no hypermedia
- The connection between collection and granules is not tightly coupled
- Specification of API is not as compact (GetCapabilities + GetDomain \approx OSDD)

QUESTIONS?

This material is based upon work
supported by the National
Aeronautics and Space
Administration under Contract
Number **NNG15HZ39C.**

Raytheon