

[Earth Observing System \(EOS\) Clearing House \(ECHO\) Working Session](#) **[1]**



Submitted by ktbaynes on Tue, 2014-01-07 11:45 Friday, January 10, 2014 - 10:30 to 12:00

Event: [Winter Meeting 2014](#) [2]

Session Type: [Workshop](#) [3]

Abstract/Agenda:

This session/discussion will start with a presentation of OpenSearch usage led by Doug Newman.

Following Doug's talk, we will discuss the past year in ECHO's activities including our participation in NASA's ISO 19115 activities, enhancing our Near-Real-Time data capabilities, revamping our order workflow and recent additions to our visualization capabilities in Reverb. If there is interest we will be having hands-on discussions about ECHO API usage.

We will finish out our session by looking forward towards the next year of ECHO activities and our future involvement with the Common Metadata Repository and discussing our plans and vision for this new system

Notes:

Maintaining the momentum of OpenSearch in Earth Science data discovery ++ - Doug Newman (NASA ECHO)

- ECHO - 6 million queries to api each year - one of the api is OpenSearch
- OpenSearch
 - o Is a collection of simple formats for the sharing of search results
 - o Earth Data discovery use case
- § Descriptor document - explains to client how to http get request
- § HTTP Response - puts together unique inventory with ID and include further search links
 - o 2 step approach
 - Earth Data OpenSearch today
 - o ESA - next gen
 - o CEOS - in CWIC including ESA
 - o ESIP Federation
 - o NASA ECHO - restful services in general
 - o Metrics says that since went to REST like interfaces

- o 10K/week in 2011 to 115K/week in 2013
- Why successful
- o Lightweight and simple
- o Standard-based
- o RESTful
- o Low entry cost
- o 'free text + spatial + temporal' satisfied 90% of all search criteria
- Metrics
- o Controlled vocab are usually important for science data discovery - this is not the case
- § 52% free text
- o Q (Thomas)- while a lot of people use free text - science key works increases ranking - because people are indexing vocabularies with free text - can abstract key works behind the scenes - affects ranking
- Maintaining success
- 1. Converge where possible
- o Three payers for OpenSearch
- § ESIP discovery cluster
- § CEOS/CWIC
- § OGC
- o Try to converge when possible
- o All agree on
- § Free_text
- § Bounding_box
- § State_date
- § End_date
- o OGC
- § Uid
- § Place_name
- § Geometry - pt, line, polygon
- § Now in ESIP best practice
- o Differ

§ OGC includes relations

§ Described_by (link to metadata) but OGC uses 'via' (because of legacy issues)

o

· 2. Free text + spatial + temporal = success

o Pro: 90% (Reverb metrics for the last year)

o Free text is not as accurate as a controlled vocabulary

o Q - really depends on the user - they don't want to do a controlled key word search and go right to the data. Controlled key words are likely from new users

§ At the moment, OpenSearch's lack of controlled vocabulary is not hurting things at the moment

o Can rank results, which does increase the relevant results

§ Pushing the ranking option in best practices

· 3. Understanding the API

o Parameter extension is good now that it is updated

o Defining "free text" search

o Thomas - can you do googly like things to define search terms

§ No - suggested associate it with a profile (ex. Lysine)

o Define 'geometry' capabilities

· 4. Additional functionality

o Result ordering - OGC is started to look into it - ex. Creation date

o Result ranking, added ECHO in Dec 2013

o Faceted search

§ Looked at how google does it... maybe look at how amazon do it... makes it easier with bin - hopefully in the next year

· How achieve these 4 goals

o Have small group of people working together

ECHO Overview and State of the System - Katie Baynes

· Metadata clearhouse for NAASA EOS (earth observing satellites)

o not data, point to data

· Reverb

- o Rever.echo.nasa.gov/reverb
- o Do order brokering - generate scripts for download for bulk retrieval
- o Services - sub-setting and re-projection
- Metadata Ingest
- o When people provide data they have a number of ways to come in, change as move to CMR (common metadata repository)
- Search and Metadata Retrieval
- Metrics (as of Nov 2013)
- o 2820 Collections, 142 million granules
- o Client api
- § Reverb
- § ASTER volcano group
- § GHRC data albums Access
- § OpenSearch Clients (CWIC)
- § GIBS (Global imagery browse system)
- § Non-self-identifying clients - choose to remain obscurity
- o Raising traffic over time
- § Seen a few spikes
- o Who is using the data changes week to week - trying to identify
- Thomas - which DAAC is your largest user - don't have one... DAACs don't query
- o There are some DAACs that have their own clients that also hit echo (Giovanni)
- o Doug - Those are just pass through for other uses - partition by data provider doesn't make sense
- o Thomas - looking for statistics of which DAAC is higher API user
- o Exclude our own queries (ex. Doug's query)
- Q in what way can you determine who the other are
- o Can pin via reverse look-up - know ip of some of the ones we get
- o Can break down by country - but usually lump into other
- Q - opensearch client ID? Has this come up
- o Fix that by the interface to the descriptor document (url) is imbedded - that is an ECHO specific thing - ACTION
- Metric (cont)

- o 12 week moving average - go now - getting faster
- o Spatial heat map - all of the bounding box point searchers using 1 degree by 1 degree - polar searchers are going to NSIDC people
- § Make assumptions of how system is being used - don't know until you look at metrics
- o How people are using reverb - crazy egg that tracks origin, time of day, day
- § People use save query - give you a bookmark
- Work on serving the right people with the right tool - which will move into the future
- o Technical interchange - on site near Goddard in Spring 2013
- § Show case and get into api & working with data partners
- § 2 tracks - development (sample code)
- o NRT (near real-time) data
- § Operation with 4 out of 5
- § Provides a death date for granules
- § Working on standardizing
- § Added a back end to ECHO - if not NRT then it is science quality - have 58 or 59 NRT (2/3 of the data)
- o Order Workflow Re-work
- § Change mock up to reverb work flow - simply sets for ordering process
- o Indexing Landsat 8 Data
- § New visualization - image view - granule level data and show gridded format to see images
- § Can see yearly and 5 and 10 yr - see what is available
- Q would this just be for LANDSAT 8 - no any granule will be displayed this way
- o Revisiting User Settings
- § Use of save query
- § Relevance sorting
- § Some want it on a map or timeline
- § Eventually will let people set preferences - above and order status
- o ISO Activities
- § Following MENDS activity
- § Learning more about community use and adoption

§ Can give api extension call and get it in the NASA flavor

- `echo@ echo.nasa.gov`
- <http://goo.gl/IRvY70> [4]

Common Metadata Repository

- URS, GIBS, EMS (metrics), discovery tools, DAACs, NRT
- o CMR is in the middle - provide a vehicle where ties systems together
- o This is not yet another metadata catalogue
- CRM builds on top of GCMD and ECHO as authoritative source
- Client api will continue to function (backwards compatible and expose the api)
- Also have a suite of standard api as well
- CMR will have resources built around it
- WHY

o Emphasis on metadata quality

§ ECHO, GCMD each have holdings - there are problems with the metadata

§ Ex. Link with a ton of html describing it

- o Goal of the crm is to clean up and fix metadata
- In metadata input - have a quality review loop
- o Before it was programmatic
- o There will be automated quality check

§ New and hasn't been seen

§ Technically invalid

§ Not good enough for Ted

- Gets put into a queue for review
- o Q talking about collection or file level

§ Collection level

- Plan
- o Diff support added to ECHO
- o GGCMMD is programmatically monitoring ECHO for new Collection and Updates - in place now
- o Assessment and correlation queue between ECHO and GCMD holding - this has started

- o Initial pass will associated ECHO 10 Collections with GCMD DIFs – STARTING
- o Science coordinators will assess difference between associated records and work from easiest to hardest with providers
- o Q Luther – is there any difference to the provider
- § Not yet – there likely will be as go forward – if producing 2 records work with you to get this sorted
- o Q – can't say there wont be any breakage in the work flow – import a small set of data – identifying work flow from DAACs to see how they are submitting – limit the breakage in workflows
- § Katie – not a breakage, but retire FTP ingest – interested in time lines – need to change to RESTful
- o Q (Luter) – question about push back
- § Last week...
 - Handling the changes
- o Science coordinate with data providers to get this change at the source
- o In cases where that isn't feasible – also looking at ingest adapters (with provider knowledge) – for issues where can't affect source metadata immediately
- o Science coordinator can make changes (with provider approval)
- Initial Metadata targets
- o LANCE (NRT), then Measures, MODAPS
- Emphasis on performance
- o Almost 4 K in ECHO + 24K in GCMD – granule 140 m granule
- o 500 k granules/week
- o MODIS 6 = 100K /day
- o Reverb Ops 18% in Oct, - there are enough clients on ECHO that there are more spikes in data – need to scale up to usable
- o GIBS/world view – asks ECHO for science granules for image... users pan & move dates... Worldview needed faster search performance
- § Trying to get that performance – currently at 3 sec ... trying for sub-second search time
- Transition
- o CMR is going to build on ECHO and GCMD capabilities
- o Initial requirements on MASI and MASII studies
- o Reconciliations piece has staretd
- o Focus on search performance category
- o Progressively replacing component until replace with CMR component – no double system

- Metadata Concept Support
 - o Sky Bristol talked about services for metadata, Lynnus for granule metadata
 - o Science metadata (collections and granules) and services (clients...)
 - o Start looking at visualization - being driven by GIBS and Worldview
 - o Parameter level is similar
 - o Future - want to be able to expand as system grows
 - o UMM - is a mapping (not a format) - tell us where things are in a given record - then can provide things in ISO 19115
 - o Then can create aggregate records
 - o Lynnus - talked about key word tagging of granule and moving in time and space - move beyond just automatic tagging of weather events, but users can tag for each other - social tagging
- How affect you
 - o Metadata providers
 - § Expect more hands on world
 - § Not need to reconcile multiple systems
 - § Most will continue to work - except FTP upload tool
 - § Current formats continue to function - except where things are missing, need to augment - will be a smooth transition
 - § Enforce controlled vocabulary
 - § Will have voice through Technical Committee
 - o Client Developer
 - § Access more metadata and of better quality
 - § Existing API will be supported (try not to break)
 - § Standards-compliant APRI will be available...
 - § More metadata will be in ISO19115
 - o End User
 - § Unified view of US holdings, regardless of api or system
 - § Higher quality and more metadata
 - § Start looking at data casting
- Q (Luther) - with existing API being support - if building to ECHO api - is going to CMR api going to be an upgrade in quality

Earth Observing System (EOS) Clearing House (ECHO) Working Session

Published on Commons (<http://commons.esipfed.org>)

- o No numeric info yet – might be peroxided – minimize the amount of overhead – don't yet have recommendations yet
- o From performance perspective, trying to be the same in many
- o Except – if currently using SOAP – discourage using this – it might not make it to CRM – it doesn't see the usage – will work with clients if they are using it (currently only CWIK for csw collector – going to OpenSearch.. then will stop usage)

Actions:

- opensearch client ID- Fix that by the interface to the descriptor document (url) is imbedded – that is an ECHO specific thing - ACTION

Attachments/Presentations:  [ECHOPresentation_ESIPWinter2014.pptx](#) [5]

 [Open Search - ESIP 2014.pptx](#) [6]

 [CMR ESIP Overview Winter 2014.pptx](#) [7]

Session Leads:

Name: [Katie Baynes](#) [8]
Organization(s): [NASA](#) [9]
Email: kathleen.baynes@nasa.gov [10]

Name: [Doug Newman](#) [11]
Organization(s): [NASA](#) [9]
Email: douglas.j.newman@nasa.gov [12]

Name: [Daniel Pilone](#) [13]
Organization(s): [NASA/ EED / Element 84](#) [14]
Email: dan@element84.com [15]

Notes takers:

Name: [Kelly Monteleone](#) [16]
Organization(s): [University of New Mexico](#) [17]
Email: krbm@unm.edu [18]

Participants:

Kelly Monteleone, Thomas Cherry, Jason Gilman, Jeffy Lee, Bently Hela, Minnie Way, Charlie Martin, Emily Worthup, Christian Alarcon, Deann Muller, Ed Armstong, Tyler Stevens, Scott Ritz, Archana Subramanian, Brett McLaughlin, Luther Lightly, Andrew Mitchell, Justin Goldstein

Creative Common License: Creative Commons Attribution 3.0 License

Accepted:

Keywords: [ECHO](#) [19]

[OpenSearch](#) [20]

[metadata](#) [21]

[CMR](#) [22]

[API](#) [23]

Source URL: <http://commons.esipfed.org/node/1996>

Links:

[1] <http://commons.esipfed.org/node/1996>

[2] <http://commons.esipfed.org/taxonomy/term/1029>

[3] <http://commons.esipfed.org/session-type/workshop>

[4] <http://goo.gl/IRvY70>

[5] http://commons.esipfed.org/sites/default/files/ECHOPresentation_ESIPWinter2014_2.pptx

[6] http://commons.esipfed.org/sites/default/files/Open%20Search%20-%20ESIP%202014_0.pptx

- [7] http://commons.esipfed.org/sites/default/files/CMR%20ESIP%20Overview%20Winter%202014_0.pptx
- [8] <http://commons.esipfed.org/node/577>
- [9] <http://commons.esipfed.org/taxonomy/term/228>
- [10] <mailto:kathleen.baynes@nasa.gov>
- [11] <http://commons.esipfed.org/node/521>
- [12] <mailto:douglas.j.newman@nasa.gov>
- [13] <http://commons.esipfed.org/node/1973>
- [14] <http://commons.esipfed.org/taxonomy/term/1306>
- [15] <mailto:dan@element84.com>
- [16] <http://commons.esipfed.org/node/430>
- [17] <http://commons.esipfed.org/taxonomy/term/297>
- [18] <mailto:krbm@unm.edu>
- [19] <http://commons.esipfed.org/taxonomy/term/568>
- [20] <http://commons.esipfed.org/taxonomy/term/288>
- [21] <http://commons.esipfed.org/tags/metadata>
- [22] <http://commons.esipfed.org/taxonomy/term/1226>
- [23] <http://commons.esipfed.org/taxonomy/term/1227>