



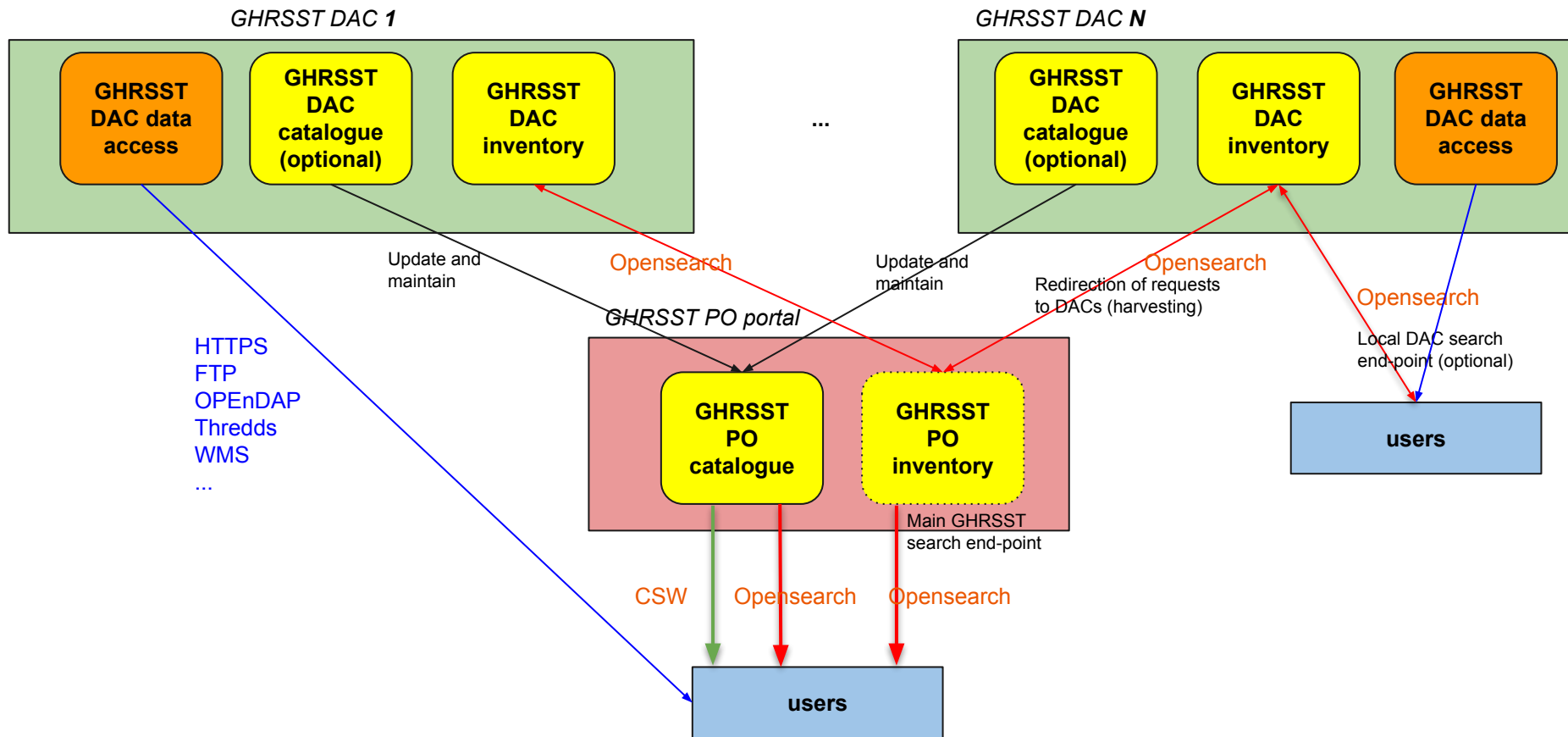
The new GHRSSST Regional/Global Task Sharing



motivations

- No single host for all datasets => different access points
- No one-stop catalogue and data access
- More providers distributing their own datasets (Eumetsat, CMEMS, Jaxa,...) without connection to a GDAC
- Somewhat outdated system with growing number of products and providers, difficult to maintain GDACs up-to-date
- More and more services (data on cloud, etc...)

R/G TS refined data discovery, search and access system





Minimum data access services

- Mandatory : HTTP or HTTPS
- Strongly recommended : FTP and DAP (Hyrax/OPeNDAP or Thredds)
- Recommended : THREDDS, WMS and WCS (for L3/L4)



Catalogue

Description of datasets (collections) and associated services or accesses

Fully centralized, hosted at GHRSSST-PO

Single up-to-date list of all GHRSSST datasets accepted by GHRSSST-PO

Main access point for users with no prior knowledge of what and where datasets are available - explicit redirection to DACs

Metadata edited and maintained by remotely by producers (dataset description) and DACs (access services)

Metadata profile currently following **ISO 19115-3** standard

Webservice for queries : **CSW** protocol

Integration of GHRSSST catalogue (or subset) in tools or other portals without duplicating metadata

Already implemented by many data centers, which does not ensure interoperability



Inventory search with Opensearch

Inventory of all granules (files) from each dataset

Searchable

Single end-point at GHRSSST-PO / local end-point at each DAC

No knowledge required by users of where data are physically stored when querying from GHRSSST-PO - federated query with implicit redirection

Several URLs returned for granules accessible at different DACs or through different access services

Search through **Opensearch** webservice

Search query arguments : dataset, time frame, bounding box

Usage by automatic scripts

Service already implemented in some data centers



Opensearch query example

https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=NEODAS-L2P-AVHRR17_L_1&startPage=1&count=1&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z&geoBox=-60,24,60,90&clientId=foo

```
<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/" xmlns:geo="http://a9.com/-/opensearch/extensions/geo/1.0/" xmlns:georss="http://www.georss.org/georss" xmlns:dc="http://purl.org/dc/terms/" xmlns:time="http://www.w3.org/2003-01-25/xmlschema11-time#">
  <title>CWIC OpenSearch Response</title>
  <updated>2019-05-31T11:45:39Z</updated>
  <title>CWIC OpenSearch Response</title>
  <author>
    <title>CWIC OpenSearch Response</title>
    <name>CEOS WGISS Integrated Catalog (CWIC) - CWIC Contact - Email: cwic-help@wgiss.ceos.org - Web: http://wgiss.ceos.org/cwic</name>
    <email>cwic-help@wgiss.ceos.org</email>
  </author>
  <id>https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=NEODAA5-L2P-AVHRR17\_L\_1&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z&geoBox=-60,24,60,90</id>
  <dc:identifier>https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=NEODAA5-L2P-AVHRR17\_L\_1&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z&geoBox=-60,24,60,90</dc:identifier>
  <opensearch:totalResults>3336</opensearch:totalResults>
  <opensearch:startIndex>1</opensearch:startIndex>
  <opensearch:itemsPerPage>1</opensearch:itemsPerPage>
  <link rel="self" type="application/atom+xml" href="https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=NEODAA5-L2P-AVHRR17_L_1&startPage=1&count=1&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z&geoBox=-60,24,60,90" />
  <link rel="search" type="application/opensearchdescription+xml" href="https://cwic.wgiss.ceos.org/opensearch/datasets/C1215196808-NOAA_NCEI/osdd.xml?clientId=foo" title="Search this resource" />
  <link rel="first" type="application/atom+xml" href="https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=C1215196808-NOAA_NCEI&count=1&startIndex=1&geoBox=-60,24,60,90&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z" />
  <link rel="last" type="application/atom+xml" href="https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=C1215196808-NOAA_NCEI&count=1&startIndex=3336&geoBox=-60,24,60,90&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z" />
  <link rel="next" type="application/atom+xml" href="https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=C1215196808-NOAA_NCEI&count=1&startIndex=2&geoBox=-60,24,60,90&timeStart=2008-09-02T00:00:00Z&timeEnd=2010-05-18T23:59:59Z" />
  <opensearch:Query role="request" cwic:datasetId="C1215196808-NOAA_NCEI" espidiscover:clientId="foo" startIndex="1" count="1" geo:box="-60,24,60,90" time:start="2008-09-02T00:00:00Z" time:end="2010-05-18T23:59:59Z" />
  <!-- Remote search completed in 9.35 seconds. -->
  <entry>
    <title>C1215196808-NOAA_NCEI:GHRSS-NEODAA5-L2P-AVHRR17_L_20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2</title>
    <id>http://cwic.wgiss.ceos.org/opensearch/granules.atom?uid=C1215196808-NOAA\_NCEI:GHRSS-NEODAA5-L2P-AVHRR17\_L\_20081210-AVHRR17\_L-NEODAA5-L2P-10dec081213\_wsst.8bit-v01.nc.bz2</id>
    <dc:identifier>http://cwic.wgiss.ceos.org/opensearch/granules.atom?uid=C1215196808-NOAA\_NCEI:GHRSS-NEODAA5-L2P-AVHRR17\_L\_20081210-AVHRR17\_L-NEODAA5-L2P-10dec081213\_wsst.8bit-v01.nc.bz2</dc:identifier>
    <updated>2018-03-14</updated>
    <author>
      <name>DOC/NOAA/NESDIS/NCEI &gt; National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce - TECHNICAL CONTACT - ; ; - Email: NODC.DataOfficer@noaa.gov - Phone: 301-713-3272 - FAX: </name>
      <email>NODC.DataOfficer@noaa.gov</email>
    </author>
    <georss:box>33.83169937133789 -94.2088394165039 87.04456329345703 45.60078811645508</georss:box>
    <georss:polygon>33.83169937133789 -94.2088394165039 87.04456329345703 -94.2088394165039 87.04456329345703 45.60078811645508 33.83169937133789 45.60078811645508 33.83169937133789 -94.2088394165039</georss:polygon>
    <dc:date>2008-12-10/2008-12-10</dc:date>
    <link title="HTTPS" rel="enclosure" type="application/octet-stream" href="https://data.nodc.noaa.gov/ghrsst/L2P/AVHRR17_L/NEODAA5/2008/345/20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2" />
    <link title="FTP" rel="enclosure" type="application/octet-stream" href="ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR17_L/NEODAA5/2008/345/20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2" />
    <link title="THREDDS OPeNDAP" rel="enclosure" type="application/octet-stream" href="https://data.nodc.noaa.gov/opendap/ghrsst/L2P/AVHRR17_L/NEODAA5/2008/345/20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2.html" />
    <link title="THREDDS(TDS)" rel="enclosure" type="application/octet-stream" href="https://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR17_L/NEODAA5/2008/345/catalog.html?dataset=ghrsst/L2P/AVHRR17_L/NEODAA5/2008/345/20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2" />
    <summary type="text">Granule metadata for C1215196808-NOAA_NCEI:GHRSS-NEODAA5-L2P-AVHRR17_L_20081210-AVHRR17_L-NEODAA5-L2P-10dec081213_wsst.8bit-v01.nc.bz2</summary>
  </entry>
  <!-- Atom response generation completed in 1ms. -->
</feed>
<!-- OpenSearch response completed in 9357ms. -->
```




R/G TS Pilot project

- Pilot project to assess R/G TS validity and implementation effort
 - 4 teleconfs since last GHRSSST meeting
 - Based on existing services to minimize efforts
 - Central catalogue [Ifremer]
 - Import
 - Editing
 - Access
 - Federated query [NOAA and PODAAC]
 - Example of query
 - Analysis of aggregated result
- Provide on-the-shelf software recommendations (esp. for opensearch query) for DAC candidates
- Discuss and suggest implementation roadmap



central catalogue - status

- Hosted by Ifremer - based on existing solution for European Project (based on GeoNetwork)
- For testing purpose : not comprehensive, not user friendly, not beautiful
- Subtasks:
 - Metadata profile for GHRSSST/satellite type of product
 - Harvesting existing profiles from PODAAC or NOAA
 - Issue with ISO version : ISO 19115-2 vs ISO 19115-3 : “translators” had to be implemented to convert from one version to another
 - Information content very different from one catalogue to another
 - A “template” metadata profile was defined for GHRSSST products
 - Manual editing through a form for GHRSSST datasets
 - Access to external partners (user login)



Catalogue test

Edition form:

<https://sextant.ifremer.fr/geonetwork/srv/eng/catalog.edit#/metadata/334265208?tab=eo-tab-description>

User catalogue view:

https://sextant.ifremer.fr/eng/Data/Catalogue#/search?_groupPublished=GHRSSST



Opensearch federated query - status

Example implemented (<https://gitlab.ifremer.fr/jp0540d/rgtsdemo/>)

```
server = GHRSSSTServer()  
result = server.search(datasetId="NEODAAS-L2P-AVHRR19_L",  
                        timeStart="2010-05-06", TimeEnd="2010-05-07",  
                        geoBox='-180,-90,180,90')  
server.display(result)
```

INFO:root:Fetching DAC: PODAAC

DEBUG:opensearch.results.fetching

<https://podaac.jpl.nasa.gov/ws/search/granule?itemsPerPage=25&bbox=-180%2C-90%2C180%2C90&startTime=2010-05-06&endTime=2010-05-07&datasetId=PODAAC-GH19L-2PS01&format=atom>

INFO:root:Fetching DAC: NOAA

DEBUG:opensearch.results.fetching

https://cwic.wgiss.ceos.org/opensearch/granules.atom?datasetId=NEODAAS-L2P-AVHRR19_L_1&count=25&timeStart=2010-05-06&timeEnd=2010-05-07&geoBox=-180%2C-90%2C180%2C90&clientId=cwicClient

AGGREGATED RESULT

- 20100506-AVHRR19_L-NEODAAS-L2P-06may100113_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100113_wsst.8bit-v01.nc.bz2

NOAA:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100113_wsst.8bit-v01.nc.bz2

HTTP : http://data.nodc.noaa.gov/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100113_wsst.8bit-v01.nc.bz2

THREDDS:

http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/catalog.html?dataset=ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100113_wsst.8bit-v01.nc.bz2

- 20100506-AVHRR19_L-NEODAAS-L2P-06may100255_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100255_wsst.8bit-v01.nc.bz2

NOAA:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100255_wsst.8bit-v01.nc.bz2

HTTP : http://data.nodc.noaa.gov/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100255_wsst.8bit-v01.nc.bz2

THREDDS:

http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/catalog.html?dataset=ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100255_wsst.8bit-v01.nc.bz2

- 20100506-AVHRR19_L-NEODAAS-L2P-06may100435_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100435_wsst.8bit-v01.nc.bz2

NOAA:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100435_wsst.8bit-v01.nc.bz2

HTTP : http://data.nodc.noaa.gov/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100435_wsst.8bit-v01.nc.bz2

THREDDS:

http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/catalog.html?dataset=ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100435_wsst.8bit-v01.nc.bz2

- 20100506-AVHRR19_L-NEODAAS-L2P-06may100614_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100614_wsst.8bit-v01.nc.bz2

NOAA:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100614_wsst.8bit-v01.nc.bz2

HTTP : http://data.nodc.noaa.gov/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100614_wsst.8bit-v01.nc.bz2

THREDDS:

http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/catalog.html?dataset=ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100614_wsst.8bit-v01.nc.bz2

- 20100506-AVHRR19_L-NEODAAS-L2P-06may100930_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100930_wsst.8bit-v01.nc.bz2

NOAA:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100930_wsst.8bit-v01.nc.bz2

HTTP : http://data.nodc.noaa.gov/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100930_wsst.8bit-v01.nc.bz2

THREDDS:

http://data.nodc.noaa.gov/thredds/catalog/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/catalog.html?dataset=ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may100930_wsst.8bit-v01.nc.bz2

- 20100506-AVHRR19_L-NEODAAS-L2P-06may101109_wsst.8bit-v01.nc

PODAAC:

FTP : ftp://ftp.nodc.noaa.gov/pub/data.nodc/ghrsst/L2P/AVHRR19_L/NEODAAS/2010/126/20100506-AVHRR19_L-NEODAAS-L2P-06may101109_wsst.8bit-v01.nc.bz2



Opensearch issues - mapping Challenges

- Spatial
 - Common attribute **bbox**: Order of lons and lats may need mapping. Both use commas as delimiter
- Temporal
 - Different attributes
 - NASA: **startTime**, **endTime** in ISO-8601
 - NOAA: **timeperiod** range in ISO-8601. Spaces and 'TO' as delimiter
 - But NCEI Onestop specifies: `&time={time:start?/{time:end?}`. More elegant.
- Dataset
 - Different attributes
 - NASA: **datasetID**. Cryptic PO.DAAC unique ID
 - NOAA: **fileIdentifier**, GHRSSST shortname?
- Results are also formatted differently
- Different opensearch results have different interpretations for the same queries (ex: time)



Lessons learned

- Standards do not necessarily guarantee complete interoperability
 - At least not without common syntax (think NetCDF vs GDS NetCDF): could take a long-time, considering also these systems are institution level
 - Alternative is translation service between services (e.g., translating bbox syntax)



DAC's responsibilities and effort - Data access

- Mandatory : HTTP or HTTPS
- Strongly recommended : FTP and DAP (Hyrax/OPeNDAP)
- Recommended : THREDDS, WMS and/or WCS (for L3/L4)



DAC's responsibilities and effort - suggested roadmap

	HTTP(S)	OPeNDAP	(S)FTP
PO.DAAC	OK	OK	No
NCEI	OK	OK	OK
Ifremer / OSI SAF	OK (protected)	OK	OK (protected)
CMEMS	?	No	OK (protected)
Eumetsat	sort of, requires calls to an API (protected)	?	No
Jaxa	?	?	OK
Star	OK	OK	OK

We need to agree and put some dates : GHRSSST XXI?



GHRSSST PO effort - catalogue

- Implement and operate central catalogue
 - Re-using existing catalogue system or implementing one
- Main tasks
 - Deployment/Reuse of catalogue service solution
 - Solve security issues : access for editing by GHRSSST producers
 - Design / creation of metadata profile, ISO 19115 compliant
 - Addition of required information for Opensearch translation
 - Editing, publication workflow (builtin) for review and decision by GHRSSST PO
 - Allow converters for existing metadata descriptions (one only for PO.DAAC?)
 - Design/formatting of catalogue search and display
 - Design of edition form (assuming remote manual edition by producers)
 - Operation, monitoring and maintenance



DAC and producer's effort - Catalogue

- No software implementation, basically mostly documentation maintenance
- Once central catalogue service is set:
 - Interact with GHRSSST PO for submission of a dataset and approval process (format compliancy, metadata)
 - Fill in web forms for dataset description (collection level metadata)
 - Maintain these metadata up to date (versions, contacts, end of production,...)



GHRSSST PO's effort - inventories and opensearch

Implement, operate and maintain the central opensearch endpoint which redirect search requests to connected DACs and assemble result

- Web service to receive and process user queries
- Translation to DACs end-points and redirection (some information to store in central catalogue) : vocabulary change and formatting, alignment of time and space criteria, on case by case basis (PODAAC, NOAA, CODA, ...)
- Translate back result and assemble
- Handle many possible error cases: time-out of a local point, non-responding, too large result, unexpected result,....
- Lots of testing
- Load issue may occur at some point in frequent usage conditions (like for any web service)
- Operation, monitoring and maintenance



DAC's effort - inventories and opensearch

- Build, operate and maintain inventory of distributed granules with spatial and search capabilities
- Provide and operate opensearch webservice to query this inventory
- Describe this service in GHRSSST-PO central catalogue
- Implemented through in-house software or on-the-shelf software
- A possible alternative is provide granule metadata to CEOS CWIC (researching pathway)



Supporting software

- Opensearch server
 - CEOS CWIC specifications:
<http://ceos.org/ourwork/workinggroups/wgiss/access/cwic/>
 - GeoServer:
<https://docs.geoserver.org/latest/en/user/community/opensearch-eo/index.html>
 - OpensearchServer: <http://www.opensearchserver.com/>
 - Data Hub Software: <http://sentineldatahub.github.io/DataHubSystem/>
 - PyCWS: <https://pycsw.org/>
 - NOAA OneStop may provide in future: <https://github.com/cedardevs/onestop/>
 - ESIP Opensearch specification implemented by PO.DAAC



Suggested roadmap

- Central Catalogue
 - Existing systems ready to host
 - Effort by pilot project to build metadata mappings and pretty much all functions required
 - Could be operating by **GHRSSST XXI (2020)**
- Opensearch
 - Pre-existing services (PO.DAAC, NOAA, CWIC, CODA), not all fully functional, status unclear
 - Other DACs need to implement and deploy such services
 - Central service for federated queries needs to be implemented
 - Minimal service could be up by GHRSSST XXI
 - Is GHRSSST XXII a reasonable date for fully functional system (= all DACs connected) ?
- Eumetsat or Copernicus to support implementation and operations of GHRSSST-PO central services