Finding Atmospheric Composition (AC) Metadata

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

Richard Strub[1], Stefan Falke[2], Ed Fiakowski[2], Steve Kempler[1], Chris Lynnes[1], Oleg Goussev[3]

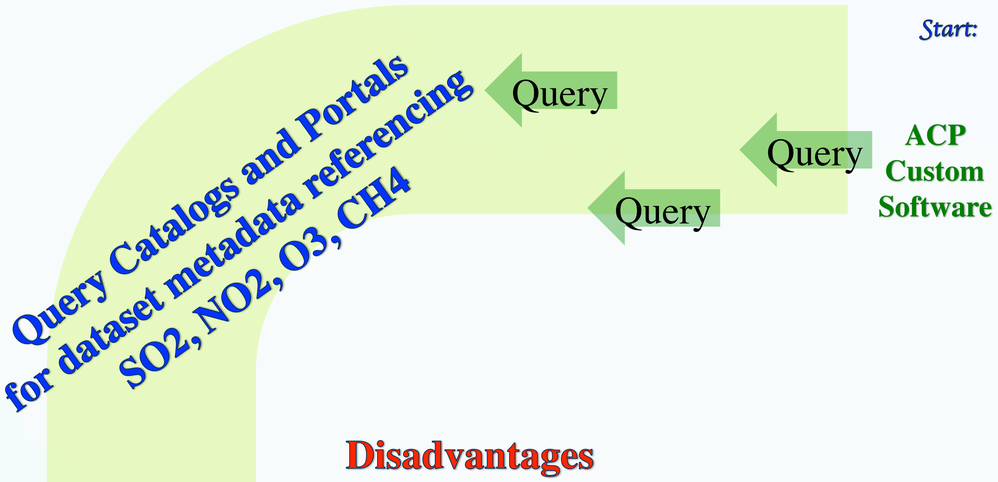
[1]GES-DISC, [2]GMU, [3]DLR

Searching the world for AC dataset level metadata

The Atmospheric Composition Portal (ACP) is an aggregator and curator of information related to remotely sensed atmospheric composition data and analysis. It uses existing tools and technologies and, where needed, enhances those capabilities to provide interoperable access, tools, and contextual guidance for scientists and value-adding organizations using remotely sensed atmospheric composition data. The initial focus is on Essential Climate Variables identified by the Global Climate Observing System – CH4, CO, CO2, NO2, O3, SO2 and aerosols. This poster addresses our efforts in building the ACP Data Table, an interface to help discover and understand remotely sensed data that are related to atmospheric composition science and applications. We harvested GCMD, CWIC, GEOSS metadata catalogs using machine to machine technologies - OpenSearch, Web Services. We also manually investigated the plethora of CEOS data providers portals and other catalogs where that data might be aggregated. This poster is our experience of the excellence, variety, and challenges we encountered.

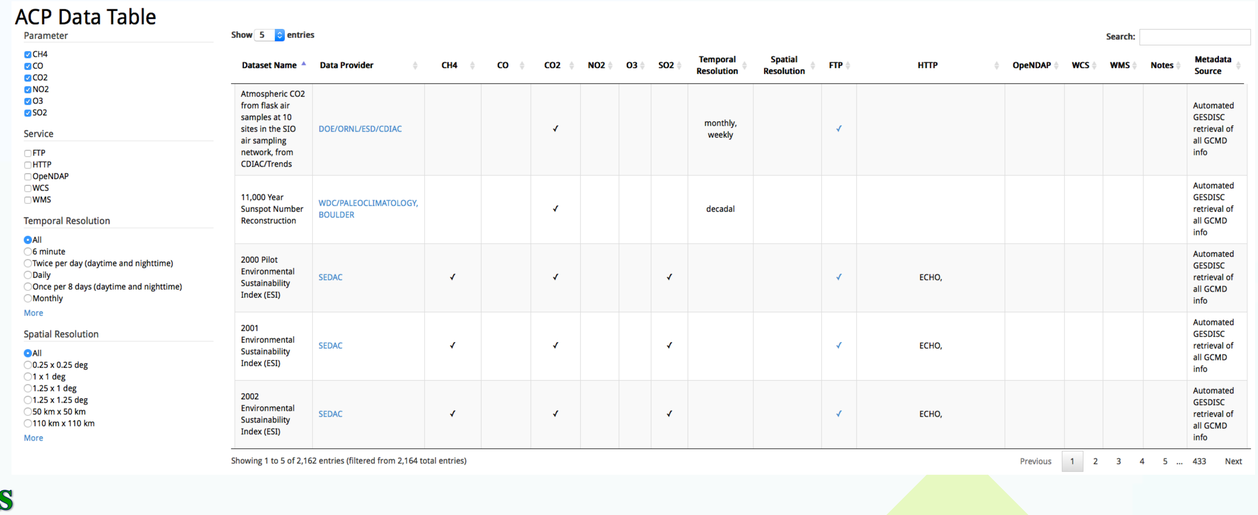
Conclusions:

1. The significant benefits that the major catalogs provide are their machine to machine tools like OpenSearch and Web Services rather than any GUI usability improvements due to the large amount of data in their catalog.
2. There is a trend at the large catalogs towards simulating small data provider portals through advanced services.
4. Populating metadata catalogs using ISO19115 is too complex for data providers to do in a consistent way, difficult to parse visually or with XML libraries, and too complex for Java XML binders like CASTOR.
5. The ability to search for IDs first and then for data (GCMD and ECHO) is better for machine to machine operations rather than the timeouts experienced when returning the entire metadata entry at once.
6. Metadata harvest and export activities between the major catalogs has led to a significant amount of duplication. (This is currently being addressed)
7. Most (if not all) Earth science atmospheric composition data providers store a reference to their data at GCMD.
8. Our experience showed that dataset level metadata search tools, catalogs and portals are constantly improving – some problems that we encountered when we started developing the ACP Data Table have been resolved by metadata providers and metadata catalog providers.



Start:

Result:



Disadvantages

Advantages

	Multiple Formats and Algorithms	Complex XML Format	Duplication	Good User Interface	2 Step Retrieval	Consistent Format	Open Search	Archives all available Metadata
	Multiple sites, different response formats etc. in each data provider's portal are difficult to deal with	ISO 19115 cannot be parsed by XML binding engines like CASTOR so parsing is left to XML libraries. Data providers interpret these complex XML formats differently and put the same data in different parts of the schema	Large Catalogs harvest and populate each other's sites leading to identical data products listed several times	A pleasant, focused user experience is easy because it is tailored to a small variety of data.	Metadata Ids can be retrieved separately from each metadata record so that a network timeout doesn't occur trying to retrieve thousands of full records.	Consistent response format across all variables	Easy Search and Retrieval	Seemingly universal availability of atmospheric composition data product metadata
GCMD – CMR/ECHO (Global Change Master Directory)	http://gcmdservices.gsfc.nasa.gov/mws/entryids/dif?query=[Data_Center:%20Short_Name=%27*%27]&format=XML				✓	✓	✓	✓
CWIC/GCMD (CEOS WGISS Integrated Catalog)	http://gcmd.gsfc.nasa.gov/KeywordSearch/OpenSearch.do?searchTerms=ozone&output=atom&count=1000&startIndex=1&startPage=1&Portal=cwic				✓	✓	✓	
FEDEO (Federated Earth Observation missions access) (ESA gateway to certain data)	http://fedeo.esa.int/opensearch/request/?httpAccept=application/atom%2bxml&recordSchema=iso&startRecord=1&maximumRecords=2000&query=ozone	✓				✓	✓	
GEOSS (Global Earth Observation System of Systems)	http://production.geodab.eu/gi-cat-StP/services/opensearch?&ct=100&st=ozone&ts=2002-01-01T00:00:00Z&te=2010-01-01T00:00:00Z	✓	✓			✓		✓
ACADIS, ANZ, ARL, BODC, CCHDO, CDIAC CNDP, EFI, EPA ESA, ESPO (NASA), GA Tech GEIA, GESDISC...	...JAMSTEC, ,JP NIPR, KOPRI, LARC, NCDC, NSIDC, ORNL, Palmer Station, SEDAC, SOLAS, UCAR, U of Miami, UNEP, USDA, USGS, US GLOBEC, US JGOFS, UTM	✓	✓	✓				

Large Centralized Catalogs

Local Provider Portals

DataBase

Results

GCMD was easy to query And contains references to almost all available AC metadata