Finding Atmospheric Composition (AC) Metadata

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 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.

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.
3. Metadata harvest and export activities between the major catalogs has led to a significant amount of duplication.
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. The query results returned by the major catalogs are not consistent and return data in different formats or根本不一致。 For example, GCMD returns the ISO 19115 XML format, which contains all variables specified in the request with no additional tags. However, it is left to XML libraries to interpret these formats differently, and too complex for Java XML binders like CASTOR.
7. Metadata catalog issues are often not visible in search engines for non-technical users.
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.

Results

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

Disadvantages

- Multiple Formats and Algorithms
  - Multiple sites, different response formats etc. in each data provider’s portal are difficult to deal with
- Complex XML Format
  - 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
- Duplication
  - Large Catalogs harvest and populate each other’s sites leading to identical data products listed several times

Advantages

- Good User Interface
  - A pleasant, focused user experience is easy because it is tailored to a small variety of data.
- 2 Step Retrieval
  - 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 Format
  - Consistent response format across all variables
- Open Search
  - Easy Search and Retrieval
- Archives all available Metadata
  - Seemingly universal availability of atmospheric composition data product metadata

Large Centralized Catalogs

- GCMD
  - CMR/ECHO
    - Global Change Master Directory
      - http://www.ceda.ac.uk/CMR/CMRIndex.html
    - GCMD – CMR/ECHO
  - CWIC/gcmd
    - GCMD (Integrated)
  - FEDEO
    - http://fedeo.esa.int
    - http://fedeo.esa.int
    - ECHO
      - http://www.echo.nasa.gov
      - http://www.echo.nasa.gov
  - GEOSS
    - http://www.geos.org
    - http://www.geos.org
    - http://www.geos.org

Local provider format

- ACADIS
  - ANZ, ARL, BDOC, CDMA, CIER, EPA, ESA, ESO (NASA)
- G4 Tech
- GESA, GESDISC

Query Catalogs and Portals for dataset metadata referencing SO2, NO2, O3, CH4

Query

Query

Query

Query

Query

Start

Result

ACP Data Table

Query

Start

Result

Database

Results

Query

ACP Custom Software

Query

Query

Query

Query

Query

Multiple Disadvantages

- Multiple Formats and Algorithms
- Complex XML Format
- Duplication

Multiple Advantages

- Good User Interface
- 2 Step Retrieval
- Consistent Format
- Open Search
- Archives all available Metadata

Richard Strub1, Stefan Falcon2, Ed Fiakowski2, Steve Kempler1, Chris Lynnes1, Oleg Goussev1

1 GES-DISC, 2 GMU, 3 DLR