WMO WOUDC UPDATE
ORM11 Part II
2021-07-20

Tom Kralidis
Geospatial and Open Data Systems
Meteorological Service of Canada
Environment and Climate Change Canada
WMO WOUDC UPDATE

• Data Centre Status
• 2019 by the numbers
• ORM10 Recommendations
• Future work
DATA CENTRE STATUS

• Meteorological Service of Canada
  – Data Management
  – Data Products and Services
  – Geospatial and Open Data

• Collaboration with ECCC Science and Technology Branch

• WMO reporting
  – SAG on Ozone/UV
  – WMO Expert Team on Atmospheric Composition Data Management
    • formerly ET on World Data Centres (ET-WDC)
  – Data Centre Interoperability (DCIO)
DATA CENTRE STATUS

• Renewed in 2015
• Current operations
  – Contributor support
  – User support
  – Continuous data improvement
    • Data ingest
    • Metadata quality assessment / correction
  – Enhancements
GAW IMPLEMENTATION PLAN – DATA MANAGEMENT

• Federation, interoperability, data access, discovery
• Open access, metadata
• WIGOS
• Data archiving, analysis centres
• Known quality, documentation
• Near real-time delivery (GTS/WIS)
• Data submission and data use
• DOIs
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad-band</td>
<td>137415</td>
</tr>
<tr>
<td>OzoneSonde</td>
<td>92261</td>
</tr>
<tr>
<td>Lidar</td>
<td>675</td>
</tr>
<tr>
<td>Multi-band</td>
<td>92041</td>
</tr>
<tr>
<td>TotalOzoneObs</td>
<td>110274</td>
</tr>
<tr>
<td>Spectral</td>
<td>278713</td>
</tr>
<tr>
<td>UmkehrN14 2.0</td>
<td>10241</td>
</tr>
<tr>
<td>RocketSonde</td>
<td>177</td>
</tr>
<tr>
<td>TotalOzone</td>
<td>82932</td>
</tr>
<tr>
<td>UmkehrN14 1.0</td>
<td>10701</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>815430</strong></td>
</tr>
<tr>
<td>Month</td>
<td>Total Files</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>January</td>
<td>1419</td>
</tr>
<tr>
<td>February</td>
<td>20558</td>
</tr>
<tr>
<td>March</td>
<td>37476</td>
</tr>
<tr>
<td>April</td>
<td>32330</td>
</tr>
<tr>
<td>May</td>
<td>1742</td>
</tr>
<tr>
<td>June</td>
<td>460</td>
</tr>
<tr>
<td>July</td>
<td>1307</td>
</tr>
<tr>
<td>August</td>
<td>514</td>
</tr>
<tr>
<td>September</td>
<td>1821</td>
</tr>
<tr>
<td>October</td>
<td>11653</td>
</tr>
<tr>
<td>November</td>
<td>2328</td>
</tr>
<tr>
<td>December</td>
<td>641</td>
</tr>
<tr>
<td>Total</td>
<td><strong>112249</strong></td>
</tr>
</tbody>
</table>
## 2019 Downloads

<table>
<thead>
<tr>
<th>Dataset</th>
<th>WAF (number of files)</th>
<th>Dataset Archive Files</th>
<th>Geospatial Web Services (number of requests)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total ozone obs</td>
<td>719,127</td>
<td>70</td>
<td>1213</td>
</tr>
<tr>
<td>Ozonesonde</td>
<td>5,303,085</td>
<td>94</td>
<td>21036</td>
</tr>
<tr>
<td>Total ozone daily obs</td>
<td>20,339,162</td>
<td>83</td>
<td>25709</td>
</tr>
<tr>
<td>Spectral</td>
<td>611,990</td>
<td>69</td>
<td>1565</td>
</tr>
<tr>
<td>Multi-band</td>
<td>274,647</td>
<td>76</td>
<td>741</td>
</tr>
<tr>
<td>Broad-band</td>
<td>434,934</td>
<td>67</td>
<td>945</td>
</tr>
<tr>
<td>Umkehr 2</td>
<td>340,118</td>
<td>65</td>
<td>389</td>
</tr>
<tr>
<td>Umkehr 1</td>
<td>329,591</td>
<td>64</td>
<td>562</td>
</tr>
<tr>
<td>Lidar</td>
<td>6,006</td>
<td>64</td>
<td>576</td>
</tr>
<tr>
<td>RocketSonde</td>
<td>15,722</td>
<td>15</td>
<td>715</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>28,374,382</strong></td>
<td><strong>667</strong></td>
<td><strong>53,451</strong></td>
</tr>
<tr>
<td>Year</td>
<td>Visits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>17,245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>59,042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>98,084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>547,352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>1,688,274</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## 2019 TOP HITS BY VISITOR DOMAIN

<table>
<thead>
<tr>
<th>Domain</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>noaa.gov</td>
<td>116,526,443</td>
</tr>
<tr>
<td>ecmwf.int</td>
<td>31,740,769</td>
</tr>
<tr>
<td>oma.be</td>
<td>15,963,146</td>
</tr>
<tr>
<td>auth.gr</td>
<td>8,753,982</td>
</tr>
<tr>
<td>nasa.gov</td>
<td>3,105,121</td>
</tr>
<tr>
<td>(not set)</td>
<td>2,808,616</td>
</tr>
<tr>
<td>jussieu.fr</td>
<td>2,706,078</td>
</tr>
<tr>
<td>kishou.go.jp</td>
<td>1,864,959</td>
</tr>
<tr>
<td>sachina.edu.cn</td>
<td>697,279</td>
</tr>
<tr>
<td>niwa.co.nz</td>
<td>473,114</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>187,522,236</strong></td>
</tr>
</tbody>
</table>
ORM10 RECOMMENDATIONS
ENHANCED LINKAGE AMONG ESTABLISHED DATA CENTRES

• Related Data Links
  Available via https://woudc.org/resources/links.php
  – European UV database (EUVDB) added in 2019
  – Other WMO WDCs

• Simple linkages
INTEROPERABILITY

• Open government
• Open data
  – Open by default
• Open software
• Libraries harvesting the Data Centre
• Federation
• System architecture presented at ET-WDC face to face meeting (2017, Kjeller, Norway)
  – Presenting (tomorrow) at Jülich Supercomputing Centre, Hörsaal, building (M. Schultz)
DATA CENTRE INTEROPERABILITY

- DCIO
- Started from GEOMS activities
- Led by NILU (Ann Mari Fjæraa)
- WOUDC, NDACC, and others
- Harmonize metadata and data exchange
SHADOZ INTEROPERABILITY

- SHADOZ submits to WOUDC
- WOUDC converts to Extended CSV using pyshadoz
- Data fed into processing cycle
- Data is available for download from WOUDC
- Data are part of metrics
SHADOZ UPDATES

• Data submission conversion / processing (2016-2017)
  – Ascension, Costa Rica, Fiji, Hanoi, Hilo, Irene, Kuala, Samoa, Natal, La Reunion, Nairobi

• Development of pyshadoz, Free and Open Source SHADOZ processing software
  – [https://github.com/wmo-cop/pyshadoz](https://github.com/wmo-cop/pyshadoz)
  – Advertised on SHADOZ website: [https://tropo.gsfc.nasa.gov/shadoz/Links.html](https://tropo.gsfc.nasa.gov/shadoz/Links.html)
NDACC INTEROPERABILITY

- NDACC makes available filelist of all data
- WOUDC processes filelist (periodically) and create file index
- NDACC data is discoverable via federated search
- NDACC data available for download from NDACC
- Data are not part of metrics
Data Search / Download

The NDACC data archive can be searched by data category; there are six ozone data categories and three ultraviolet (UV) radiation data categories. The ozone datasets for total column ozone include total ozone and total ozone observations and the vertical ozone profile includes lidar, ozone sondes, and Umkehr N-value and C-Umkehr. The UV datasets for UV irradiances include broadband, multiband, and spectral.

To search and download data, select the dataset and observation time period. Optionally, draw your map extent of interest and then hit search. All available data for that time period will be displayed.

For more details on how to use this page, please view the How to Use guide.
NDACC INTEGRATION

• Discussions with NDACC (June/July 2018)
• Co-presentation with Jeannette Wild (NDACC DHF Curator) at the NDACC SC Meeting (September 2018)
  – History of (numerous) efforts (since 2002)
  – Pros/cons of previous approaches
NDACC INTEGRATION

• Mirroring
  ✓ Consistent data formats
  ✓ Contributes to WOUDC metrics / reporting
  × Station lookup variations
    × WOUDC standardized on GAW/OSCAR
  × Derivations, calculations, interpretations, PI involvement
  × Managing updates to data (versionitis)

• Distributed Search
  ✓ Data managed closer to data centre source
  ✓ Reduces data centre data duplication
  × Inconsistent data formats
  × Integration effort still on the user
NDACC INTEGRATION

• Data Centre Interoperability project (DCIO)
  – Since 2008
  – Harmonized dataset metadata
  – Concept of peering
  – Data discovery

• WOUDC/NDACC believe the most sustainable way forward is to support distributed searching
  – Evolution of DCIO project
  – Reduce problems associated with data duplication
  – Authoritative single source
  – Some discussions on piloting common formats (2018) such as HDF
EUBREWNET INTEROPERABILITY

- EUBREWNET makes available filelist of all data
- WOUDC processes filelist (periodically) and create file index
- EUBREWNET data is discoverable via federated search
- EUBREWNET data available for download from EUBREWNET
- Data are not part of metrics
Data Search / Download

Data updates and notifications

The WOUDC data archive can be searched by data category: there are six ozone data categories and three ultraviolet (UV) radiation data categories. The ozone datasets for total column ozone include total ozone and total ozone observations and the vertical ozone profile includes lidar, ozonesonde, Umkehr N-value and C-Umkehr. The UV datasets for UV irradiance include broadband, broadband and spectral.

To search and download data, select the dataset and observation time period. Optionally, draw your map extent of interest and then hit search. All available data for that time period will be displayed.

For more details on how to use this page, please view the how to use guide.

Select Dataset, Station, Instrument, Time Period

Dataset

EUBREWNET

Country | optional

Station | optional

- Abu Dhabi (397)
- Academy of Athens (440)
- Aosta (479)
- Arosa (635)
- Belgrano II (515)
- Buenos Aires Observatorio (651)
- Casablanca (156)
- Copenhagen (309)
- Davos (591)
- El Arenosillo (213)
- Fairbanks (105)
- Goddard Space Flight Center (447)
- Graciosa (527)
- Hobi (302)
- Herdecke-Kraakow (690)
- Izana (300)
- Jokkmokk (404)
- La Perva (NaN)

Search  Reset

Set Your Map Extent

How to Use: Interactive Map

Leaflet | Use data CC-BY-NC-SA by OpenStreetMap | Tiles by MapSaver
WOUDC OSCAR/GAWSIS INTEGRATION

• [https://gawsis.meteoswiss.ch/GAWSIS](https://gawsis.meteoswiss.ch/GAWSIS)

- WOUDC
  - Ozone/UV data/metadata
  - Station metadata
  - Observation metadata
PROVIDING DATA IN SEVERAL ACCEPTED STANDARD FORMATS

• WMO SAG O3/UV Recommendation (2019)

• (13) In principle, data providers should only have to submit data in a single format. In practice WOUDC should continue to pursue the maximum achievable interoperability and data exchange and discoverability in co-operation with other major ozone databases
WOUDC DATA FORMAT SUPPORT

• Data is submitted as Extended CSV format
• Web Services provide various formats
• Quality assessment available in the data validator
• Summaries are available for application datasets
  – Daily summary master file: https://woudc.org/archive/Summaries/TotalOzone/Daily_Summary
DATA VALIDATION

- Deeper metadata checks
- Makes backend processing logic available to end user
- Provides suggestions
PROVIDING DATA IN SEVERAL ACCEPTED STANDARD FORMATS
DOI IMPLEMENTATION

• [https://woudc.org/about/data-policy.php#dois](https://woudc.org/about/data-policy.php#dois)

• Levels of granularity
  – First order (Ozone, UV)
  – Dataset level
  – TODO: Stations (from GAWSIS/OSCAR)
STATION DOI IMPLEMENTATION

• A DOI for individual stations should be implemented at WOUDC

• Best implemented via GAWSIS/OSCAR
  – GAWSIS creates DOIs
  – WOUDC integrates them on website

• Being discussed at ET-ACDM
OPEN DATA POLICY

• WMO Data Use Policy: https://woudc.org/about/data-policy.php
• Free and unrestricted exchange of data
• GAW Data Use Policy
• Use appropriate citation
PRO-ACTIVE COMMUNICATION BETWEEN DATA CENTRES AND DATA PROVIDERS
CONTRIBUTOR CONTACT VALIDATION

• Implemented September 2017
• Annual notice sent to all contributors to validate their contact information
CONTRIBUTOR DATA SUBMISSION REMINDERS

• [1] Have contributed data within last 10 years  
  – AND
• [2] No data received within the last 6 months
• Reminder notification sent to all contributors who meet both [1] and [2]
• The process runs monthly
• Contributor will be notified until they do not meet [2]
• 10 years denotes active contributors and stations
• Process can be configured as required; multiple messages, intervals between messages, cc list, bcc list, etc.
• Record of notifications archived
DATA SUBMISSION NOTIFICATIONS

• Implemented December 2017
• Contributors notified via email
  – Data Submission Confirmation (nightly)
    • List of submitted files to ftp.woudc.org contributor account
  – Data Processing report (weekly)
    • Total files received
    • Number of passed files
    • Number of manually repaired files
    • Number of failed files
    • Summary of Manual fixes
METADATA QUALITY CORRECTION

• Metadata qa/qc
• Manual fixing of metadata for data submissions (in parallel to contributor feedback)
• Bulk corrections (ongoing) in consultation with various stakeholders
  – ECCC/STB (V. Fioletov)
  – AUTH (A. Bais, K. Garane, M. Koukouli)
FUTURE WORK
WOUDC 2.0

• Simplified archive (data registry concept)
• Updated data API
• Simplified UI
• Updated infrastructure / deployment
Data Search and Download

The WOUDC data archive can be searched by data category; there are six ozone data categories and three ultraviolet (UV) radiation data categories. The ozone datasets for total column ozone include total ozone vertical ozone profile includes lidar, ozonesonde, Umkehr N value and C-Umkehr. The UV datasets for UV irradiance include broadband, multiband and spectral.

To search and download data, select the dataset and observation time period. Optionally, draw your map extent of interest and then hit search. All available data for that time period will be displayed.

For more details on how to use this page, please view the How to Use guide.
The World Ozone and Ultraviolet Radiation Data Centre (Woudc) is one of six World Data Centres which are part of the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO). Woudc contains ozone and UV data measured by instruments located on ground-based, shipborne or airborne platforms. Data are subject to the Woudc Data Policy (https://woudc.org/about/data-policy.php).

Terms of service

Woudc Data Use Policy

GET / Landing page
GET /collections: Collections
GET /conformance: API conformance definition
GET /openapi: This document
GET /processes: Processed

Stac: Spatial/Temporal Asset Catalog

Projects: Connection to projects in the Woudc Data Registry Search Index.
GET /collections/projects: Get Woudc Data Registry Projects metadata
GET /collections/projects/items: Get Woudc Data Registry Projects items
GET /collections/projects/items/{featureId}: Get Woudc Data Registry Projects item by id
GET /collections/projects/queryables: Get Woudc Data Registry Projects queryables

Datasets: Connection to datasets in the Woudc Data Registry Search Index.
GET /collections/datasets: Get Woudc Data Registry Datasets metadata
GET /collections/datasets/items: Get Woudc Data Registry Datasets items
Thank you!

- Website: https://woudc.org
- Guidebook: https://guide.woudc.org
- Code: https://github.com/woudc
- GitHub: https://github.com/woudc
- Wiki: https://github.com/woudc/wiki