WOUDC Report

2014 ET-WDC meeting

WOUDC Renewal

- Fulfill GAW Data Centre ToR…address gaps identified in ET-WDC audit
- Ozone and UV SAG requirements…user requirements
- Web accessibility and usability standards…GoC requirement and growing international requirement
- FTP performance and stability
- Streamline manual operational level of effort…processing workflow
- GAW interoperability objective…standards based web services
- Support international standards (OGC/ISO) for discovery, access and visualization
- Modernize technology footprint…open source software

Environment Canada   Environnement Canada

Canada

Renewal Objectives

- Enhance the <u>data submission</u> mechanism and infrastructure
- Effective and efficient <u>data management</u> approach
- Modernize <u>data access</u> mechanisms
- Improve accessibility and usability on the <u>website</u>

Environment Canada   Environnement Canada

Canada

Renewal Phases

	Phase 1 (Foundaton)	Future Phases
Data Submission	• Enhance performance and stability of FTP services (submission and access)	• Implement new Ozonesondes data contribution format • Migration from FTP to WAF • Submission of level 0 data • Update data submission guidelines • New XML submission format
Data Management	• Enhance data management lifecycle and workflows: relational database, reduce manual level of effort in processing data, manage data history, logging, & notifications • Qa – format & descriptive metadata	• Enhanced management of Interpretive (WIGOS) metadata…including calibration • Qa – plausibility and consistency checks on data…beginning with Ozonesondes • UMKEHR derived data • Data Management Metrics…including outstanding contributors
Data Access and Use	• Web Accessible Folder (WAF) • Geospatial web services (WMS &WFS)…support interoperability • Geospatially enabled • Discovery Metadata (CSW)…WIS-DCPC	• Data in other data centres: NDACC, EUVDB, SHADOZ • Migration from FTP to WAF • Renew systems for creating Ozone and UV data products • Restricted access to level 0 data
Website	• Meet international web accessibility and usability standards (WCAG 2.0) • Improve navigation	

Environment Canada   Environnement Canada

Canada

System Overview: Architecture
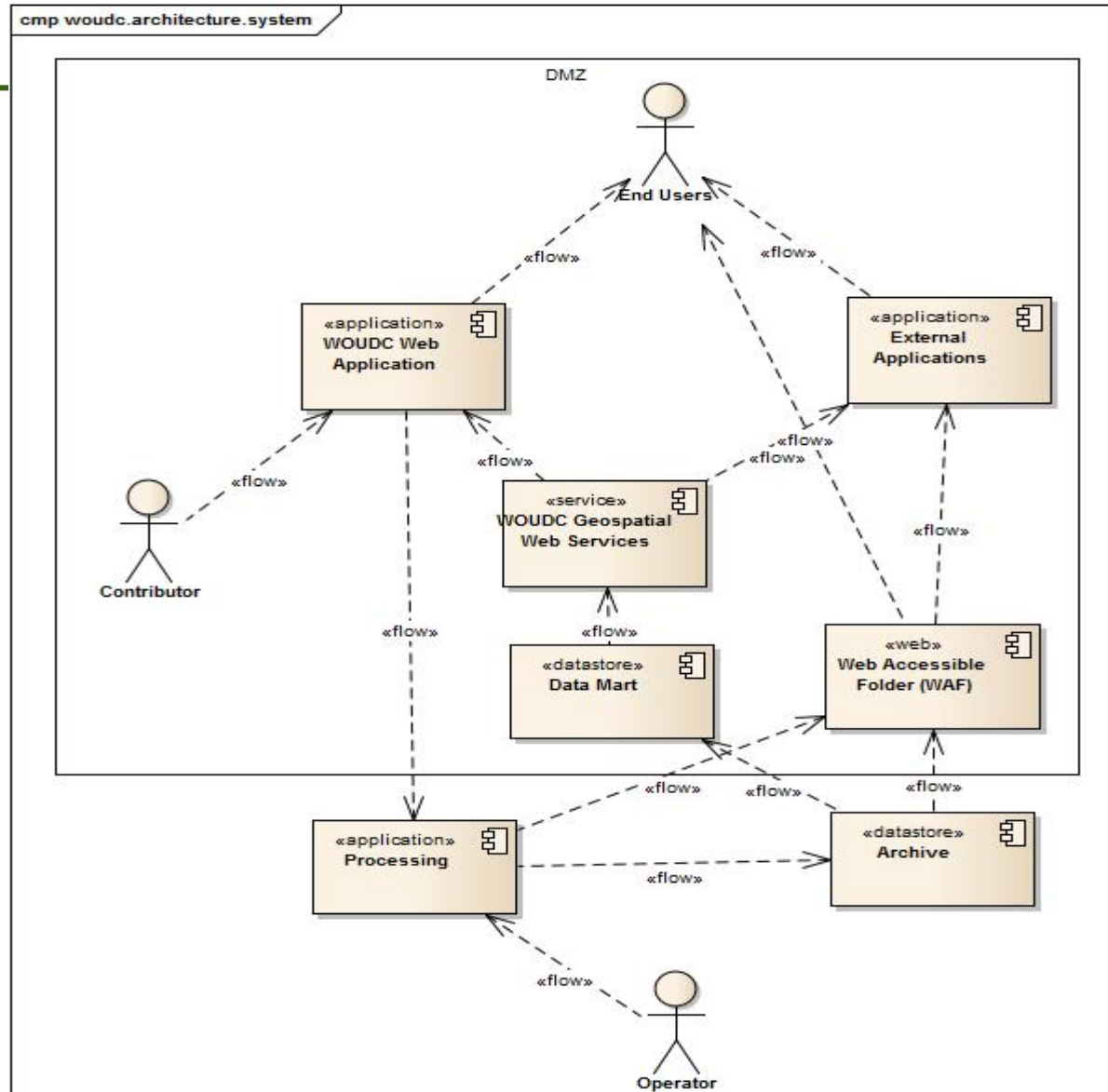
Software: Standards Based and Open Source

- Backend Processing System
  - Python

- Archive Database / Datamarts
  - PostgreSQL
  - PostGIS: geometry handling

- Geospatial Web Services
  - MapServer (WMS, WFS, SOS)
  - pycsw (CSW, SRU, OAI-PMH)

- Web Application / UI
  - Web Experience Toolkit, PHP
  - jQuery, Leaflet
  - Selenium Web Automation testing

Environment Canada   Environnement Canada

Canada

Taxonomy

- Here are all the attributes of a WOUDC dataset that uniquely defines itself, in the order they are assigned:
    1. **Organization** (JMA, EC …)
    2. **ISO Topic Category** (ClimatologyMeteorologyAtmospheric, …)
    3. **Data Category** (Ozone, UV, …)
    4. **Data Type** (Observation, …)
    5. **Data Set** (Total ozone – daily, Total ozone – monthly. UV irradiance, Vertical Ozone profile (VOP), …)
    6. **Instrument Class** (Brewer, Broadband, Dobson, Multi-broadband, Ozonesonde, Spectral)
    7. **Instrument Type** (ECC, …)
    8. **Instrument Model** (5a, 6a, …)
    9. **Data Level** (1.0, 2.0)
    10. **Data Format Type** (ASCII, …)
    11. **WOUDC Station ID** (STN007, …)
    12. **Observation Datetime** (YYYMMDDhhmmss)
    13. **History** (1.0, 1.1, 1.2, …)

- Example:
    - The string below uniquely represents the original vertical ozone profile observation from JMA's STN007 on January 10, 2014 at midnight using ECC-6a instrument model:
    - /jma/climatologyMeteorologyAtmospheric/ozone/observation/vop/ozonesonde/ecc/6a-1.0-ascii/stn007/20140110000000/1.0
    - Used for filing and processing of incoming and ext CSV files…also for data discovery and search

Environment Canada   Environnement Canada

Canada

Data Metrics

Data Submission
- # of data files contributed for each network
- # and % of total of active and inactive stations for each network…timeline
- # and % of total of corrected files submitted

Data Management
- Processing timelines
- # and % of total files that get Qa
- # and % of total files that get flagged by Qa

Data Access
- # of FTP access visits and # of files downloaded
- # of files accessed through web services

Website
- # of website visits by year since 2000
- # of emails received

Environment Canada   Environnement Canada

Canada

Targeted Renewal Milestones

- Phase 1 (foundation: Processing and archiving system, Web Services, website, WAF)
  - Development release Jan, 2014
  - UAT Mar, 2014 (EC Feb and external Mar)
  - Operational release Apr, 2014
- Phase 2 (additional functionality and features) Sept, 2014
  - File Transfer Protocol (FTP) to Web Accessible Folders (WAF) transition period Apr. 1 – Sept. 30.
- Additional releases as required.

Environment Canada   Environnement Canada

Canada

WOUDC Ozone Data Products

- Total Ozone Maps: Global Total Ozone Maps, Ozone Maps for the Northern Hemisphere, Ozone Maps for the Southern Hemisphere, Ozone Maps Archive, Recent Ozone Maps Individual Data Sources and Ozone forecast Maps

- Time Series Graphs: Ozonesonde plots, Total Ozone plots

- Data summaries and Reports: Annual data report, Total Ozone Summary Data: daily and monthly means, Umkehr data by algorithm, Zonal mean ozone from ground-based instruments, IONS data and mapping plots

Environment
Canada

Environnement
Canada

Canada

Collection and access to level 0 data

- Brewer (B-files)
  - Level 0 data submitted by 7 of 135 stations
  - Currently stored but not accessible externally
  - BDMS?
- Dobson (R-values)
  - Level 0 data submitted by 4 of 164 stations
  - Currently stored but not accessible externally
- Ozonesondes
  - Some level 0 data will be submitted as part of the new Qzone sondes format
- Issue – Not all stations contribute level 0 data and level 0 data contributed isn't easily accessible. Is level 0 available from calibration centre?

Environment Canada   Environnement Canada

Canada

Discovery/Calibration metadata

- Data Sponsorship Statement (DSS) methodologies and practices of the data originators

- website & reports
  - IOS Inc. website – Brewer
  - Ad-Hoc Dobson Committee website
  - Dobson Spectrophotometers report
  - Varotsos report

- Does this meet requirements? Can we make available calibration information from the calibration centres?

Environment Canada   Environnement Canada

Canada

Agenda 4.1 - WOUDC Station Types

- GAW has two station types, namely Global and Regional GAW stations. Additionally, Contributing stations are also listed in GAWSIS because they submit data to the GAW World Data Centres.

- Global stations shall be designed in relation to the global requirements which are to provide for data required to address environmental issues of global scale and importance. [WMO Technical Regulations, B.2.4.2.1] The current GAW strategic plan describs these Essential Characteristics of a Global Station.

- Regional stations shall be designed primarily to address regional aspects of global environmental issues and environmental problems of regional scale and importance. [WMO Technical Regulations, B.2.4.3.1] According to the current GAW strategic plan Regional Stations fulfil these ten Essential Characteristics.

- WOUDC has 2 Station Types: STN and SHP

- There are only 2 Mobile stations (both SHP) which are registered to 3 Agencies.

- But there is no data for any SHP stations.

-

Project	Agency	country	Station Type	Woudc Station ID	Station Name	Data Class	Data Category	instrument	instrument model
IONS	NOAA - Climate Monitoring and Diagnostics Laboratory	USA	SHP	440	R H BROWN RESEARCH SHIP	Ozone	OzoneSonde	ECC	2Z
	NASA Goddard Space Flight Center	USA	SHP	440	R H BROWN RESEARCH SHIP	Ozone	OzoneSonde	ECC	2Z
	Polish Academy of Science	POL	SHP	188	POLISH RESEARCH SHIP	Ozone	TotalOzone	Filter	M-83

Environment Canada   Environnement Canada

Canada

Agenda 5.2 WOUDC Data Catalogue

- Ozone Data sets: Total ozone – daily, Total ozone – monthly, Vertical Ozone profile
- Ozone Instruments: Brewer, Dobson, Ozonesonde
- UV Data sets: UV irradiance
- UV Instruments: Broadband, Multi-broadband, Spectral

Environment Canada   Environnement Canada

Canada

Agenda 5.3 - digital object identifier (DOI)

- used to uniquely identify an object
- Should we use a service like DataCite to assign Digital Object Identifiers (DOI) to manage citations?
- **What is DataCite?**
  We are an international organisation which aims to:
  - establish easier access to research data
  - increase acceptance of research data as legitimate contributions in the scholarly record, and to
  - support data archiving to permit results to be verified and re-purposed for future study.
  - We have a Managing Agent (currently the German National Library of Science and Technology, TIB) along with regional Members.
- What is a Digital Object Identifier (DOI)?
  - The Digital Object Identifier (DOI®) System is for identifying content objects in the digital environment. For further information, see our website or the DOI entry on Wikipedia.

Environment Canada

Environnement Canada

Canada

Agenda 5.3 - CF standard names for GAW variables

Environment Canada   Environnement Canada

Canada

Agenda 5.8 - Related data centres

One data centre - WOUDC

- Ingest – SHADOZ provides data for ingest
- Mirroring - Network for the Detection of Atmospheric Composition Change (NDACC)
  - Approach: initial harvest, update as required and make available on WOUDC FTP in Ext CSV format…current focus is on Ozonesondes (2160 NDACC format, ~20K files)

Decentralized data centres

- Links - European UV Database (EUVDB).
  - Exploring mirroring/data management options with FMI
- Web services – standard access to decentralized data centres

Environment Canada   Environnement Canada

Canada

Agenda 5.9 – role of data centre in near-real-time data submission

- Guidelines for Reporting Total Ozone Data in Near Real Time: http://www.wmo.int/pages/prog/arep/gaw/documents/FINAL_GAW_193.pdf
- Europe?

Environment Canada   Environnement Canada

Canada

Agenda 6.1 - Quality Assurance

- Data collection:
  - Standard Operating Procedures (SOP) provide direction of instruments, observation practices and processing, and quality assessment (Qa)
  - Operate calibration data centers and complete comparative analysis
  - Level 0 to level 1 processing software and Qa software…different versions
  - Historical data quality assessment for key stations – fed back to contributors
- Data centre:
  - WOUDC guidelines for submitting and accessing data
  - Check submitted data for necessary format elements and the availability of comprehensive metadata and reject the submission of data that do not meet these formal criteria.
  - Perform plausibility and consistency checks on submitted data, flag data problems, and provide feedback to the stations, when necessary. Ozonesonde pilot (range checks of ozone partial pressure during ascent & checks of pump temperature).
  - Descriptive metadata – sensor history, calibration history…
- Users
  - Feedback…send to contributors.
  - Analytics software…homogenized data

Environment Canada   Environnement Canada

Canada

Agenda 6.2 Data flagging schemes

Flag Value	English Abbreviation	French Abbreviation	English Description	French Description
-10	R	R	Suppressed	Réprimer
-1	M	M	Missing	Manquant
0	E	E	Error	Erreur
10	D	D	Doubtful	Douteux
15	S	S	Suspect	Suspect
20	I	I	Inconsistent	Incohérent
100	A	A	Accepted/Passed	Accepté
101	N	N	Qa test(s) not performed	Test non exécuté
110	U		Optimized range test not performed because associated element's value unavailable for use	valeur associée non disponible

Environment Canada Environnement Canada

Canada

Agenda 6.2 Data flagging schemes

Qa flag values currently used by the DMS in XML decoded outputs or products:
<u>Note</u>: highlighted rows are only available in the SWOB (surface Weather OBservation) product and are the result of "NativeQC" provided in the incoming message by the data provider prior to DMS quality assessment (Qa) being performed. Non highlighted rows are the Qa flags that can be generated by the DMS Qa system after decoding.

<u>Qa Flag Definitions:</u>

Suppressed – the data provider has indicated that the data value is not to be used or published due to instrumentation or environmental issues which have compromised the data.

Missing – the "value" of an element is null, blank/space, "MSNG" (from the decoder), a code representing missing, or could not be derived.

Error – Typically physically impossible values, or beyond the capability of the sensor to detect.

Doubtful – physically possible, but at or exceeding typical extreme vales.

Suspect/Warning - the data provider has indicated that the data <u>may be</u> unreliable or is outside of nominal limits

Inconsistent – as a result of an inter-variable compatibility or consistency check, the value was found to be inconsistent with what is expected. May not be possible to determine which variable is suspect without considering other Qa results.

Accepted – passed the test(s), but not necessarily "good".

Qa Test not performed – the requirements to run the test were not met (e.g. associated element missing or not of sufficient quality to proceed)

Environment Canada   Environnement Canada

Canada

Agenda 7.1 - Data Metrics

<u>Data Submission</u>

- # of data files contributed for each network
- # and % of total of active and inactive stations for each network…timeline
- # and % of total of corrected files submitted

<u>Data Management</u>

- Processing timelines
- # and % of total files that get Qa
- # and % of total files that get flagged by Qa

<u>Data Access</u>

- # of FTP access visits and # of files downloaded
- # of files accessed through web services

<u>Website</u>

- # of website visits by year since 2000
- # of emails received

Environment Canada   Environnement Canada

Canada

Agenda 7.2 - Outstanding Contributors

- WOUDC guidebook asks data to be submitted on a monthly basis or sooner if possible
- Issue – There are many stations that haven't contributed in more than 1 year
- Letter is to be sent from Manager of WOUDC and Ozone and UV SAG chairs to contributors

Time Period	#
< 1 year	60
1 – 2 years	10
2 – 3 years	17
> 3 years	61
Registered but never reported	20

Environment Canada   Environnement Canada

Canada