Minutes from the Expert Team on World Data Centres (ET-WDC)

1-3 October 2019, NASA Langley, Hampton VA, USA

WORLD METEOROLOGICAL ORGANIZATION

GLOBAL ATMOSPHERE WATCH

TABLE OF CONTENTS

OPENING OF THE MEETING

Welcome by Rosemary Baize, the Deputy Director for Research and Mission Science in NASA Langley Research Center.

- Data management critical for NASA, excited to host this meeting
- Langley has 1800 collaborators, 1500 contractors
- Earth science predominant
- TEMPO mission delivered
- CLARREO Pathfinder (CPF) has passed important milestone
- Airborne campaign for Aeolus validation using lidar
- Airborne campaign Firex-AQ, ACTIVATE
- SAGE, etc
- Need to get results out for applications, many missions will be housed at Langley data center
- Make data available to the people.

Stoyka Netcheva, WMO Secretariat

- Thanked participants for their support of WMO data management, which is very important.
- Metadata sharing, discoverability, access to improve.

1. INTRODUCTION OF PARTICIPANTS

Gao Chen

- Science panel of observational data

Ryan Stauffer (technical), Debra Kollonige (data archiver, took over from Jacquie Witte)

- SHADOZ project
- SCOPE project, AQ in Gulf of Mexico

Kjetil Torseth

- Research Director at NILU
- Acid rain, EMEP, EBAS database
- Head of EMEP CCC
- ACTRIS data center
- Stratospheric research NADIR (no longer active)
- Collaboration with EEA

Richard Eckman

- Programme manager at NASA Langley and Washington for several missions
- Chair on WMO atmospheric variables team

Makhan Virdi

- Science data outreach center
- Host lots of data from aircraft, sondes, ships
- Works with Gao and Nate James

Judd Welton

- Aerosol, clouds
- MPLNet project manager
- Co-chair of GALION
- SDS-WAS Americas Steering Board
- Member of ET-WDC
- Wants to build GALION data center

Enrico Fucile

- WIS branch
- Metadata and data representations
- Used to be at ECMWF leading observations team
- Expertise with data formats and metadata, operational aspects
- WMO moving towards Earth Systems approach, try to harmonize data formats and increase interoperability

Tom Kralidis

- MSC EC, WOUDC
- Geospatial and Open Data Systems (GODS)

Atsuya Kinoshita

- WDCGG
- Senior scientist,
- Involved with WDCGG since 1999

Rosemary Baize

- Research Director, responsible for data center for Langley
- Nasa encouraged to reach out to non-traditional data users
- Think long-term about data mgt for large datasets, cloud vendor lock-in

Keichii Sato

Ted Habermann (from Colorado)

- Metadata standards for airborne data centers
- Used to be at NOAA and involved in WIS
- ISO, ICARTT

Nate James

- Data systems engineer at Nasa Goddard, EOSDIS, core capability
- 12 discipline oriented data archives, Nate is the engineer for atmospheric science data center at Langley

Jörg Klausen

- Scientific advisor at MeteoSwiss
- Chair ET-WDC, Co-chair TT-WMD

2. REPORT BY CHAIR OF ET-WDC, OBJECTIVES OF MEETING, APPROVAL OF AGENDA *(Joerg Klausen)*

Background, purpose and objectives of the meeting.

To understand what progress is made by WDC and networks and if not why, challenges, how to do things better, agree on vision for the future.

Items 7&8 on Agenda are open for ideas, suggestions, where you fit, what part of the plan and what actions each participant identifies with.

In reference to interoperability – we not only understand mapping code lists but it is open for input and broad ideas.

Item 9 Expected Draft actions to follow up.

Outcome action items will be uploaded on website.

Work of the team set up under OPAG EPAC Technical Commission of Atmospheric Sciences (CAS) with focus on maintaining and improving the World Data Centres, interoperability of data centres with GAWSIS and its transformation into a centralized catalogue of GAW observations and repository of WMO metadata of composition measurements, increased visibility of those data based on continuous innovation and implementation of technology advancements.

Meeting aims to increase cooperation between data providers, networks and links to data archives and together move towards achieving the goals of WMO for better interpretation and integration of supporting information for improved services.

The importance of physical and chemical parameters of atmosphere in focus area are reflected in GAW Implementation Plan with specific objectives related to data management which need to be updated.

Finally, how to achieve progress as team in harmonized way based on common objectives.

Discussion

- PBL and similar variables have no clear home in GAW but also nowhere else, while they are very important for GAW.
- [Suggestion Chair] discuss later in part where we will try to define variables.

AI2019-1: Bring ancillary variable PBL height (and related variables) to the attention of SSC and discuss how they could fit into the GAW data management (Chair, November 2019).

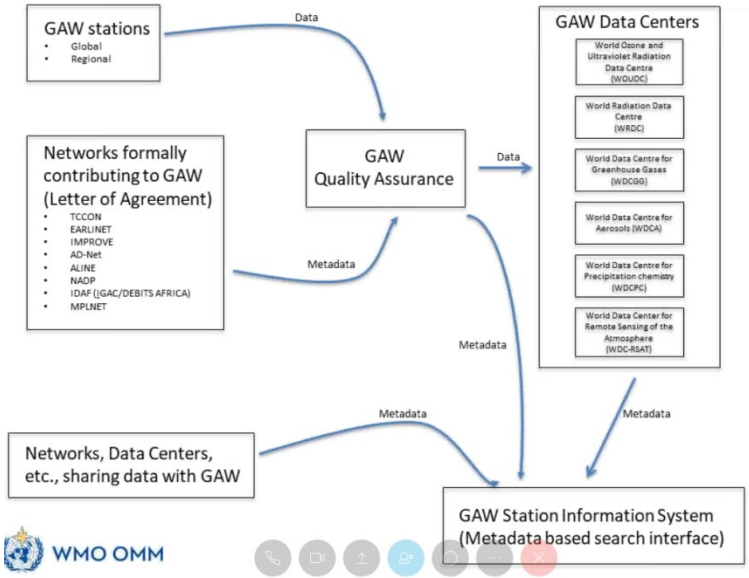
3. RELEVANT OUTCOMES OF WMO CG-18 AND EC-70

(*Stoyka Netcheva*)

Cg-18

1. 5 goals approved
   a. Serve societal needs
   b. Enhance Earth system ob and predictions
   c. Advance targeted research
   d. Close capacity gap
   e. Strategic alignment of WMO structure and programmes for effective policy- and decision-making.

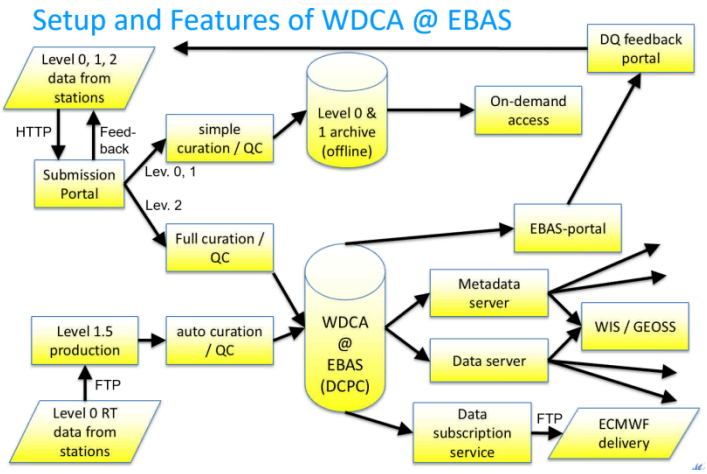
2.



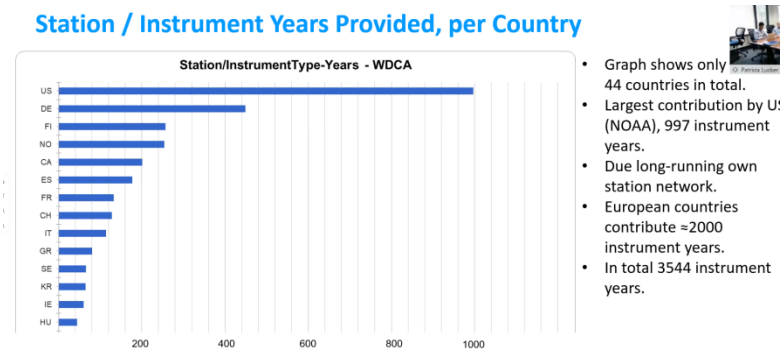
- Key priorities
- Actions in GAW IP
  - A-DM-xxx
  - Not too much progress on record, but will take stock after this meeting.

4. REPORTS BY GAW WDCS: ACHIEVEMENTS, STATUS AND PLANS

Doc 4.1 - WDCA

- 

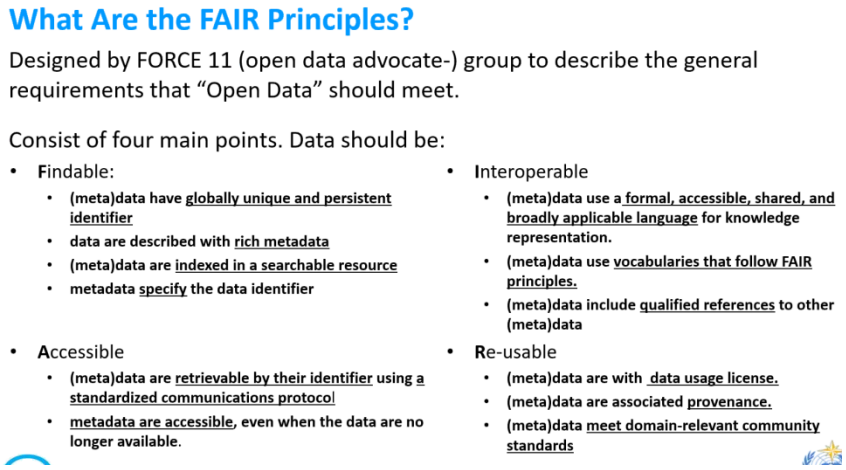
- Data levels agreed throughout Europe
- Continuous effort to develop tools to register data
- Mentions free DAQ s/w for stations wanting to participate in RT delivery
- Ca. 150 instruments reporting each year

- 

- More KPIs in presentation
- FAIRness and open data present challenges for data archives
- Access to data have different motivations and funding streams

Assisting data originators in submitting data is big task involving development of tools for checks, different initial metadata to move to FAIR principle –rich metadata, low rep and engagement outside Europe, invitations for submissions and follow ups sent.

Reporting templates are extensive for different data levels, check files and clean them, carry out QA and check consistency of data files. Relational database can support any kind of data – different platforms, measuring methods, instruments, different data heavy metals, photochem products, can combine different type of data while retrieving them from portal, feedback is incl. and information for issues on submitted data if problem exists. NRT for operational forecast already developed. Based on FAIR principles to interface with WIS, GEOSS, Vocabulary– to verify consistency with WMO. Needed harmonized vocabulary and to be based on where data are used.

- 

Discussion

For aerosol applications shown statistics by WDCA covers a small part of aerosol data. Most of aerosol data are outside GAW – for example, AERONET and LIDAR data and are not contributing for various reasons. Half of the contributors are not part of NASA. All existing aerosol data to be put in one place is difficult. What limits integration? Need rational solution to integrate huge suite of valuable measurements. Topic collection? To find a way to harmonize including use of DOIs, data system, metadata standard?

- Enrico Fusile: WMO working on a new data guide that will include the FAIR principles (which can be interpreted and implemented in different ways).
- Ted Haberman: what formats are emitted by the EBAS metadata server?
  - Delivering to WIS, working on WIGOS format.
  - Vocabularies are still a major issue, have advanced quite far with CF, WMDR still a challenge.

European infrastructure project is a challenge. DOIs are considered for final products. Data to be open but their use documented by DOIs, licenses, open licenses to be done in a proper way and kept.

Appears there are different terminologies, different meaning of different levels. Data levels consistency within GAW and WMO.

AI2019-2: Consider WMDS code table 7-06 (Level of data), cf. https://github.com/wmo-im/wmds/blob/master/tables_en/7-06.csv and create issues if necessary. (all, Dec 2019).

Doc 4.2 - WDCRG

- WDCGG (June 2015)
  - Total 1076/1137 datasets, the latter number reflects parallel observations at some locations
  - CO: 208/376,
  - NOAA flask samples represents 589 datasets (CO=89, VOCs=500)
  - Remainder (~172 datasets): mainly EMEP data which has been downloaded by JMA to populate WDCGG
- WDCRG Sept 2019 (status 2018 in brackets)
  - Total 2281 datasets (1923)
    - 83 stations (63)
    - 35 countries (28),
    - 129 components (126)
  - Historic data (-> 2013):
    - 1216 datasets
    - 49 sites
    - 22 countries
    - 111 components

- Request to re-submit to get data conforming "with our metadata standard".
- Issue with NOAA flasks not resolved. NOAA VOC flask data not imported yet, because NOAA doesn't report data.
- NILU mostly working with INSTAAR, not with NOAA directly.

WDCRG Challenges –different standards; training and capacity development needs, how to do different parts of data management; historical data what to do with standards – new metadata, new standards, issue with data versions to not replace new /better version with old one, how to do verification on versions, how to make those available to WIGOS without having duplications. Concern is also conversions and risk of leaving out valuable data. Developed already scripts for importing data but it is challenge to harvest all data. WDCs are for long term archiving –for eternity and challenges related to do that. For old datasets - try to ensure the best data come to DC – and it is an obligation of DC to maintain history of dataset, need contact person and that is a challenge for historical data.

Doc 4.3 - WDCGG

WDCGG Statistics (Data providers/Contributors)

Contributors (Data providers) — Total 69

JMA	HMS	GERC
NOAA	IAFMS	NIES
AEMET	IGP	NILU
AGAGE	IMKIFU	METRI
AICH	INRNE	NIWA
AIST	INSTAAR	NMH
ANSTO	IOEP	ONM
ARSO	INPE	OSAKAU
BAS	ISAC	PolyU
BMKG	ITM	RIVM
CHMI	KMA	RSE
CMA	KMD	SAIPF
CSIRO	KSNU	SAWS
DMC	KUP	SHIZU
IAA	LA	TU
DWD	LAMP	UBAG
ECCC	LSCE	UBAA
EMA	MGO	UNIURB
Empa	MMD	UMLT
ENEA	MRI	UYRK
FMI	NAGOU	UNIVBRIS
FRA	NEDO	JAMSTEC
HKO	VNMHA	IIA

Station — Total 205

GAW Global	32
GAW Regional	112
GAW Contributing networks	1
GAW Other elements	27
Mobile	33
REGION I	12
REGION II	36
REGION III	7
REGION IV	35
REGION V	15
REGION VI	57
ANTARCTICA	10
MOBILE	33

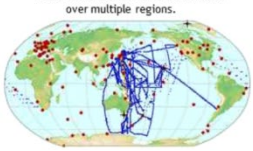
Gas species — Total 56

CO_2, CO_2[C-13], CO_2[O-18]	
CH_4, CH_4[C-13], CH_3D	
N_2O, SF_6, SO_2F_2, NF_3, COS	
Halocarbons	2
CFCs	6
HCFCs	4
HFCs	11
PFCs	4
Halon	3
Reactive Gas	10
Other Gas	1
Radionuclide	3
TIC	

Country/Territory — Total 56

REGION I (Africa)	10
REGION II (Asia)	13
REGION III (South America)	5
REGION IV (North and Central America)	5
REGION V (South-West Pacific)	6
REGION VI (Europe)	25
ANTARCTICA	7
MOBILE	6

Note: There are countries that extend over multiple regions.



How to find/use data (for users)

- Download observation data -

Download list

- ✓ Download from multiple observational sites is possible on the new WDCGG website.
- ✓ User ID registration is required to download data.

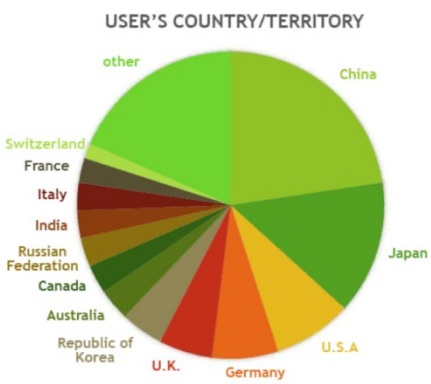
My Favorite

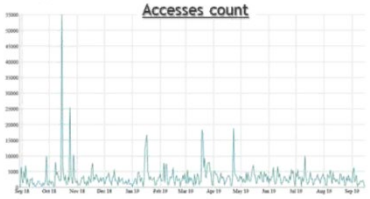
- ✓ If you often look at specific data, please use My Favorite.

When you download, the download information is sent to the data provider(s).

Statistics on WDCGG users

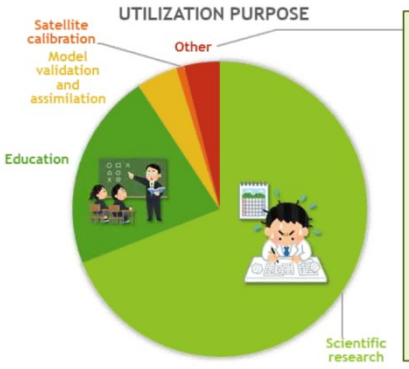
- Who and how many access -





Statistics on WDCGG users

- How used -



Other purpose

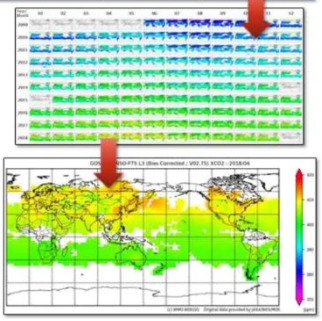
- ✓ Internal and public research on climate change and megatrends
- ✓ Data to use within my third year dissertation
- ✓ I have been Interested in Climate Change for the last 10 years.
- ✓ Personal interest in climate change
- ✓ Airborne validation
- ✓ News reporting

..... and so on.

Provision of satellite data since March 2019

Satellite	Organization	Gas Species	Version * : the latest version	Updated Date	Gallery/Metadata
GOSAT	NIES (National Institute for Environmental Studies)	CO_2	0053-9001-1001-08-08-9999 -2019-03-19-0900*	2019-03-19	view

- ✓ WDCGG began online provision of CO_2 observation data from Japan's Ibuki Greenhouse gases Observing SATellite (GOSAT).
- ✓ Monthly global map of column-averaged CO_2 mole fractions can be seen in the Gallery/Metadata.
- ✓ Downloadable CO_2 data are L2 column volume (SWIR) in daily (combined by month) HDF5 Format.



DL	Gas Species	Organization	Type	Period	Filename	Size
☑	CO_2	NIES	daily	2009-04-23 - 2009-04-30	SWIRL2CO2_200904_V02.75.tar	4.18MB
☑	CO_2	NIES	daily	2009-05-01 - 2009-05-31	SWIRL2CO2_200905_V02.75.tar	5.61MB
☑	CO_2	NIES	daily	2009-06-01 - 2009-06-30	SWIRL2CO2_200906_V02.75.tar	15.66MB
☑	CO_2	NIES	daily	2009-07-01 - 2009-07-31	SWIRL2CO2_200907_V02.75.tar	18.96MB

- Now also hosting OCO-2 data

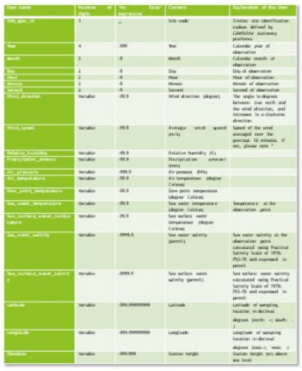
Currently working plans

Publication of meteorological data

- ✓ WDCGG started collecting meteorological data records as environmental information for each observation station from this year again.
- ✓ Publication method is undecided yet.

Expansion of satellite data

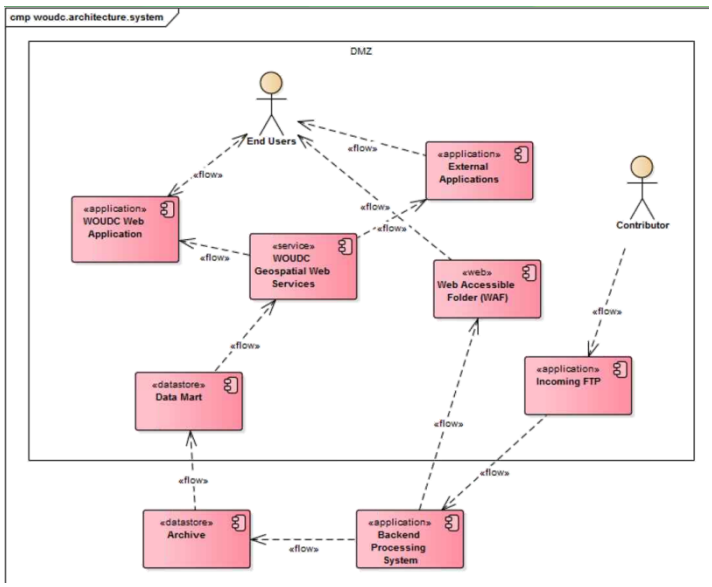
- ✓ WDCGG plans to continue improving its services for the collection, archiving and distribution of satellite data worldwide, including for GOSAT-2 (the successor to GOSAT).
- ✓ In addition, OCO-2 data will be included.



- 

Under Harmonized metadata: information and clarification needed on variables, required format, data originator, DOIs-is planned for each observation data, uncertainty will be included in future, next: exchange with GAWSIS using API. Find harmonized checking and machine reading.

Doc 4.4 - WOUDC

- 

Number of Files 2018

Broad-band	123147
OzoneSonde	90117
Lidar	675
Multi-band	91649
TotalOzoneObs	100804
Spectral	269217
UmkehrN14 2.0	9889
RocketSonde	177
TotalOzone	81608
UmkehrN14 1.0	10522
TOTAL	**777805**

Downloads 2018

Dataset	WAF (number of files)	Dataset Archive Files	Geospatial Web Services (number of requests)
Total ozone obs	2993589	30	1162
Ozonesonde	266959796	31	8977
Total ozone daily obs	9991490	38	7261
Spectral	999597	25	1226
Multi-band	385929	21	538
Broad-band	253428	23	973
Umkehr 2	327136	25	353
Umkehr 1	354592	23	417
Lidar	6212	25	421
RocketSonde	16174	22	281
Total	**282,287,943**	**263**	**21,609**

Website Visits 2018

Year	Visits (<= 30 min)
2015	17245
2016	59042
2017	98084
2018	547352

- Made Open by default, no login etc. resulted in increased visits and use.

Top Hits by Domain 2018

Domain	Hits
ecmwf.int	281487149
noaa.gov	41143661
oma.be	8906628
auth.gr	3185643
nasa.gov	2946676
kishou.go.jp	1694876
(not set)	1492231
archive.org	759360
chonbuk.ac.kr	461577
knmi.nl	304746
...	
total	**343804972**

Enhancements

Updates

- Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST)
- Global and zonal total ozone variations estimated from ground-based and satellite measurements
- HTTPS (w/ HTTP -> HTTPS redirect)
- Station Data Summaries
- Non-standard format support tools
  - SHADOZ
  - Umkehr

Interoperability

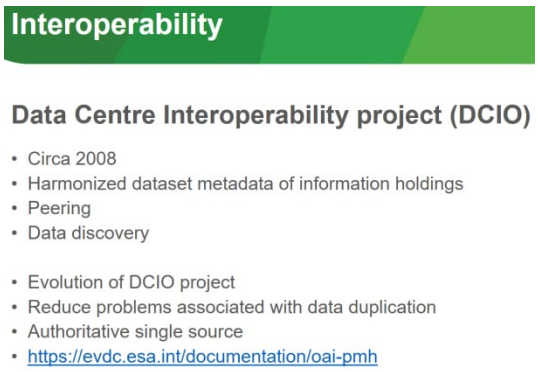
Key Drivers: Alignment

- World Meteorological Organization
  - Weather, Climate, Water
- Public Access: Canadian Open Data
  - WeatherCAN: Mobile Weather App
  - Canadian Centre for Climate Services (CCCS)
- Government of Canada Geospatial
  - Federal Geospatial Platform (FGP)
- Beyond: GEOSS, etc.

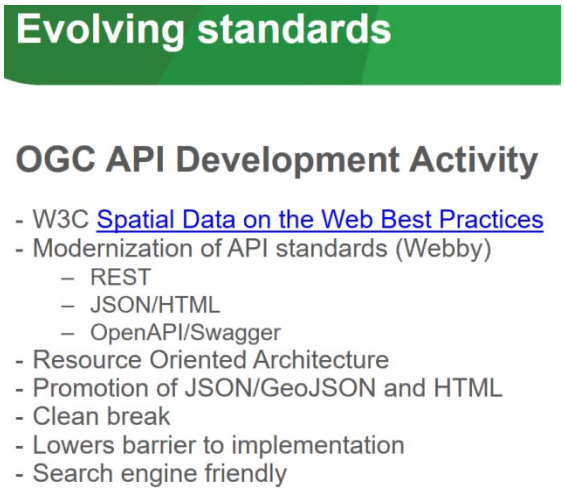
Interoperability

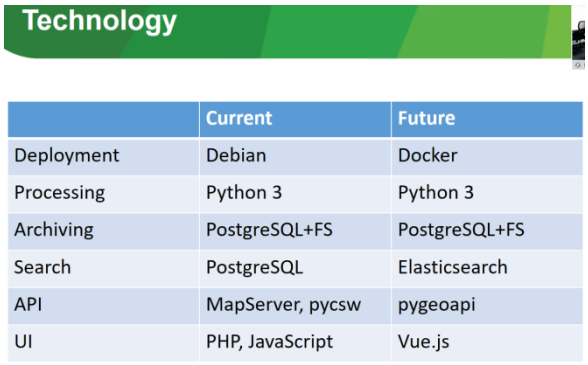
Core Principles

- FAIR
- Standards
- Open Geospatial Consortium (OGC)
- World Wide Web Consortium (W3C)
- International Organization for Standardization (ISO)

- 

NDACC 'integration' by 2020

- 

- 

	Current	Future
Deployment	Debian	Docker
Processing	Python 3	Python 3
Archiving	PostgreSQL+FS	PostgreSQL+FS
Search	PostgreSQL	Elasticsearch
API	MapServer, pycsw	pygeoapi
UI	PHP, JavaScript	Vue.js

Developments since introduction of new platform used contributors input and SAG direction. Data cleansing challenge. Delay in submission or not reporting improved using WMO contractor. The search is made friendly, anonymous which increased visitor numbers. Build on FAIR, standards, open, ISO standard. Request from O3 UV SAG for station DOIs. Metadata is part of data file and is extracted from files. WOUDC linked 9 datasets with GAWSIS. Work on linking with NDACC- SHADOZ and Eubrewnet in progress. Build for finalized data. NRT of Canadian O3 sondes through GTS.

Discussion

- How much should data archivers assist data providers in providing correct data?

Doc 4.5 - WRDC

Head of DC submitted presentation but not participating, not presenting. Not clear exact progress or issues or plans.

To follow up on commitment and progress.

Doc 4.6 - WDCPC (*Chris Lehmann*)

- WDCPC currently managed by Van
- Workplan submitted to NOAA
  - Serve as intake for TAD data
  - Validate data
  - Store data at University of Illinois
  - No replication of existing regional data centers
  - Provide links to these other programmes
  - Fairly well harmonized efforts through SAG TAD.

AtmDep has not taken off since suggested after working on Global Atm Deposition Assessment. WDCPC Precipitation chemistry data center is in exploratory phase, transitioning to new home and management. The current stage is to submit transition plans to NOAA including DB structure, data flow. At this point the work plan is more like a wish list as to what we hope to achieve. Considered now intake of WDCPC data from assessment, standard templates, validation of data. No clear strategy for global coverage, access to global datasets, what is role of Regional centers? Agreement from 2014 is to serve not duplicate Regional Data centers but have archival of data. Standardized – metadata cross walk, seeking guidance for consistent format and by setting it up and can move to WIGOS now. Need formats and how to do metadata exchange. Which one to use? Collect /provide links to orphaned data is challenging as need to find a contact person. Metadata availability – how to get it in WIGOS? WIS? Infrastructure exits but not running at this point. Need to consider harmonization and interaction with other centers.

Discussion

- EANET, EMEP has plans on how to exchange metadata with GAWSIS, how about for NADP, CaPMon, DEBITs.
- NILU has compiled datasets for assessments and keeps relationship with CaPMoN, SAG. Not sure how to set and be searchable and discoverable by Precip Center.
- Some have simple/basic others more complex metadata and might take years to develop while the intention is to preserve developed infrastructure.

Doc 4.7 - WDC-RSAT

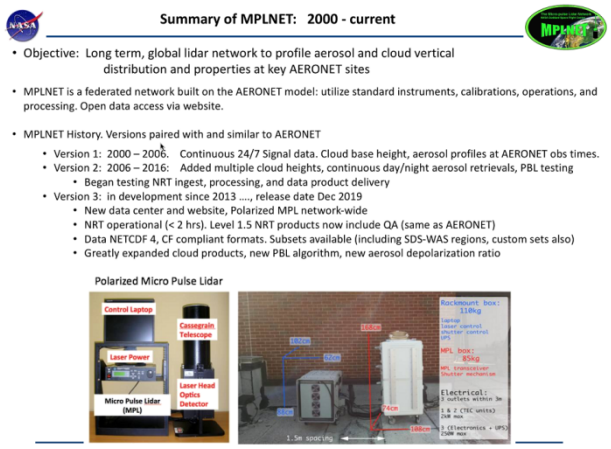
Not represented. To follow up on commitment and progress, status of facility in GAW.

AI2019-3: All WDCs to (re-)establish/confirm metadata exchange with GAWSIS-OSCAR/Surface. Consult with MeteoSwiss and https://library.wmo.int/doc_num.php?explnum_id=5844 (Section 3.12) for guidance and present timeline. (all WDCs, 10 Nov 2019)

5. CONTRIBUTING PROGRAMMES, OTHER DATA ARCHIVES

Doc 5.1 – MPLNet

- Nasa's lidar network, put at key Aeronet stations



- Aeronet reaches 30 yrs, MPLNet reaches 20 yrs, bound to stay, so needs to become more operational.

WMO GAW Aerosol Lidar Observation Network (GALION):
A lidar network of networks organized through the GAW program, and is composed primarily of the world's leading lidar networks. Each is an official contributing network to GAW (or soon will be).

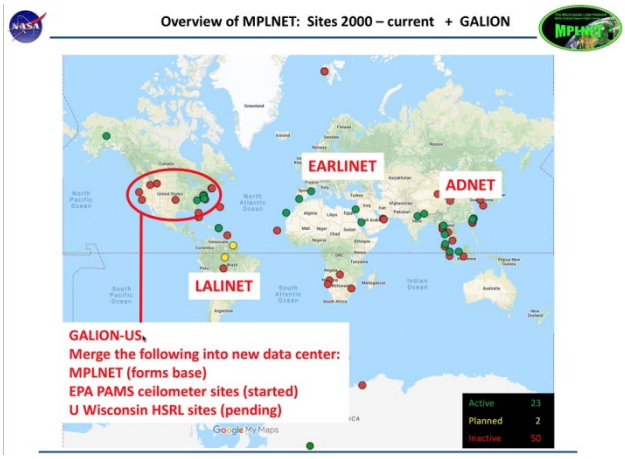
GALION Networks:
EARLINET
AD-NET
CIS-LINET
LALINET
CORALNET
CREST
MPLNET (global)
NDACC (global)

GALION Co-Chairs:
Gelsomina Pappalardo (CNR IMAA)
Ellsworth J. Welton (NASA)

Work Groups:
Calibration, QA/QC, processing/products, applications, data center

Map is a few years old at this point …

- CORALNET no longer active

- 

- **Overview of MPLNET:** Version 3 Product Suite

Detailed information on V3 Products: mplnet.gsfc.nasa.gov/product-info/

V3 Product	Descriptions
NRB	Lidar signals; volume depolarization ratios; diagnostics
CLD	Cloud heights; thin cloud extinction and optical depths; cloud phase
AER	Aerosol heights; extinction, backscatter, and aerosol depolarization ratio profiles; lidar ratio
PBL	Surface-Attached Mixed Layer Top and estimated AOD
Product File Formats	
Formats	MPLNET V3 products are NETCDF 4, CF compliant files. Subsets for each product may be selected to reduce file sizes.

-

Product Levels	Availability	Calibration	QA Screen	Ancillary Input
L1_NRB L1_CLD L1_PBL L1_AER	Automated Browse: Near Real Time Download: Next Day *	intial, ongoing field calibrations	none	GEOS5 Forecast NRT, reprocessed next day with GEOS5 Assimilated, AERONET L15 AOD
L15_NRB L15_CLD L15_PBL L15_AER	Automated Browse: Near Real Time Download: Next Day *	intial, ongoing field calibrations	L15	GEOS5 Forecast NRT, reprocessed next day with GEOS5 Assimilated, AERONET L15 AOD
L2_NRB L2_CLD L2_PBL L2_AER	upon request †	intial, ongoing field calibrations, post calibration, additional‡	L2	GEOS5 Assimilated, AERONET L2 AOD

* Near real time data can be provided to site partners and forecasting/modeling centers
† L2_AER products subject to availability of L2 AERONET data
‡ Additional L2 calibrations may include corrections for instrument temperature and manual inspection of data

MPLNET: Data Communications

In 2018 NASA declared all data communications must be secure (SFTP, FTPS, HTTPS)

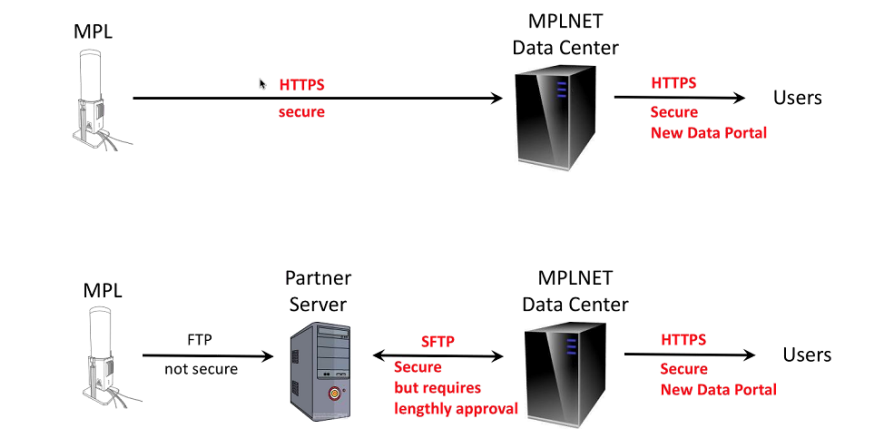
MPLNET Incoming

- Raw data files from network instruments
- Custom binary format from manufacturer (open)
- Previously utilized combination of FTP (primary, push) and SFTP (pull from some partners)
- 2019 converted to a web data portal (HTTPS) for all incoming data
  - Security features
    - Instruments assigned a passcode
    - Uploaded data files are whitewashed and content checked
  - We provide assistance for some partners who need to push files to us

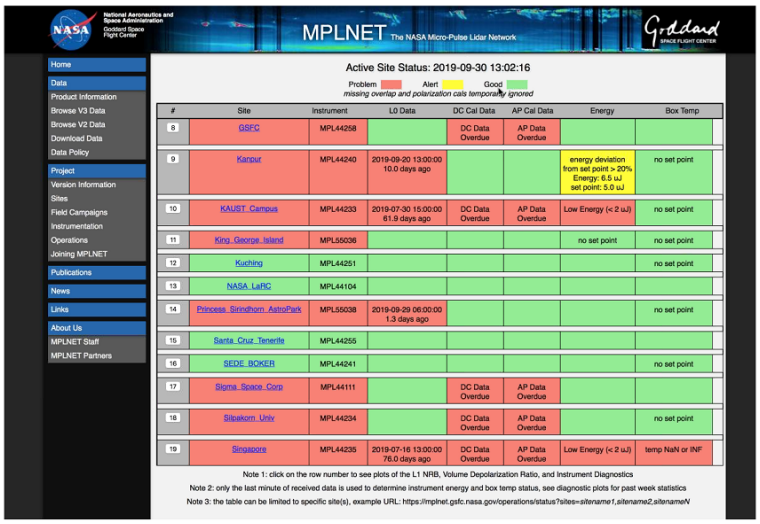
MPLNET Outgoing

- Data products delivered to user community
- Data products have always been open access, no username or password
  - Data quality same as AERONET project, clearly stated on web & data portal
- Webserver converted to HTTPS in 2016
- Version 3 outgoing data portal redesigned in 2018
  - Copied format utilized by NASA Global Modeling and Assimilation Office (GMAO)

MPLNET: Data Communications



MPLNET Network Management: Site Status & Automated Alert System



-

MPLNET Product Variables vs GCOS ECV Proposed List

NRB Product
NRB Lidar Signal
Volume Depolarization Ratio
Lidar Co-pol Signal
Lidar Cross-pol Signal
Diagnostic Variables

CLD Product
Cloud Heights * (but base missing)
Cloud Phase
Cloud Optical Depth *
Cloud Fraction Profile (column only)

AER Product
Aerosol Height
Aerosol Optical Depth *
Sunphotometer & Lidar
Aerosol Backscatter
Aerosol Extinction
Aerosol Depolarization Ratio
Aerosol Lidar Ratio

PBL Product
Mixed Layer Height
Mixed Layer AOD

-

IPET-OSDE OSCAR/Requirements Workshop Essential Climate Variables (ECV)

(this is a work in progress ….
…. Some differences with satellite GCOS ECV team)

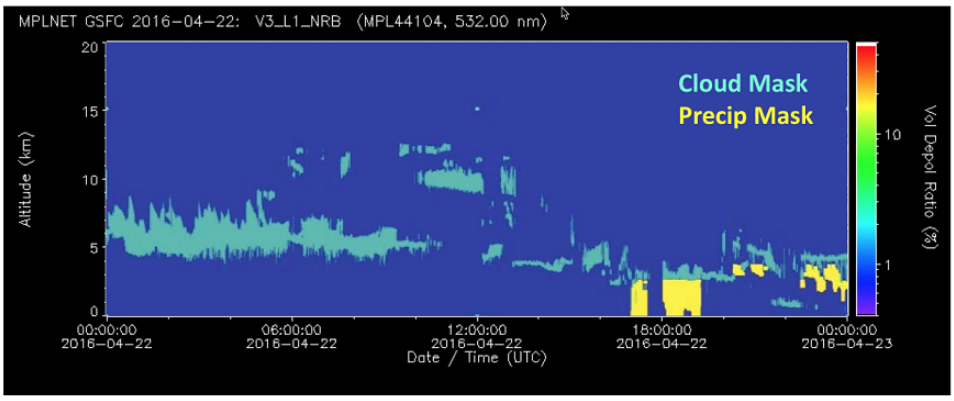
Cloud Cover (column fraction)
Cloud Top Height
Cloud Top Temp
Cloud Optical Depth
Cloud Liquid water path
Cloud ice water path
Cloud drop Re

Aerosol Optical Depth
Aerosol Single Scattering Albedo
Aerosol Layer Height
Aerosol Extinction Coeff (diff between lidar & in-situ)
Aerosol Size Distribution
Aerosol Composition
Aerosol Refractive Index

MPLNET Research to Ops: aerosol-cloud interactions

Detection of light precipitation (below typical radar detection capability)

- Contribution to wet deposition for frequent light rain events
- aerosol processing at cloud base
- Series of papers published on estimating rain drop size & rain rate



Lolli, et al., JTECH, 2013.
Lolli, et al, JTECH, 2016.
Lolli, et al, Remote Sens., 2018.

Series of papers on proof of concept
Rain drop size & rain rate

Next paper on operational precipitation detection and mask

MPLNet and GALION had been developed as projects and including all functions QC, processing, reprocessing and not easily merged. Measured, available parameters have not been included in GCOS only half of the variables have been listed as essential climate variables ECV. Is there a constraint in the number of variables? Now seems to be review phase of Climate Nomenclature of observed quantities, code of variables published, units. How to overcome limitations related to variables and in particular specific for LIDAR measurements which are used in few AQ applications/ models/ forecasts? Those models/applications have their own specific requirements/formats and have been met on one to one basis. Those variables have not been fully included/reflected in variables definitions produced by (RRR) groups - example PBL height, mixed layer height, aerosol extinction. It seems there are no common definitions of variables. Need to compare what variables are included in GAWSIS, WIGOS, OSCAR/requirements. To review if variables could have been entered under different names. Conversion factors might be provided. A review process of GCOS/WIGOS variables could be used for feedback and suggestions to add missing ones? Protocols and mechanisms for adding variables are in place with WMO involving SAG and review process. NRT processing is in place and automated at 1.5 version, there is no post calibration through GALION data center. NRT data are customer oriented data provision used for forecasting and validation purposes in AQ. Networks are moving away of ftp as NASA new policy for secure communication is implemented and does not allow ftp. Have common approach on level /version data – any convention or guidelines? (→ Consider **AI2019-2**)

Where GALION, AERONET fit within GAW?

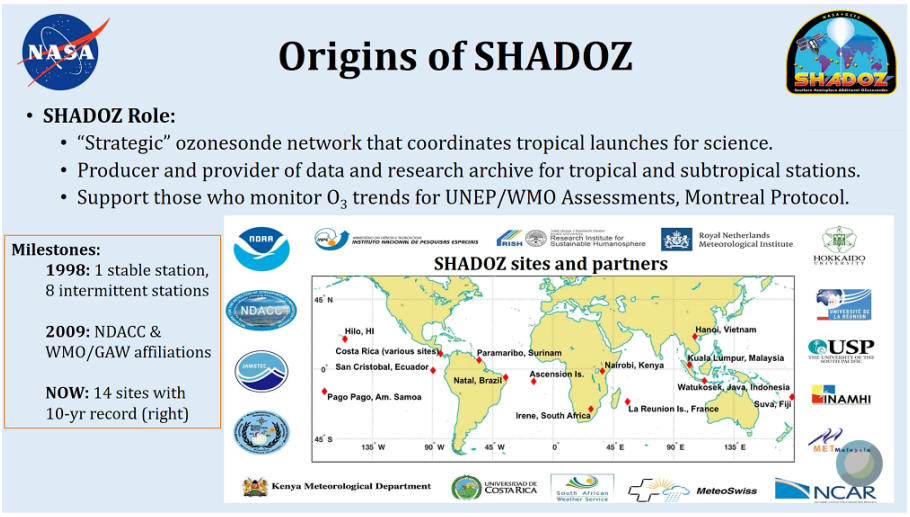
LALINET – can the name be updated?

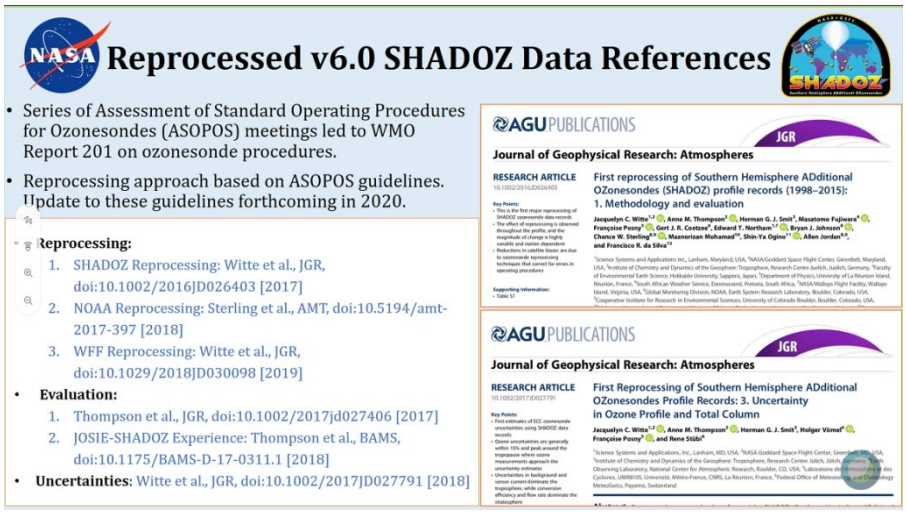
AI2019-4: Send example of metadata to be encoded in WMDR XML to MeteoSwiss (Judd Welton, asap). Judd Welton will review information on WIGOS metadata and elements-variables, prepare BUFR metadata transfer to be available with data.

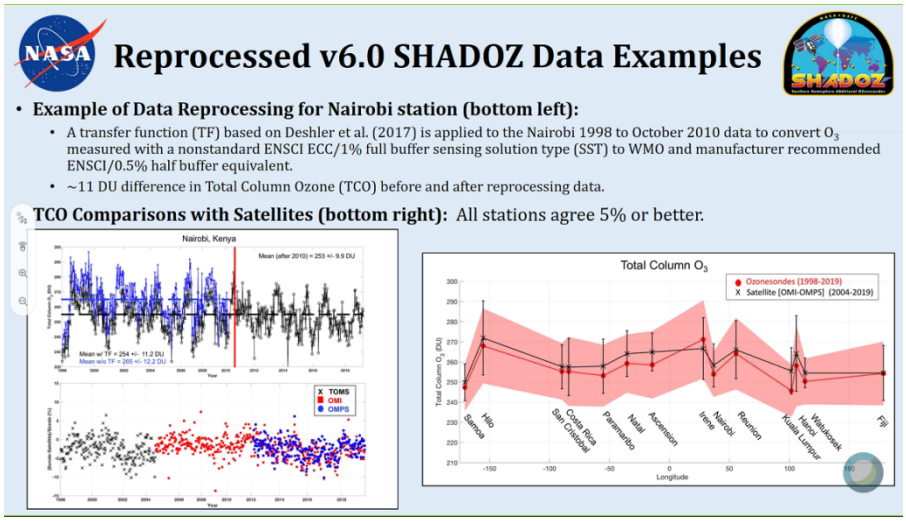
Discussion

- Tim Haberman: CF attribute convention for attribute discovery.
- Tim Haberman: THREDDS server by UNIDATA.

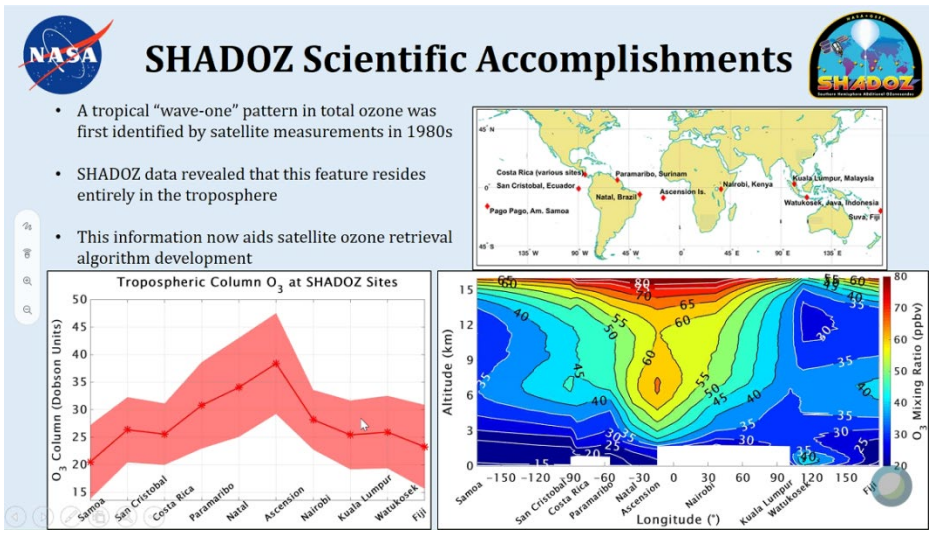
Doc 5.2 – SHADOZ

- 

- Data files are text files, which is what users want
- 

- 

- 

- 

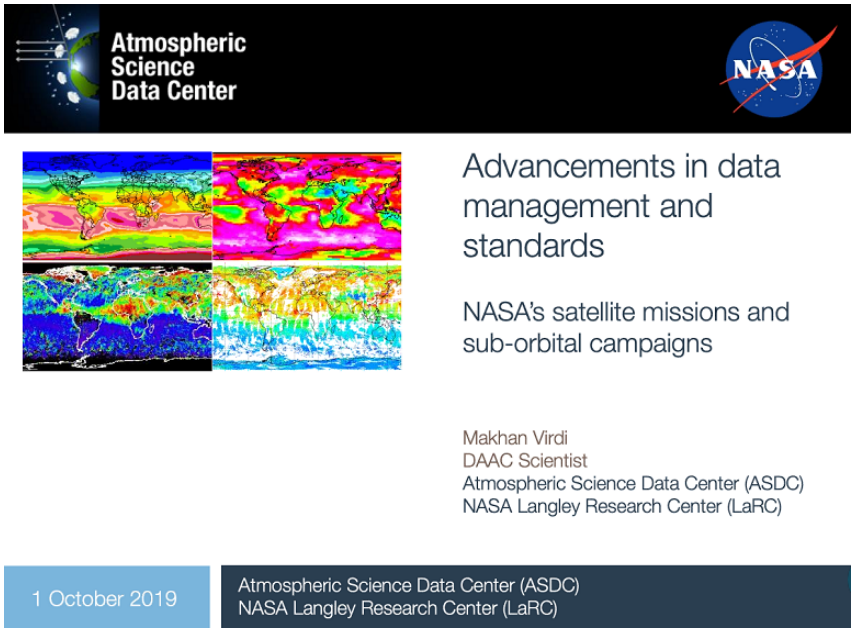
- Maximum explained by biomass burning, as it resides in the troposphere.
- SHADOZ data are submitted to WOUDC every now and then.

- WOUDC doesn't support uncertainties yet, though!

Data format are txt files, record of metadata in data with essential and desired fields, filled in through check list, flight prep information, established and QA, QC parameters. Included is uncertainty of measurements. Required fields are used for validation. Data reside with NDACC under NOAA transitioned to NASA management. MetaData connected(ing) with WOUDC and trough that to WIGOS. Continuous effort for standardization, reprocessing and homogenizing data within ozonesonde community for years results in updated versions of data. Most recent review of SOP to be completed this year and to be submitted for publication under WMO. Finalized Data are submitted/linked to WOUDC through once a year update. Version 6 is to be submitted. No correction factor is provided for TOC in SHADOZ.

Doc 5.3 – GCW Not represented?

Doc 5.4 (Other) NASA data centres

- 

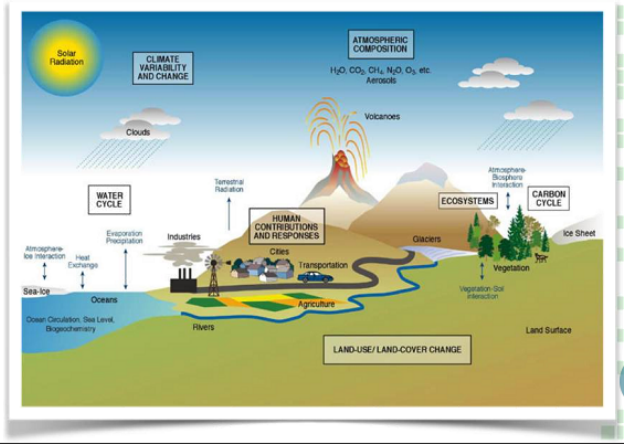
Atmospheric Science Data Center

NASA Earth Science

3

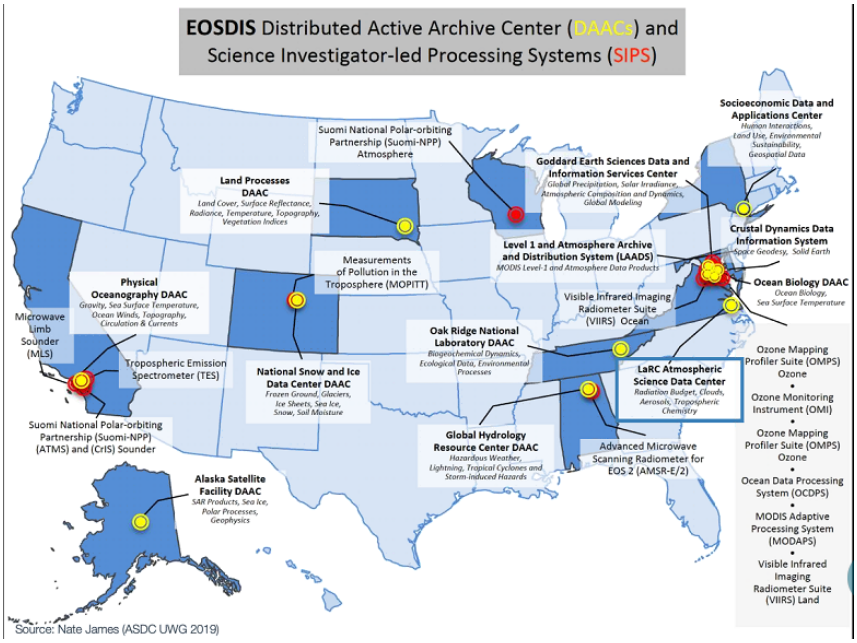
EOSDIS Comprises Data of the Whole Earth System

Atmosphere
Winds & Precipitation
Aerosols & Clouds
Temperature & Humidity
Solar radiation
Ocean
Surface temperature
Surface wind fields & Heat flux
Surface topography
Ocean color
Cryosphere
Sea/Land Ice
Snow Cover
Land
Cover & Usage
Soil Moisture
Topography & elevation
Temperature
Human Dimensions
Population & Land Use
Human & Environmental Health



Source: Nate James (ASDC UWG 2019)

- Valerie Dixon ("Nasa metadata queen")



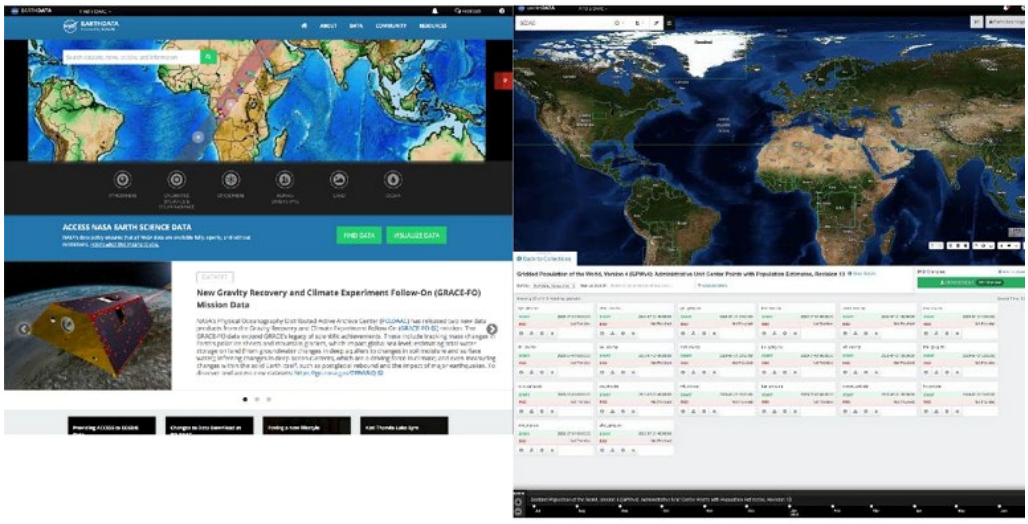
Source: Nate James (ASDC UWG 2019)

- All Nasa DAACs are searchable via the Common Metadata Repository

Atmospheric Science Data Center

NASA Earth Science

6

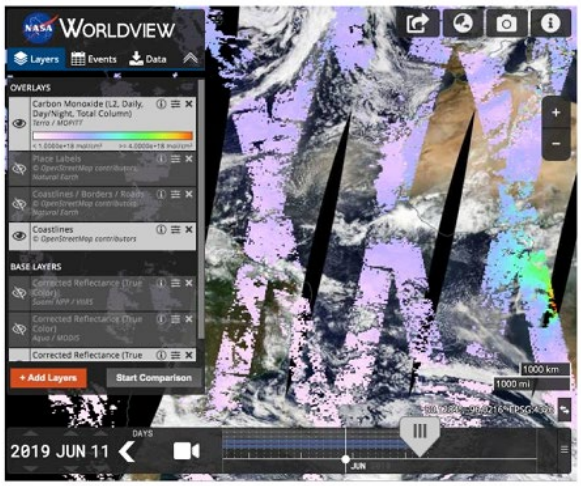


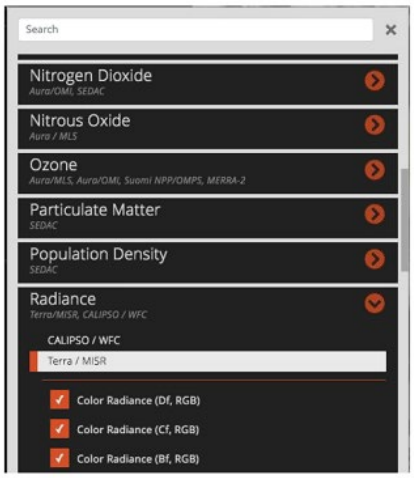
Source: Nate James (ASDC UWG 2019)

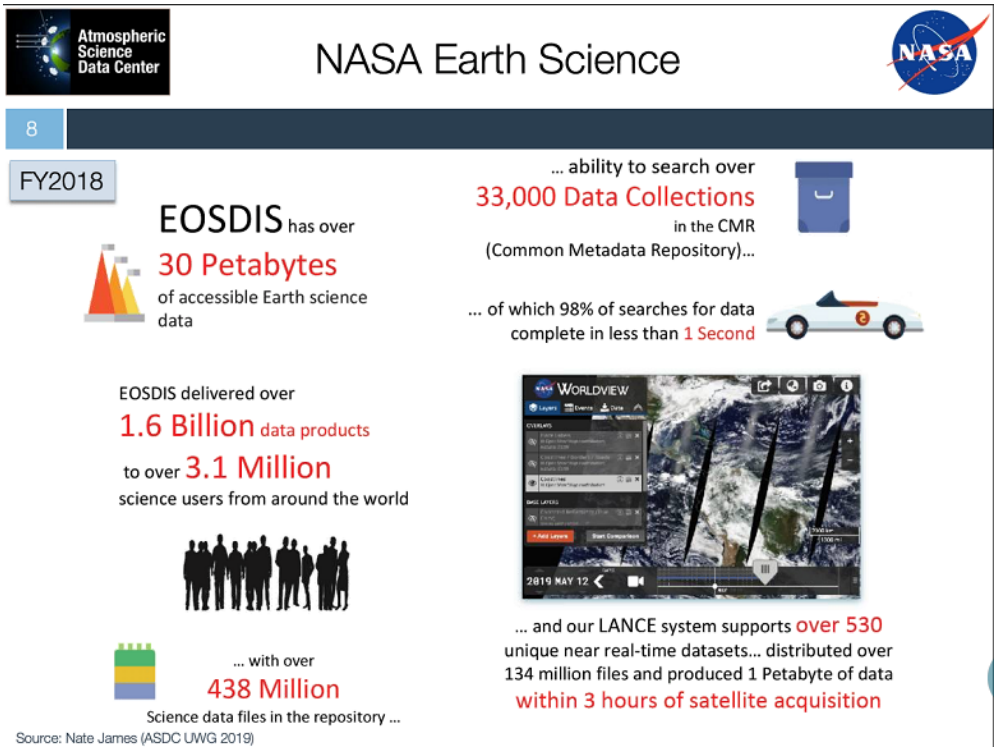
Atmospheric Science Data Center

NASA Earth Science

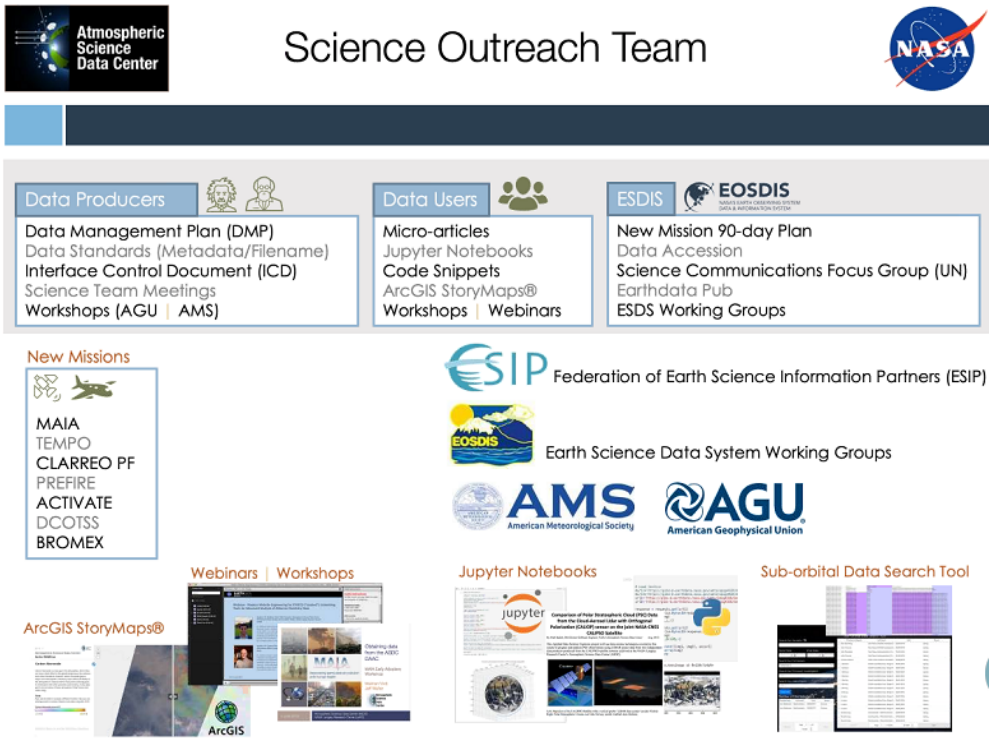
7

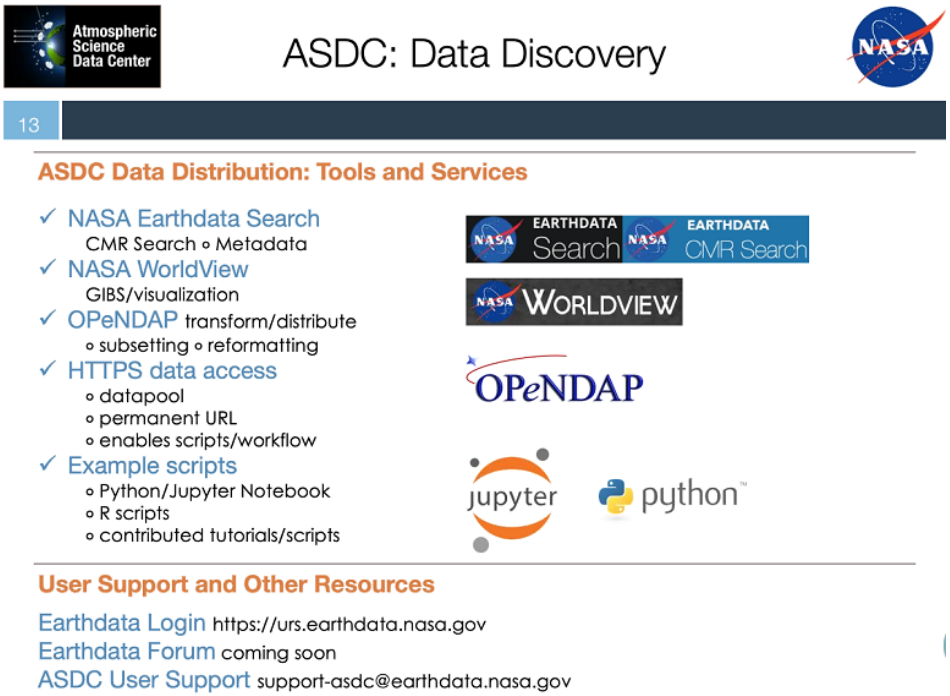




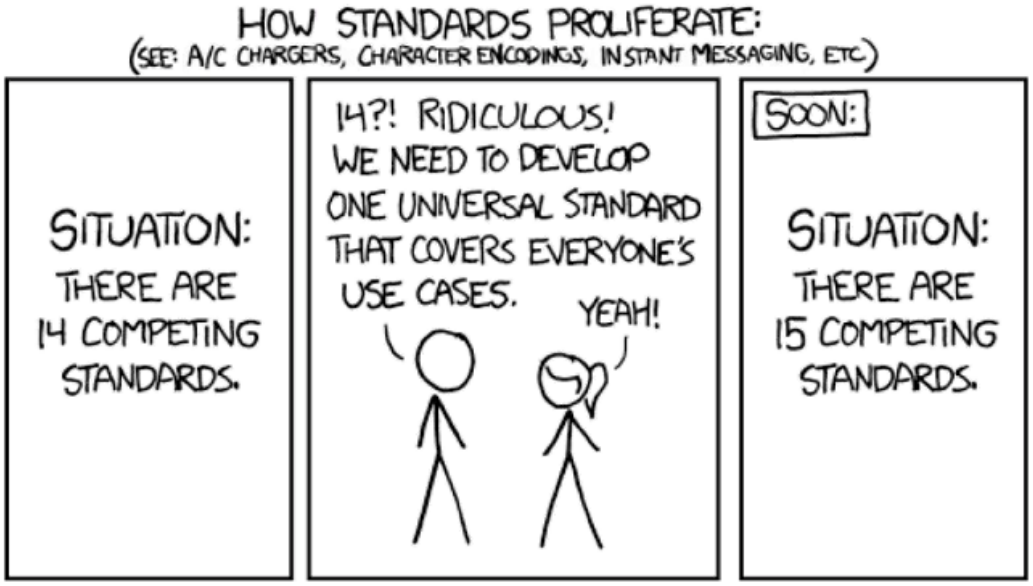
- 

- A collection is a set of data of similar observations
- Most cloud activities use Amazon, but Nasa tries to be vendor-agnostic

- 



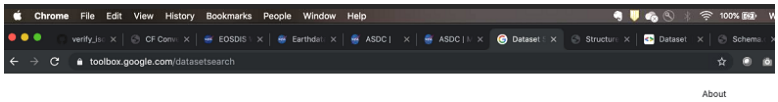
- OPeNDAP / THREDDS to allow sub-setting; requirement is a compliant data format, e.g. netCDF

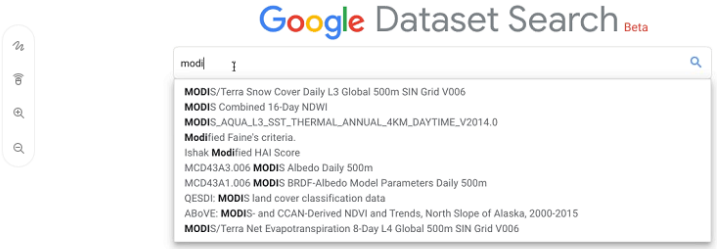


source: https://xkcd.com/927/

Metadata Standards and Data Files Formats

- Data Files
  - File Naming Convention
  - File Formats
    - netCDF, HDF, GRIB
    - ICARTT, CSV
  - Tools
    - GrADS
    - Panoply, HDFView
    - Custom: R, Python/GDAL
    - GIS Tools
- Data Collection
  - Digital Object Identifier (DOI)
  - Landing page for DOI collection
    - Mark up of landing page in JSON-LD (Linked Data)

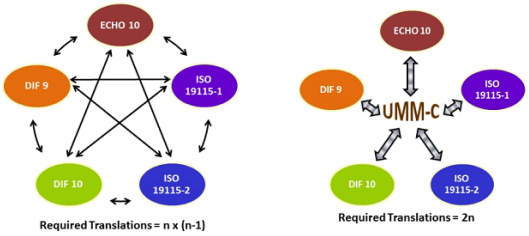




- Investigate JSON-LD

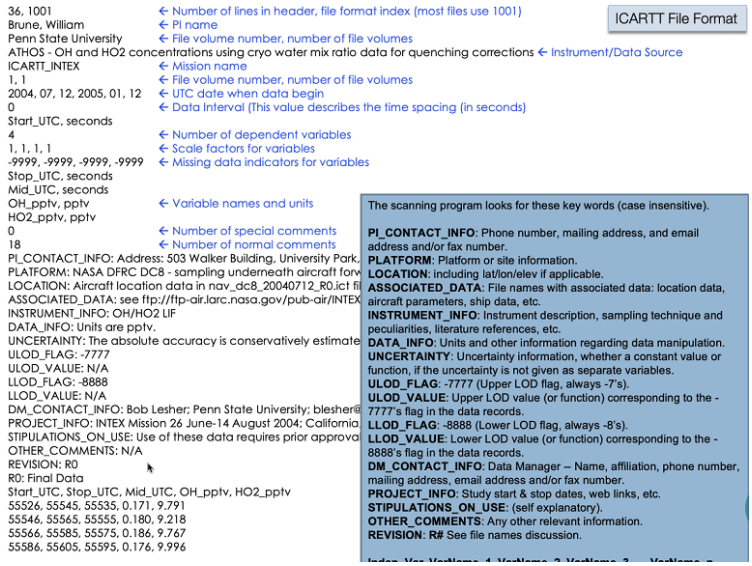
Metadata Standards and Data Files Formats

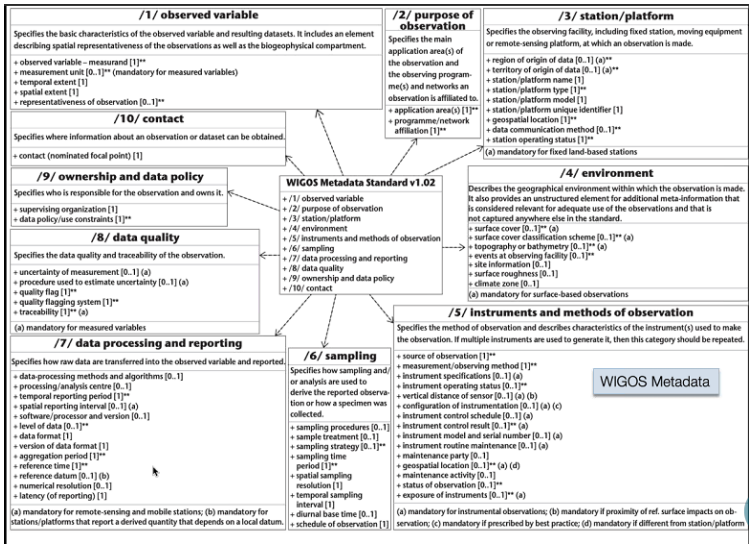
- Standards
  - Climate and Forecast (CF) Metadata (v 1.8)
  - ESIP Attribute Convention for Data Discovery (ACDD)
  - NASA's Universal Metadata Model (UMM)
  - ICARTT metadata



Source: https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr/umm

- UMM-C: Universal metadata model



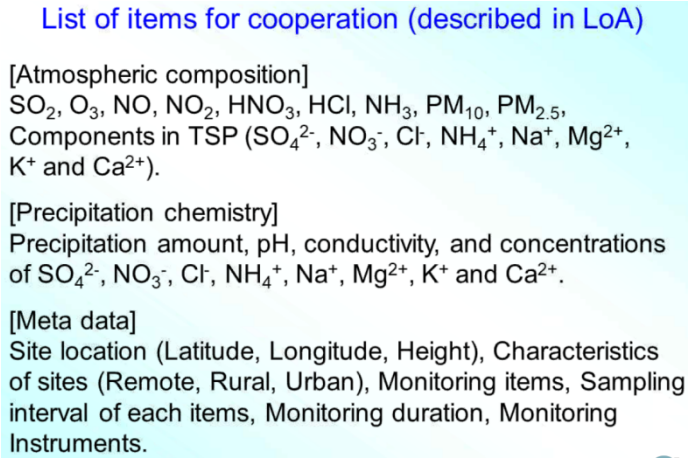


NASA DB consolidates 12 centers. Some centers do data processing. Clouds are not only used for depositing but also for some analytics. Google and Amazon work on providing tools for big data. Potential issues: Vendor locking. Taking/downloading data from the cloud is costly that's why processing is done on the cloud. If data are free how do we limit users' downloads? Develop tools for search and discovery of airborne data. Options built to download subset of info or datasets, NCFS(netCDF) compliance for discoverability, cross walk to link 1 center to another, Html node embedded and part of metadata looks like google scholar, CF compliance, data discoverability -ACDD. Instruments and stations undergo a standardized process, one metadata standard. NASA metadata have mininum essential /requirement fields for metadata in common metadata standard UMM model (online documented), also ISO compliant metadata; DOIs in NASA not to be used due to character of networks and participants.

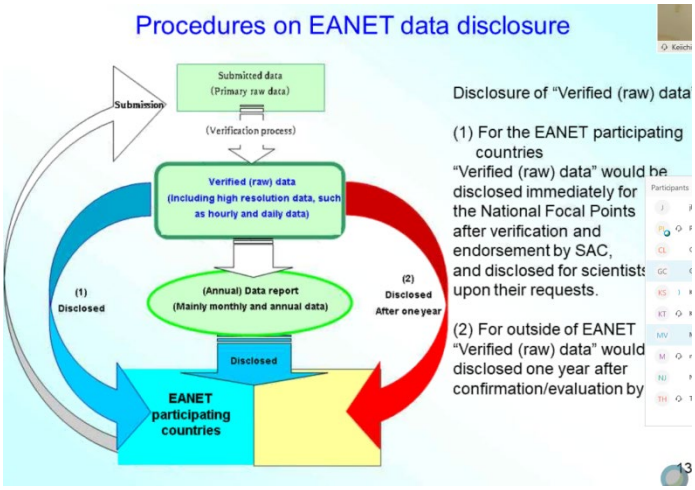
Clouds provide a new kind of analysis analytics on the cloud data. Amazon focused on facilitating analysis, defining data and metadata at production level by defining minimum requirements for metadata.

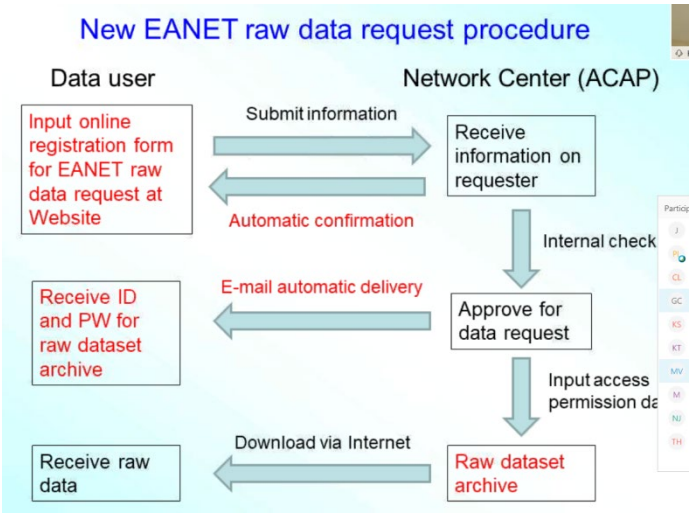
Doc 5.5 - EANET (*remotely Keichii Sato*)

- LoA with WMO signed in August 2018



- https://doi.org/10.5194/acp-2018-1283 (2019)



- 

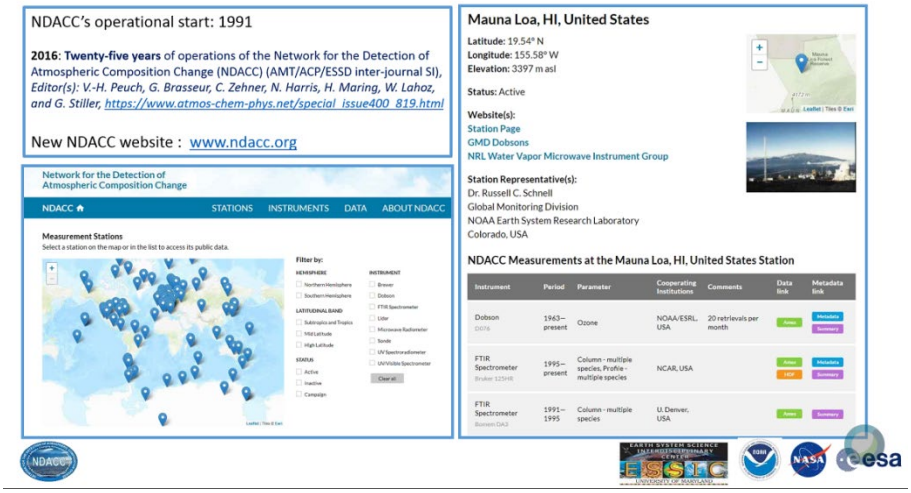
- Data report and QA/QC reports are openly available w/o login
- Data requests require registration
- Investigating use of DOI, but finds it challenging for now
- Currently, about 60 requests for data since April 2019
- Information is needed on how to exchange information with GAWSIS, metadata and in terms of the characteristics and specifics of included data (concentrations, different sampling frequencies, sampling periods are used for different data, its own citation requirements for data users, DOI ..) and how to link with WDCs/WDCPC (precipitation chemistry). How could the flow be organized? Potentially use this link to inform about 10th International Conference on Acid Deposition.

Discussion

- Enrico Fucile: Requirement for DOIs not supported in WIS, but registration in WIS would be possible
- Data access still fairly restrictive
- Vocabularies inline with WMDS?
- Enrico Fucile: Is there any plan to register data in WIS? WIS decided not to use DOIs. A lot of things implemented in EANET are within WIS framework and will not be a problem to register in WIS. Need analysis on what is done, what is missing, how to be done by different DBs not only EANET.

Doc 5.6 - NDACC *(presented by Gao Chen on behalf of Martine de Mazière)*

- NDACC is the standard for SAGE science team for validation purposes

- 

- **Facts & Perspectives about NDACC Data Handling Facility (DHF):**
  Oldest data record in the NDACC archive: Sept, 1966: Boulder Dobson #091

  Currently > 140,000 files in the NDACC data archive

  Over 1 million file downloads so far in 2018

  Files may be in NASA Ames or GEOMS HDF format, or both See

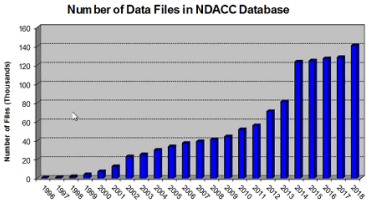
  Agreement with WOUDC for synchronisation of ozone data

  Perspectives

  DHF is moving from NOAA to Nasa Langley

  Data will get a DOI (via NILU/EVDC), a data license, and data policy will be revised correspondingly
  - data licenses envisaged are CC0
    CC-BY-SA (4.0)
    CC-BY-NC-SA

  Move towards FAIR data



- Home / Data / Data Formats

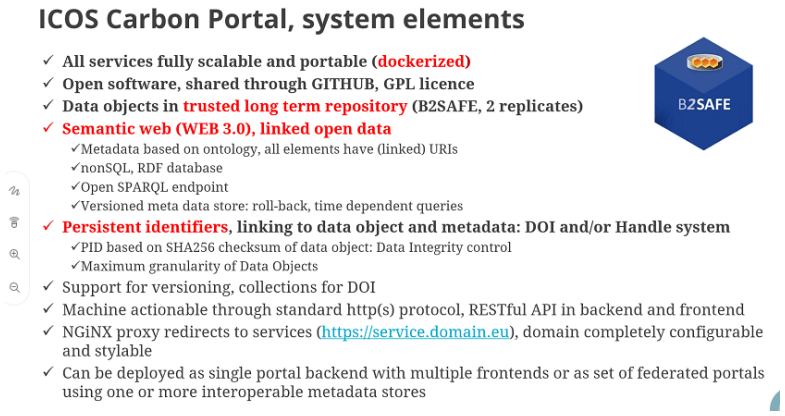
Data Formats
NDACC accepts data in ASCII Ames or GEOMS compliant HDF4 formats depending on instrument type as follows:

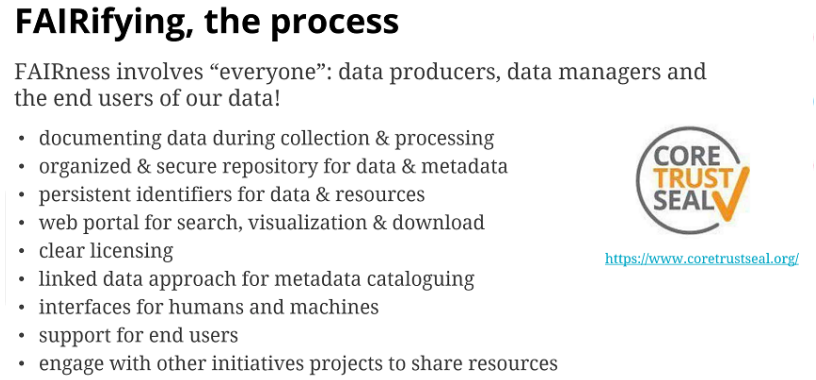
Format Type	Relevant Instruments	Documentation
ASCII Ames	Dobson/Brewer FTIR (total column only) Lidar Microwave Ozonesonde Spectral UV UVVis (total column only)	File Reading Software Format Checking Software Gaines & Hipskind: Format Specifications NDACC Header line NDACC Data Quality Flag NDACC Filenaming NDACC Variable Recommendations Ozonesonde Guidelines
GEOMS/HDF	FTIR Lidar Microwave UVVis DOAS	Generic Earth Observation Metadata Standard (GEOMS) HDF to netCDF Conversion Tools Network of Remote Sensing (NORS) NORS Data User Guide NORS Uncertainty Budgets Additional NORS Documents

- Gao Chen (NASA Langley) taking over the data management for NDACC with a long-term perspective.

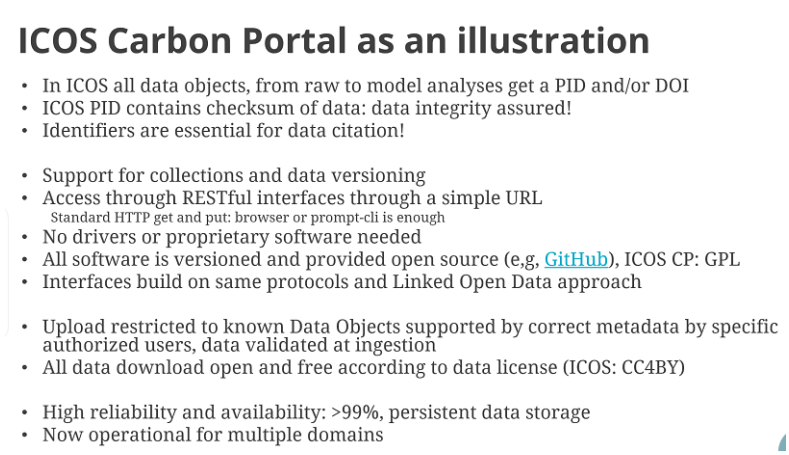
Doc 5.7 - ICOS-CP

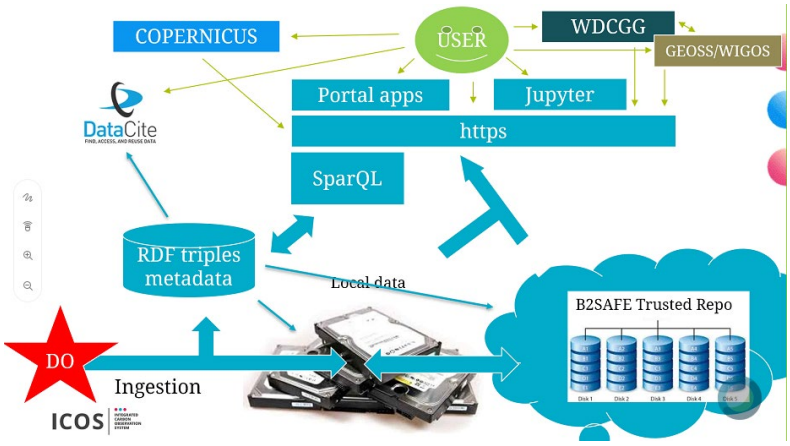
- Dockerized
- RDF DB w/ SPARQL endpoint
- Open-source, GitHub
- Maximum granularity of data objects
- Support versioning

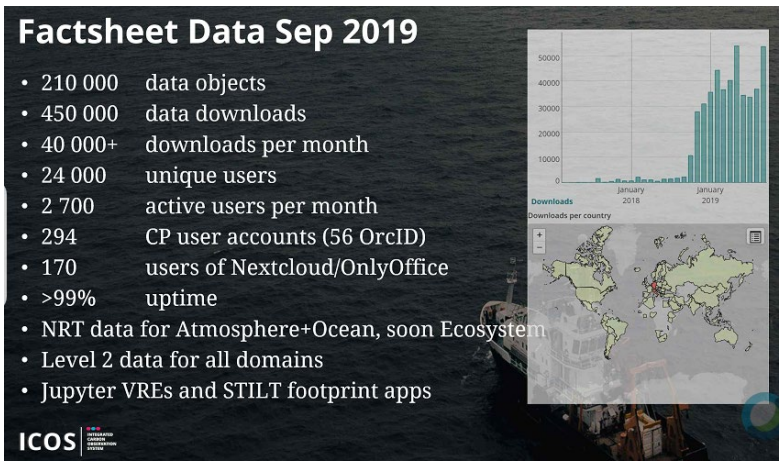


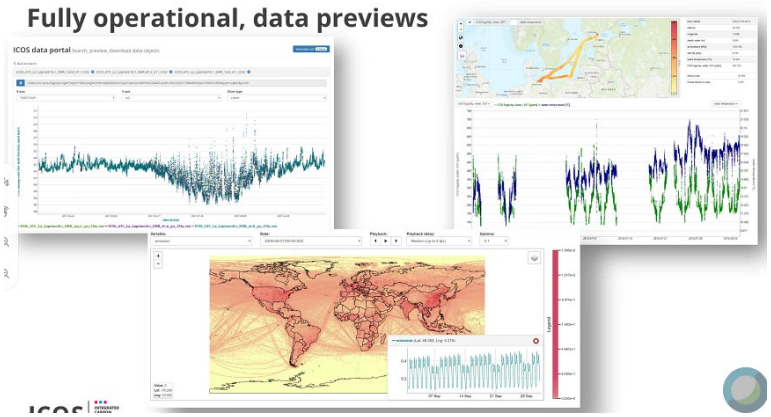


- Life-sciences most advanced in terms of FAIR



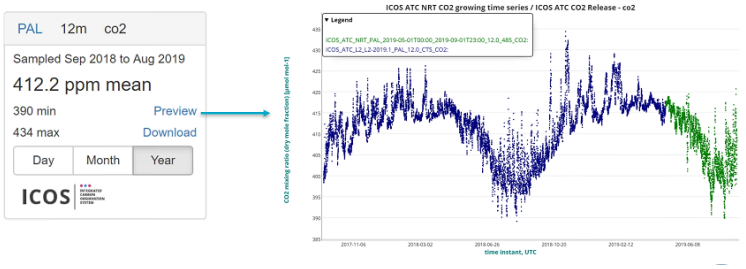
- 
COPERNICUS
USER
WDCGG
GEOSS/WIGOS
Portal apps
Jupyter
DataCite
https
SparQL
RDF triples metadata
Local data
B2SAFE Trusted Repo
DO
Ingestion
ICOS

- 
Factsheet Data Sep 2019
• 210 000 data objects
• 450 000 data downloads
• 40 000+ downloads per month
• 24 000 unique users
• 2 700 active users per month
• 294 CP user accounts (56 OrcID)
• 170 users of Nextcloud/OnlyOffice
• >99% uptime
• NRT data for Atmosphere+Ocean, soon Ecosystem
• Level 2 data for all domains
• Jupyter VREs and STILT footprint apps
ICOS
Downloads
Downloads per country

- 
Fully operational, data previews
ICOS data portal

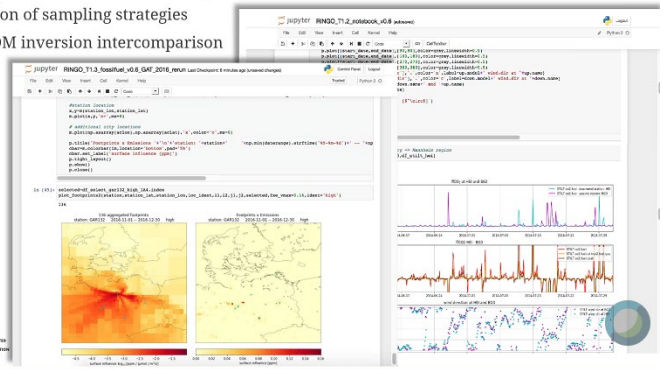
Example dashboard servlet

Station L2+NRT data average for dissemination, e.g. as servlet in news web site:
https://data.icos-cp.eu/dashboard/?stationId=PAL&valueType=co2&height=12



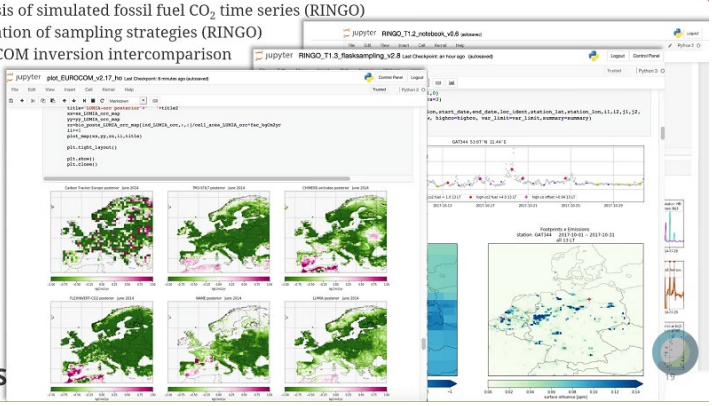
Interactive analysis tools for model results & data

- Analysis of simulated fossil fuel CO_2 time series (RINGO)
- Evaluation of sampling strategies
- EUROCOM inversion intercomparison
- …

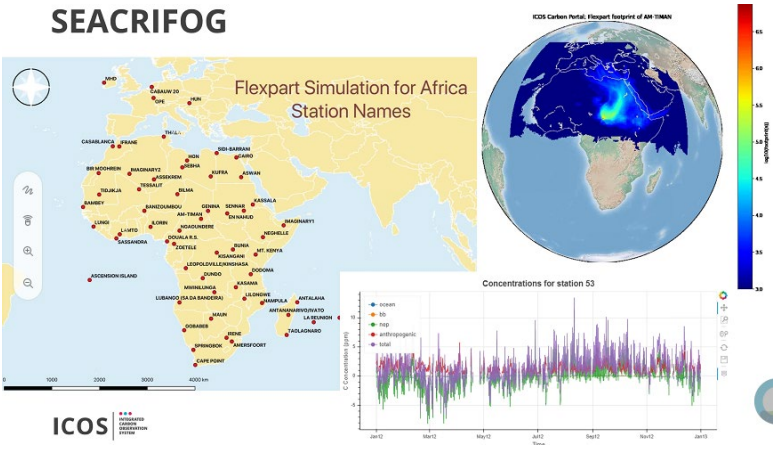


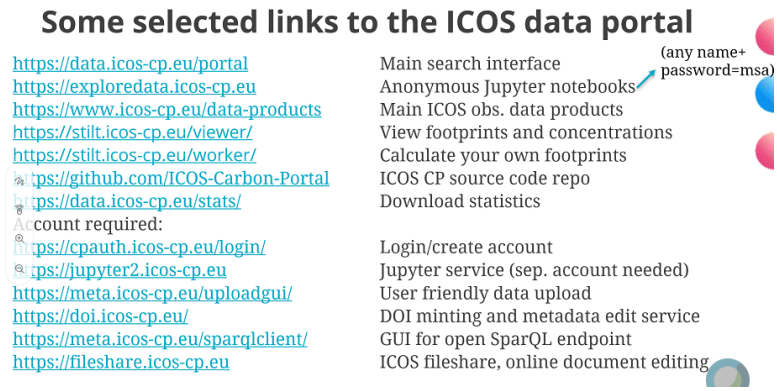
Project-specific Jupyter notebooks

- Analysis of simulated fossil fuel CO_2 time series (RINGO)
- Evaluation of sampling strategies (RINGO)
- EUROCOM inversion intercomparison
- …



SEACRIFOG



- 

ICOS

Used identifiers, licensing, DOIs, Datacite, based on FAIR principles coretrustseal.org, including visualization, various stats, preview, download and advance data products and science applications.

Discussion

- Q: What data formats are accepted, dry mole fraction only, or also others?
- A: ICOS highly standardized
- Q: Are files transformed into a format that is more suited for visualization via web services?
- A: Internal binary format for use in data cubes, most performant; so 2 data.

6. THE BIGGER PICTURE: WMO DATA MANAGEMENT AND HOW GAW CAN/SHOULD CONTRIBUTE

Doc 6.1 - Earth System Prediction: data for services

(Greg Carmichael, Vincent-Henri Peuch, Frank Dentener, Oksana Tarasova)

- GAW monitoring and services (GHG bulletin, etc.), but more and more adding enhanced services through measurement-model integration.

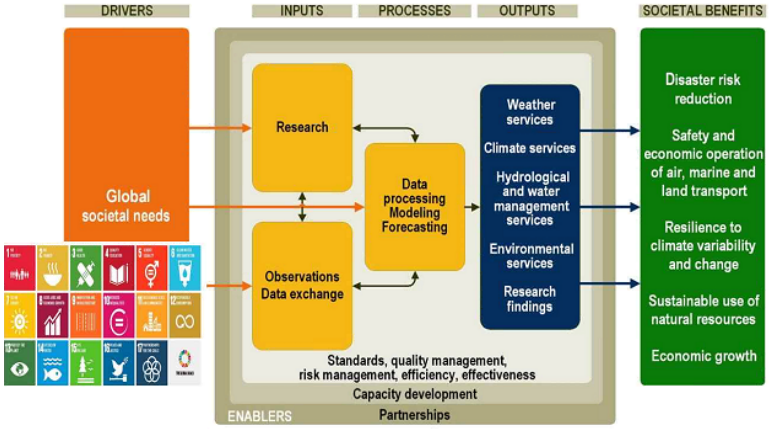
Elements integrated in GAW

- •Observations
- •Quality assurance
- •Data management
- •Modeling and analysis
- •Joint research
- •Capacity building
- •Outreach and communications

WMO OMM

Promote a "value chain" from observations to services

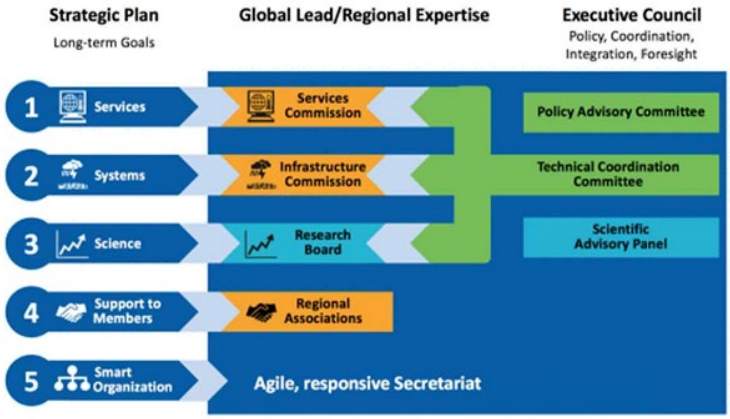
Overarching Objective - Improve Prediction Capabilities via Incorporating/Integrating Composition, Weather and Climate



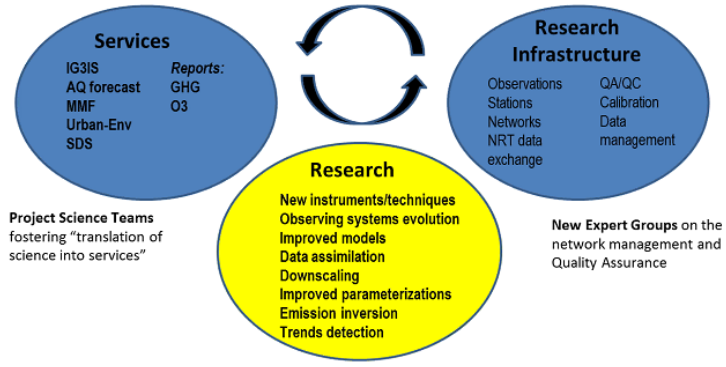
Seamless Prediction *Across all Relevant Temporal and Spatial Scales* (***GDPFS***)

- IG$_3$IS, Measurement-model fusion, health, integrated urban services
- Reform of WMO asks for:
  - Better service delivery in response to societal needs
  - Enhance Earth system observation and prediction
  - Advance targeted research
  - Capacity building
  - Alignment of structure and way of working.

Structure of WMO constituent bodies



GAW Program Elements: Aligning with new WMO structure



Project Science Teams fostering "translation of science into services"

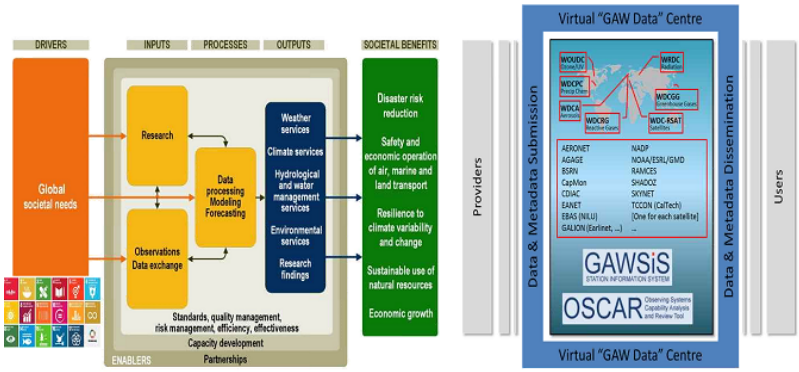
New Expert Groups on the network management and Quality Assurance

Scientific Advisory Groups more focusing on the scientific aspects

WMO OMM

Overarching Objective - Improve Prediction Capabilities via Incorporating/Integrating Composition, Weather and Climate

Earth Systems Modeling Approach



Seamless Prediction *Across all Relevant Temporal and Spatial Scales* (***GDPFS***)

Promote value chain from observation to service. Improve prediction capabilities via incorporating composition, weather and climate.

Cross-cutting thematic projects to develop new services:
Integrated Global Greenhouse Gas Information System - IG^{3}IS (support of climate services); Measurement-model fusion for total deposition (support of the ecosystem assessment and food security); MAP-AQ (support of the health sector); Contribution to the integrated urban service.
Reform goals are better service to societal needs, enhanced Earth system observations and prediction, advance targeted research, capacity building, alignment of the structure to increase efficiency and effectiveness.

- ET-WDC meeting an important input to SSC meeting in November
- Use of data to further develop models

Discussion

- Ted Haberman: What identifiers are we going to be using to track 'objects' (data granules, stations, etc.) in this data flow?
- New data flows due to low-cost sensors.

Doc 6.2 - Introduction to WIS and WIGOS: organize data for services

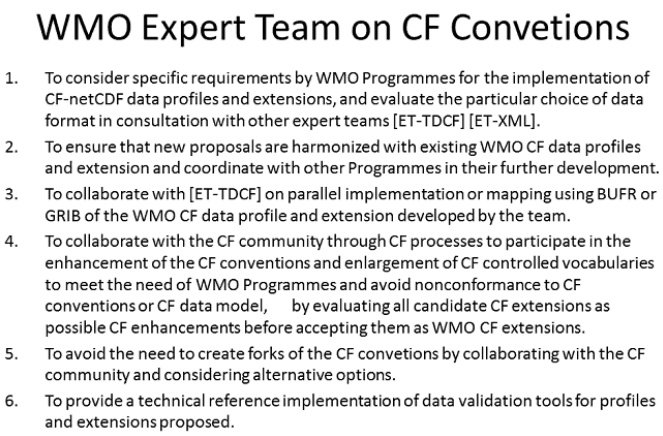
- WWW = GTS + GOS + GDPFS
- Changing to WIS + WIGOS + Seamless GDPFS
- WIS moving towards WIS 2.0
- WIGOS tools
  - OSCAR, open access and controlled for changes
  - WDQMS
- WIGOS components go beyond GOS, including GAW, WHOS, GCW, GOOS, GCOS, GTS
- WIS is an operational system
- WIS 2.0 principles
  - Adopt web technologies and leverage best practices and open standards
- https://www.ietf.org/, W3C, OGC, ISO
  - Use URLs to identify all objects in WIS
  - Use public networks, e.g. Internet
  - Provide web services and/or API
  - NC and DCPC should provide data reduction services to make big datasets more easily accessible
  - Use open standard messaging protocols based on publish-subscribe message patterns, AMQP, MQTT
  - Real-time distribution to cache/store messages for 24-hours
  - Adopt direct data-exchange between provider (aggregator) and consumer
  - Phase out routing tables

- Catalogue should describe both data and services
- Catalogue should be searchable on public search engines.

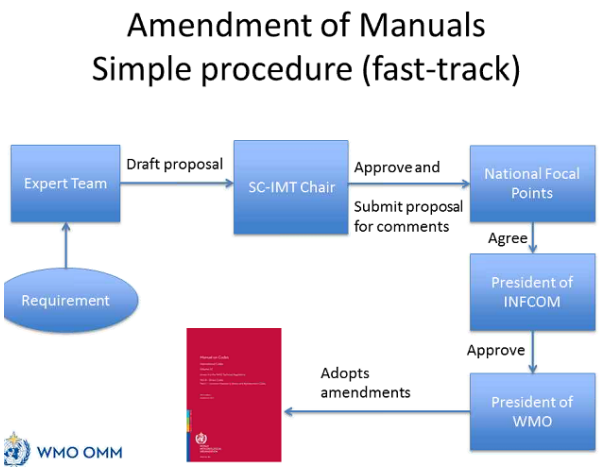
Discussion

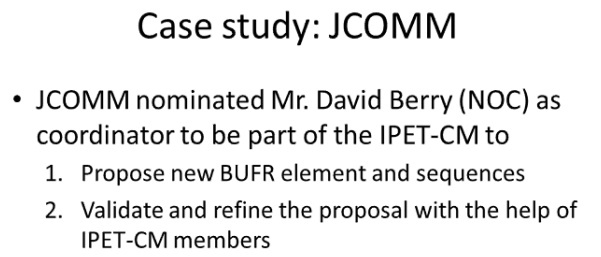
- Judd Welton: should not use Google to search for scientific data.
- Enrico Fucile: search takes place in the GISC portal, but the right granularity is a challenge.
- Ted Haberman: questions about Google are interesting, Google dataset search is the beginning of a more structured search, we need interoperability, understanding and trust; WIGOS standard has some points in that direction, but it is very old! We require more cutting-edge.
  - WIGOS is nice, it is built on ISO 19156 but uses 19115 MD, which allows only one name, more recent ISO standards allow more names; result quality part points to 19157 which is fairly old.
  - Science community has many requirements and capabilities to driving this forward.
- Uptake and penetration in the global community is a huge challenge, there is no silver bullet.

Doc 6.3 - Data and Metadata exchange in WMO – need for long-term archives, rapid delivery

- WMO 306
  - Vol I.1 (frozen, obsolete)
  - Vol I.2 (easy to maintain)
- BUFR, GRIB
- Table driven
  - Vol I.3 (sustainability problems)
- ML model driven, ISO, OGC
- Mainly IWXXM, difficult to maintain
- WMO CF-netCDF workshop 19-20 Sep 2019
  - Conclusion: need for an ET on CF Conventions
  - Representation from GCW, Satellite community, radar.
  - 

- Ted Haberman: need to include the ACDD conventions

- 

- 

Discussion

- JW: Many contributors don't know what to do with this, but for GALION, trying to link in. Data format is netCDF, for metadata need some stability.
- GC: some frustration with CF because it doesn't contain much of what he needs; there is a need for a coordinating body.

Doc 6.4 - Archiving and disseminating model data and products

Vincent-Henri Peuch - AQ forecasting

- CAMS
  - Operated at ECMWF delivering a range of services from AQ to atmospheric composition
  - Prototype for where many of the continents are going
  - Relying heavily on NWP infrastructure, synergies with KNMI, MF
  - All operations are integrated in the traditional data flow of ECMWF NWP
  - Some products are disseminating some of the products through the WIS mechanism
  - New services making use of cloud computing, for example, climate data store.
- SAG-APP
  - Help operationalize data flow; what is the best mechanism for future services?
- Avoid single point of failure, coordinated de-centralized approach: methodologies provided centrally, actual data flow de-centralized and from the providers directly.

- Mentions finalizing the contract with MeteoSwiss for GAW4CAMS soon
- Global AQ forecasting, including WHO and UNEP
    - Define framework for global AQ forecasting
    - Enable operational services based on observations.

Frank Dentener – TAD

- Recognized that services are missing for deposition products
- SAG TAD currently defining the objectives for a product on TAD
- Much behind in comparison with AQ forecasting products
- Make better use of data, maps and numerical data.

7. WORKSHOP "CONNECTING THE DOTS"

- Review objectives of GAW Implementation Plan
- Identify common objectives
- Assess readiness to support operational and research needs
- Establish individual action items for 2020-2021.

The following discussion themes for the workshop were considered along with reported and expressed concerns, issued, need of guidance, direction or consensus.

Questions for discussion and brainstorming

- Impact of WMO reform on WMO perspective
- Cloud's role in service delivery
- Earth System approach
- Discoverability
- Targeted service and NRT delivery

KPI what to use

- Indicator for overall health of GAW or WDC?
- Progress to connecting with WIGOS
- Interoperability – how to evaluate progress

Who and how to set priority and what has to be done?

- Where the direction will be coming from? WMO DC tasks and directions are coming from SAGs, OPAG EPAC SSC, Technical Commission of Atmospheric Sciences (CAS) such as linking NDACC ozone data with WOUDC.

The participants agreed on the following workshop discussion topics due to commonality and the need of being addressed in order to achieve progress in priority areas.

Network integration

- How can Aeronet and GALION be somehow integrated?

- KPIs, which ones should be reported, which ones should be provided centrally from OSCAR, which ones by individual data archives?

Data interoperability

- Common data format? How to improve the notion of a 'uniform user experience'?
- SHADOZ vs WOUDC vs NDACC vs GAWSIS: metadata and data management
- Registration of data with WIS.

Metadata

- Is something missing in the WMDR? How can it be improved? Who is willing to contribute?
- WIGOS metadata exchange with GAWSIS
- UMM-C vs WMDR
- DOIs vs WIGOS Identifiers: What are the requirements? Identifiers.org

Publish-subscribe mechanisms, AMQP, MQTT
New data flows due to low-cost sensors.

Concerns raised during discussions or where clarity needed

GAW Station Information System (GAWSIS) serves as a database of the GAW stations and other stations contributing to GAW, and provides links to the data in WDC and contributing data centres. Its role is to serve data discovery and provide information on GAW stations and measurements of contributing networks. WMO Observing System Capability Analysis and Review (OSCAR) system is tool to integrate GAWSIS, WDCs, data centres of contributing networks and other data centres in WIGOS and facilitate data exchange across WMO programmes and disciplines.

WIGOS is not intended for publishing data, WIS is. - Shall a URL be included to find the data or not? GAWSIS-OSCAR/Surface implements the WMDS which supports the specification of a URL for every deployment. WIGOS has approved list of variables with definitions under review now (essential climate variables which is related to development of Climate observing System and services are a part). Other missing items could be added to this list following proper protocol. OSCAR has fixed vocabularies and the definitions are important to ensure what needed is captured under vocabularies

Discussion has already taken place in task team (TT-WMD) if licenses should be included in WIGOS.

AI2019-5: Verify existing recommendation on licenses from TT-WMD, inform ET-WDC (Jörg Klausen, Nov 2019) and suggest additional codelist entries (all WDCs, Dec 2019).

How to connect to GAWSIS, how to interconnect for different maturity networks and centres?

Work is in progress on linking data centres/contributing networks NDACC, Eubrewnet and WOUDC could use this experience and guide future/potential DCIO and basic metadata exchange, discoverability. Will help having developed architecture, framework, API. WIS- uses URL for resources, based on open standards and public telecoms, web based with API interface, to have direct data provider-consumer exchange, no tables no bulletins, will have catalogues of metadata for data and services.

WMO Manual of codes exists and updated. Uses CF convention – netCDF. Also used by satellite community. Process in place for Amendment of Manuals. More WMO references at the end of the document.

GAW is research programme, data flow specific and large range of info. Not significant part is in NRT. Some participate, some not. Ongoing work with Technical Commission on evolution of observing systems, looking at chemical data consistency, data uncertainty.

Ongoing is the work on RRR related to atmospheric composition. Each application has different requirements making it challenging. Integration of new, more data coming from low-cost sensors and how to use it is a new challenge.

SAG App chair looks at how atmospheric composition is used for services.

Vocabularies are currently reviewed by a joint GAW/GCOS task team (lead: Richard Eckman), possible avenue is mapping vocabularies.

It is not clear who should provide guidance on vertical distribution of properties – which SAG? Aerosols, O3, RG - each in their field of expertize? Input is gathered through communities and experts and used for standardization. Sometimes input is given by invited experts but not included and might be missed. If something is missed, review and feedback process on publications and following protocols could be used in order to add pieces that have been missed (github.com/wmo-im/wmds).

WIGOS metadata standard should be used to talk to WMO and hand shake.

Important is the creation of a Framework where people can address their needs. National Focal Points exists, however, their involvement with GAW is sometimes poor.

If standard changes too much, it is not good, but we also need mechanism/flexibility, duration of standard when to change the standard and how to change in order to evolve and who will propose, make decisions.

Key for metadata is WIGOS compatibility: need to have agreement on names and metadata. A lot of contributing networks are not (yet) compliant with WMDS, how to integrate?

Mapping metadata is a huge effort depending on different users and different requirements but conversion scripts could make conversions on units.

Contributing networks are developed to serve clear user requirements from the start. Those requirements might be different from WDC in a significant way.
Consider NASA approach - UMM extensible metadata model for mapping between CMR-supported metadata standards which are mapped centrally to UMM. Can it be used to link with GAWSIS/OSCAR and how?

How people/ centres can change and apply one agreed metadata and data format which can be processed by WDC. Add data structures for access and all can benefit from the new dataset that is in the system?

LIDAR network tried to use CF, ICAP model community does not hold for CF metadata NASA uses CF, frequent change of CF versions is a problem.

Metadata description of uncertainties need to be defined. Metadata sometimes are not part of the data or reside at different levels (RT, NRT, science). Mechanism for exchange are different for different streams, different needs. Metadata standards should not be forgotten or missed when data are exchanged.

Archiving and history
Versions of data and metadata and related documentation need to be preserved for the future. History of metadata is important and needs to be included in long-term archiving functions. OSCAR has been built on this approach – history of data and time stamp are exported to user. Specific events on site can be documented and added to the history of the station in OSCAR. Updating data requires updates in OSCAR and this is the responsibility of the data provider. Data and metadata should go together somehow but WIGOS is challenged to keep history development.

8. CONTROLLED VOCABULARIES

- *Doc 8.1 Variables according to WMDS*
- *Doc 8.2 Variables according to OSCAR/Space, Requirements*
- *Doc 8.3 Methods according to WMDS*
- *Doc 8.4 ESDS Atmospheric Composition Variable Standard Names Working Group Report*

9. WAY FORWARD IN ATMOSPHERIC CHEMICAL COMPOSITION DATA MANAGEMENT IN SUPPORT OF RESEARCH AND SERVICES

Doc 9.1 Draft based on

- NRT document
- GAW reports on data management (e.g. Summary of Kloten meeting, 2015),
- Big data essay (Hov et al., 2016; submitted to Executive Council - Sixty-Eighth Session (EC-68))
- GAW Implementation Plan itself

DCIO/GEOMS (20') – (*Kjetil Torseth, Tom Kralidis)*

EVDC a contract with ESA, operated at NILU

Not an open data archive

Data stored according to GEOMS metadata

EVDC – ESA Atmospheric Validation Data Centre http://www.esa.int

metadata harvesting and display of information:

Already implemented:

- EARLINET – harvest and display of metadata, non-GEOMS

- CloudNet – harvest and display of metadata, non-GEOMS

In progress:

- EUMETNET – harvest of metadata, non-GEOMS

- WOUDC – harvest and display of data, non-GEOMS

Currently direct mirror of data:

AVDC (hereunder SHADOZ, WOUDC) NDACC – harvest and display of data, GEOMS

Campaign data from e.g. Sentinel-5P, GEOMS

More information at https://evdc.esa.int/documentation/oai-pmh/

EVDC – ESA Atmospheric Validation Data Centre

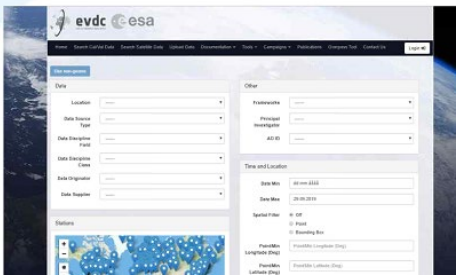
Harvesting through API:

https://dcio-ng.evdc.nilu.no/oaicat/ (non-GEOMS)

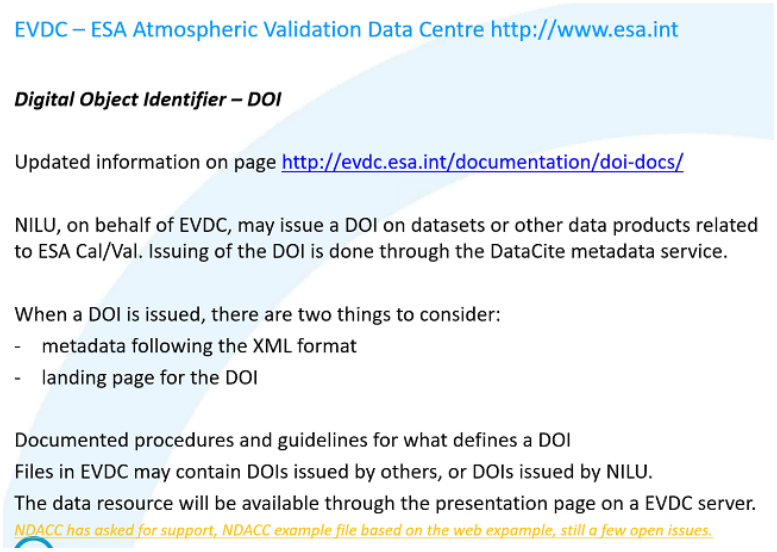
https://dcio.evdc.nilu.no/oaicat/ (GEOMS)

Cal/Val search on web:

merged system for GEOMS and GEOMS data in preparation



Similar metadata portal as ACTRIS


EVDC – ESA Atmospheric Validation Data Centre http://www.esa.int

Digital Object Identifier – DOI

Updated information on page http://evdc.esa.int/documentation/doi-docs/

NILU, on behalf of EVDC, may issue a DOI on datasets or other data products related to ESA Cal/Val. Issuing of the DOI is done through the DataCite metadata service.

When a DOI is issued, there are two things to consider:
- metadata following the XML format
- landing page for the DOI

Documented procedures and guidelines for what defines a DOI
Files in EVDC may contain DOIs issued by others, or DOIs issued by NILU.
The data resource will be available through the presentation page on a EVDC server.
NDACC has asked for support, NDACC example file based on the web expample, still a few open issues.

Discussion

- SAG O_3 had asked WOUDC to inject NDACC, but Tom Kralidis was not in favour of creating duplicate datasets
- WOUDC looking into using OAI for exchange
- GEOMS is being used, e.g. by TOLNet
- **AI2019-6**: Gao Chen and Tom Kralidis will follow-up on DCIO/GEOMS exchange mechanisms and inform the next ET-WDC telco
- Input from Tom Kralidis (via E-Mail)
  https://evdc.esa.int/documentation/oai-pmh/
  https://evdc.esa.int/documentation/geoms/

DOI (30')

Data providers are not in agreement at what granularity DOI's should be minted.
Minting a DOI requires being able to produce the data
https://blog.datacite.org/tracking-the-growth-of-the-pid-graph/

Alex Vermeulen: DOI is a specific use-case for handles
Who makes the decisions on the open questions regarding DOIs?
Alex designated and accepted to inform and involve ET-WDC in work of ENVRI-Fair
https://www.rd-alliance.org/groups/persistent-identification-instruments-wg

Input from Martin Schultz (via e-mail)
How about using handle.net PIDs instead of DOI's for this purpose?
See http://www.handle.net/ and https://eudat.eu/services/userdoc/pids-in-eudat.

Is there consensus yet on how detailed the community list should be? I am a bit worried that the effort may lead to a registry which is either too broad to be useful in measurement-centric applications, where detail matters, or too

specific in more general applications, where it is more important to capture the type of an instrument, and perhaps the manufacturer, but not all details.

Is the group aware of the US EPA collections of "allowed" instruments? See https://www.epa.gov/measurements-modeling/collection-methods, and as one example (criteria air pollutants) https://www.epa.gov/sites/production/files/2019-08/documents/designated_reference_and-equivalent_methods.pdf
Input from Ted Habermann (via E-Mail)

NRT

WDCGG

- No NRT capabilities

WOUDC

- O_3 sondes from Canadian stations, CREX, send to GTS
- Level 0 (raw data), password protected
- UV made available
- Limited to MSC
- No requirements articulated by SAG
- Discussion:
  - CREX format not recommended and not supported by any of the NWPs.

WDCA, WDCRG

- Fully developed system for WDCA,
- EBAS - GTS connection in place (but not operationalized)
- O_3 sonde service for ECMWF.

MPLNet

- In the past, NRT delivery all test cases, PBL height, aerosol profile information for ECMWF
- South Pole uses lidar as a ceilometer for in and out flights
- EPA Taiwan accesses data in NRT, use unclear
- Version 3
  - delivers within 1 hour, through https, open, unrestricted
  - Quality-screened within 1 day for the public, limit is the assimilated meteorology
  - Custom packaging on a password protected access portal for forecast centers
- ICAP a primary user.

NDACC has some NRT capability under CAMS contract.

Discussion on 'NRT document'

- Document approved by SSC OPAG EPAC based on recommendation by SAG APP. The basic approach of enabling stations to deliver NRT data directly to the GTS is non-negotiable.
- NILU is in opposition to establishing new NRT-services not complying with the already existing services operated by WDCA and WDCRG
- Currently, there is no [GAW4CAMS] contract, and given the delays in the past, it is completely uncertain if/when a contract will be agreed on time to be able to develop and implement something useful.

Time allocated for the workshop and document draft was used to further discussions on agreed topics which were found important to advance understanding of the current stage, capabilities, possible avenues and consensus on the path forward for WMO data management related to atmospheric composition to support research and services. Participants agreed to continue their research, comment and make recommendations on those topics, report on progress made under the GAW Implementation Plan and assist in drafting the document as intended for the last day of the meeting.

Can GALION be discoverable through web entrance from the GAW website, with its two geographically specialized primary nodes of identical metadata and discovery information including a common site database. This site database will be used to provide routine updates to GAWSIS working and learning from experience of other NASA projects and centers. Need to find what is missing, what is needed and to be added in metadata by looking for example at the project of linking NDACC and WOUDC. What files formats are/can be supported and how to improve discoverability. Need to have WIGOS as unifier and simplifier for user service and delivery. Connection to OSCAR is not a problem and MeteoSwiss provides assistance and guidance (cf. **AI2019-4**).

Linking data center, networks and WIGOS:

AI2019-7: Review of project/programme level of variables' names consistency (lead? , deadline?)

WMDR XML Schema for aircraft data looks good and needs to find where there is a need for a change and what. Continue work on this platform will increase the presentation of poorly represented aircraft data in GAWSIS, WIGOS. Need to move closer to mapping with WIGOS.

AI2019-8: How to present metadata/ changing latitude, longitude, altitude for non-stationary platform for GCOS, WIGOS. NCAR experience to be looked at? (Gao Chen, Enrico Fucile?, deadline?)

How to represent uncertainties – clear guideline needed: absolute, %, formula, value, what to use for a day? Can a citation from a publication on the topic in metadata be

used? Requirements are not clear. Metadata description of uncertainties need to be defined. (TT-WMD, 2020)

AI2019-9: TT-WMD and TT-OD, together with OSCAR/Surface team at MeteoSwiss/WMO to draft a document on representation of uncertainties in WIGOS metadata (TT-WMD, 2020).

DOIs should they be site/station specific, instrument specific, campaign/data collection specific, dataset specific, only for finalized data or growing collections. Data providers are not under the same understanding on when, where and how DOI is to be assigned. DOIs provide information on value, reference and use of data for different projects and applications but there is also sensitivity for highly needed data for narrow limited need and funding those. NDACC receives DOIs for data through NILU, WDCGG assigns DOI to each observation data. Granularity is not clear from provider nor from the user perspective. Request from O_3 UV SAG is for station DOIs. DOIs supports different versions. Another aspect comes from archiving functions and preserving all information. DOIs for final data products is also used/considered. Collected information and requirements from providers and users (not only limited to science use, regional, global …) can be provided on WIS DOI working group as input and considered as part of WIS evolution.

AI2019-10: Alex Vermeulen to inform ET on what has been considered and how with respect to DOIs (Alex Vermeulen, next telco of ET-WDC). Feed into WIS DOI WG (Enrico Fucile, 2020).

WIGOS ID system is doing exactly the same thing as DOI and has room for adding instrument information. One way to go could be: we know and use a unified structure which includes versions and develop handlers to link WIGOS ID with DOIs for data.

Measurements/instruments intercomparability is another issue. How to do those for different types of instrumentation, methods, etc. (depends on WMO Reform and future QC/QA SAG).

NRT no capability in WDCGG. GALION had NRT data processing and submission for forecasting different levels at different timescales from 1.5 hours for level 1.5 and longer. Auxillary data are used to finalize the data for LIDARs. NDACC has a contract for NRT data. O3 sondes of SHADOZ: NRT not possible due to the nature of data collection - research data archive, not an operational character. WOUDC: Some O_3 sondes level 0 had been paired with GTS data submissions for Total Ozone and UV forecasting. UV is NRT. NILU: some opportunities exists and delivers to ECMWF. There is no global NRT TOC streamlined contribution. Those are preliminary data based on NRT character and need for service. Need clear understanding of what is needed.

AI2019-11: Enrico Fucile to connect Tom Kralidis with people for NRT (Enrico Fucile, Nov 2019).

MPLNet/GALION to produce a few examples for WMD records for Lidar (Judd Welton, timeline to be announced; cf. **AI2019-4**).

AI2019-12: Collaborate with Gao Chen (Tallnet) to produce WMDR records for these networks for the GALION landing page and for GAWSIS/OSCAR (Judd Weldon, Gao Chen).

AI2019-13: Translate metadata formats, MetaDAL project, look into UMM-C model (Makhan Virdi, Tom Kralidis).

AI2019-14: Review WMD code tables (SHADOZ), deliver data to WOUDC, WOUDC to generate the WMDR records to reflect links to WOUDC as well as SHADOZ; same is true for NDACC (Debra Kollonige/Ryan Stauffer, Tom Kralidis, 2020).

AI2019-15: Find consensus on variable names, on the programme level, work with Makhan Virdi for representation of aircraft observations in WMDR (Judd Weldon, Marham Virdi)

Representation of uncertainty, particularly for aircraft data (ICARTT allows 1-2 lines to document this).

Enrico Fucile: Some duplication of efforts perceived; landscape is very fragmented; WIS and WIGOS overlap, which may or may not be useful. Metadata change with time, and that needs to be reflected.

Vocabularies

- Aerosol variables in GCOS/GAW
- Code list 7-06 on data levels (refer to EBAS, NASA)

AI2019-16 (all, November 19): Report progress on GAW Implementation Plan items (10 November) [mostly completed]

Actions in the GAW Implementation Plan	*Specific activities, contributors, timelines?*
A-DM-1. Establish and implement a federated data management infrastructure including GAW Data Centres, data centres of contributing networks, and GAWSIS that enables interoperable data discovery and access mechanisms.	Next ET-WDC physical meeting to derive work plan (late summer mtg., Q1/2018).
A-DM-2. Improve open access to data and comprehensive metadata including calibration histories of ground-based, aircraft and satellite observations for the primary GAW variables.	Review/confirm data policy. Private sector requirements may be at variance with fair-use policy. Calibration history is embedded in WIGOS metadata standard. Document calibrations (data providers, TT-WMD)
A-DM-3. Harmonize GAW data management activities with the WIGOS framework, in particular with regards to metadata documentation.	WMDS provide framework is implemented in GAWSIS. Interaction between data providers and archives needed (see also A-DM-1 activities).
A-DM-4. Develop and promote support of data archiving and analysis centres that address the needs of applications and service delivery.	Strengthen and expand capacity of existing infrastructures to address these needs. Users and archives need to interact to formulate requirements and specifications. Facilitate remote sessions annually between data archives and user communities (stakeholders) for improved interaction.

Actions in the GAW Implementation Plan	*Specific activities, contributors, timelines?*
A-DM-5. Ensure that data collected and archived by WMO/GAW WDCs and archives of contributing networks are of known quality, adequate for their intended use and documented comprehensively.	Work with WIGOS metadata standard documentation. Establish plausibility checks at data centres. Services like CAMS should feed back to the observation providers. Establish methodology for users to feedback to providers on quality issues. Archives make provision to accept quality flags.
A-DM-6. Promote delivery of those variables pertinent to air quality and forecasting in NRT, using WMO GTS/WIS as it evolves into an open, decentralized and node-oriented structure. Continue to seize opportunities to expand the NRT delivery services for GAW variables.	In Europe, CAMS is supporting and promoting NRT services. CAMS is engaging in contracts with ICOS, ACTRIS, GAW and EMEP. Satellite NRT delivery via direct broadcast could be expanded.
A-DM-7. Develop data submission and data use procedures with the inclusion of uncertainties with the GAW data products, making it possible to select and use data in accordance to the criteria set out by the RRR process.	Implement at GAW data centres where not yet employed. The WIGOS metadata standard supports this activity. (as soon as possible)
A-DM-8. Continue to make best efforts towards program-wide adoption of digital object identifiers (doi) for GAW datasets to facilitate proper recognition of the data contributors in scientific analyses and reports and it will also allow for better monitoring of the actual data use.	DOIs are finding acceptance in the community.

10. SUMMARY OF ACTION ITEMS

AI2019-1: Bring ancillary variable PBL height (and related variables) to the attention of SSC and discuss how they could fit into the GAW data management (Chair, November 2019).

AI2019-2: Consider WMDS codetable 7-06 (Level of data), cf. https://github.com/wmo-im/wmds/blob/master/tables_en/7-06.csv and create issues if necessary. (all, Dec 2019).

AI2019-3: All WDCs to (re-)establish/confirm metadata exchange with GAWSIS-OSCAR/Surface. Consult with MeteoSwiss and https://library.wmo.int/doc_num.php?explnum_id=5844 (Section 3.12) for guidance and present timeline. (all WDCs, 10 Nov 2019).

AI2019-4: Send example of metadata to be encoded in WMDR XML to MeteoSwiss. (Judd Welton, asap). Judd Welton will review information on WIGOS metadata and elements- variables, prepare BUFR metadata transfer to be available with data.

AI2019-5: Verify existing recommendation on licenses from TT-WMD, inform ET-WDC (Jörg Klausen, Nov 2019) and suggest additional codelist entries (all WDCs, Dec 2019).

AI2019-6: Gao Chen and Tom Kralidis will follow-up on DCIO/GEOMS exchange mechanisms and inform the next ET-WDC telco.

AI2019-7: Review of project/programme level of variables' names consistency. (Gao Chen and Tom Kralidis, deadline?)

AI2019-8: How to present metadata/ changing latitude, longitude, altitude for non-stationary platform for GCOS, WIGOS. NCAR experience to be looked at? (Gao Chen, Enrico Fucile?, deadline?)

AI2019-9: TT-WMD and TT-OD, together with OSCAR/Surface team at MeteoSwiss/WMO to draft a document on representation of uncertainties in WIGOS metadata. (TT-WMD, 2020).

AI2019-10: Alex Vermeulen to inform ET on what has been considered and how with respect to DOIs (Alex Vermeulen, next telco of ET-WDC). Feed into WIS DOI WG (Enrico Fucile, 2020).

AI2019-11: Enrico Fucile to connect Tom Kralidis with people for NRT (Enrico Fucile, Nov 2019).

AI2019-12: Collaborate with Gao Chen (Tallnet) to produce WMDR records for these networks for the GALION landing page and for GAWSIS/OSCAR (Judd Weldon, Gao Chen).

AI2019-13: Translate metadata formats, MetaDAL project, look into UMM-C model (Makhan Virdi, Tom Kralidis).

AI2019-14: Review WMD code tables (SHADOZ), deliver data to WOUDC, WOUDC to generate the WMDR records to reflect links to WOUDC as well as SHADOZ; same is true for NDACC (Debra Kollonige/Ryan Stauffer, Tom Kralidis, 2020).

AI2019-15: Find consensus on variable names, on the programme level, work with Makhan Virdi for represenation of aircraft observations in WMDR (Judd Weldon, Marham Virdi).

AI2019-16 (all, November 19): Report progress on GAW Implementation Plan items by (10 November) [set up Google Docs].

---

11. INFORMATION AND RESOURCES

Role of identifiers and what to use for discoverability of data, FAIRfying, from discoverability to share, use, service
https://evdc.esa.int/documentation/oai-pmh

FAIR principle coretrustseal.org
Identifiers.org –Catalogue of Identifiers.

WMO new website:
https://community.wmo.int/activity-areas
https://community.wmo.int/activity-areas/wigos
https://community.wmo.int/activity-areas/wmo-information-system-wis

WIGOS Metadata Standard (WMO-No. 1192).
https://library.wmo.int/doc_num.php?explnum_id=3653
http://www.wmo.int/schemas/wmdr/1.0RC6/documentation/WMDR_ModelAndSchemaGuide.pdf
Manual on Codes (WMO-No. 306), Volume I.3 - WMO Library

https://etrp.wmo.int/course/view.php?id=146
https://codes.wmo.int/wmdr
WMO Technical Regulations; WMO-No. 49, Volume I, Part I – WIGOS and the Manual on WIGOS

Link to GAWSIS and requirements
https://gawsis.meteoswiss.ch/GAWSIS/#/

Persistent identifiers for instruments
Home page of the RDA group working on instrument identifiers.
https://www.rd-alliance.org/groups/persistent-identification-instruments-wg
http://identifiers.org/
http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery_1-3
http://www.geoscienceontology.org/

Using handle.net PIDs instead of DOI's? See http://www.handle.net/ and https://eudat.eu/services/userdoc/pids-in-eudat

UMM
https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr/umm

US EPA collections of "allowed" instruments: https://www.epa.gov/measurements-modeling/collection-methods,

one example (criteria air pollutants) https://www.epa.gov/sites/production/files/2019-08/documents/designated_reference_and-equivalent_methods.pdf

https://www-air.larc.nasa.gov/missions/TOLNet/
https://www.ietf.org/
https://github.com/wmo-im/wmds

Ted Habermann on FAIR metadata on the DataCite blog
https://blog.datacite.org/metadig-recommendations-for-fair-datacite-metadata/

DCIO and GEOMS
https://evdc.esa.int/documentation/oai-pmh/
https://evdc.esa.int/documentation/geoms/

GLOSSARY OF ACRONYMS

BUFR
Binary Universal Form for the Representation of meteorological data, ... 20

CMR
Common Metadata Repository - https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr, ... 44

DC
Data Center, ... 3

DCIO
Data Center Interoperability - https://evdc.esa.int/documentation/oai-pmh/, ... 42

DOI
Digital Object Identifier, ... 30

ICAP
International Cooperative for Aerosol Prediction - http://icap.atmos.und.edu/, ... 44

netCDF
Network Common Data Form - https: //en.wikipedia.org/wiki/NetCDF, ... 43

OPAG EPAC
Open Programme Area Group on Environmental Pollution and Atmospheric Chemistry, ... 3

UMM
Unified Metadata Model - https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr/umm, ... 44

ANNEX

PROVISIONAL AGENDA

Tuesday (09:00 – 17:00)

	Welcome by the Deputy Director for Research and Mission Science in NASA Langley Research Center (15') practical arrangements (10')	Baize Chen Netcheva
1.	Introduction of participants (30') - *Doc 1.1 – List of participants (to be completed during meeting)*	All
2.	Report by chair ET-WDC, objectives of meeting, approval of agenda (15') - *Doc 2.1 – Report by chair* - *Doc 2.2 – Provisional agenda (this document)*	Klausen
3.	Relevant outcomes of WMO Cg-18 and EC-70 (20') - *Doc 3.1 – Outcomes Cg-18 and EC*	Netcheva
	Coffee break (10:30-11:00)	
4.	Reports by GAW WDCs: Achievements, status and plans (20' each)	
	- *Doc 4.1 WDCA*	Tørseth
	- *Doc 4.2 WDCRG*	Tørseth
	- *Doc 4.3 WDCGG*	Kinoshita
	Lunch break (12:30-14:00)	
.	Reports by WDCs (20' each)	
	- *Doc 4.4 WOUDC*	Kralidis
	- *Doc 4.5 WRDC*	Tsvetkov
	- *Doc 4.6 WDCPC*	Lehmann
	- *Doc 4.7 WDC-RSAT*	Meyer-Arnek

Coffee break (15:30-16:00)

5. Contributing programmes, other data archives (120')

- *Doc 5.1 MPLNet* — Welton
- *Doc 5.2 SHADOZ* — Kollonige
- *Doc 5.3 GCW* — Godoy
- *Doc 5.4 (Other) Nasa data centres* — Virdi
- *Doc 5.5 EANET* — Sato
- *Doc 5.6 NDACC* — Chen for De Mazière

Wednesday (09:00 – 18:00)

5. Contributing programmes, other data archives

- *Doc 5.7 ICOS*

6. The bigger picture: WMO data management and how GAW can/should contribute

- *Doc 6.1 Earth System Prediction: data for services* — Hov/Carmichael (remote)
- *Doc 6.2 Introduction to WIS and WIGOS: organize data for services* — Fucile
- *Doc 6.3 Data and Metadata exchange in WMO – need for long-term archives, rapid delivery* — Fucile
- *Doc 6.4 Archiving and disseminating model data and products* — (?CAMS, DLR, Vermeulen, GDPFS, FutureEarth?)

Coffee break (10:30-11:00)

7. Workshop "Connecting the dots" — All

- *Review objectives of GAW Implementation Plan*
- *Identify common objectives*
- *Assess readiness to support operational and research needs*
- *Establish individual action items for 2020-2021*

Lunch break (12:30-14:00)

Workshop wrap-up
- *Present individual action items* All

Coffee break (15:00-15:30)

8. Controlled vocabularies

- *Doc 8.1 Variables according to WMDS* *Klausen*
- *Doc 8.2 Variables according to OSCAR/Space, Requirements* *Klausen*
- *Doc 8.3 Methods according to WMDS* *Klausen*
- *Doc 8.4 ESDS Atmospheric Composition Variable Standard Names Working Group Report* Chen

9. Way forward in Atmospheric Chemical Composition Data Management in support of research and services

- *Doc 9.1 Draft based on*
  - NRT document
  - GAW reports on data management (e.g. Summary of Kloten meeting, 2015)
  - Big data essay (Hov et al., 2016; submitted to Executive Council - Sixty-Eighth Session (EC-68))
  - GAW Implementation Plan itself

_____