

Minutes from the telecom held on 11 March 2020 by the Expert Team on World Data Centres (ET-WDC)

v0.1 (Draft)

Agenda

1. Welcome and Approval of minutes from the meeting in Hampton (Jörg, 10')
2. Outcome of Data policy meeting at WMO, 11-13 February (Jörg, 10')
3. Use of DOIs/PIDs in ICOS and applicability to WDCs (Alex Vermeulen, 60')
4. Review of action items with upcoming/past deadlines (Stoyka, 15')
5. AOB, Tour de table (All, 20') for next meeting
6. Next meeting (Jörg, 5').

Participating:

- Jörg Klausen
- Alex Vermeulen
- Markus Fiebig
- Tom Kralidis
- Atsua Kinoshita
- Kjetil Torseth
- Debra Kollonige
- Nate James
- Makhan Virdi
- Judd Welton
- Enrico Fucile
- Oystein Godoy
- Christopher Lehmann?
- Stoyka Natcheva (Rapporteur)
- Drasko Vasiljevic

Excused

- Vincent-Henri Peuch

Notes

1. Welcome and Approval of minutes from the meeting in Hampton (Jörg, 10')

WMO hired a consultant, Drasko Vasiljevic, former employee of the European Centre for Medium-Range Weather Forecasts, to work with WDC and contributing networks to advance harmonization of metadata, discoverability and access to data on atmospheric composition. He will be contacting and working with different members of the group.

Please respond to his requests for information and engage in discussions to make a good progress.

Minutes of Hampton meeting are approved, thanks to Stoyka for hard work.

2. Outcome of Data policy meeting at WMO, 11-13 February (Jörg, 10')

JK informed on presentation [0] at WMO ad-hoc meeting on data issues and policies 11-13 February (<https://community.wmo.int/meetings/ad-hoc-meeting-data-issues-and-policies>). JK represented the atmospheric composition data issues and policies at meeting session dedicated to "Status and adequacy of data exchange and data policy for other application areas". Meeting aimed to establish the status of the data policies and to provide preliminary advice to Policy Advisory Committee (PAC) for a progress report to be presented at seventy-second session of Executive Council (EC) for a review of the data policies.

A sub-group on Data issues and Policy will be established during the join session at EC meeting to look at how data policies. JK suggested at the meeting that those should be adjusted if necessary and be inclusive not only of NHMS but other stakeholders in terms of policy. Licensing of data was also included in the discussions.

Q: Who will be part of this group. E Fucile and 2 other persons from WMO (Dimitar Ivanov, public and private sector engagement) Secretariat. Members will be approved at meetings of EC and Congress. Enrico will be able to provide more information.

A: ET-WDC will need to agree on how to feed in this process.

3. Use of DOIs/PIDs in ICOS and applicability to WDCs (Alex Vermeulen, 60')

AV presented "PIDs for WMO" [1]: Why, who needs them? Long lasting identifiers to digital resources with link which is maintained. Different approaches had been used to ID data sets and to give visibility to data originators, provenance, traceability with their specific advantages and disadvantages; describing teams, organizations or data; automated or manual, operate at different stages of Data Management, versions, releases, workflow; create citations, links, services. Features and strengths of Handle system, Datacite, Collections including collection of collections. Examples and details demonstrated. DOIP overview presented.

Q: Who is behind DOIs, PIDs? Is overarching guidance existing and can we rely on this? What is expected to be operational in long-term? Adapting standards is tedious work with little influence on the market.

A: Options that had been explored in the past by communities include European Open clouds data, Open source and what we could learn from those.

Q: Sounds like there is limited synergy with Open Data (and PIDs), it will be interesting to see how they interact.

Using POIs prevents duplication of data. They reside in their depositories and make federated data. They still have their own interface but use data from other depositories while each repository has its own control over license of data and downloads.

AV: We have different players for one data centre /data set with different requirements and systems. Those might have different solutions using one or the other or

combinations of PIDS/DOIs and using links for applications to access the data when it is not the primary repository.

JK: Data duplication is a fact and theoretically is not needed. How best to reference this original source? What Alex described looks like good solution. Others might also exist. There is no Match to Match and reformatting in this mechanism. Data centers require submission of data which might have different format of data and metadata and create interfacing issues. In some cases if match to match or mapping metadata is relatively easy to do it might be the way out and such interface could be created and used. In systems do not have match interfaces. When such is built it involves proof and time.

KT/MF: In the ACTRIS project, NILU presents the concept of federated collection with DOIs which by design should be flexible.

Adopting data FAIR principals by GAW might be long-term process (5-10 years) but could be achieved step-by-step with short and long term goals. It does not deal with duplication but if access is enough you do not need to duplicate and duplication could be avoided. It might not address everything but solution might be found later. Specific applications might need to do manipulations on the data or combine them with other data and for that reason create a copy (duplicate) and this will need to be cleared. Metadata contain information on different versions.

Duplications could come from requirements of different character: Governments require data to be stored on national archives, same data might need to be submitted to ICOS and to WDC under contributing agreement – 3 places or more. Solution might be to replicate the original version or create a copy. Duplication could be noted in metadata or look at integrity suffix.

Everything could be locally done or centralized or in between. Both principles are good some issues could exist and be discussed – like address should point to the authoritative source or integrate the address in metadata.

The goal of the group is not to look at how the things are organized today as this is changing fast and WIS system is moving towards 2.0 version with fully distributed system and interoperability of the data.

NJ: As we archive products, it is important to handle persistency carefully in long term and properly maintain DOIs which are critical for persistency. If duplication exists it should be known where data are preserved for long-term and archive maintains associated DOIs. Records and DOIs should be kept at the same place within repositories and archives with appropriate maintenance. Our approach should meet requirements for National Records but also look outside of them. National, World/Global requirements are overlapping and we/WMO centers are not authoritative source. How can we simplify this and make progress? Global DC and developed countries develop their own things but leave others or keep what we have. Balance is needed including interfaces and opportunity to grow and develop. We need to find stable system. WDC are not flexible now. People do measurements, process data and use data for different things. We are already behind moving forward or not.

EF: What worked well in the weather complex and could be used as an example is that all data requirements were driven to serve the Numerical Weather Prediction. Having clear and strong driver will put things in focus. For atmospheric composition there is a

lack of use cases, requirements. Using FARE principals we need to define what we measure and which systems need it then the rest (requirements) will come from processing centers. We need to keep in mind the global picture, act quickly as move is towards big data processing. Most important is to find a/the driver.

AV: when it comes to atmospheric composition (and ICOS in particular) we have examples and many cases to show but we also have different users and applications with different requirements, different cycles such as agriculture and assessment and difficulty comes when we try to meet them all. Some of those are very specific like using/requiring data on profiles which are not even part of GAW. Forecast/assimilation/Trends analyses – summarize few of those ?

OG: We need to know the ecosystem of the data centres and how we could organize ourselves. Where to go? Why on ICOS and not in the weather archives? Try to follow WMO specifications. WIS is open to different formats and moves to be more inclusive. FAIR might not be enough and more is needed because we plan to adopt the matrix and include open data format and metadata. Those need effort to create if objective does not exists and difficult to drive the process.

JK: Propose for the next meeting to develop action items related to the presentation and examples, or invite other experts to give a talk or clarification and try to agree. If we do not know what is driving and where we go it is difficult to make progress in harmonious way and it is a costly and complex process. We, as WMO team, need to agree where to put the emphasis, how to develop WDCs in future and propose action items for our path forward.

SN will update ppt on the progress and make corrections where needed. Will request information from Gao and Tom on few points and will send a meeting invitation for 28 to reserve the timeslot on everybody's calendar.

4. Review of action items with upcoming/past deadlines (Stoyka, 15')

Postponed for lack of time

5. AOB, Tour de table (All, 20') for next meeting

Postponed for lack of time

6. Next meeting

28 April 13:00 to 15:00 CET.

References

[0]

[https://wmoomm.sharepoint.com/sites/wmocpdb/eve_activityarea/WMO%20Information%20System%20\(WIS\)_97452102-7575-e911-a98e-000d3a44bd9c/SG-DIP/SG-DIP-1/Presentations/12-2-2020/SGDIP1_Geneva_12Feb2020_Klausen.pdf?CT=1587473673230&OR=ItemsView](https://wmoomm.sharepoint.com/sites/wmocpdb/eve_activityarea/WMO%20Information%20System%20(WIS)_97452102-7575-e911-a98e-000d3a44bd9c/SG-DIP/SG-DIP-1/Presentations/12-2-2020/SGDIP1_Geneva_12Feb2020_Klausen.pdf?CT=1587473673230&OR=ItemsView)

[1] alex_vermeulen-wmopid.pdf