

SeaDataCloud 1st Annual Meeting

Athens, Greece

18-19 October 2017

Comments from the Advisory Board

Kenneth Casey – NOAA (USA)

Pascal Derycke – EMODnet (Belgium)

Mathieu Ouellet – DFO (Canada)

Peter Pulsifer – NSIDC (USA)

George Petihakis – HCMR (Greece)

WP 3 – Training of data providers

Consider engaging an online learning resources specialist to optimize experience for those who can't attend physically

WP 4 - Communication / Dissemination

Pointed out during meeting and duly noted :
open ASBL legal entity terms of reference to
potentially accept non-European countries

SDC would benefit from broader communications,
beyond the oceans community (e.g. polar data
organizations). The project is an excellent
reference implementation that others could and
should learn from.

WP 5 – Metadata and Data Catalogues

Direct link between CDI and CSR very good

Growth of CDI encouraging; but what about scalability ? Should start considering.

Purpose and scope of EDIOS catalogue not clear; perceived overlap between EDIOS and EDMED and EDMERP

Consider EDIOS catalog clean up; or clear descriptions /definitions of what goes in those three catalogues (EDMED, EDMERP and EDIOS) and periodically review contents.

Legacy of E prefix in catalogue names: EDMERP, EDIOS.. European, the contents in these catalogues have global coverage.

Opportunity : Crossref, DataCite and ORCID recently announced the formation of an Organization Identifier Working Group, perhaps EDMO should apply as start-up host organization

WP 6 – Core and Advanced Services

Excellent idea to have « robot shopper » testing services every day; essential in this day and age of service-based solutions

Noted that metrics will be collated and analyzed in annual metrics reports. Could this be more frequent, perhaps with less analysis? More frequent reporting would support more responsive system tuning.

WP 7 – Scientific committee

SC could and should be an integral part of the User Centered Design process if this is not already the case.

This would involve ensuring the connection between SC and systems designers and developers, not simply making SC reports available.

WP8 - Governance of standards and development of common services

Support the decision not to forbid netCDF 3.
Backward compatiability is important

At the same time, encouraging netCDF-4

Consider netCDF format for towed data ? Current
SDN formats / standards only allow for
monotonic profiles.

Consider publishing other linked open data
formats (e.g. Turtle, JSON-LD) to enhance
usability

A synoptic view of the vocabularies and ontologies in use by the SDC community would be useful. Difficult to easily see which vocabularies/ontologies are in use.

This may result in unnecessary duplication of vocabularies as system grows and connects to global system.

Support engaging in standardization level with the global community. SDC must be aligned with broader global community.

Is there a strategy for the use of unique persistent identifiers at the data and sensor level?

WP 9 - Developments of upstream services

Positive developments resulting in less 'clicks' for users to download data from several sources

Integration of external datasets from international : be mindful of aggregating aggregators. Potential for overlap and confusion, presenting user with exaggerated list of datasets for similar or same data.

Important to track data version, there may be cases where different versions (different QCs, calibration) of same data co-exist in the cloud. Use of persistent identifiers may be a solution

Should check out NOAA OneStop prototype (www.ncdc.noaa.gov/onestop/#) for comparison; 2-tiered discovery level (soon to be moved to production)

WP 9 - Developments of upstream services cont.

General question: does SDC definition of “marine” include life sciences and biological aspects of marine? If yes, then would better connections to OBIS, GBIF etc. be valuable? This is not prominent in the presentation(s).

WP 11 - Development, update and publication of data products

Many products presented; avoid situation of
« product looking for users »

-->User engagement through web form to collate suggestions for next wave of products.

Use the SDN discovery interfaces to document their availability

Allow users to do their own computations and products.

Make code available to reproduce existing products, and let users adapt them to different settings using the VRE

Conclusive Remarks

Team seems strong and well organized!



Users should drive requirements; start with user stories, build, then seek feedback

->User-centered design<-

More enthusiasm / emphasis about the computing and research capabilities offered by the cloud

IT development should receive similar care as the data we manage : modern methods and best practices (Agile..), good documentation, use of GitHubs, ideally a SDN Framework for IT development. This will result in easier code reuse and will add value to the project.