Pilot Blue Cloud: Discussion note for workshop

Issue

- Deliver the pilot blue cloud by 2020
- identify gaps and possible next steps forward.

Political state of play

- The European Commission aims to deliver a European Open Science Cloud by 2020 (*European Cloud Initiative* communicated by the Commission on 19 April 2016). In the planned timeline the Commission, as of 2017, will work with Member States to connect the priority European research infrastructures to the EOSC. Rapid progress is expected – among other reasons – because the Open Science Cloud is – also - a federation of existing and future Thematic Clouds.

- At the 2016 G7 S&T Ministerial, Commissioner Moedas referred to one such thematic cloud, when he proposed the development of a pilot Blue Cloud – containing data from seas and oceans - as a pilot initiative of the European Science Cloud, in support of the G7 'Future of the Ocean Initiative'. The G7 Ministers stated in their Communiqué that they support the development of an initiative for enhanced global sea and ocean observations required to monitor inter alia climate change and marine biodiversity.
- On 10 November 2016 Commissioner Vella and Vice President Mogherini adopted a joint Communication for international ocean governance which refers explicitly to 'strengthening international ocean research and data'.

Situation

- A pilot Blue Cloud will be a good example because it will include very different data types, even if the applications will be limited in a first phase. The pilot Blue Cloud can be an early mover because it will build on a number of already existing infrastructures, projects and initiatives to store and share data in the marine environment including: EDMED (FP3), SeaDataNet and SeaDataCloud, ODIP, EMODnet or ARGO. As such its establishment will be a great opportunity to demonstrate the EOSC, contributing at the same time to the G7 initiative for a global ocean observing system. This is not a coincidence. The EOSC's FAIR principles are particularly important for the fields of oceanographic and marine research, considering that:

1) Seas and oceans are obviously interconnected and they do not have hard boundaries.

2) Seas and oceans are highly complex systems and can only be understood with a holistic research approach that integrates – among others - physical, biological, biogeochemical and societal data.

3) The high costs associated with collecting data in this challenging and large environment make the principle of 'collecting data once and using it as many times as possible' particularly relevant.

4) Continuous improvement in forecasting requires cooperation in observations, open access to data and the capabilities of performing real time computations.

5) Only 20% of relevant data is hosted in national oceanographic data centres. Other data are in programme data centers or in industries. One of the challenges for the cloud will be to bring all other relevant data together.

Why?

- The challenges and opportunities generated by the advent of Big Data are well known. By some indicators more data is being generated every week than in the last millennia.[1] While this figure is smaller for research data, it is certainly true that Big Data Science is already a well-established reality, particularly in this area.

- The field of blue research is very much part of this trend and it is certainly "big science".[2] At the same time blue research, and improved forecasting in particular, will require – more and more - to merge data and scientific expertise from different disciplines.

- This state of affairs raises several practical challenges:

- The large amount of data will need to be effectively transmitted and stored and be made open access, often in real time. This is particularly true for data products delivered by observations such as the ones of the Global Argo Network (floats), EuroGOOS and satellite-based earth observation systems (such as COPERNICUS) that are continuously acquiring information.
- One of the prerequisites for a cloud is the open access, including interoperability of data. Among others, GEOSS has launched an Interoperability for Weather, Ocean and Water project that supported the development of the GEOSS Information System exploiting cloud possibilities.[3]
- Bringing data from in situ and remote measurements together needs further developments.

[1] OECD, STI Policy Note, data-driven innovation for Growth and Well-being, October 2016.
[2] Attachment 2 to Tsukuba Communiqué
[3] Big Data challenges in building the Global Earth Observation System of Systems, Nativi Stefano et al. 2015.

- Analysing it will require ever growing high performance computing capabilities. The challenge for data analysis of large data sets is not only scale, but also speed, the time value of information is critical (particularly in forecasting).
- Increasingly, such analyses will be multi-disciplinary and they will require the combination of different ocean key variables among themselves as well as with other indicators (for instance socioeconomic and governance indicators), this will require fully integrated platforms where data and expertise are shared in real time. This is quite different from other applications where the data sets are large, but more homogeneous.
- Cloud technologies will be crucial in addressing these challenges; also here it is crucial to learn from past and ongoing efforts such as large European Research Infrastructures that have already developed clouds and cloud services.

- EOSC and the pilot Blue Cloud promise to unlock the potential of data driven innovation for the blue economy. There are growing legitimate demands that research funded by the EU should impact positively the lives of its citizens by fostering innovation. The Blue Cloud should also be a tool to shorten the time span between research and innovation, in line with the guidelines of the Commission Communication on *A digital single market strategy for Europe*.[4]

- For blue researchers and scientists, cloud computing can provide a powerful and flexible platform for analysis and data storage. Large datasets can be stored in the cloud and accessed from any location with connectivity, solving the issue of data storage and facilitating collaborations. High performance computing will also be increasingly distributed with the rise of analytical capabilities on the cloud. This means that HPC will be acquired as needed without the necessity of

[4] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0192&from=EN

creating and supporting duplicated infrastructure, allowing saving precious resources.[5]

What is it for?

- The Blue Cloud therefore could soon become an important tool for scientists and policy makers and will support the addressing of pressing issues such as global climate regulation, change and impact, marine biodiversity, sea-level rise and the impact of human activities on marine ecosystems.

- Furthermore, the Blue Cloud will enhance the growth of the blue economy through the improvement of our understanding of what resources exist in the oceans and the seas, and how they can be harvested sustainably.

- In line with the expected EOSC implementation, progresses on a pilot Blue Cloud will require to:

- build on and leverage already available resources. Given the expected time-frame, the pilot Blue Cloud will be built with existing observations and data.

- data considered should already be stored and processed according to FAIR principles; capitalising on the long preparations and investments by the Commission and other bodies on issues like cataloguing, data sharing and interoperability.

- build on past and ongoing cloud related projects such as GEOWOW, SeaDataCloud, EVER-EST and BlueBridge.

- Engage with the blue stakeholders to develop a specific list of needs from selected users.

[5] 'A Primer on Cloud Computing', in Cloud Computing in Ocean and Atmospheric Sciences, 2016

Possible Next Steps (for discussion)

- realize a pilot blue cloud together with other partners, for example GEO/GEOSS or IOC/IODE.

- set the contents of the pilot blue cloud to existing essential ocean variables and other variables needed for monitoring inter alia climate change and marine biodiversity.
- develop pilot cloud applications to use the cloud, for example to forecast the impact of climate change.