The future data infrastructure

ALVARO LORENZO 06 JUILLET 2022 14H06

1. Is the concept of DAC or GDAC appropriate for the new datasets

for certain types of data, e.g. Biology, there are infrastructures and data flows setup that could be expanded to deal with NRT data. Duplicating the effort and the resources it needs seems not a great idea

Yes to both. GDAC most appropriate for that data type. Central place to advertise data held elsewhere

GDAC by variable type sounds more appropriate for some data (PCO2, taxonomy)

connections with specialized centers are necessary but I would not call them 'GDACs' for gliders
— PIERRE TESTOR

what are alternatives? Necessary to disseminate data in different thematic data centers (e.g. ecotaxa)

Ability to combine different types of data measured in the same time should be ensured. — ANONYME

2. What should be the role of GROOM RI regarding these new data types

Develop a road map for integration into the European Data Management Infrastructures

Best practices for data sharing

Offer single access

archive?

sharing the resources to avoid duplication of effort

3. What should be the role of GROOM RI in the European Data Infrastructure Landscape

Coordination/ High level management

Participate in interoperability exercises with other RIs

Providing support for data mangement (BP, data management services)

Guarantee the link between type of data (in the case of multiple data).

Address and propose solutions and services for datas not handled by other RIs

a node well connected to the different stakeholders

4. Near Real Time Data. Can we do more with a dedicated infrastructure?

Depends on the type of data, this is not straightforward for Biological data (e.g. taxa identification from images that require validation from experts)

EMODnet Biology would be ready to ingest NRT imagery plankton data, if not validated. The question is: Is it more interesting to have NRT not validated data OR delayed validated data

I guess we need both — PIERRE TESTOR

real time corrections

5. Managing New datasets in delayed/recovery mode. What are the challenges?

linking the data together

Lack of Image repositories and Imagery metadata catalogues linked.

same thing for acoustics — PIERRE TESTOR

Specific expertise on the data types not easily transferable/acquired

delayed mode data sometimes not shared with DAC

Reducing the time between NRT and delayed mode data being available

define data and metadata format for each new data type.

volume - data transfer to DACs/GDAC

metadata collation (engineering variables)

The amount of data

transfer of expertise