

Newsletter of the Digital Earth Project

Contributions of HMGU HelmholtzZentrum München

This newsletter presents the projects of the institutes of the HMGU being involved in activities within the Show Cases or Work Packages of Digital Earth.

DEEP Hybrid DataCloud

Keiichi Ito & Wolfgang zu Castell

HelmholtzZentrum München - German Research Center for Environmental Health (HMGU)

Deep Hybrid-DataCloud (DEEP) supports ML/DL (machine learning and deep learning), parallel post-processing of very large data, and analysis of massive online data streams over distributed e-Infrastructures in the European Open Science Cloud (EOSC). It came out of a European project aiming at preparing a new generation of e-infrastructure that harness latest generation technologies, supporting deep learning and other intensive computing techniques to exploit very large data sources. The project concluded in April 2020. It provides tools and web interfaces to deploy develop and use containers in a Cloud (remotely located hardware). The codes and tools are available through GitHub repositories and its own Marketplace. The system is intended to facilitate research and collaboration dealing with large data and requiring high computational speeds. Three main benefits from the user perspective are listed below.

1. **Train at scale** can be achieved by browsing and collaborating through the marketplace and submitting newly developed models to the EOSC resources while training and monitoring their status are observable.



Figure 1: DEEP Marketplace

2. **The Development environment** is a Jupyter Lab environment to develop your model in the remotely deployed Docker container. Automatically generated directories and files are also available to facilitate Python module creation.



Figure 2: Data Science Template (directory and file template)

3. **Deploy as a service (DaaS)** to serve your models through a common HTTP endpoint, with automated deployment there is an API.

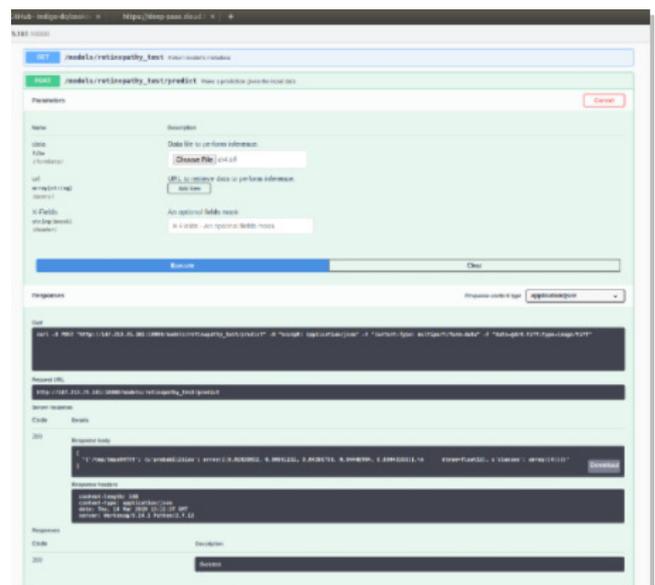


Figure 3: Dashboard for container deployment

- 4. A web interface is implemented to deploy your container in a cloud (remotely distributed hardware)

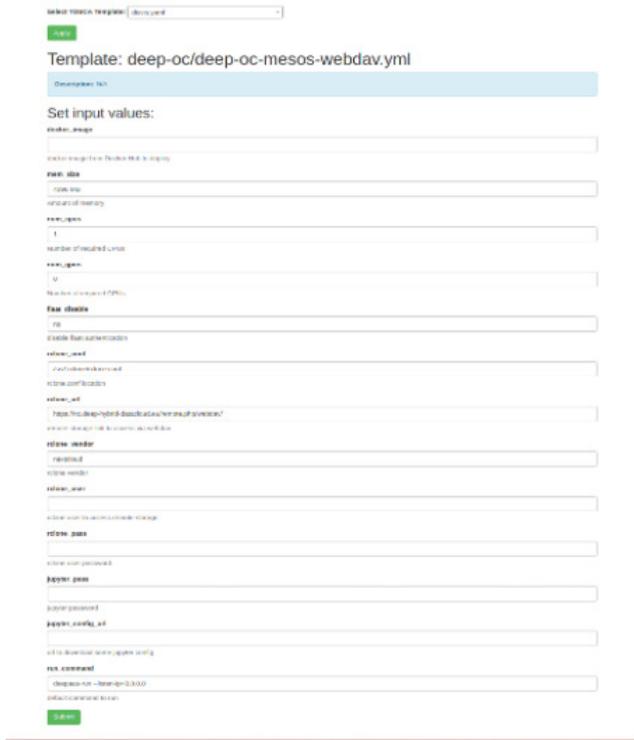


Figure 4: DEEPaaS API (application interface)

Microbial Community Explorer Tool

Carlos Garcia-Perez & Wolfgang zu Castell
 Helmholtz Zentrum München - German Research Center for Environmental Health (HMGU)

The microbial communities can be found in different environments, such as human and animal gut, oceans, rivers, plants, etc. Therefore, it is crucial to understand the interaction and function between these microbial populations. In order to perform this type of analysis, we have designed a visual explorer tool for microbial communities, with four linked plots: An interaction-network where the user can see the interaction of bacteria communities through a graph, a clustering of samples based on abundances through a Multidimensional/Non-Multidimensional scaling which is linked to the network graph, a bar plot of the top relative abundance bacteria based on the clustering of the abundances and a taxonomic composition plot. Moreover, the tool incorporates the metadata as an interaction menu for fast filtering and thus update the views. Finally, the tool can perform analyses at different taxonomy levels for a comparison within the same session.

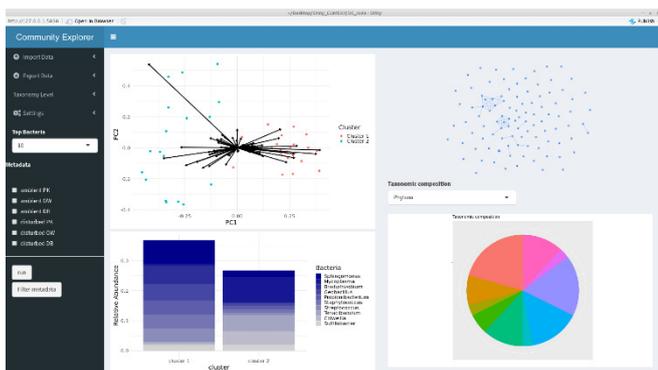


Figure 5: General view of the Community Explorer Tool

In general, the Community Explorer Tool (CET) (Figure 5) is a tool that can be used for any type of microbial datasets by speeding the analysis and gain insight of the data. The study from Wegner (Wegner et al., 2013) focuses on the difference of individual host-associated microbial communities based on population and individual genetic differentiation, and to identify the bacterial taxa in those communities that react similarly to disturbances stress in individuals and host populations, in order to determine if the disturbance promote the increase of opportunistic and potentially pathogenic strains. The information collected is from three oyster beds in the northern Wadden Sea (Figure 6) Diedrichsenbank (DB), Oddewatt (OW) and Puan Klent (PK).

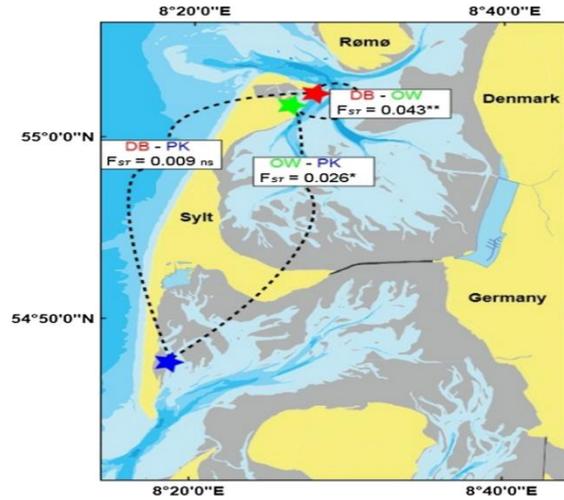


Figure 6: Oysters bed localization. Original image from (Wegner et al., 2013).

Figure 7 shows three of the views that provide the same information but in a visual way as the analysis made in the study. The CET groups the three oyster beds (DB, OW and PK) into two groups (Cluster 1 and Cluster 2) shown in Figure 7a.

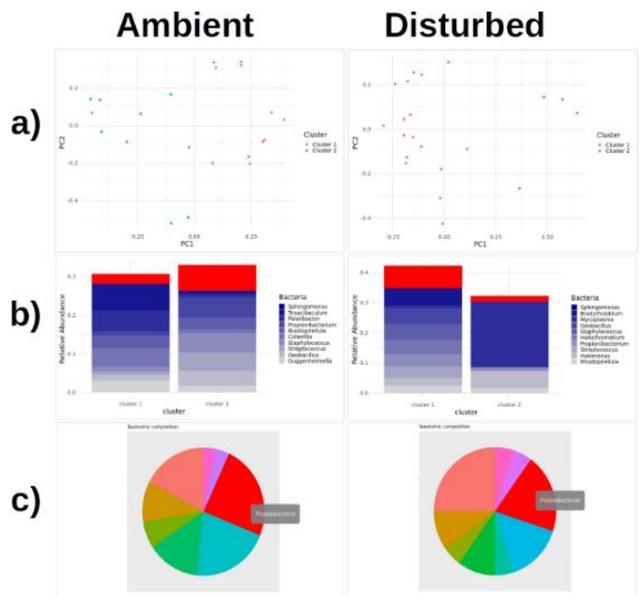


Figure 7: Taxonomic composition and taxonomy abundance.

The study analyzes the taxonomic composition for both groups. The group without the perturbation (ambient) is in the left column, it shows the most abundant taxonomic composition at the phylum level (Figure 7c) were: Proteobacteria, Planctomycetes, Fusobacteria, Actinobacteria and Firmicutes. And at the genus level (Figure 7b) were:

Sphingomonas, Tenacibaculum, Polaribacter, Propionibacterium and Colwellia. In contrast, the right column shows the group disturbed. The comparison shows that the group is also abundant in Proteobacteria as the ambient group, however in the group is less abundant, the rest of the most abundant phylum was: Actinobacteria, Planctomycetes, Bacteroidetes and Firmicutes. The difference with the ambient group is found when analyze the genus of the group were Mycoplasma is the most abundant followed by Sphingomonas, Bradyrhizobium and Halomonas. These visual results of the CET are consistent with what was reported in the study, in which the Mycoplasma genus in the disturbed group is highlighted. The views were obtained through the interaction between the CTE and the user using the filters on the left menu of the CTE (Figure 5). Figure 8 shows the comparison of the bed PK for the ambient and disturbed groups. In the upper part of the figure, two groups are shown (Cluster 1 and Cluster 2).

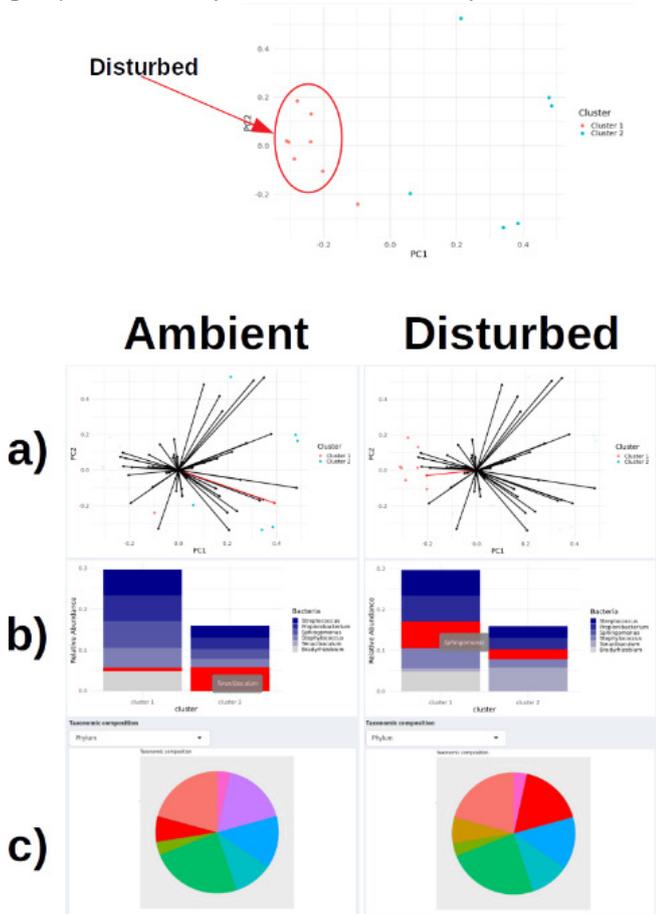


Figure 8: Shared taxonomy comparison.

Cluster 1 represents the disturbed group while the other represents the ambient group (we have removed the arrows to clearly show the groups). One of the analyses performed in the study was the comparison of taxonomy to determine which taxa were shared by the two groups. In this case we compared both PK (ambient and disturbed), Figure 8a shows that ambient group has a greater influence of Tenacibaculum (red arrow to the right), in contrast the disturbed group contains several taxa that have a strong influence on the group, such as Streptococcus, Propionibacterium or Sphingomonas (red arrow to the left). The same information can be seen in Figure 8b, by means of these linked views the selection of a bar is automatically reflected in the view above (Figure 8a), and the proportion of phylum shown in Figure 8c, Bacteroidetes (ambient) and Proteobacteria (disturbed) is also highlighted as a reaction of bar selection. We have shown that the CET is a tool that fits well in the context of the DE project and can be beneficial to the different groups within the project.