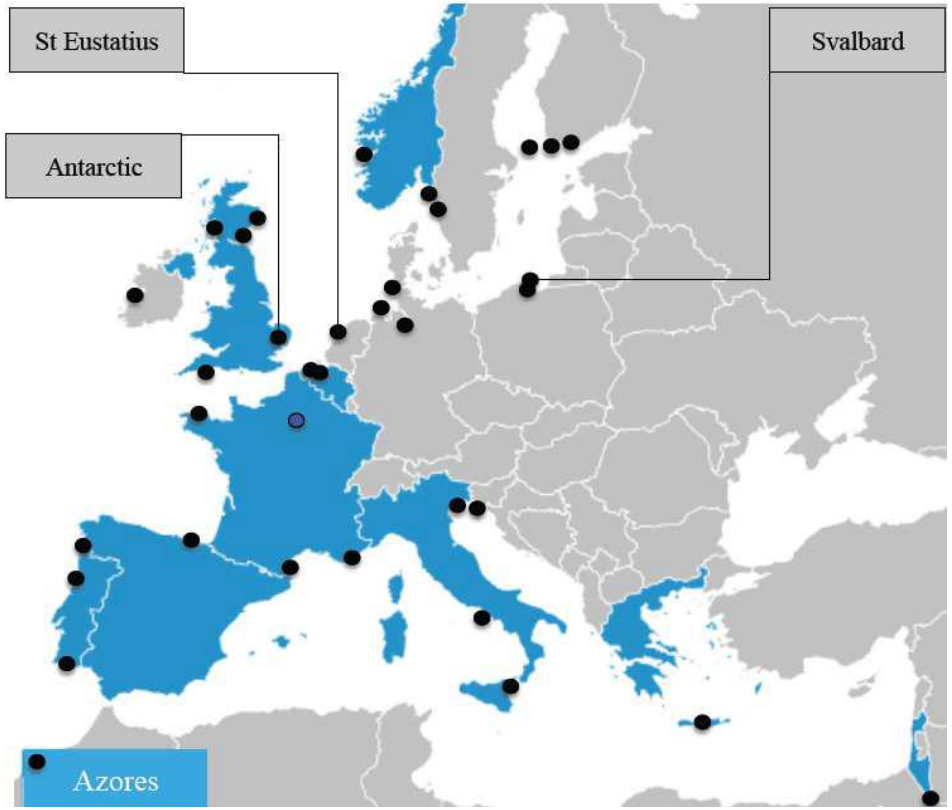
WP4 – NA2 Improving virtual access to marine biological stations data, information and knowledge

Klaas Deneudt – Flanders Marine Institute (VLIZ)

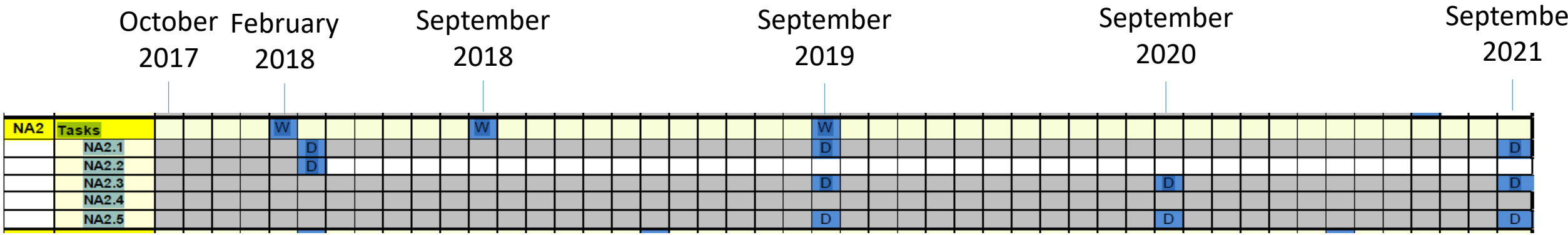
Objectives

- Improve virtual access to marine biological stations data, information and knowledge by:



- Shaping and Assemble Plus community **data management plan**
- Improving **interoperability** with related e-infrastructures
- Installing an **on-line system** to provide access to marine biological stations **information**
- Improving **virtual access to data resources** from Assemble Plus stations, with a specific focus on **genomic** and **long term biodiversity** data series
- Offering a **virtual analysis platform** as a service for genomic and biodiversity data analysis

Timeline


October 2017
February 2018
September 2018
September 2019
September 2020
Septembe 2021
NA2
Tasks
NA2.1
NA2.2
NA2.3
NA2.4
NA2.5
W
D

Work package number	NA2	Lead beneficiary				VLIZ	
Work package title	Improving virtual access to marine biological stations data, information and knowledge						
Participant number	17	10	20	16	8	1	12
Short name of participant	VLIZ	HCMR	MBA	MPIMM	UGOT	UPMC	SZN
Person/months per participant:	**34.5**	**18**	**8**	**9**	**4.5**	**4.5**	**3**
Participant number	14	9	4	11	15	18	21
Short name of participant	CCMAR	UPV/EHU	UH	HUJI	AWI	SAMS	NERC-BAS
Person/months per participant:	**3**	**2**	**1.5**	**1.5**	**1.5**	**1.5**	**1.5**
Participant number	23	2	5	3	6	7	**19**
Short name of participant	MSS	NIB	IOPAN	NIOZ	UG	NUIG	USTAN
Person/months per participant:	**1.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Start month							

NA2.3: **Multidisciplinary datasets inventory and archival**: UPMC(x3), NIB, NIOZ, UH, IOPAN, UG, NUIG, UGOT, UVIGO, EHU, HCMR, HUJI, SZN(x2), UiB, CCMAR (x2), AWI, MPIMM, VLIZ, SAMS, USTAN, MBA, NERC, MSS

NA2.4: **LT biodiversity data series & genomic data**: UPMC, UH, EHU, HCMR, HUJI, SZN, CCMAR, AWI, VLIZ, SAMS, MBA, NERC, MSS

Objectives

Improve virtual acces
- Shaping an ASSEM
- Improving interope
- Installing an on-line
- Improving virtual a

long term biodiversity data series;
- Offering a virtual analysis platform as a service for genomic and biodiversity data analysis.

Task NA2.1. Design Data Management Plan and interoperability with related e-infrastructures (1/2)

Task Leader: VLIZ; participants HCMR, MBA, MPI.

Deliverables: D NA2.1a,b, c: Data Management Plan (M6, M24, M48)

A data management plan (DMP) will be drafted as a reference document for the ASSEMBLE Plus community. This document will outline how the **research data** collected or generated within the **JRAs and as part of the TA** will be handled during the ASSEMBLE Plus project and will remain in force after the end of the project.

The DMP will build on current best practices at the national, EU and international levels, and recommendations and guidance from the ICSU World Data System (ICSU-WDS), Research Data Alliance (RDA) and the International Oceanographic Data Exchange (IODE). The DMP will define the methodology and standards to be applied, whether and how these data will be shared and/or made open, and how they will be curated and preserved.

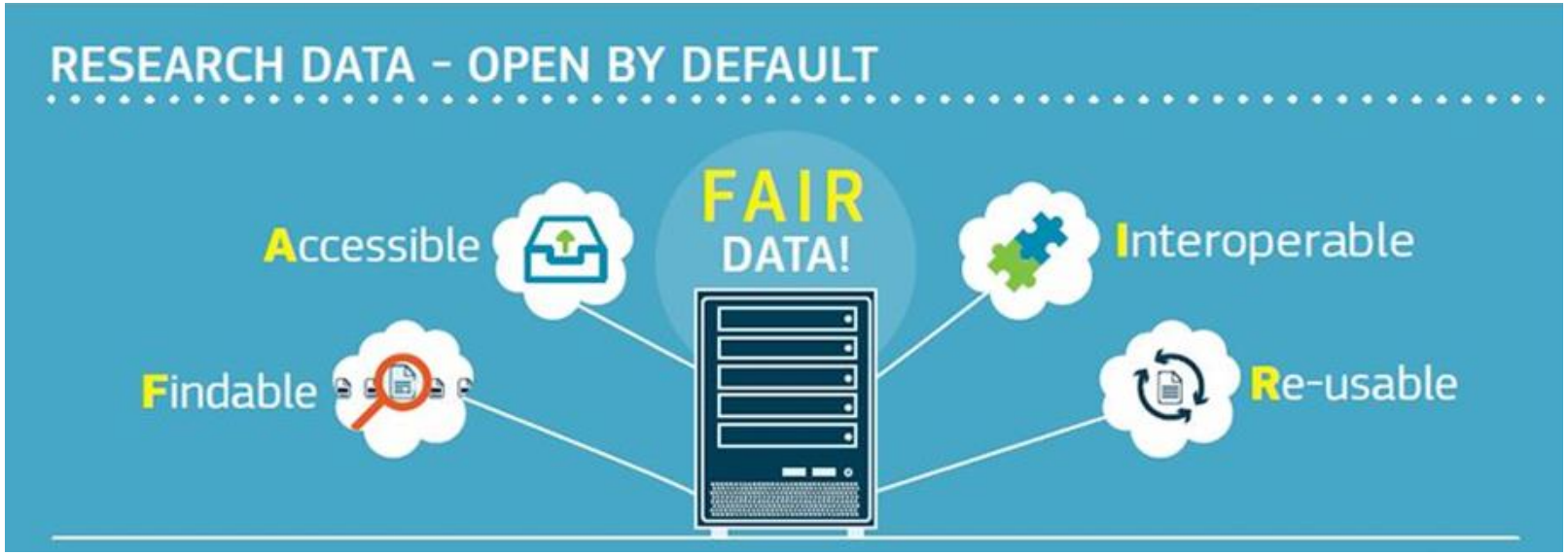
What is a DMP

- **What is a data management plan?**
  - Document defining what data will be generated, how data will be curated and preserved and whether and how data will be shared or made available.

- **Why do we need it?**
  - H2020 requirements
  - Essential part of planning research activities

What is a DMP

- **Elements of a Data Management Plan**
  - Framework and scope
  - Data Collection
  - Documentation and Metadata
  - Storage and Backup
  - Selection and Preservation
  - Data Access, Sharing and Rights
  - Responsibilities and Resources



Task NA2.1. Design Data Management Plan and interoperability with related e-infrastructures (2/2)

Task Leader: VLIZ; participants HCMR, MBA, MPI.

Deliverables: D NA2.1a,b, c: Data Management Plan (M6, M24, M48)

During a **DMP workshop** organized early in the project (month 5), data experts from related international and European e-infrastructures will be invited to discuss applicable standards and interoperability issues for the data collected by the marine biological stations. A specific session will be dedicated to the genomic observatories data protocols. The DMP will not be static, but rather will evolve and gain substance throughout the project.

An initial, mid-term and final version will be provided as deliverables from this WP (D NA2.1 at M6, M24, M48).

Task NA2.2 Creation of the Knowledge Transfer Platform

Task Leader: VLIZ.

Deliverable: D NA2.4: Knowledge Transfer Platform Online (M6)

An online **Knowledge Transfer Platform** will be implemented as part of the ASSEMBLE Plus web portal. This online system will become a central tool for provision and exchange of information of researchers and research groups, their expertise and scientific output in terms of publications with relevant stakeholder groups. The online system will include two modules:

1. Knowledge Outputs module will provide the infrastructure to provide users with access to an unrestricted, user friendly and searchable database of **Knowledge Outputs** which may be relevant to them, **as provided for by NA3**. Outputs from TA and from research developed at each marine station will be submitted through a customized form which includes:

2. Publications module will be integrated as a **publication catalogue and archive in the web portal**. This module will serve as an open archive for publications resulting from the ASSEMBLE Plus project. The use of the Knowledge Transfer Platform will be registered and reported using in-built tools (D NA2.4).

ASSEMBLE +

ASSOCIATION OF EUROPEAN MARINE BIOLOGICAL LABORATORIES

This platform supports communication and knowledge exchnage among stakeholders, scientists, policy makers and the public.

Featured Knowledge Output

Recent publications

- **Heesch S, Cho GY, Peters AF, Le Corguille G, Falentin C, Boutet G, Coedel S, Jubin C, Samson G, Corre E, Coelho SM, and Cock JM**. A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. In *New Phytologist*, 188: 42–51

Most viewed Knowledge Output

ASSEMBLE +

ASSOCIATION OF EUROPEAN MARINE BIOLOGICAL LABORATORIES

Search

KO types

Book/Review
Multimedia
Product
Report
Publication (7)
...

Potential End Users

Education & Training
Environmental managers & monitoring Industry
Policy makers / decission makers
Scientific community
...

Publications

- **Zingone A, Forlani G, Percopo I, & Montresor M**. Morphological characterization of *Phaeocystis antarctica* (Prymnesiophyceae). In *Phycologia*. doi: 10.2216/11-36.1.

- **Reid EL, Worthy CA, Probert I, Ali ST, Love J, Napier J, Littlechild JA, Somerfield PJ & Allen MJ**. Coccolithophores: Functional Biodiversity, Enzymes and Bioprospecting. In *Marine Drugs*. doi:10.3390/md9040586.

- **Ota S, Kudo A & Ishida K**. Gymnochlora dimorpha sp. nov., a chlorarachniophyte with unique daughter cell behaviour. In *Phycologia*. doi: 10.2216/09-102.1.

- **Kher CP, Doerder FP, Cooper J, Ikonomi P, Achilles-Day U, Küpper FC, Lynn DH**. Barcoding *Tetrahymena*: discriminating species and identifying unknowns using the cytochrome *c* oxidase subunit I (cox-1) barcode. In *Protist*. doi: 10.1016/j.protis.2010.03.004

- **Harnstrom K, Ellegaard M, Andersen T, Godhe A**. Hundred years of genetic structure in a sediment revived diatom population. In *Proceeding of the National Academy of Sciences*. doi:10.1073/pnas.1013528108

- **Hagino K, Bendif EM, Young JR, Takano Y, Probert I, Horiguchi T, de Vargas C & Okada H**. New evidence for morphological and genetic variation in the cosmopolitan coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) from the coxIb-ATP4 genes. In *Journal of Phycology*. doi: 10.1111/j.1529-8817.2011.01053.x

- **Bendif EM, Probert I, Hervé A, Billard C, Goux D, Lelong C, Cadoret JP, & Véron B**. A taxonomic reassessment of the Pavlovophyceae (Haptophyta). In *Protist*. doi: 10.1016/j.protis.2011.05.001.

Currently 234 publications from ASSEMBLE Marine in the Integrated Marine Information System (IMIS)

IMIS

log in

Publicaties | Instituten | Personen | Datasets | Projecten | Kaarten

Records pp | Sorteren | Selecteren | RSS feed | Help

mandje[?] (0): toevoegen | legen | toon

Geen records geselecteerd

1 2 3 4 5

Records 1 - 10 van 234

zoekopdracht: nieuw - terug

Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida
(2013)
. *Proceedings of the National Academy of Sciences of the United States of America 110(13)*: 5247-5252. https://hdl.handle.net/10.1073/pnas.1221259110

Influence of ocean acidification on a natural winter-to-summer plankton succession: first insights from a long-term mesocosm study draw attention to periods of low nutrient concentrations
Bach, L.T.; Taucher, J.; Boxhammer, T.; Ludwig, A.; The Kristineberg KOSMOS Consortium; Achterberg, E.P.; Algueró Muñiz, M.; Anderson, L.G.; Bellworthy, J.; Büdenbender, J.; Czerny, J.; Ericson, Y.; Esposito, M.; Fischer, M.; Haunost, M.; Hellemann, D.; Horn, H.G.; Hornick, T.; Meyer, J.; Sswat, M.; Zark, M.; Riebesell, U. (2016)
. *PLoS One 11(8)*: e0159068. https://hdl.handle.net/10.1371/journal.pone.0159068 [download pdf]

Isolation of clonal cultures of endosymbiotic green algae from their ciliate hosts
Achilles-Day, M.; Day, G. (2013)
. *Journal of Microbiological Methods 92(3)*: 355-357. https://hdl.handle.net/10.1016/j.mimet.2013.01.007

Unexpected high genetic diversity at the extreme northern geographic limit of *Taurulus bubalis* (Euphrasen, 1786)
Almada, V.C.; Almada, F.; Francisco, S.M.; Castilho, R.; Robalo, J.I. (2012)
. *PLoS One 7(8)*: e44404. https://hdl.handle.net/10.1371/journal.pone.0044404 [download pdf]

The identification and management of pain, suffering and distress in cephalopods, including anaesthesia, analgesia and humane killing
Andrews, R.; Darmaillacq, S.; Dennison, N.; Gleadall, G.; Hawkins, P.; Messenger, B.; Osorio, D.; Smith, J.; Smith, A. (2013)
. *Journal of Experimental Marine Biology and Ecology 447*: 46-64. https://hdl.handle.net/10.1016/j.jembe.2013.02.010

Intact cluster and chordate-like expression of ParaHox genes in a sea star
Annunziata, R.; Martinez, P.; Arnone, I. (2013)
. *BMC Biology 11*. https://hdl.handle.net/10.1186/1741-7007-11-68

Pattern and process during sea urchin gut morphogenesis: The regulatory landscape
Annunziata, R.; Perillo, M.; Andrikou, C.; Cole, G.; Martinez, P.; Arnone, I. (2014)
. : 251-268. https://hdl.handle.net/10.1002/dvg.22738

Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae
Arslan, D.; Legendre, M.; Seltzer, V.; Abergel, C.; Claverie, M. (2011)
. *Proceedings of the National Academy of Sciences of the United States of America 108(42)*: 17486-17491. https://hdl.handle.net/10.1073/pnas.1110889108

Tidal controls on trace gas dynamics in a seagrass meadow of the Ria Formosa lagoon (southern Portugal)
Bahlmann, E.; Weinberg, I.; Lavric, J.V.; Eckhardt, T.; Michaelis, W.; Santos, R.; Seifert, R. (2015)
. *Biogeosciences 12(6)*: 1683-1696. https://hdl.handle.net/10.5194/bg-12-1683-2015 [download pdf]

Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples
Balzano, S.; Marie, D.; Gourvil, P.; Vaulot, D. (2012)
. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology 6(8)*: 1480-1498. https://hdl.handle.net/10.1038/ismej.2011.213

Geen records geselecteerd

1 2 3 4 5

Records 1 - 10 van 234

Records pp | Sorteren | Selecteren | RSS feed | Help

mandje[?] (0): toevoegen | legen | toon

Bekijk details van record

Top

Task NA2.3 Set up virtual open access entry point to data resources (1/3)

Task Leader: VLIZ; participants: HCMR, MBA, MPI (and all other partners).

Deliverables: D NA2.2a, b, c: Virtual Access hits to ASSEMBLE Plus data resources (M24, M36, M48)

The central entry point that allows users to access common ASSEMBLE Plus data resources (from JRAs, NA and TA) will be **created based on existing components**.

The datasets module of the Integrated Marine Information System will be plugged into the ASSEMBLE Plus website and will allow users to query and select data resources based on a set of relevant search criteria. In order to create a digital inventory of the data resources available at the ASSEMBLE Plus stations, **each of the partners will describe their datasets in a digital catalogue, using an online form.** The scope, characteristics, state and accessibility of the data will be described following common standardized formats compliant to ISO19115 metadata standard. In case the data is already accessible through local online databases a web link to the existing interfaces will be included in the dataset description.

Data will be deposited into a central ASSEMBLE Plus data repository for archival purposes.

Task NA2.3 Set up virtual open access entry point to data resources (2/3)

Task Leader: VLIZ; participants: HCMR, MBA, MPI (and all other partners).

Deliverables: D NA2.2a, b, c: Virtual Access hits to ASSEMBLE Plus data resources (M24, M36, M48)

Data infrastructure components that are specific to the genomics observatories data and that were developed during the FP7 MicroB3 project **will be revitalized and connected to the entry point** to meet the GOs data requirements. This includes links to other publicly available virtual environments, services, standards registries, and data repositories (e.g. Moorea Biocode, the biocode commons, the GSC, workflows developed during MicroB3, ENA-EBI data repository, etc.). Data gathered in JRA1 (Genomics Observatories) will be developed into standardized and user friendly formats and incorporated in a “one stop shop” virtual access entry point to GO data and services developed by MicroB3. In addition to providing access to, the portal will provide information on community data policies and standards, experimental protocols/logs, and the outcomes of benchmarking exercises.

Task NA2.3 Set up virtual open access entry point to data resources (3/3)

Task Leader: VLIZ; participants: HCMR, MBA, MPI (and all other partners).

Deliverables: D NA2.2a, b, c: Virtual Access hits to ASSEMBLE Plus data resources (M24, M36, M48)

Data that can be brought into the public domain will be prepared for data publication. A technical quality control will verify that all required information is included. Each of the datasets proposed for publication will be made citable and will be labeled with a digital object identifier (DOI). All gathered information will be checked, improved and validated during a common data workshop (M12) to which all providers will be invited.

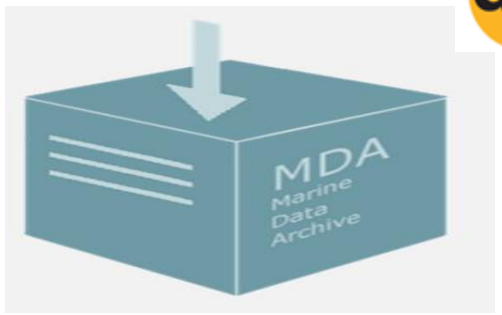
The use of the virtual open access platform will be registered and monitored using built-in tools. Data availability and "hits" will be assessed by the Advisory Board and reported as a yearly deliverable to the EC (D-NA2.2a, b, c at M24, 36, 48).

Data archiving & publication service

IMIS

Integrated Marine

Information System



DataCite

ICSU
WORLD DATA SYSTEM

creative commons

Task NA2.4 Improve data access and standardization of genomic and long term marine biodiversity observation (1/2)

Task leader: VLIZ with contributions from HCMR, MBA, MPI, all genomic & biodiversity observatory stations.

Most marine biological stations collect and manage **historical times series of biodiversity data**. However, many valuable, historic datasets still remain inaccessible to the larger scientific community. The stations that manage these biodiversity data series and the **genomic observatory stations** will be invited to dedicated sessions during the **common data workshop (M12)**, during which they will receive **training in providing access** to these types of data.

This includes documenting, annotating, archiving, quality controlling and dealing with the IPR issues through existing licensing approaches (e.g. use of moratorium periods and accreditation by producing data citations and data papers). By M18 the described data series will be made accessible by the stations to the task team for final validation and incorporation in the VA platform. A contact point of the task team will be ready to assist the stations with any data related technical issues.

Task NA2.4 Improve data access and standardization of genomic and long term marine biodiversity observation (2/2)

Task leader: VLIZ with contributions from HCMR, MBA, MPI, all genomic & biodiversity observatory stations.

In order to prepare the data for correct downstream application, data collected using different sampling protocols needs to be annotated and standardized using common formats and vocabularies on used instruments, sampling protocols and analysis methodologies. To ensure interoperability with related e-infrastructures, standardization of the data will be performed according to existing community based standards.

The full range of **biodiversity and genomics observatory data will be made compliant with data formats and standards**, such as those available from EMODnet, LifeWatch, OBIS and GBIF as well as European data pathways developed within MicroB3 and Elixir. Moreover, the standardization process will take into account the appropriate international standards, such as those by the Genomics Standards Consortium (GSC, http://gensc.org/), Barcode of Life Data Systems (www.boldsystems.org), and UNITE (https://unite.ut.ee).

Task NA2.5 Set up virtual platform for data analysis

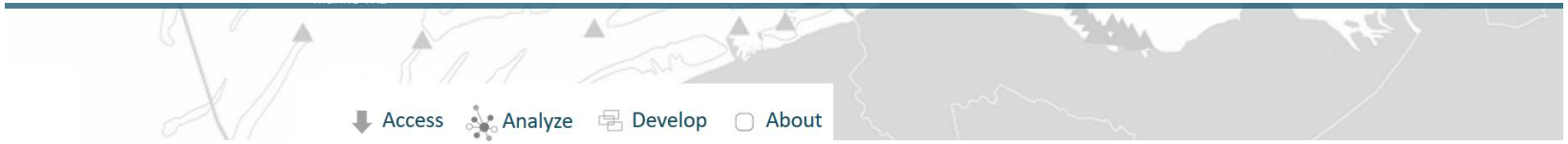
Task leader: HCMR with contributions from VLIZ, MPI, and UGOT.

Deliverables: D NA2.3a, b, c: Virtual Access runs to analysis platform (M24, M36, M48)

The **standardized data from long-term biodiversity and genomics observatories will be connected to a virtual analysis platform**. This platform will be an online portal (based on R web services and Taverna workflow systems) that enables the user to select from available datasets and perform predefined processing and analysis workflows on the selected data. Examples of workflows offered by the platform are existing outlier detection, visualization, statistical analysis and index calculation **workflows for biodiversity observation data and bioinformatics pipelines** (QIIME based demultiplexing, quality filtering, OTU picking, etc.). Depending on the type of workflow, output will be displayed as maps, reports and/or data tables. An authentication system will allow users to log on to the platform and access the pre-installed workflows.

A **workshop (M24)** will be organized to allow the data providers to **test and validate the analysis platform**. During the workshop a **webinar** will be prepared for teaching and stimulating the further use of the platform.

Use of the virtual analysis platform will be monitored using built-in tools. Number of users and “runs” per user will be assessed by the Advisory Board and reported on a yearly deliverable to the EC (DNA2.3a, b, c at M24, 36, 48).

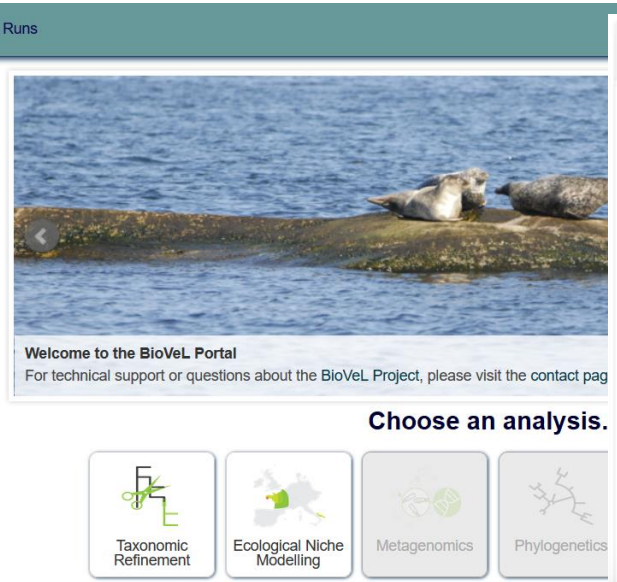


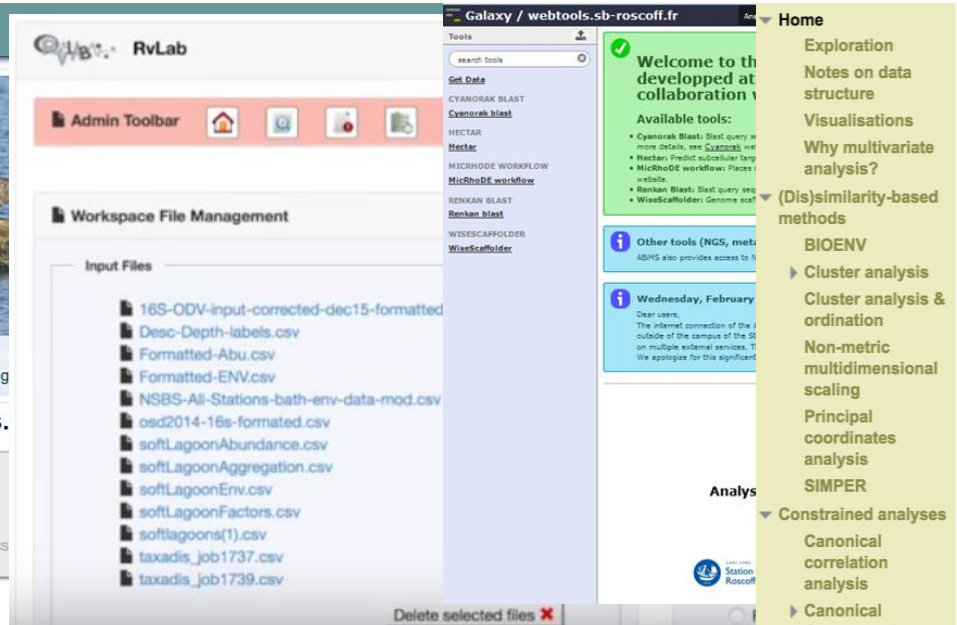
Access

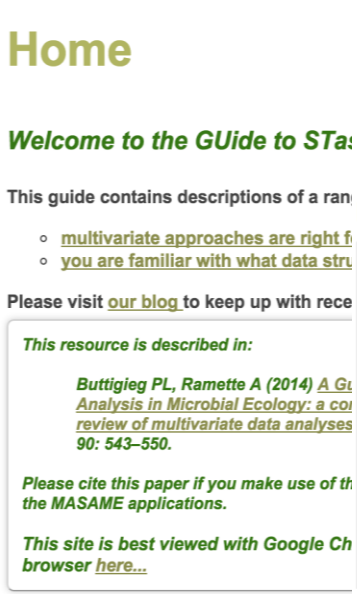
Retrieve and access data resources holding marine biodiversity and ecosystem data. A range of data systems offering data on species names, traits, distribution and genes.

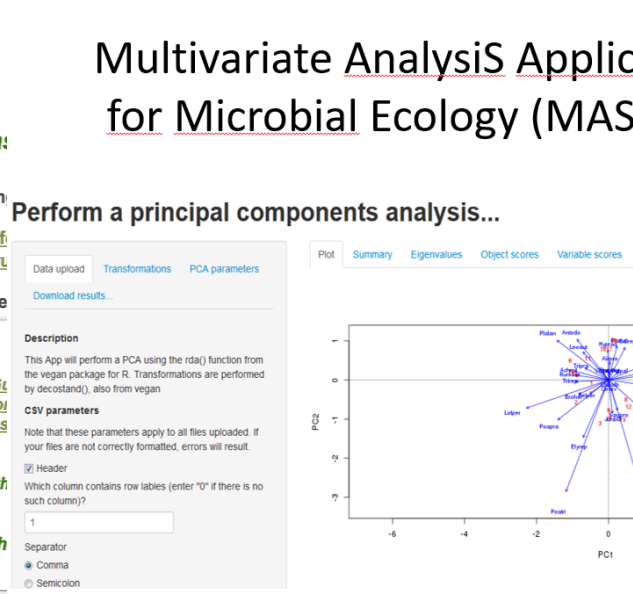
Analyze

Online tools that facilitate data analysis of marine biodiversity and ecosystem data. Analysis is performed on data from known data resources and/or data uploaded by the user.

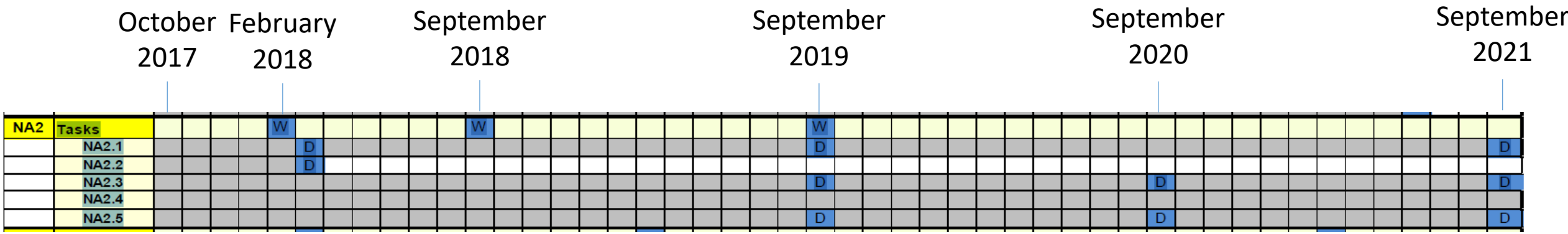








Deliverables



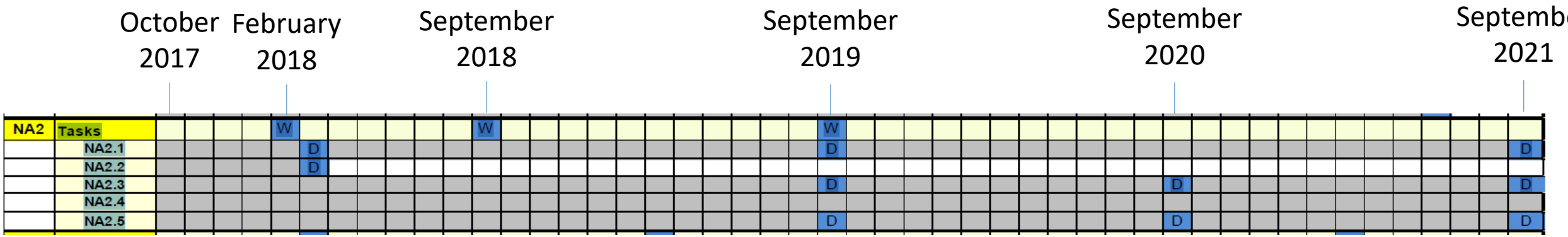
Task NA2.1: **D NA2.1a,b, c: Data Management Plan (M6, M24, M48)**

Task NA2.3: **D NA2.2a, b, c: Virtual Access hits to ASSEMBLE Plus data resources (M24, M36, M48)**

Task NA2.5: **D NA2.3a, b, c: Virtual Access runs to analysis platform (M24, M36, M48)**

Task NA2.2: **D NA2.4: Knowledge Transfer Platform Online (M6)**

Workshops



Task NA2.1: **DMP workshop (M5: February 2018)**

Task NA2.4: **Data workshop on biodiversity and genomic data (M12: September 2018)**

Task NA2.5: **Data analysis workshop (M24: September 2019)**

NA2 working group session (9:00 – 11:00)

- Discussion on:
  - Implementation
  - Timing
  - Workshops
  - Action points

ASSEMBLE+
ASSOCIATION OF EUROPEAN MARINE BIOLOGICAL LABORATORIES

Thank You!