

ASSEMBLE



ASSOCIATION OF EUROPEAN MARINE BIOLOGICAL LABORATORIES EXPANDED

Acronym: ASSEMBLE Plus

Title: Association of European Marine Biological Laboratories Expanded

Grant Agreement: 730984

Deliverable D4.4

First virtual access hits to ASSEMBLE Plus data resources 09 - 2019

Lead parties for Deliverable: VLIZ

Due date of deliverable: M 24 [30/09/2019]

Actual submission date: M 24 [26/09/2019]

All rights reserved

This document may not be copied, reproduced or modified in whole or in part for any purpose without the written permission from the ASSEMBLE Plus Consortium. In addition to such written permission to copy, reproduce or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright must be clearly referenced.



GENERAL DATA

Acronym: **ASSEMBLE Plus**

Contract N°: **730984**

Start Date: **1st October 2017**

Duration: **48 months**

Deliverable number	D4.4
Deliverable title	First virtual access hits to ASSEMBLE Plus data resources
Submission due date	30/09/2019
Actual submission date	26/09/2019
WP number & title	WP4, Improving virtual access to marine biological stations data, information, and knowledge
WP Lead Beneficiary	VLIZ
Participants (names & institutions)	Katrina Exter (WP leader NA2), Flanders Marine Institute (VLIZ); Georgios Kotoulas (WP Leader JRA1), Hellenic Center of Marine Research (HCMR); Christos Arvanitidis (NA2 partner, LifeWatch), Hellenic Center of Marine Research (HCMR); Dan Lear (NA2 partner, EMBRC WGEI), Marine Biological Association (MBA); Klaas Deneudt (NA2 partner, EMBRC WGEI), Flanders Marine Institute (VLIZ); Ivaylo Kostadinov (NA2 partner), Max-Planck Institute for Marine Microbiology (MPIMM)

Dissemination Type

Report	<input checked="" type="checkbox"/>
Websites, patent filling, etc.	<input type="checkbox"/>
Ethics	<input type="checkbox"/>
Open Research Data Pilot (ORDP)	<input type="checkbox"/>
Demonstrator	<input type="checkbox"/>
Other	<input type="checkbox"/>

Dissemination Level

Public	<input checked="" type="checkbox"/>
Confidential, only for members of the consortium (including the Commission Services)	<input type="checkbox"/>



Document properties

Author(s)	Katrina Exter
Editor(s)	Katrina Exter
Version	1

Abstract

This is the first reporting on the virtual open access entry point to the data resources of ASSEMBLE Plus (Task NA2.3 of WP4), created for the 24-month point in the project. This document explains what the access point is, what data resources it includes, future plans, and reports on the hits to the site.



1. Introduction	5
2. The data resources being accessed	5
2.1 The Transnational Access data resources	5
2.2 The JRA data resources	6
2.3 The long-term and genomics data resources	7
2.4 The publications.....	9
3. The virtual open access entry points	9
3.1 The datasets access point	10
3.2 The publications access point	11
4. Hits to the virtual open access entry points	12



1. Introduction

This deliverable is to report on the *virtual open access entry point to the ASSEMBLE Plus data resources*: what the data resources being accessed are, the platform via which they are accessed, the future plans, and a reporting on the hits to the site.

Setting up this access point is the work of WP4 Task NA2.3, and the data resources being accessed are largely the Type 1 and 2 ¹data discussed in the Data Management Plan (WP4 Task NA2.1; D4.2), which includes the data being gathered in WP4 Task NA2.4 (Long-term biodiversity and genomics observations).

2. The data resources being accessed

The ASSEMBLE Plus data resources that are being accessed via the open access entry point are:

- Data produced by the users of the Transnational Access programme during their TA visits
- Data and data products produced by the Joint Research Activities
- Data records and datasets gathered from our ASSEMBLE Plus marine stations and partners in Task NA2.4: Long-term marine biodiversity and genomics observations
- ASSEMBLE-Plus related publications

2.1 The Transnational Access data resources

The “TA data resources” are the datasets that are created by the users of the TA programme during their visit to the ASSEMBLE Plus marine station(s). The management of these data is explained in the Data Management Plan (Task NA2.1; D4.2): the users are requested to archive their data in the Marine Data Archive ([MDA](#)) and catalogue them in the Integrated Marine Information System ([IMIS](#)), where they are included in the “ASSEMBLE Plus collection”. The datasets linked to the records are required to be open access within two years after data collection.

The MDA and IMIS are two VLIZ data systems, and as part of the remit of its Data Centre, assistance is given to the TA users in the archiving and cataloguing process. Guidelines are also provided on the ASSEMBLE Plus [FAIR data management webpages](#). The FAIR expectations are there explained: for these TA data the emphasis is on the **F**indable (creating a suitable metadata record in IMIS), and **A**ccessible and **R**e-useable (open access at least after two years from data collection, and obtainable via a direct download link). Advice about creating **I**nteroperable, i.e. standardised, datasets is given on the ASSEMBLE Plus webpages, however a full curation of the interoperability of these datasets is beyond the resources available.

¹ Type 1: data generated by a public research project carried out at a member stations and originating from ASSEMBLE Plus user access provision and where the project covers the operating costs. Type 2: ASSEMBLE Plus partners’ institutional data that are not necessarily generated in the context of ASSEMBLE Plus, but are explicitly listed by the partners as part of the service offer.



The individual TA datasets will mostly be limited in scope as they are research projects that are usually smaller parts of a larger whole, and run for at most a month. The scope of the *topics* covered by the TA part of the data collection will, however, be wide.

At present, only five TA datasets have been added to the ASSEMBLE Plus collection (despite over 100 projects having taken place to date). *Uptake of the archiving and cataloguing is very low*, and responses to emails are rare – possibly most TA users “forget” this commitment. (Call 1 TA users should be excused here because the FAIR requirements were not made very clear in the application process for that call.) Enforcement of this is difficult. However, we are considering replacing the submission of a DMP by the TA users with an actual archiving and cataloguing process, and this would most likely improve the uptake.

2.2 The JRA data resources

This data arising from the Join Research Activities will be varied in size and scope.

- **JRA 1 Genomics Observatories.** The motivation for this JRA is to foster the application of genomics technologies at Long-Term Ecological Research Network (LTER) sites. The project encompasses: populating and verifying databases of taxonomic reference barcodes; harmonising meta-barcoding standard operating procedures (SOPs) across the consortium; and inter-calibration of classical biodiversity data and genomics data. The final objective is the establishment of a distributed Genomics Observatory across the partnership and beyond, of which the data will be available for virtual access (VA). A large part of the data resources arising from JRA1 will come from Ocean Sampling Day ([OSD](#)), currently to include OSD2014, 15, 18, and 19. Data from the ASSEMBLE Plus contribution to the Automated Reef Monitoring Systems project ([ARMS](#)) will also form part of the JRA1 data resources.
- **JRA 2 Cryopreservation of Marine Organisms.** This JRA will address a constraint in the exploitation of marine genetic and biological resources, namely the current paucity of capability to preserve these resources *ex-situ* with a guaranteed genetic, phenotypic and functional stability. The JRA will develop robust, reproducible cryopreservation methodologies for various life-stages of a range of marine macro-organisms and currently cryo-recalcitrant microorganisms. This JRA will collect best current practises and create new protocols from laboratory experiments in the cryo-preservation of marine organisms.
- **JRA 3 Functional Genomics.** This JRA will address the need to establish links between genomic information and phenotypes of marine model species, by developing small-scale functional genomic approaches for several marine models for the generation of Genetically Modified Marine Organisms (GMOs). This JRA will be largely dedicated to transferring established techniques for the generation of genetic resources, and where necessary adapting those techniques, to model organisms for which these techniques have not yet been applied.
- **JRA 4 Development and Standardisation of On-site Instrumentation for Experimental Marine Biology and Ecology.** The aims of this JRA are (i) to produce detailed technical specifications for biological resource centre infrastructure and experimental facilities; (ii) to produce best practise guidelines for future cross-consortium implementation of standardised experimental systems and associated infrastructure. This JRA will collect technical design specifications of experimental systems and associated infrastructure, with the aim of improving the service provision of future instrumentation.



- **JRA 5 Scientific Diving.** The goal of this JRA enable a standardised employment of emerging or breakthrough diving technologies. The aim is to improve diving-based science delivery by improving the use of emerging technologies. The data collected by this JRA will allow the building of a common service and will generate a wider and more diverse user group of this type of data.

Management of the JRA data is described in the DMP (Task NA2.1; D4.2): they are to be archived in the Marine Data Archive ([MDA](#)) and catalogued in the Integrated Marine Information System ([IMIS](#)), and to be included in the ASSEMBLE Plus collection in IMIS. As these data are created within ASSEMBLE Plus, it is expected that the process of making the data FAIR will run smoothly: Findable is the responsibility of the JRAs (to create the metadata records) and VLIZ (to assist in the process and curate the results); Accessible is the responsibility of VLIZ (as the owners of the MDA and IMIS); and Interoperable and Re-useable are the responsibility of the JRAs.

Most of these data will be open access:

- **JRA 1.** All metadata will be open and freely available from the day of creation. Most of the remaining data will be open access at publication. Sensitive data (e.g. endangered species/habitats) may have access restricted to members of the consortium and the EC. All Ocean Sampling Day and ARMS data will be open access.
- **JRA 2,3.** Once the protocols have been developed, these (as refereed publications or ASSEMBLE Plus reports) and the laboratory data leading to them (JRA 2) will be provided with open access.
- **JRA 4.** Basic information on equipment/infrastructure will be open access. Data embargo for part of the data due to the potential creation of a consultancy service within the EMBC infrastructure is being considered.
- **JRA 5.** The final data products (curated images, 3D models, environmental data from sub-tidal buoys) will be provided with open access. Provision of the raw image files with open access is still under discussion because of their large number and file sizes, and their limited re-use potential compared to the curated products

At the time of writing, no JRA data records in IMIS have yet been created, but this is as expected as the data are the result of complex experiments and data-gathering processes. The data from OSD14 (JRA1) have been archived in the MDA, and metadata records are expected to be added to the ASSEMBLE Plus collection before M27.

2.3 The long-term and genomics data resources

At the time of writing, the bulk of the ASSEMBLE Plus data collection consists of data records from the ASSEMBLE Plus marine stations/partners and which were already catalogued in IMIS: these were added in bulk to the ASSEMBLE Plus collection. These records are for data that were collected as part of international and national projects, monitoring observations, data for specific projects (e.g. theses), and so on. The bulk of the datasets were taken from the 1980s onwards, with the oldest being from 1570.

There are currently 503 data records in the ASSEMBLE Plus collection from our marine stations: of these, 183 are defined as long-term data series (defined as having more than 2 years of data collecting activity), of which 161 are specifically long-term ecological data series (aka LTEDS). The



figure below shows the range of topics that are covered by these datasets, as measured by the keywords included in their IMIS data records. For the records concerning biological datasets, the most common topics are: fish, plankton, benthos, ecology and biodiversity, invertebrates, macroalgae. For the records concerning non-biological datasets the most common topics are: water composition, fisheries, physical records (water currents, etc), pollution, dredging, coastal studies. The scientific scope of these datasets is clearly wide.

The data are owned by the marine stations themselves, not by ASSEMBLE Plus, and so there is no requirement that they are open access. About half the data records (and 2/3 of the LTEDS) state that they are open access (e.g. [CC BY](#) or “unrestricted”), however a great deal (~70% of LTEDS) do not actually have a direct download link in the IMIS record. Curation of these data records is therefore an important part of Task NA2.3.

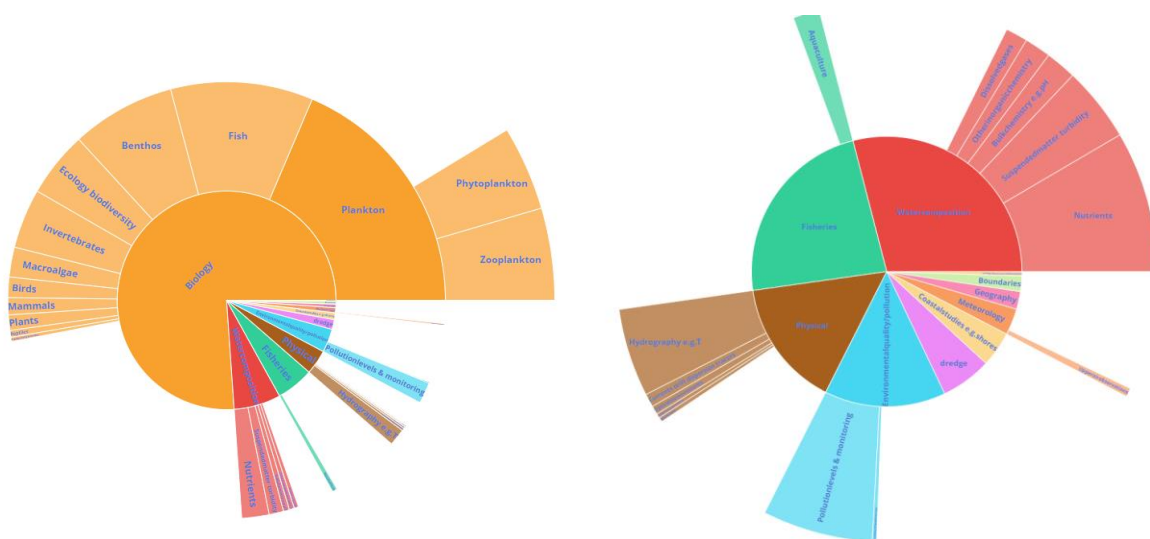


Figure 1 Graphic displaying the topics covered by the ASSEMBLE Plus data collection. Left are the areas covered by all datasets, right is for the non-biological datasets. This interactive figure is provided as part of the virtual open access point (see Sec. 3): clicking on any part of the pie will zoom in on its associated sub-topics (it is otherwise difficult to read the text). The active figure can be found [here](#).

A FAIR Data Management [workshop](#) took place at VLIZ in June 2019. All stations with records in the LTEDS part of the ASSEMBLE Plus collection were invited, as were partners from JRA1 to discuss curation of the OSD and ARMS data. Eventually, 20 representatives from 12 marine stations attended. As an outcome of the workshop, the processes to jointly curate the data records were agreed: VLIZ will guide the process, station by station, while the inputs will come from the stations themselves. Emphasis will first be on the **Findability** (the completeness of the metadata records in IMIS), then on **Accessibility** and **Re-usability** (providing the data download link as part of the record and ideally making most open access). Interoperability will only be addressed for the data that will be included in WP4 Task NA2.5 (*Set up virtual platform for data analysis*). For this final step, assistance from EurOBIS and EMODnet (based on VLIZ) will be enlisted.



2.4 The publications

Publications that are linked to ASSEMBLE Plus are included in the data resources that we provide access to via our website. These publications are collected in the IMIS publications catalogue: depositors are required to send the citation, DOI, and sometimes PDF of the publication to add them to our collection. There are currently only 7 ASSEMBLE Plus publications in the collection, but we have included an additional 234 publications from [Assemble Marine](#) (grant agreement nr. 227799), which was a precursor project to ASSEMBLE Plus and also operated under EMBRC.

An interactive graphic with an overview of the topics included in this collection at present is shown in Fig. 2: marine genomics, environmental impact, biodiversity, climate change, oceanography are the most common topics.

It is a requirement of the TA and JRA programmes that all refereed publications are open access. JRA publications can call on ASSEMBLE Plus resources to pay for this (where absolutely necessary), TA users cannot. It is made clear to TA users that a condition of accepting ASSEMBLE Plus funding is that any refereed publication that is based on their TA data must be open access. In order to broaden the range of journals that TA and JRA researchers can publish in, the ASSEMBLE Plus Open Repository was created: PDFs of the pre-print can be deposited by the authors publishing in so-called “green” access journals, and anyone accessing the publication via our collection can download the PDF to read, but not to distribute. Therefore, the aim is that all ASSEMBLE Plus-related publications accessed via our publications collection can be downloaded directly from the collection. In reality, uptake of adding their publications to our collection and publishing as open access by TA users is slow, and enforcement of these points is difficult.

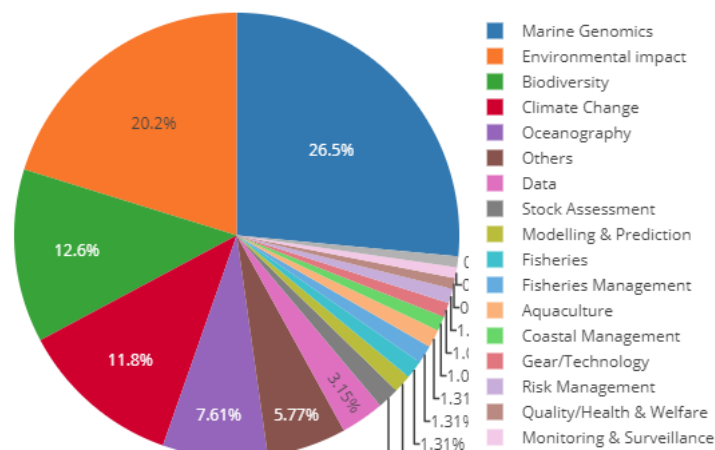


Figure 2 Topics covered by the ASSEMBLE Plus publications collection

3. The virtual open access entry points

The “virtual access entry points” for the ASSEMBLE Plus collections are gathered and introduced on [a page](#) on the ASSEMBLE Plus site. The data resources that are linked on this page are:

- The ASSEMBLE Plus [datasets collection](#) (Task NA2.3, 2.4)
- The [virtual research environments](#) (Task NA2.5)
- The ASSEMBLE Plus [publications collection](#) (Task NA2.2)



- Guidelines for [FAIR data management](#) in ASSEMBLE Plus for the TA users, the JRAs, and the marine stations
- [Internal reports](#)

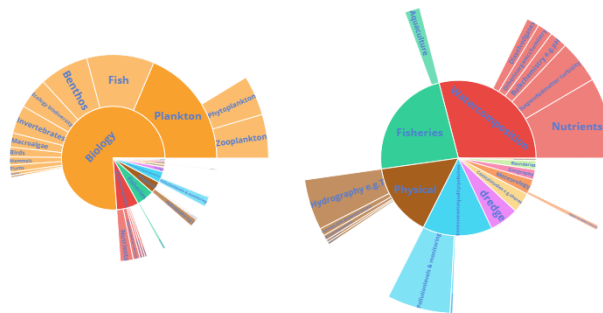
Of interest to this report are the access point to the *datasets collection* and the *publications collection*.

3.1 The datasets access point

The datasets access point consists of a landing page where the ASSEMBLE Plus data collection is summarised, and a link to the page where the data collection can be browsed and read. A screenshot of the landing page is shown in Fig. 3. An overview of the collection is given, including the interactive pie-chart that is described in Fig. 1. Clicking on the orange bar at the top of the page leads to the catalogue browse page, which is shown in Fig. 4.

SEARCH THE ASSEMBLE PLUS DATA COLLECTION →

The geographic scope of our datasets reflects the locations of our marine stations: from the North Sea around Europe to the Mediterranean, and further afield to the Antarctic and Caribbean. The themes covered by our datasets are broad: ecology and biodiversity, water composition, physical parameters, fisheries, pollution, and many more. *Click on the sunburst plots below to see the themes that are included in the collection* (image on the left show the biological themes; image on the right shows all other themes; clicking at the edge of the central circle/slices will zoom in on those themes; clicking on the words will send you to the search page with a search on that keyword).



Datasets from the [marine stations](#) that participate in ASSEMBLE Plus have been gathered together in our [ASSEMBLE Plus data catalogue](#). Some of these datasets are open access, for others you can send an email to the contact address to request access. Two child collections – the Long-Term Ecological Data Series (LTEDS) and the LTEDS: biological – have been created from these collected datasets.

Figure 3 Datasets landing page



Dataset search

Keyword:

Collection:

[FILTER](#) [RESET](#)

507 records found

11 ... Last

Study of the life cycle evolution of the Mediterranean holopelagic medusa *Pelagia noctiluca*: embryonic development and metamorphosis
Citation: Kraus, Y.; Leclère, L.; Department of Evolutionary Biology, Faculty of Biology, Lomonosov Moscow State University (MSU); Russia; Villefranche-sur-mer Developmental Biology Laboratory (LBDV); France; (2019): Study of the life cycle evolution of the Mediterranean holopelagic medusa *Pelagia noctiluca*: embryonic development and metamorphosis. <http://www.assembleplus.eu/information-system?module=dataset&dasid=6310>

[ABSTRACT](#) [MORE INFO](#)

Genetic and experiments results information related to the clades of the colonial ascidian *Botryllus schlosseri* collected from four sites in Roscoff
Citation: Reem, E.; Israel Oceanographic and Limnological Research (IOLR); Israel; (2019): Genetic and experiments results information related to the clades of the colonial ascidian *Botryllus schlosseri* collected from four sites in Roscoff. <http://www.assembleplus.eu/information-system?module=dataset&dasid=6308>

[ABSTRACT](#) [MORE INFO](#)

Figure 4 Datasets collection browse page

Browsing the ASSEMBLE Plus data collection uses an IMIS API: the collection(s) to browse are hardwired in the API, and various search filters can be added. Currently one can filter on the collection (All, Long-term, and Long-term ecological) and free-text keyword. A list of search results is returned, each with a title and 2 links: clicking allows one to read the abstract or to open the IMIS metadata record. This metadata record includes the following information:

- Title, data creator contact details, citation, access rights (e.g. licence)
- Abstract and longer description
- Keywords that describe the scope of the collection, these being entered by the record creator via a drop-down ASFA listing or as free text
- The geographic, temporal, and taxonomic coverage
- Parameters of the data
- Information about the contributing agency, and any other links the data creator provided
- Completion status and information about any related datasets

As the ASSEMBLE Plus collection is quite wide in scope – scientific, temporal, geographic, parameter space, data types – a planned future development is to improve the filtering offered on the search page, and to categorise the search results to allow users to select on a useful range of topics. This is a development that is planned within VLIZ for its IMIS data system, and we will therefore benefit from these developments.

3.2 The publications access point

The publications access point consists of a landing page where the ASSEMBLE Plus publications collection is summarised, and a link to the page where the collection can be browsed and publications can be downloaded. An interactive graphic with an overview of the topics included in



this collection is given on the landing page, as was shown in Fig. 2. Clicking on a link on the landing page leads to the catalogue browse page, which is shown in Fig. 5.

Publication search

243 records found

1 2 3 4 5

Achilles-Day, U.E.M.; Day, J.G. (2013). Isolation of clonal cultures of endosymbiotic green algae from their ciliate hosts. J. microbiol. methods 92(3): 355-357. <https://hdl.handle.net/10.1016/j.mimet.2013.01.007> MORE INFO

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

Figure 5 Browse the catalogue of ASSEMBLE Plus publication

Browsing the ASSEMBLE Plus publications collection also uses an IMIS API: the collection(s) to browse are hardwired in the API, and a set of search filters can be added. Currently one can filter on the collection (All, ASSEMBLE Plus, Assemble Marine), KO type (publication, book, case study, modelling/software, prototype, exploitable result, services) and free-text keyword. A list of search results is returned, with the title of each publications displayed and one or two links. Clicking allows to read the IMIS record or (for open access publications) to directly download the PDF. The IMIS record includes:

- Title, author, DOI
- Access constraints, link
- Author-entered keyword
- Author list

A planned future development is to improve the filtering offered on the search page, to allow a filtered search in TA and JRA parts of the collection. We may also benefit from any VLIZ developments on their API, e.g. to allow a better categorisation of the search results.

4. Hits to the virtual open access entry points

This datasets and publications access pages have been available since March 2019 (M18). The number of unique visits to the landing pages since then are 60 and 27, respectively. Europe forms the largest origin of visits, about 15% come from outside Europe. The number of hits to our data records since then is 333, with however 1/3 of these hits occurring during and after the FAIR data management workshop and probably arising from people working on their own stations's data records. For the publication records, the number of hits is 103.



It should be taken into account that these access pages are not linked to any other pages on the ASSEMBLE Plus website (although they are linked in the main menu), and it is expected that the number of hits to our collections will increase once: (1) the JRAs add their datasets and publications, (2) the JRA stories are added to the website (expected by M27): where the JRA developments are explained in non-technical terms and links to associated data and publications are included, (3) links to the TA datasets and publications are added to the “TA success stories” on the website.

