



# **Deliverable D2.3**

## **Integration of programmatic web interface into BRC websites**

**Date: May 31, 2019**



**HORIZON 2020 - INFRADEV**  
**Implementation and operation of cross-cutting services and solutions**  
**for clusters of ESFRI**

**Grant Agreement number: 654008**

**Project acronym: EMBRIC**

**Contract start date: 01/06/2015**

**Project website address: [www.embric.eu](http://www.embric.eu)**

**Due date of deliverable: 31/05/2019/ month 48**

**Dissemination level: Public**

## Document properties

<b>Partner responsible</b>	MBA
<b>Author(s)/editor(s)</b>	Guy Cochrane, Peter Harrison, Annalisa Milano, Nils P Willassen and Adelino Canário
<b>Version</b>	1

## Abstract

Marine bioresources are fundamental for scientific research and biotechnological R&D and therefore having access to the available information related to a particular organism will increase the probability of successful research and reduce costs. This deliverable documents in the form of a “point of entry” guide taxonomic- and collection-related through searches across bioinformatics data resources typical of marine. The intention is that this documentation be embedded (dynamically if possible) into the web sites operated by marine bioresource collections.

# Table of Contents

---

<b>Table of Contents.....</b>	<b>4</b>
<b>1 Introduction.....</b>	<b>5</b>
<b>2 Accessing marine bioresource-related data from collection web sites .....</b>	<b>6</b>
<b>2.1 Documentation to be included in Biological Resource Centre websites.....</b>	<b>6</b>

# 1 Introduction

---

Marine bioresources are fundamental for scientific research and biotechnological R&D and therefore having access to the available information related to a particular organism will increase the probability of successful research and reduce costs.

Task 2.1 had as objective to map existing data resources of Biological Research Centers (BRCs) in EMBRIC, evaluate to the extent to which these should be linked, and review external data resources that could be linked to these. The aim was to identify options for user-friendly searching of the linked data resources and for presenting search results. The final aim was to integrate the programmatic service interface developed in WP4 into websites of contributing BRCs.

Towards the objectives, partner DMSZ led development of “BacDive” (<https://bacdive.dsmz.de/>), a database providing organism-linked information of bacterial biodiversity - taxonomic classification, morphology, physiology, cultivation, origin and natural habitat for more than 50,000 bacterial and archaeal strains.

WP2 in collaboration with EMBL-EBI has taken initial discussion with people in algal collections, the Coimbra collection of algae (ACOI) and the Culture Collection of Algae and Protozoa (CCAP), to recruit cases studies to develop search services around the WP2 programmatic service (see D4.3 report). However, for multiple reasons this was not possible in reasonable time. The alternative was agreed that a set of taxonomic- and collection-related searches, typical of those that might be asked by a user of marine bioresources, be documented centrally. This documentation has been provided in a central repository (<https://github.com/enasequence/dcat-biological-resource-centres-docs>), such that the developers of bioresource collection web sites can rapidly, and in a synchronised way, present the content to their users.

## 2 Accessing marine bioresource-related data from collection web sites

---

Taxonomic- and collection-related searches across bioinformatics data resources and typical of marine users have been documented in the form of a “point of entry” guide. The intention is that this documentation be embedded (dynamically if possible) into the web sites operated by marine bioresource collections.

The documentation has been made publicly available in Github (<https://github.com/enasequence/dcat-biological-resource-centres-docs>) in “markdown” format (<https://github.com/enasequence/dcat-biological-resource-centres-docs/blob/master/docs/Access%20to%20bioinformatics%20data%20for%20users%20of%20Biological%20Resource%20Centres.md>) for presentation in bioresource collection web sites and in PDF (<https://github.com/enasequence/dcat-biological-resource-centres-docs/blob/master/docs/Access%20to%20bioinformatics%20data%20for%20users%20of%20Biological%20Resource%20Centres.pdf>) for direct reading.

The current version of the documentation is also provided in this report.

We finish our work on this deliverable with an invitation to collection web sites to embed the documentation on their respective sites.

### 2.1 Documentation to be included in Biological Resource Centre websites

---

The following text is taken from the central documentation at <https://github.com/enasequence/dcat-biological-resource-centres-docs>

## Access to bioinformatics data for users of Biological Resource Centres

- **Access a summary of all DNA and RNA sequences based on a taxonomic name** is provided from the [European Nucleotide Archive](#): starting from the [Advanced Search page](#), select the “Taxon” domain and enter a taxonomic name at any rank, leaving the checking the “include subordinate taxa” box unchecked. Browse through to view the result and select the “Portal” tab if it is not already showing. This provides a summary of all held sequence data with clickable links to browse further into these data. See the example results for family *Prochloraceae* [here](#) and the example of the query search [here](#).
- **List assembled/annotated sequences linked to a given culture collection:** using the [cross-reference search](#) of the [European Nucleotide Archive](#), select a collection of interest, such as the Culture Collection of Algae and Protozoa (CCAP), from the “xref” pull down menu and choose “sequence” from the “target” pull down menu.

- **List proteins linked to a given organism:** using the Universal Protein Resource (UniProt), go to the home page [here](#). Go to the search bar in the UniProt banner at the top of the page and click on the dropdown to the left of the search box and select your dataset. The default is UniProtKB. Click on 'advanced' to the right of the search box and select term type "organism" and enter "*Pardachirus*", then hit the search button. This provides a summary of all held proteins isolated from the genus *Pardachirus* with clickable links to further browse these data, including different *Pardachirus* species such as *Pardachirus pavoninus* and *Pardachirus marmoratus*. See the example result for *Pardachirus* [here](#).
- **Access all genomes from marine microbial species** is provided from the MarRef databases. The MAR databases are [MarRef](#), [MarDB](#) and [MarCat](#). Starting from the [MAR databases](#) home page, select one of the three databases to search for specific genomes or attributes, or [BLAST](#) nucleotide or protein sequences against the content in the MAR databases. The MAR databases, contextual and/or sequence data resources can be downloaded from the MAR Download page which can be found [here](#).
- **List marine microbes based upon taxonomy:** from [MarDB starting page](#) or [MarRef starting page](#) select "Browse" and by using the filter panel on the right side of the page select the taxa of interest on a phylum, family or genus. Browse through the results and select the entry of interest.
- Select one specific microbe: from the from [MarDB starting page](#) or [MarRef starting page](#) and select "Browse" - type the microbe of interest in the "Search" box e.g. *Verrucosipora maris* and the contextual data for the microbe will be provided. Activate the entry by clicking on the entry name either under "MMP ID", "full scientific name", "strain", "type strain", "geographic location" or "collection date". One will get access to all available information for the microbe such as taxon, isolate, host and pathogenicity, assembly, annotation, phenotype and secondary metabolites. See an example results for *Verrucosipora maris* [here](#).
- **List all marine microbes with compound entries in ChEBI or ChEMBL:** from the Browse page in [MarDB](#) and [MarRef](#) enter ChEBI or ChEMBL in the "Search" box on the left side. The results is a list of genomes with links either to ChEBI or ChEMBL and the example results can be seen [here](#).
- **View compound predicted secondary metabolites pathways in a particular genome:** from the genome page e.g. [Verrucosipora maris](#) click on the "Secondary metabolites" curtain. Information about AntiSMASH secondary metabolite clusters and types will be available. Link to the result page is available [here](#).