



OceanEcology

Development of MEDIN Compliant PSA Export Functionality within ABACUS

Ref: OEL_MEDABA0919_TCR

Prepared for



Funding Body:	MEDIN
Contact:	Clare Postlethwaite
Organisation:	OceanEcology Limited
Report Title:	Development of MEDIN Compliant Export Functionality within ABACUS
Report Number:	OEL_MEDABA0718_TCR
Recommended Citation:	Ocean Ecology Limited (2019). Development of MEDIN Compliant Particle Size Analysis Export Functionality within ABACUS. Report No. OEL_MEDABA0919_TCR, 14 pp

Version	Date	Description	Author(s)	Reviewer(s)
01	03/10/2019	Draft	Joseph Turner, Elliot Carter	Ross Griffin

CONTENTS

1. INTRODUCTION	4
2. METHODS	6
2.1. Development Process	6
2.2. Import Functionality	6
2.3. Sample Recording	7
2.4. Exports	8

SUMMARY

Over the last two years Marine Environmental Data and Information Network (MEDIN) partner Ocean Ecology Limited (OEL) have been developing a novel web-based tool (ABACUS) in partnership with Peninsula Data Solutions (PDS) for recording, managing and archiving the large volume of marine sample data generated by OEL's NE Atlantic Marine Biological Quality Control (NMBAQC) scheme participating laboratory.

In 2018, a collaboration between OEL, PDS and The Archive for Marine Species and Habitats Data (DASSH) was successful in applying for funding through MEDIN's Small Data Archiving Project initiative to develop a MEDIN compliant macrobenthic data export functionality within ABACUS with the aim of increasing the flow of data into MEDIN. This project was successfully completed in 2018 (report available [here](#)) and now allows for DASSH validated exports of macrobenthic data to be created and downloaded with just a few clicks, skipping error-prone and slow manual processes. These exports are created as excel sheets aligned to the most up to date MEDIN Grab and Core Data Guidelines and include all general and detailed metadata populated using SeaDataNet, ICES, and BODC other controlled vocabularies.

In 2019, a collaboration between OEL, PDS and DASSH was successful in obtaining further funding through MEDIN's Small Data Archiving Project to development this export functionality further. This second phase of development has focused on adding the ability to export MEDIN compliant particle size analysis (PSA) data either independent of or alongside corresponding macrobenthic data. This work included a significant update to the export functionality to accommodate changes made to the MEDIN Grab and Core Data Guideline templates in June 2019 alongside extensive testing and validation through liaison with DASSH using real-world survey datasets. ABACUS has also more recently benefited from the addition of a batch import functionality combined with a validation tool meaning that a variety of data types can now be imported in bulk into the database including macrobenthic and PSA datasets, metadata (e.g. controlled vocabularies), stations and samples.

1. INTRODUCTION

Many thousands of marine samples are collected and analysed on an annual basis to satisfy statutory monitoring commitments (e.g. Water Framework Directive (WFD), Habitats Directive) and conditions of marine licences granted for marine activities. These include seabed samples that undergo macrobenthic, particle size distribution (PDS) and chemical analysis, water samples analysed to monitor planktonic communities and scientific trawl samples to assess fish and other mobile species. Despite the requirement for these analytical processes to be conducted by laboratories participating in recognised quality control schemes (e.g. the NE Atlantic Marine Biological Quality Control (NMBAQC) scheme), there remains fundamental issues surrounding the recording of non-standardised marine biological data. These issues stem from inter-analyst and inter-laboratory variability in sample analysis methodologies, recording practices, species naming, the use of taxonomic qualifiers and so on. This is thought to be a contributing factor for wide spread mis-interpretation of trends in marine biological communities which, in some cases, can have serious consequences for both Statutory Nature Conservation Bodies (SNCBs) and private sector organisations legally obligated to conduct robust marine ecological monitoring.

To address this, Ocean Ecology Limited (OEL) have developed the web-based data management application 'ABACUS' that has been established to act as a platform for marine scientists to record, quality assure, store and export standardised marine sample data in line with internationally recognised data standards (e.g. the Marine Environmental Data and Information Network (MEDIN), GEMINI, ISO) (see Figure 1). In 2018, a collaboration between OEL, PDS and The Archive for Marine Species and Habitats Data (DASSH) was successful in applying for funding through MEDIN's Small Data Archiving Project initiative to develop a MEDIN compliant macrobenthic data export functionality within ABACUS with the aim of increasing the flow of data into MEDIN. This project was successfully completed in 2018 (report available here) and now allows for DASSH validated exports of macrobenthic data to be created and downloaded with just a few clicks, skipping error-prone and slow manual processes. These exports are created as excel sheets aligned to the most up to date MEDIN Grab and Core Data Guidelines and include all general and detailed metadata populated using NERC, SeaDataNet, ICES, BODC and other controlled vocabularies. The exports generated also include a cover page describing how the data has been generated and how it can easily be archived with DASSH for safe keeping and dissemination (if permitted).

As standard practice, OEL collect multiple sediment samples at each sampling station visited during its seabed sampling surveys which are retained for subsequent laboratory analysis. Normally, up to three of these samples undergo full macrobenthic analysis to produce the datasets discussed above. The fourth (but sometimes sub-samples of one or all of the macrobenthic samples) undergo Particle Size Distribution (PSD) analysis to a) understand the relationship between sediments and the associated macrobenthos and b) to establish sediment type for subsequent habitat/biotope mapping. These samples are processed at OEL's laboratory using both dry sieving and laser diffraction techniques and are, on occasion, also analysed to establish Total Organic Carbon (TOC). As PSD analysis can be used to inform marine licence applications and some projects comprise of large numbers of samples, it is particularly important to record the resulting data in a standardised manner to ensure consistency and comparability between subsequent surveys and existing datasets. As such, additional functionality for both recording and exporting PSA data has been added to ABACUS to standardise, streamline and better manage the PSD analysis process. This development allows PSD data produced to be exported in a MEDIN compliant format for rapid archiving with DASSH (and/or other DACs) alongside the biological data. This work, described in detailed in the following sections, has included a significant update to the export functionality to accommodate changes made to the MEDIN Grab and Core Data Guideline templates in June 2019 alongside extensive testing and validation through liaison with DASSH using real-world survey datasets. ABACUS has also more recently benefited from the addition of a batch import functionality combined with a validation tool meaning that a variety of data types can now be imported in bulk into the database including macrobenthic and PSA datasets, metadata (e.g. controlled vocabularies), stations and samples.

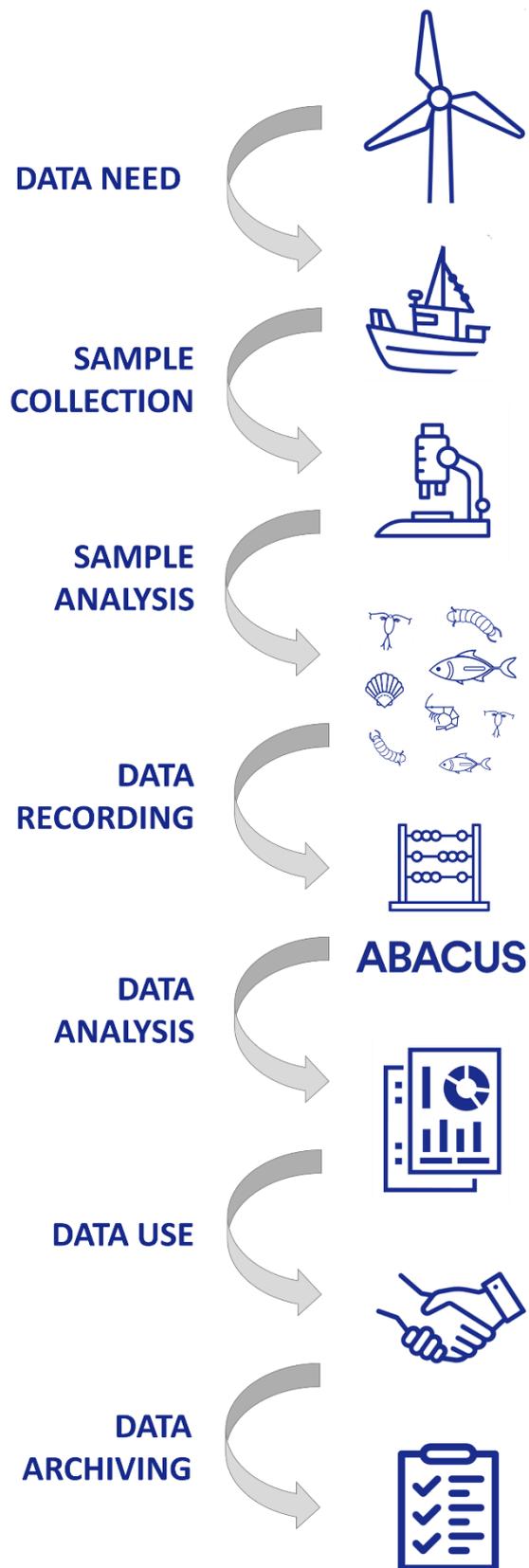


Figure 1. The role ABACUS plays in OEL's data management process and work flow.

2. METHODS

2.1. Development Process

ABACUS has been developed using the latest Microsoft technologies (ASP.NET CORE, MVC, C#, Microsoft Azure SQL Database) and is encrypted using industry standard SSL and HTTPS. The PSA functionality was developed with the intention of keeping the analysis process as simple as possible for users while maintaining visibility of data at each stage of the process. This included developing functionality to allow users to copy and paste laser diffraction outputs (from Excel files) directly into an ABACUS webpage which automatically converts and parses the results drawing attention to any missing values and eliminating the need to perform error-prone transcription tasks. Maintaining visibility of data at each stage of the process has been achieved by clearly presenting data at all calculation stages, such as laser normalisation, as web pages. Calculations are initiated by a viewing the web page which provides transparency on the results of the calculations and acts as a quality control feature. Once all calculation stages are marked as complete the sample data can be exported. A mapping table exists in the database to map particle sizes to classes in several classification scales (Wentworth, Phi, Half Phi). This allows the same result data from samples to be simultaneously exported in three different classifications providing consistent results as well automatically calculating other attributes such as Folk class.

As part of the development of the updated export functionality, a number of demonstration versions were tested by sediment scientists at OEL's laboratory using real-world samples that allowed for the process to be fine-tuned and optimised. A variety of other developments are underway with the aim of making the ABACUS application available to other organisations through a licencing model in 2020. As part of this, the development team have pledged to ring-fence a percentage of any licence fees for OEL/PDS to join MEDIN's consortium of sponsors in recognition of the integral role the MEDIN data standards have played in the development of ABACUS' database architecture. Furthermore, a percentage of licence fees will also be used as donations to the World Register of Marine Species (WoRMS) in recognition of the criterial role that the WoRMS webservice plays in recording of standardised macrobenthic species data.

2.2. Import Functionality

It is now also possible to add data to ABACUS in bulk using the imports section of the application. This is achieved by creating a new import batch. An import batch is a collection of rows/records (e.g. a batch of 20 stations for a project) to be imported into a specific area of the website (e.g. stations). Currently import features exist for six areas of the application; Metadata¹, Stations, Samples, Chemical measurements, Macrobenthic matrices and PSA datasets. Currently there are two methods to begin a new import batch; 'Generate Blank Rows' and 'Load from File' (using a provided excel template). The general sequence of steps for both methods is as follows:

1. Start a new batch, using one of two methods
 - a. Generating a specified number of blank rows
 - b. Loading data from an Excel file
 - i. Reviewing log of file load attempt
2. Bulk edit (multi row/record editing)
3. Validation
4. Import

¹ Populated using NERC, SeaDataNet, ICES, BODC and other controlled vocabularies

The addition of this bulk import functionality now allows users to rapidly import large volumes of validated data into ABACUS at once vastly reducing the time required to set up new surveys / projects. A detailed guidance document has been compiled to guide users through the process which will be made available within the ABACUS application.

2.3. Sample Recording

Like macrobenthic samples, PSA samples can now be tracked through the key processing stages including log in, addition of sample information, empty sieve data entry, raw dry sieve data entry, raw laser upload data and laser merging (see Figure 2). Once recorded and merged (if required) the PSA outputs require a series of validation steps to be completed before finalisation to sense check the calculations, workings, normalisation and data merging.

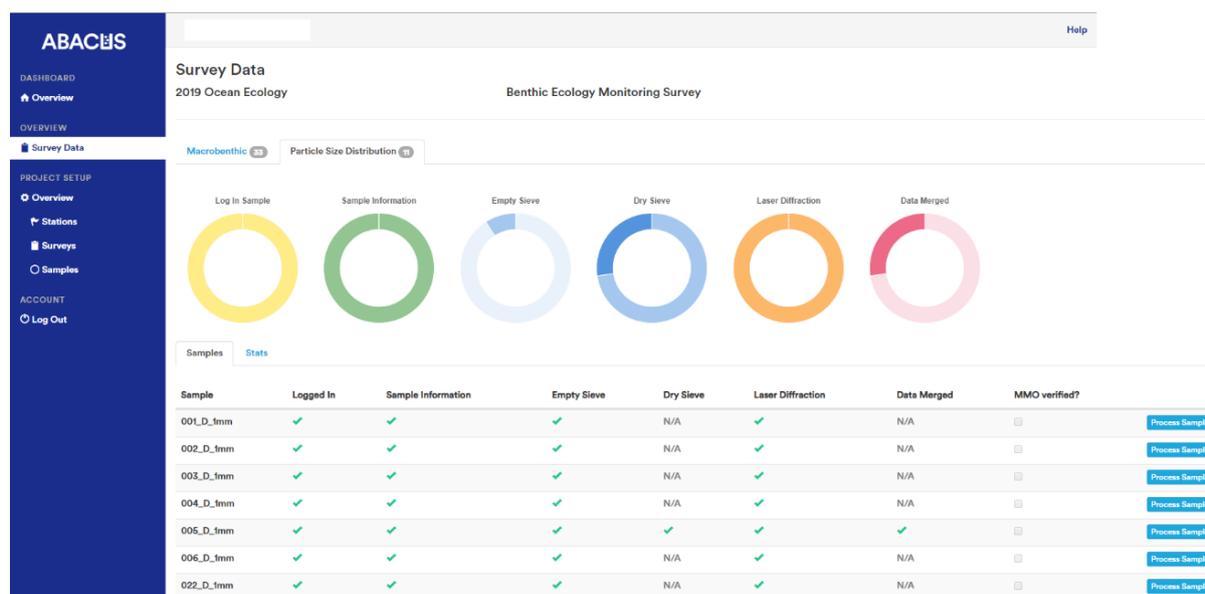


Figure 2. PSA sample tracking page within ABACUS.

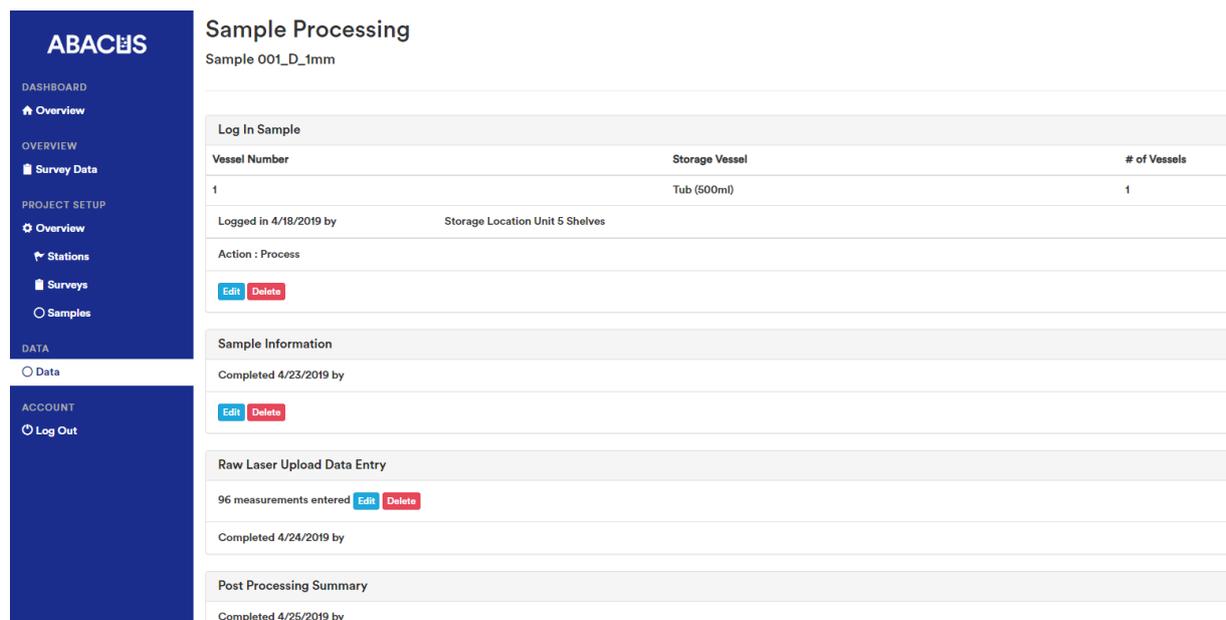
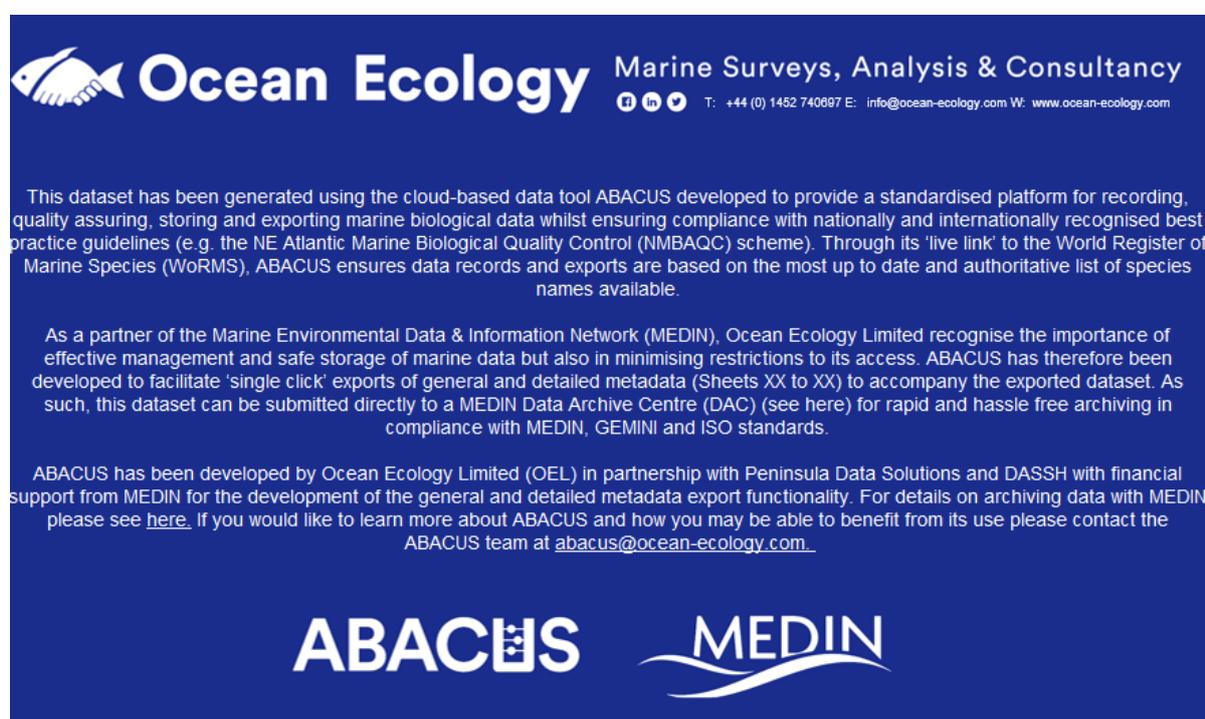


Figure 3. PSA sample processing page within ABACUS.

2.4. Exports

As with macrobenthic sample data, DASSH validated PSA data exports can now be created and downloaded with just a few clicks, skipping error-prone and slow manual processes (see Figure 5). As the exports are aligned to the latest version of the MEDIN Grab and Core Data Guideline template this includes all corresponding general and detailed metadata and any corresponding macrobenthic data (if also recorded).

OEL do not always hold ownership of the data outputs its laboratory produces and is therefore frequently required to submit the datasets to its clients who may not feel it necessary to archive its data with DASSH (or other DACs). To simplify this process for the data owners, the exports generated from ABACUS include a cover page describing how the data has been produced and how it can easily be archived with DASSH for safe keeping and dissemination (if permitted) (see Figure 4). Furthermore, a Data Permission Agreement for Archiving and Dissemination has also been compiled based on that used by DASSH which OEL intend to disseminate to its clients when submitting its laboratory outputs (Appendix 1). This will act to quickly establish which PSA and macrobenthic datasets can be archived with DASSH (and other DACs) immediately and which may require restrictions in use or embargos to be imposed if archived.



Ocean Ecology Marine Surveys, Analysis & Consultancy
T: +44 (0) 1452 740897 E: info@ocean-ecology.com W: www.ocean-ecology.com

This dataset has been generated using the cloud-based data tool ABACUS developed to provide a standardised platform for recording, quality assuring, storing and exporting marine biological data whilst ensuring compliance with nationally and internationally recognised best practice guidelines (e.g. the NE Atlantic Marine Biological Quality Control (NMBAQC) scheme). Through its 'live link' to the World Register of Marine Species (WoRMS), ABACUS ensures data records and exports are based on the most up to date and authoritative list of species names available.

As a partner of the Marine Environmental Data & Information Network (MEDIN), Ocean Ecology Limited recognise the importance of effective management and safe storage of marine data but also in minimising restrictions to its access. ABACUS has therefore been developed to facilitate 'single click' exports of general and detailed metadata (Sheets XX to XX) to accompany the exported dataset. As such, this dataset can be submitted directly to a MEDIN Data Archive Centre (DAC) (see here) for rapid and hassle free archiving in compliance with MEDIN, GEMINI and ISO standards.

ABACUS has been developed by Ocean Ecology Limited (OEL) in partnership with Peninsula Data Solutions and DASSH with financial support from MEDIN for the development of the general and detailed metadata export functionality. For details on archiving data with MEDIN please see [here](#). If you would like to learn more about ABACUS and how you may be able to benefit from its use please contact the ABACUS team at abacus@ocean-ecology.com.

ABACUS **MEDIN**

Figure 4. Preamble provided with all MEDIN compliant data exports from ABACUS describing how the data was generated and the process in which it can be easily archived with a MEDIN Data Archive Centre (DAC).

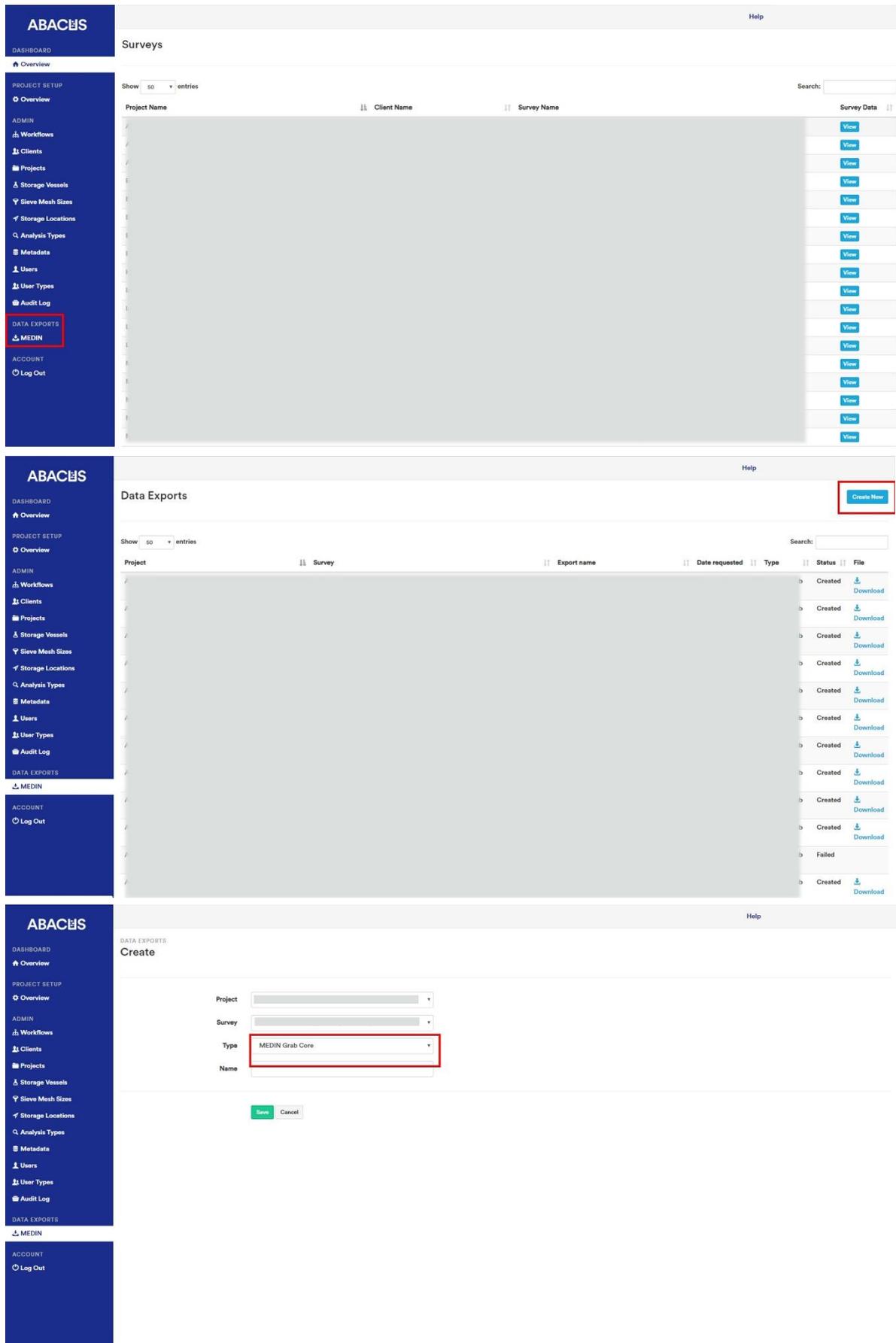


Figure 5. MEDIN export process in ABACUS. Redacted (grey areas) for confidentially purposes.

Appendix 1



Data permissions agreement for archiving and dissemination of data

OceanEcology Limited (OEL) are a partner of the Marine Environmental Data and Information Network (MEDIN) and recognises the importance of the effective management and safe storage of marine data. In order to help facilitate this, OEL have developed a cloud-based tool ('ABACUS') for recording and storing data generated during our sample analysis. A key criterion was that this facilitated hassle free archiving of data in line with internationally recognised data standards (e.g. MEDIN). ABACUS can provide standardised, 'single click' exports of general and detailed metadata to accompany the exported dataset. This can then be submitted to a MEDIN Data Archive Centre (DAC) for rapid and hassle-free archiving in compliance with MEDIN, GEMINI and, ISO standards.

This form gives OEL permission to archive data with a DAC (e.g. The Archive for Marine Species and Habitats Data (DASSH)) following the end of the project or following a certain time period once the project ends.

STORAGE / ARCHIVING

Data/images submitted to a DAC will be archived according to MEDIN templates. Nothing in this agreement constitutes a transfer of Intellectual Property Rights.

The DAC will hold and process data/images for the following purposes:

1. To provide a digital archive for marine biological data of both species and habitats.
2. To provide a digital repository for marine biological imagery (still and video).
3. To transfer marine biological data into a standard format and progress the data to the DASSH archive.
4. For the purposes of informing government decision making and policy development.

PUBLICATION

OEL and DACs are committed to making marine biological data Findable, Accessible, Interoperable and Reusable (FAIR) and as such will publish data in standard formats online. Depending on the level of permission from the data provider the DAC will:

1. Provide discovery, view and download facilities for datasets via an online catalogue of both metadata and data.
2. Create and publish metadata to the Marine Environmental Data and Information Network (MEDIN) Metadata Discovery Portal, and
3. Progress metadata and data in standard format to other biodiversity data aggregators including the National Biodiversity Network (NBN), Ocean Biogeographic Information System (OBIS) and the Global Biodiversity Information Facility (GBIF)

All data published to third parties will include the name of the original data provider and be shared either using the Open Government Licence² or the Creative Commons CC-BY license³ meaning any reuse of the data must include acknowledgement of the original data provider.

DATA PROTECTION

The collection and storage of personal data, as defined in the Data Protection Act (1998), the General Data Protection Regulations (GDPR) and any subsequent amendments, needs explicit written consent from the subject of that data. This information will be held and processed by the DAC only for the following purposes:

1. To maintain the contact details of data providers, owners of intellectual property rights and original recorders of records & where appropriate display their name(s) alongside their data.
2. To comply with international metadata standards for the collection of data.

Consent to store personal data can be amended or withdrawn at any time by email.

AGREEMENT

a) I confirm that I consent (on behalf of myself / my organization) to the storage and display of the data specified in the Appendix (overleaf) for the purposes set out in the statements above.

Please check applicable statement:

- Metadata and data archived by the DAC, published on the DAC website, and made available to data aggregators **under Open Government License**.
- Metadata and data archived by DAC, published on the DAC website, and made available to data aggregators **under CC-BY License**.
- Metadata archived and disseminated on the MEDIN metadata portal. Data archived by DAC **but not disseminated**.

Embargo

If you would like OEL and the DAC to hold the release of your data, please indicate below the date from which you would like your data released. If this field is not filled in the dataset(s) then no embargo will be applied.

I would like my data to be made available from this date:

b) I agree to the DAC recording and processing personal information (name and contact details) about myself / my organization.

I understand that this information will be used only for the purposes set out in the statement above and my consent is conditional upon the DAC complying with its duties and obligations under the Data Protection Act and General Data Protection Regulations.

² <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

³ <https://creativecommons.org/licenses/by/3.0/>

Name [please print]

Email

Organization [if applicable]

Signature

Date