Epibiota Quality Assurance Framework Guidance Document



Contents

[Epibiota Quality Assurance Framework Guidance 1](#_Toc67902672)

[1. Project Background 2](#_Toc67902673)

[2. Epibiota Standards 2](#_Toc67902674)

[a) Epifaunal Imagery Purposes flowchart 2](#_Toc67902675)

[3. Quality Assurance Framework (QAF) workflow overview 3](#_Toc67902676)

[4. Epibiota proforma standards 4](#_Toc67902677)

[a) Species/ Habitat Verification Proforma 4](#_Toc67902678)

[b) Habitat Mapping Proforma 5](#_Toc67902679)

[c) Monitoring Proformas 5](#_Toc67902680)

[5. Guidance for how to use the proformas 6](#_Toc67902681)

[a) Field Conditions 6](#_Toc67902682)

[b) Usage guidelines 6](#_Toc67902683)

[6. Quality Assurance Framework Tools 7](#_Toc67902684)

[a) Full QA 7](#_Toc67902685)

[b) Dataset Comparison Tool 9](#_Toc67902686)

## Project Background

The Quality Assurance Framework project has been set up through the Big Picture Group to help standardise the collection of epifaunal imagery data. The aim of this is to improve consistency between data collected from different sources for more common use and sharing in the future.

The first step to achieve standardisation was to compile information about all of the different uses and applications of marine imagery data across all the different sectors. This information was attained by the collection of information on the rationale and approach taken by a large variety of organisations for imagery data. This was then followed by more in-depth interviews with members to ascertain the details about their data collection, use and applications.

Once common uses or “purposes” were ascertained the next stage of the project was to create tools to help standardised approaches be undertaken for the analysis of benthic imagery. The purposes identified had varied collection approaches but similar analysis methods. The focus of this project is therefore on standardising analysis outputs rather than the collection methodology. This is more effective and useful as each different acquisition method is already designed to collect data most efficiently for the specific purpose for which the data is collected.

This guidance document has been produced to inform use of the following QAF products:

* Epibiota proformas
* QAF form checks
* Comparison tool

A separate guidance document is available for the Epifaunal Identification Protocol (EIP).

## Epibiota Standards

A Standard Selection flow chart has been created to readily identify the correct standard to use for your imagery data (see below).

### Epifaunal Imagery Purposes flowchart

Figure 1 details the 6 primary standards that have been defined for the different applications and usage of marine benthic imagery. Note use of imagery for creation and groundtruthing of habitat maps falls within the Full Seabed Characterisation purpose.

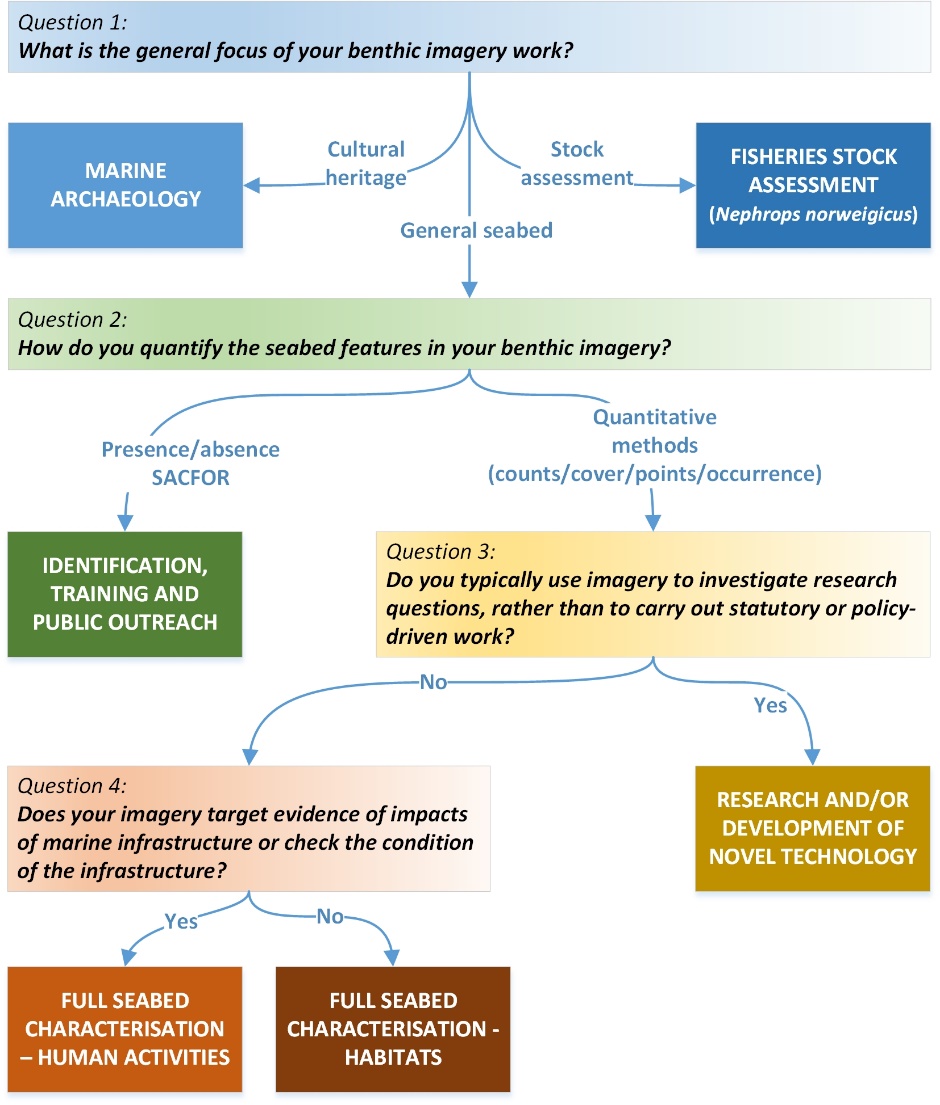


Figure 1 Purposes flowchart detailing which imagery standard is most appropriate for different uses and approaches for imagery data collection.

## Quality Assurance Framework (QAF) workflow overview

To build on the creation of Imagery purposes, three standards have been created based on three purposes for this work;

1. Species/ Habitat Verification (previously Full seabed characterisation – Human Activities)
2. Habitat Mapping (a derivative of Full Seabed Characterisation – Habitats)
3. Habitat Monitoring (a derivative of Full Seabed Characterisation – Habitats)

The quality assurance framework is built around these three standards and complies with MEDIN data standards. Four stages of data analysis and submission have been identified. Figure 2 details these and highlights the QAF products and quality checks applicable to each stage.

The four stages are:

1. **Selection of Imagery Standard**
2. **Selection of Proforma**
3. **Data Checks on Proformas** 
   1. Taxa data checks within the EIP
   2. External Data Verification
4. **Final Data checks and sign off.**

A picture containing table

Description automatically generated

Figure 2 Quality Assurance Framework stages, products and checks.

## Epibiota proforma standards

Please use the flowchart below to identify the most appropriate imagery proforma for your imagery. One proforma should be populated per survey or cruise.

More details on each standard are given in the corresponding standards sections below.

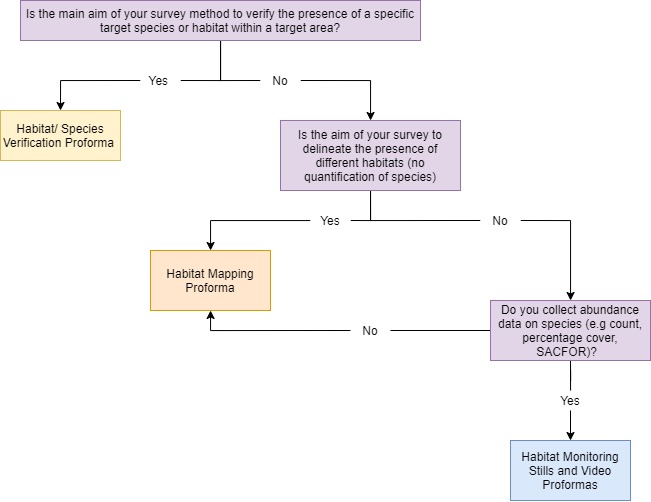


Figure 3 Standard Selection Flowchart to identify the most applicable Proforma for the imagery collected.

### Species/ Habitat Verification Proforma

This proforma is used where only presence/absence data is recorded for species or habitats to for example verify the presence of a protected habitat or species within a cable laying route. This proforma is **Video data only**. The proforma is made up of the following sheets;

* **Read Me** – Spreadsheet usage information and guidance.
* **Pre-Survey Metadata** – High level cruise and gear metadata for the survey data that has been collected. This sheet should contain a separate column for each imagery gear type used on the survey. This is high level metadata on the whole cruise.
* **Video Analysis Form** – Form for collating the information on the locational information of segments where target species are present.
* ***Hidden Tab* LookUp\_Tables** – A hidden sheet that contains all of the drop-drown lists within the spreadsheet.

### Habitat Mapping Proforma

This proforma is used to collect information on the presence of different habitat types from both still and video imagery data. The proforma should be used for habitat mapping/groundtruthing, habitat classification and habitat segmentation. Information on a variety of different habitat classifications such as OSPAR, VME and Annex 1 habitats is collected. This proforma contains the following sheets;

* **Read Me** – Spreadsheet usage information and guidance.
* **Pre-Survey Metadata** - High level cruise and gear metadata for the survey data that has been collected. This sheet should contain a separate column for each imagery gear type used on the survey. This is high level metadata on the whole cruise.
* **Stills Analysis Form** – Sheet to collect image specific information on habitats alongside locational information.
* **Video Analysis Form** – Form for collecting information on differing habitat segments and habitat delineation alongside locational information for these segments.
* ***Hidden Tab* LookUp\_Tables** – A hidden sheet that contains all of the drop-drown lists within the spreadsheet.

### Habitat Monitoring Proformas

This standard has two proformas; a stills analysis proforma and a video analysis proforma. These proformas should be used where you collect quantitative data from imagery such as taxa count or percentage cover data. These proformas are more detailed and as such have more fields than the other standards. Some of the fields may not be applicable in all situations; yellow highlighting of fields in the proformas indicate they are conditional or optional requirements.

#### Stills Analysis Proforma

This proforma contains the following sheets;

* **Read Me** – Spreadsheet usage information and guidance.
* **Pre-Survey Metadata** - High level cruise and gear metadata for the survey data that has been collected. This sheet should contain a separate column for each imagery gear type used on the survey. This is high level metadata on the whole cruise.
* **Stills Analysis Form** – Sheet to collect image specific information on habitats alongside locational information.
* **Stills Fauna Abundance Count** – This sheet contains a species matrix for count data. This matrix contains WoRMS fields and Aphia ID fields alongside morphological identification fields and VME Indicator fields.
* **Stills Abundance (%, SACFOR)** - This sheet contains a species matrix for percentage cover or SACFOR data. This matrix contains WoRMS fields and Aphia ID fields alongside morphological identification fields and VME Indicator fields.

#### Video Analysis Proforma

This proforma contains the following sheets;

* **Read Me** – Spreadsheet usage information and guidance.
* **Pre-Survey Metadata** - High level cruise and gear metadata for the survey data that has been collected. This sheet should contain a separate column for each imagery gear type used on the survey. This is high level metadata on the whole cruise.
* **Video Analysis Form –** This sheet contains video segment information
* **Video Fauna Counts –** This sheet contains a species matrix for count data. Every column should represent a habitat segment. This matrix contains WoRMS fields and Aphia ID fields alongside morphological identification fields and VME Indicator Fields.
* **Video SACFOR –** This sheet contains a species matrix for percentage cover or SACFOR data. Every column should represent a habitat segment. This matrix contains WoRMS fields and Aphia ID fields alongside morphological identification fields and VME Indicator Fields.

## Guidance for how to use the proformas

### Field Conditions

All the proformas have 3 different categories of fields within them. These different categories are highlighted in different colours throughout the proformas. The colour schemes and categories are listed below. It is important to ensure the conditions of all the fields are met ahead of any submission of data into the QAF tools.

Table 1: Details of the conditions of the fields within the proformas

|  |  |
| --- | --- |
| Mandatory Fields | Mandatory fields must be populated to comply with the NMBAQC QAF tools and to submit data to MEDIN. These fields are highlighted in red within the proformas. |
| Conditional Fields | Conditional fields within the proformas are identified by a yellow highlighting. These fields should be completed should the information for these fields be available. This category covers any information highlighted as Conditional or Optional by MEDIN standards. |
| Auto-populated Fields | Any field which has any formulae or links to other fields will be identified as white or blank within the proformas. These fields must be left blank and not edited. Where possible, to reduce time needed to populate the proformas, formulae have been added to auto-populate fields. |
| Optional Fields | Optional fields are fields which are not within MEDIN guidelines but contain useful information about the images/video data. Populate these fields if you have the data but these are not needed for any of the tools. |
|  | Fields that have the hashed lines through them indicate fields that it may not be able to populate at this stage of the analysis process. Please only populate these fields if the information is available and if not, leave blank. |

**Please note, if you are not planning on populating a specific column please DO NOT DELETE THE COLUMN. To aid useability of the spreadsheet please feel free to hide any fields you aren’t planning on populating but do not delete them. If you delete a field the TOOLS WILL NOT WORK EFFECTIVELY and will report multiple errors.**

### Usage guidelines

Once you have selected the most relevant standard for your work and the corresponding proformas for that standard, follow the next steps to complete the proforma.

#### Pre-Analysis

1. When populating the proformas, initially complete all Pre-Survey Metadata fields. These fields can be completed prior to survey work and collection of the imagery data. Populate a Discovery Level Metadata entry for MEDIN Data Archiving within the DASSH portal. [https://www. dassh. ac. uk/medin\_metadata/login](https://www.dassh.ac.uk/medin_metadata/login)
2. The next sheet in the proforma to populate is the Stills Analysis form.
   1. This form collates imagery level metadata such as geographical positioning of all the images and the date/time when the images were taken. This form also collects data on methodology used for the collection of the imagery (Columns A-R). These fields can all be populated prior to analysis either during or directly after sampling.

#### During Analysis

The Abundance matrices can be populated at this stage however if analysis is being completed using an annotation software then it may be better to populate these fields post-analysis.

#### Post-Analysis (if using annotation software)

After all images and video data have been analysed the abundance matrices should be populated (e.g. with exports from annotation software).

At this stage the proformas can be submitted through all the QAF tools to check their completeness (see Section 6).

## Quality Assurance Framework Tools

Three tools have been created to aid with the verification of the proformas to meet QAF requirements and MEDIN guidelines.

The three tools have been created to provide recommendations for the data to aid in completeness.

The errors that are flagged within these checks are to act as the start of an audit trail and an explanation can be provided alongside the outputs to explain deviation from proposed approaches.

**All tools for the QAF can be found here;** [**https://www. dassh. ac. uk/qaf/**](https://www.dassh.ac.uk/qaf/)

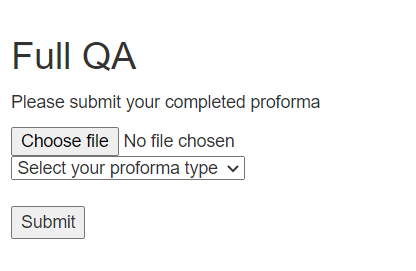
The three tools that have been developed are;

1. Full QA
2. Dataset Comparison
3. EIP Checker (stand-alone tool)

### Full QA

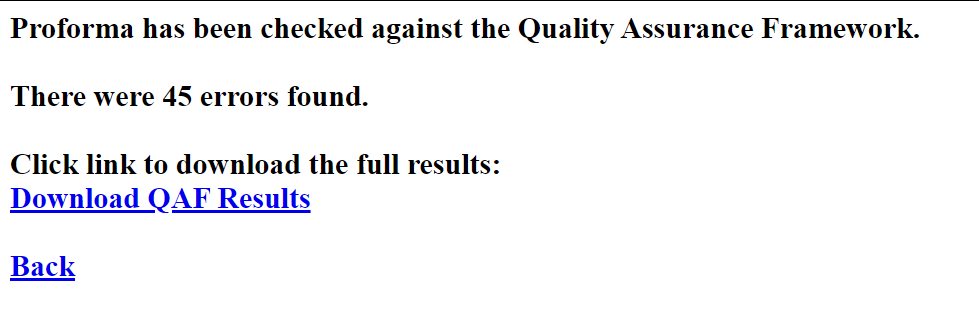
This tool has been created to verify final datasets once analysis of the data has been completed. This tool completes a set of completeness and verification checks that assess whether entries within all mandatory fields are present and whether the data within the fields is compliant with the requirements of that field, for example depth values are all greater than 1.

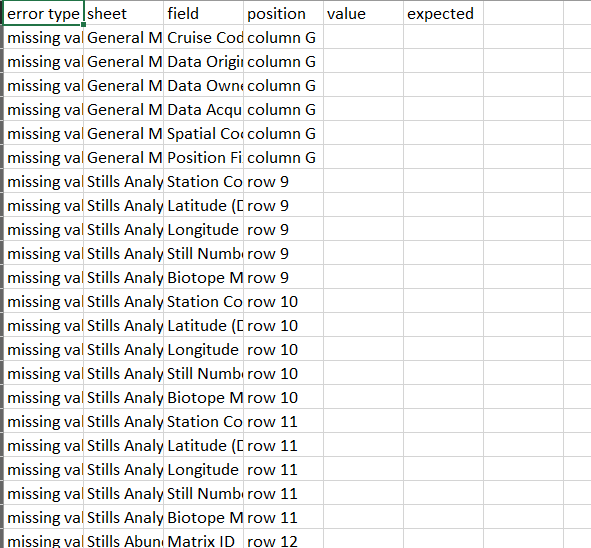
To use this tool, upload your completed proforma as an MS Excel spreadsheet to the tool and select which proforma standard has been used. Once you have submitted your data the tool will start the checks in the background. Please note that the webpage will not change except for a statement at the bottom of the page stating the proforma is ‘being Validated’. If your dataset is large the tool may take a while (several minutes) to run through the spreadsheet. Please **Do not navigate off of the page** as this will stop the tool working.



Once the tool is finished you will be redirected to a new webpage showing the following information. You can then download the full results file which will download a csv file containing a record of all the errors.

For example:



When you open the csv file it will look like this 

This can form part of an audit trail within which all amendments or explanations as to why they aren’t needed can be written against the corresponding entry. The proforma can then be ran through the tool again once edits have been made.

These checks include a check against the Epifaunal Identification Protocol (EIP), which provides a recommendation for taxonomic identification based on corresponding image/ video quality categories. These checks are referred to as “taxon identification” checks within the tool and check for spelling errors, alongside identification errors. More information on the EIP and guidance for using it can be found on the NMBAQC website alongside the current version of the EIP spreadsheet.

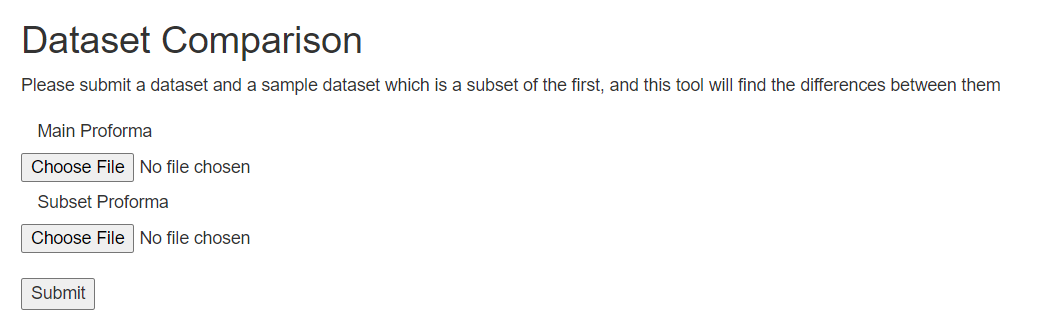
### Dataset Comparison Tool

This tool has been created to compare results of two analyses of the same data (e.g. as carried out for external quality assurance or a ring test).

The tool works by comparing the two datasets and checking for any variations between them.

The checks include a comparison of abundance values and assessment of the factors of difference between the two datasets. For example, if analysis A identifies 6 individuals and analysis B identifies 9 individuals a difference of 30% would be flagged.

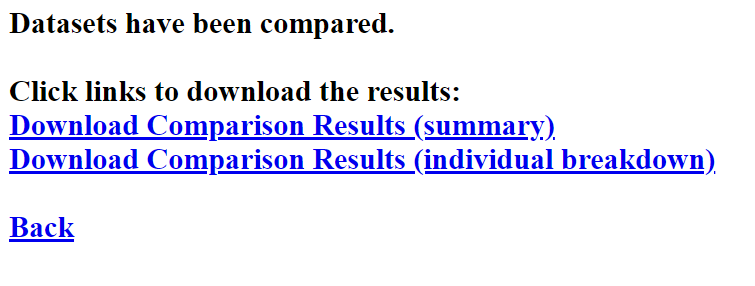
The comparison tool also performs the EIP checks described above.

The comparison tool looks like this; 

Please ensure that the secondary/audit analysis proforma is uploaded as the subset proforma and the primary/original analysis is uploaded as the main proforma.

The tool will then run in the background and as with the QAF tool, the screen will not change to show that the tool is working except for a small line saying that the tool is working at the bottom of the page. Please **Do not navigate off of the page** as this will stop the tool working.

Once the checks are complete the page will be redirected to this screen:



You can then select to view the summary results which will provide some summary statistics from the analysis.

The more detailed analysis will be given in the individual breakdown which will show exactly where any discrepancies exist and any differences in abundance or percentage cover values.