University of Glasgow

Data Management Plan template for PGR students

1. Overview	
Student name	Calum Young
Supervisor name	Sofie Spatharis
Project title	Investigating biofouling in seaweed aquaculture using eDNA diagnostics
Funder & award number	Kilchoan Crown Estate
Project Summary	Development and application of species-specific eDNA diagnostics to monitor and understand the timing, distribution, and ecological drivers of biofouling in North-East Atlantic seaweed farms, aiming to support early warning systems and sustainable aquaculture practices.

2. Data
What types of data will be collected or created?
Environmental DNA (eDNA) sequences from seawater samples to detect epibiont presence. Microscopy data including plankton counts and blade colonisation assessments. Environmental parameters such as temperature, salinity, light intensity, and hydrodynamic data (e.g., current velocity). Species-specific primer sequences and associated bioinformatics outputs (e.g., alignment files, sequence quality reports). Metadata including sampling locations, dates, methods, and specimen IDs.
What formats will you use?
FASTA, FASTQ – raw and processed DNA sequences. CSV, XLSX – plankton counts, environmental data TXT, JSON – metadata and configuration files (e.g., sensor settings). TIFF, JPEG, PNG – for microscopic images or infographics. R or Python scripts – for data processing and analysis workflows.
How much data will you collect?
eDNA sequencing data: Approx hundreds of gigabytes, depending on the sequencing depth. Environmental sensor data: Thousands of individual readings per farm, accumulating to a few gigabytes across all sites. Microscopy and visual survey data: Dozens of spreadsheets and image sets, totalling a few GB. Potential estimated total volume of 100–500 GB over the full project.

3. Documentation
How will the data be documented and described?
Datasets will be accompanied by metadata files detailing Sampling location, date, and time, Sample type (e.g., eDNA, plankton, blade), methods used, Taxonomic identification methods and confidence scores (for DNA barcoding), data processing steps and scripts plus relevant README files when needed. Code scripts and analysis workflows will be version-controlled (GitHub) and annotated with comments enable reproducibility.
Are there any standards for this in your field of research?
MIxS (Minimum Information about any (x) Sequence) – for sequencing metadata (GenBank)

4. Ethics and Intellectual Property
Who owns the data in your project?
University of Glasgow
Detail any ethical, legal or commercial considerations relating to your research data
eDNA primer designs and workflow pipelines could have future IP value
How will these concerns be dealt with?
Following University of Glasgow's intellectual property and data governance policies

5. Storage and organisation
How will the data be named, organised and structured?
Using consistent naming conventions including project name, location, date (YYYYMMDD), sample type, and version. E.g. Kelpcrofting_20220615_eDNAv1.csv
How will the data be stored for the duration of the project?
Onedrive + GitHub + Laptop + External Drive + MVLS server
How will the data be backed up during the project?
As above
Does access to the data need to be controlled for the duration of the project?
Yes
Who has the right to access the data during the project?
Myself, project supervisors, and authorised collaborators from partner farms.

6. Deposit and long-term preservation
Which data should be retained long-term?
Final datasets used in publications (eDNA sequences, environmental data, microscopy results) and Primer sequences, analysis scripts, metadata files.
How long will data be retained for?
10 years
Where will the data be archived at the end of the project?
Enlighten: Research Data (University of Glasgow repository) and GenBank (Sequence data)
What formats will the data be archived in?
CSV, JSON, FASTA, FASTQ, TXT, PDF (for protocols), and R/Python script file

7. Data sharing
Is any of the data suitable for sharing?
Yes. All non-sensitive final datasets, sequences, and protocols will be made available.
How will the data be shared?
Enlighten / GenBank
Who should be able to access and use the shared data?
Researchers, aquaculture stakeholders, policy makers etc

8. Implementation
Who is responsible for implementing this plan?
Supervisors and I
How will this plan be kept up-to-date?
Annually review to ensure coincides with institutional policy
What actions are necessary to implement this plan?

Maintained data documentation, file organisation and complete metadata for all datasets.
Ensure regular backups and version control.
Archive and publish final datasets upon publication.
What training or further information are needed to implement this plan?
Guidance from the university's Research Data Management support team