JISC

UNIVERSITY OF EXETER

Open Exeter

Developing Research Data Management Policy at Research Group Level: A Case Study with the Marine Renewable Energy Group

Ian Ashton, Hannah Lloyd-Jones & Andrew Cowley

July 2013

Contents

Background

The Marine Renewable Energy research group[1] (the "Group") is led by Dr Lars Johanning[2] and is based on the University of Exeter's Penryn campus. The Group involves approximately 20 staff, including collaborators from Biosciences[3] and the Environment and Sustainability Institute (ESI)[4] and forms part of the College of Engineering, Mathematics, and Physical Sciences (CEMPS). The Group's current research interests include hydrodynamics and marine operations, resource assessment, marine policy, offshore reliability and the environmental impacts of offshore renewable energy.

Marine Renewable Energy decided to develop a group-level research data management policy to ensure that the data it uses are secure, will be reusable in the future and can be shared easily amongst collaborators. The policy work was accompanied by a review of the way in which the Group store data. This work has been supported through the Bridging the Gaps initiative[5] and was led by Dr Ian Ashton[6] in conjunction with others in the Group.

Open Exeter[7] was a JISC-funded project which examined how research data is managed across the University of Exeter. The outputs of the project included an institutional policy on research data management (RDM), a repository for the long-term storage of completed research data, Open Research Exeter (ORE)[8] and training and guidance for researchers and research support staff on research data management.

The Open Exeter team set up a Task and Finish Group[9] which developed institutional-level policy on research data management and Open Access for researchers and postgraduate research students (PGRs).[10] The Task and Finish Group has also written a recommendation report for the implementation of this policy and a sustainable research data management service at the University of Exeter.[11]

---

[1] http://emps.exeter.ac.uk/renewable-energy/research/
[2] http://emps.exeter.ac.uk/renewable-energy/staff/lj233
[3] http://biosciences.exeter.ac.uk/
[4] http://www.exeter.ac.uk/esi/
[5] http://www.exeter.ac.uk/btg/
[6] http://emps.exeter.ac.uk/renewable-energy/staff/iga202
[7] http://as.exeter.ac.uk/library/resources/openaccess/openexeter/
[8] https://ore.exeter.ac.uk/repository/
[9] http://as.exeter.ac.uk/library/resources/openaccess/openexeter/exeterembeds/policydevelopment/
[10] For the researcher policy see: http://hdl.handle.net/10036/4280; for the PGR policy see: http://hdl.handle.net/10036/4279
[11] http://hdl.handle.net/10871/11682

During spring and early summer 2012 Open Exeter conducted over 50 interviews with researchers and research support staff as part of a Data Asset Framework survey[12] and contacted Dr Ian Ashton and Dr Helen Smith[13] to take part in an interview to find out how they manage research data. Dr Gareth Cole, the project's Data Curation Officer and Hannah Lloyd-Jones, the Advocacy and Governance Officer, met with Dr Ashton and Dr Smith in May 2012 and decided to explore the possibility of working together to develop a group level research data management policy for Marine Renewable Energy. This case study looks at how the policy and the procedures were developed and implemented and puts forward recommendations for how other research groups could go about designing their own research data management policies.

---

[12] https://ore.exeter.ac.uk/repository/handle/10036/3689
[13] http://emps.exeter.ac.uk/renewable-energy/staff/hcms201

Policy Development

Key Objectives

The Marine Renewable Energy Group identified the following reasons for developing group-level best practice guidance for research data management:

- To ensure that data are secure.
- To facilitate data sharing amongst the Group.
- To facilitate controlled access to data for collaborators.
- To maximise the usefulness of the data for members of the Group and collaborators.
- To enable future re-use of data.

In order to meet these objectives, the Group reviewed their current methods of data storage and proposed a consistent cross-Group approach to how data should be stored. These procedures and guidelines were accompanied by the development of a higher-level data management and Open Access policy.

Data Storage Review

The Marine Renewable Energy Group produces a large amount of very varied data, which are routinely shared between the Group and both internal and external collaborators. Prior to this project, data were stored on personal computers and storage systems, leading to a disparate data collection on a diverse collection of machines, with various levels of security and documentation.

The project started with a review of data storage. This was carried out through discussions with research staff and an initial meeting to establish how staff handle data. Certain key issues were raised through this process;

- Central data storage was commonly used to deposit specific data sets on request, for access by an individual researcher. Several requests for slightly different data sets could therefore lead to the same data being stored centrally several times in slightly different formats.
- All research staff involved in data collection and storage had put in significant work to develop routines to process, quality control, and store data. These were found to be effective and to contribute to excellent research outputs. However, in many cases, these procedures were individual; they were lacking in sufficient documentation for someone else to repeat the steps and obtain a workable data set. In some cases, this meant that a duplication of effort was required when another researcher wished to use the data. In certain cases, without the knowledge of the individual research staff, the data could not be used at all.

- The security and back-up routines that were implemented were highly variable. Many researchers had highly effective back up routines, although they commonly used personal hardware. In most cases, however, back-up versions of data were not stored at multiple sites.

As well as the excellent examples of data processing and storage, there were first-rate examples of data sharing. For example, a regional wave model operated by the Group automatically draws in data from a range of internal and external sources as inputs into the model, and outputs are automatically stored on network storage systems. This provided a good example of what the project aimed to achieve; a fully accessible data storage system that facilitates collaborative working with internal and external partners.

A Collaborative Project

The development of research data guidelines was conducted as a group. Following the data storage review process, it was necessary for all research staff in the Group to define the objectives of the guidelines. Researchers are by their nature very individual, and data are usually gathered at great expense, provides competitive advantage, and is a major component of achieving research goals. Therefore, understandably, research staff were found to be protective of 'their' data.

Through discussions, the Group agreed that the data sets held do not belong to an individual, whilst the individuals who gather the data should retain a strong sense of personal responsibility about the management of the data collected. The Group agreed at this stage that the researcher who puts in the initial effort of data collection, processing and consolidation, should be formally acknowledged in any subsequent research output. This agreement was included in the data policy documentation.

It was also agreed that systems should be put in place to provide a framework for the secure and consistent storage of data, which can be easily reused by research staff. For this, it was necessary to make the data processing and storage routines consistent throughout the Group and centralise all data storage.

Finally, the Group agreed that a consistent and accessible data storage system would facilitate data sharing and collaboration, therefore maximising the research potential of data sets captured and increasing the quality and quantity of research outputs for the Group.

Throughout the project, group discussions were held and used to define the goals of the project as well as to undertake concrete tasks such as the identification of data sets. These meetings also ensured that research staff remained informed of updates and their opinions heard and taken on board.

Policy Review

The Group's development of procedures and policy was informed by a review of relevant institutional and funding body policies on research data management. The Group's funding comes from diverse sources including the Engineering and Physical Sciences Research Council (EPSRC) and Natural Environment Research Council (NERC), the European Regional Development Fund, the Technology Strategy Board, and the European Commission as well as through collaborative partnerships with private businesses.

The relevant institutional policies included the Open Access and Research Data Management policy, the Data Protection Policy, the Code of Good Practice in the Conduct of Research, the Information Security Policy and the IT Regulations.

The policy and procedural documents produced by the Group were also reviewed by staff from the University's IT Governance and Compliance division, as well as by the Open Access and Data Curation team.

Technical

A key outcome of the initial discussions was the requirement for an adequate level of central storage to accommodate all the data sets collected by the Group. The final solution was a combination of local storage, and the institutional repository, separated by active and archive data sets. At all stages, the Open Exeter team provided guidance and support.

Active data sets were termed as those where data are continually being received. All raw data are received to local storage. The storage available on site consists of network attached storage. There are two devices, one having a 6x2TB raid 5array and the other having a 6x4TB raid 5 array. This achieves a total storage area of 9.1TB + 15TB with room for expansion incorporating 2 more 4TB hard drive. Further increase in storage is possible through use of expansion boxes with additional capacity of up to 40TB.

A development aim is to make this storage accessible to the Group and collaborators in real time, with controls applied to limit access for each user. Access to the data is currently restricted through Linux file permissions conforming to the

data policy and procedure, with a view to enabling Active Directory group permissions in the future.

The principal limitation of the local system was the ability to provide a secure back-up, preferably on a separate site. Securing this would require cross-college, or institutional agreements, which were raised in initial discussion with Open Exeter. Open Exeter were well placed to advise on institutional level systems and supported the integration of the institutional repository ORE into the Group data policy. This allowed the Group to make use of this excellent resource, providing not only secure long-term storage and back-up, but also an opportunity to share data through the established repository interface.

Policy Content

Initially two documents were produced; one that referred specifically to the Group's data centre and the other a policy document, which had been informed by the UK Data Archive's document Recommendations: Strategies for Research Centres. At this stage, the policy document contained six elements: data management strategy, data storage and back-up, data classification, data sharing, destruction and data review. However, both documents contained detailed, specific information that would need to be updated (for example, the names of people with certain responsibilities) and it was felt that it would be clearer to split the information into one policy document which set out policy principles and another which provided more detail about procedures and responsibilities. The content and scope of the documents were also modified in response to issues which arose when creating the central storage system.

The final version of the policy document (see Appendix 1) is divided into six sections that reflect the Group's research data management and Open Access priorities:
1. Roles and responsibilities
2. Data storage, documentation and back-up
3. Data review and deletion
4. Data sharing and archiving
5. Open Access to research papers
6. Policy revision

The procedural document follows the same structure, with the exception of an additional section about specific procedures for project setup and closing. Below is a summary of the content of each section. See Appendix 1 for the full document.

Roles and responsibilities

This section sets out the roles and responsibilities of those involved in research data management; the Group Leader, the Principal Investigator, the Data Supervisor, the Data Manager and the Computer Development Officer (CDO).

Data storage, documentation and back-up

This section looks at how the Group will store data and other useful documents in the Data Management Resources Library. It describes the folder structure and content and the components of the data log and recommends methods of file naming. Data access methods are described as well as the back-up situation.

Data review and deletion

Procedures for data review and deletion are described in this section.

Data sharing and archiving

This section describes how data will be classified and shared according to its classification and funder and University requirements. All completed data sets will be registered with the University's data repository. When possible these will be deposited on the University's data repository or in a discipline specific repository for long-term archiving.

Open Access to research papers

All research papers should be published on Open Access as soon as publisher restrictions will allow. Researchers should always comply with funder policy on Open Access publishing.

Policy revision

The Marine Renewable Energy Group policy and procedures on research data and Open Access will be reviewed annually.

Implementation

A Data Supervisor was appointed to work with the departmental Computing Development Officer (CDO) to help with the implementation of the data management procedures. The policy states that there will be an annual review of data management procedures. Therefore, it was decided that the Data Supervisor should be chosen from staff employed on a permanent contract.

The Data Supervisor, in conjunction with the CDO, established a folder structure for the local storage which included a standard structure for each individual data set to be added.

Individual research staff with responsibility for data sets were identified as those who had most experience with the data collection and processing. These members of staff were named Data Managers and were introduced to the folder structure, and given a briefing on the data policy. Individual meetings were scheduled between the Data Supervisor and the Data Managers to add the data sets to the central storage system, following the procedures detailed in the procedural document, and as described below.

It is the responsibility of the Data Supervisor, supported by the CDO to help Data Managers through this process to ensure coherence in the storage of the data sets and to highlight problems or issues that arise from different data sets (e.g. too large, or a very large number of files). Data Managers remain responsible for their data sets, and are encouraged to use the facilities provided to optimise their potential for research.

Data Consolidation

The first step towards consolidating data focussed on securing the Group's data sets on central storage and in the institutional repository, Open Research Exeter (ORE). For this, each Data Manager was asked to provide all raw data and processing software, with documentation detailing the procedures for reading the raw data. Commonly, this was a set of raw data, with proprietary software and manuals.

This first stage served multiple purposes. Firstly it allowed all Data Managers to establish their access to ORE, and to the central storage, and secondly it meant that all data held by the Group were now stored centrally and securely. This met a key objective established at the project outset. This procedure also enabled the CDO and the Data Supervisor to review the amount of data storage available against the quantity of data held by the Group.

Development of data resources

The development of data resources is an on-going process. The aim is to encourage Data Managers to develop their central storage to provide a useful data resource. The outcome of the consolidation phase did not, in all cases, yield data sets that are accessible to other researchers or a third party. In many cases, a detailed knowledge of the data sets, and specialist software are required to process the raw data (software is routinely stored with the data if required). A key recommendation being pursued is the provision of a single, base data set, which is processed and annotated with relevant metadata. This is usually the first processed data set from the raw data, but it could have been subject to basic quality control. The aim is to provide an accessible data set, which is more suitable to future use. As such, it will be suitable for repeating results (if necessary), sharing with collaborators, or open access. Concurrent work is the development of an on-line data access portal, and it is these data sets that will be offered to collaborators, either direct from central storage, or through ORE for archived data. A further consideration is the provision of supplementary data such as reports or publications with relevance to the data set, which will also help with the future exploitation of the stored data sets.

Other data processes that are in development relate to the real-time, or short term sharing of live data. There are already several successful examples of this within the Group, collaborating with other institutions and global businesses using real-time data links, and these have often fallen foul of central computing regulations within the university setup. The aim here is to develop a single, established portal for real-time data sharing. The Group identified particular stepping stones to this goal. Currently underway is work to set up automated data upload systems to ensure all data are up-to-date and feeding into the central storage. This will be supported by development of the on-line portal to view real-time data, as well as the download of historical data sets.

Conclusions

Marine Renewable Energy now has a functioning data storage system with automated access for internal staff. Procedures are in place for setting up new data sets, data reviews, and sharing data. On-going work will demonstrate this system to external collaborators, and extend the access systems to allow controlled off-site access. Initially, this will be through an on-line log-in system, but now that the groundwork to provide a consistent storage system has been done, spatial (GIS) interfaces, and direct links to national databases are possible, and being actively explored.

The projected outcomes leave the Group with a coherent, secure central storage system for live projects, which is fully integrated with the functionality of ORE. The system promotes collaborative working, and the optimisation of data sets both now, and in the future. Critically, it prepares the Group for Open Access, with flexible processes and procedures to allow different levels of control according to the requirements of the data set. In addition, it supports the diverse range of research undertaken by the Group, and demonstrates professional and effective working in collaborative projects. Furthermore, the on-line presence will promote the significant capabilities and resources that the Group has built and cement its reputation as a leading centre for field work and applied engineering in the renewable energy sector.

The Groups' policy and procedures have already been shared with other research groups at the University of Exeter who are interested in developing this type of document. With increasing funder focus on data transparency and reuse, the collection of policy templates and providing guidance on group-level data management policy are likely to become key tasks for the University's Open Access and Data Curation Team[14] over the next few years. The Open Exeter project has allowed the Team to develop their skills and experience in this area.

---

[14] http://as.exeter.ac.uk/divisions/crs/ae/open_access_&_data/

Recommendations

- Research group level policy development should be collaborative and include consultation with all members of the research group as far as possible. Feedback from the research community should be listened to; participation in policy development can give researchers a sense of ownership and make the policy implementation phase easier.

- It can be helpful to separate out the principles of a policy from the nitty-gritty of procedures; thus those who don't wish to read a longer, more detailed document can understand the main points quickly and refer to the procedural document only when necessary.

- Local research data management policies should be updated to reflect changes in institutional, funder and ethical, legal and commercial guidelines and these should be considered during policy development.

- Consider institutional as well as local and discipline-specific solutions. For example, if your institution provides a data repository, would it be better to use this for the long-term storage of data, rather than local storage or should data sets be stored in a discipline-specific repository?

- Decide on the scope of the policy; different research groups have different priorities – for example, a Psychology-based group would probably be more concerned with ethical and legal issues to do with working with human participants. It may be worth concentrating first on priority areas and rolling out a more comprehensive policy at a later date.

- Try to balance the amount of detail in the procedural document with respecting researchers' working habits. For example, is it necessary for all researchers to use the same system to name files?

- Work out an estimated timetable for policy and procedure development but be flexible to reflect changing circumstances if necessary.

- Consider the relationship between guidelines for individual projects and research group policy.

- Tailor RDM policy and procedures to the support available to your research group. For example, a group with a dedicated Computing Development Officer may be able to put into place more bespoke solutions than a group without this support.

- Listen to researchers' concerns and make sure they are clearly addressed in the policy and procedural documents.

- Provide support for the initial transition. Staff may not have time to do tasks such as consolidate and transfer old data sets to a central storage system, as they are busy with current and future work, and rarely have the time to look backwards.

Appendix 1: Marine Renewable Energy Group Policy on Research Data and Open Access

Marine Renewable Energy Group Policy

on Research Data and Open Access

The Marine Renewable Energy Group has developed a Group-level research data management and Open Access policy to ensure that the data it uses are secure, will be reusable in the future, and can be shared easily amongst collaborators, as well as to ensure compliance with funders' and the University's data management requirements. The policy also clarifies the Group's commitment to Open Access to research publications. More detailed information regarding procedures and responsibilities can be found in the document "Marine Renewable Research Data and Open Access Procedures".

7. Roles and responsibilities

i) The Group Leader, in conjunction with the Principal Investigators of research projects, is responsible for good practice in research data management within the Group.

ii) The Group Leader will assign a Data Manager to each data type that the research group collects, creates or uses as well as a Data Supervisor who will be responsible for coordinating research data management within the Group.

iii) The Group's Computing Development Officer will be responsible for the setup and maintenance of the research data storage system, technical developments, and responding to technical problems as required.

8. Data storage, documentation and back-up

i) The Group will store live and completed data sets as well as documentation about this data and other useful documents in the Data Management Resources Library.

ii) All data stored in the Data Management Resources Library will be backed up, documented and named as per the accompanying document to facilitate future re-use.

9. Data review and deletion

i) The Group's data assets will be reviewed annually.

ii) Data will be only be deleted in accordance with the processes documented in the accompanying document and the relevant funder and University policies regarding data deletion.

10. Data sharing and archiving

i) Data will be classified, and shared according to its classification and funder and University requirements.

ii) All completed data sets will be registered with the University's data repository. When possible these will be deposited on the University's data repository or in a discipline specific repository for long-term archiving.

11. Open Access to research papers

i) All research papers produced by members of the Group will be published on Open Access as soon as publisher restrictions will allow.

ii) Researchers will always comply with funder policy on Open Access publishing.

iii) Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.

12. Policy revision

i) This policy and the procedures documented in the accompanying document will be reviewed annually.

ii) They may also be revised more frequently in response to issues that arise and in response to changes in funder and relevant University policies.

UNIVERSITY OF
EXETER

Appendix 2: Marine Renewable Research Data and Open Access Procedures

Marine Renewable Research Data
and Open Access Procedures

Summary

Each research project within the Group that generates data will be expected to use the data centre. This procedural document describes how data is treated and outlines the roles and responsibilities of those involved.

Datasets are attributed to research projects, although a single project may have multiple datasets. At the outset, the University's legal requirements must be defined according to the funding agreement. Where a dataset spans multiple projects, a clear indication of which parts of the data are governed by which set of legal requirements must be provided.

Datasets are monitored in terms of the rate of data accumulation, which must be predicted early on in the conception of the project. This is used to predict hardware requirements, which must be provided through project budgets. Potential Principal Investigators (PIs) are encouraged to consult with the Data Supervisor and Computing Development Officer (CDO) at project conception in order to reflect this requirement in proposals. The Library's Open Access and Data Curation team are also available to advise on data management plans.
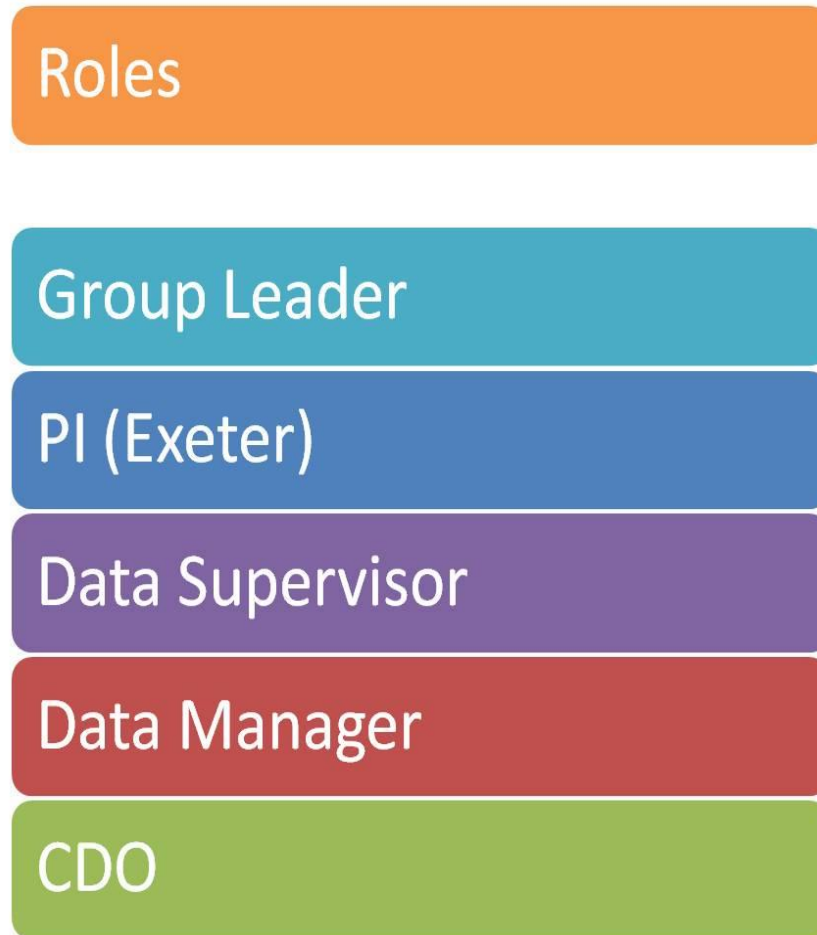
The Data Management Resources Library operates in conjunction with an external contacts database and a publications database. These are described in separate documents.

Contents

1. Role and responsibilities

Figure 1: Roles

i) The Group Leader, in consultation with the PI of the research project at the University of Exeter (UoE), will make key decisions about research data management. These include:

   a. Access to live and completed research data for external parties, and researchers in other groups within the University in accordance with funder and University policy and respecting commercial and confidentiality issues

   b. Proposed development and changes to the public representation of data, or access procedures to the storage centre.

   c. Deletion of any raw data in accordance with funder and University policy.

   d. Changes to assigned personnel and responsibilities.

   e. Changes to the data storage policy and procedures.

The current Group Leader is **Dr Lars Johanning.**

ii)   The Data Supervisor will be responsible for:

a. Monitoring the database, including alerting the CDO and relevant data managers of any issues including accumulation of data at a rate faster than anticipated, and ensuring compliance with these procedures.

b.   Setting up the Data Management Resources Library and keeping this up to date.

c. Updating the Publications Library, including uploading new material and managing content (preventing duplication etc.) and ensuring that this material can legally be stored. In the case of articles published in peer-reviewed academic journals or conference proceedings by researchers in the Group, these should be published on Open Access as soon as publisher restrictions will allow. See section 5 "Open Access to research papers".

d. Managing the contact database for commercial partners.

e. Assisting in setup and closing live projects (see section 7).

f. The annual review of each dataset with Data Managers.

g. Liaising with the University's Records Manager

The current Data Supervisor is **Andrew Vickers.**

iii)   The Data Manager will be responsible for:

a. Storing data, updating the central log and maintaining the contributions folder as per section 2 "Data storage, documentation and back-up".

b. Assisting the Data Supervisor in the annual review as per section 3 "Data review and deletion".

c. Setup, operation and closing live projects (see section 7).

The current data managers are:

Sensor	Filetype	Data manager
Wave Buoy	.pff – 3 files per deployment	Ian Ashton
DWADCP	.000 3 files per deployment, mast, slave and merged	Abdessalem Bouferrouk
ADCP	.000 1 file per deployment	David Parish
SWMTF	7 files per 10 minutes	David Parish
DMAC	.TDMS 1 per deployment	Andrew Vickers
CPOD	.cpi	Matt Witt
AMAR	.dat files	Matt Witt
RBR	.hex files	Matt Witt
BRUV	.mts2	Matt Witt

iv) The Computing Development Officer (CDO) will help with the setup and maintenance of the research data storage system, technical developments, and responding to problems as required. The CDO is responsible for:

    a. Ensuring interlinks between sources and storage.

    b. Development of communication methods, and integration of new data sources.

    c. Development of storage facilities.

    d. Maintaining access for users, whilst ensuring security of data.

    e. Ensuring interlink between storage and web server.

    f. Developing a storage and access solution for a publications library.

    g. Development of secure logins to define access rights for internal and external users.

The current Computing Development Officer is **Andrew Cowley.**

2. Data storage, documentation and back-up

i)   Data Management Resources Library (DMRL)

The data management resources library on the P drive will contain a set of folders, one for each sensor or data source.

a) Hardware

Data storage is provided by a NAS box containing 6 2TB hard drives, using RAID 5 for storage security. RAID 5 requires one hard drive for parity, leaving 10 TB un-formatted space for storage. Once the drives are formatted the size of the drives is reduced to 1.862 TB, which results in 9.31 TB of storage.

Future hardware will be designed in conjunction with the active research projects. It is expected that investments in the hardware will be funded by individual projects as required.

b) Data storage, file naming and file structure

Figure 2: DMRL Structure
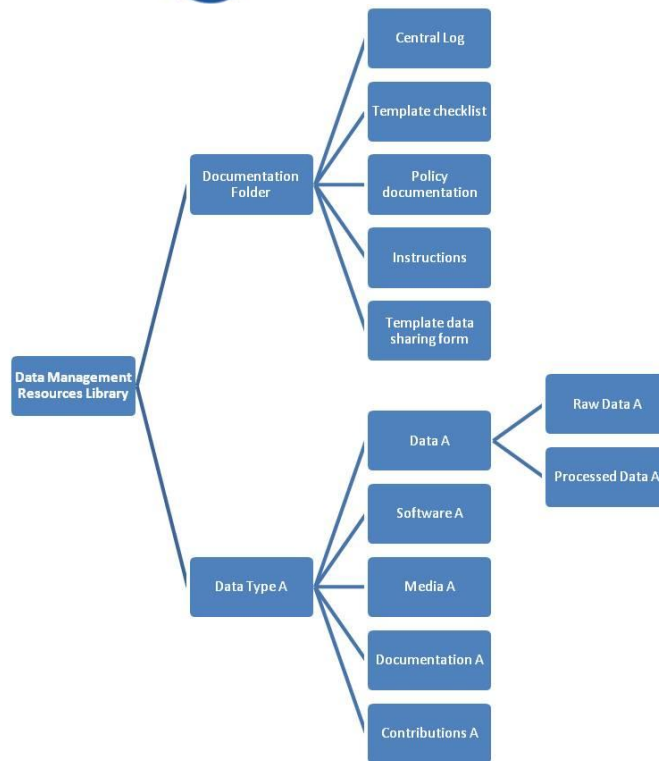
Data managers will store their data in one folder per data type on the P drive. All information and data associated with the data source will be stored in the folder. This should allow a user to require no extra information when downloading data in a usable format. Within each folder will be 5 sub-folders entitled:

- Data
- Software
- Media
- Documentation
- Contributions

This structure is fixed, although subfolders can be used. File names should include a structure of a) data type (e.g. WB for wave buoy), b) date (format YYYYMMDD) and, c) any relevant specification. Where this is not possible (e.g. for raw files generated by the sensor), a detailed description of file names should be kept, and possibly a log containing the date of generation. File naming should be checked with the Data Supervisor prior to implementation, and will be monitored at review.

The DMLR will also contain a central documentation folder comprised of:

- A central log. This will be in the form of a spreadsheet with the fields:

    1. Instrument name (e.g. model name of sensor)

2. Data manager

3. Data ownership and reference to the legal obligations.

4. Users with data access rights

5. Classification(s) of data

6. Predicted rate of data accumulation (see setup procedure, section 7)

7. Predicted date of end of project

9. Actions relating to this dataset (e.g. review, additions, deletions etc…)

- Template checklist for sharing document

- Template data sharing form

- Instructions on using the data centre

- Policy documentation including funders' and institutional policies and model data management plans for research proposals.

i. Data

The data folder will have two subfolders:

- Raw data - raw files collected from the equipment. A readme file is permitted but no other files will be present in this folder

- Processed data – The processed data will be used to receive any data that has been subject to automated processing routines, or through manual processing. It is important to document the process in a readme file. Duplicate data should be avoided in this folder.

ii. Software

The software folder will include any software that is required (or useful) for the analysis of the raw data. This will include the program required to derive the data in the processed data folder from the raw data.

iii. Media

The media folder includes all pictures and videos of the equipment or other relevant pictures and videos. Presentations that refer to the specific data in the data source folder should be included. Publications based on the dataset should be uploaded to the publications library and assigned to the relevant dataset(s).

iv. Documentation

The documentation folder includes all documentation for the equipment, and specifications. Documentation must be stored with the data to ensure that it can be used

effectively, and re-used in the future. Documentation should be in line with the metadata fields required for the institutional repository

It must include the University's legal agreement with regards data accumulated during the project. It should also include a checklist, derived from this agreement, which can be used to ensure that data sharing complies with the University's legal obligations (see appendix C for template document).

Where appropriate, it should also include the following:

- Text documents explaining how data are created, their content, structure, and details of file naming protocol as well as definitions of specialist terms and acronyms used.

- Name, model and description of instrument

- Clarification of dates (format YYYYMMDD) and times, including the time-zone

- Calibration certificates.

- Any documentation relevant to data creation methodology (e.g. user manuals).

- Position of instrument

- Sampling frequency

- Input to test (lab equipment)

- Other details of instrument setup (perhaps a setup file)

- Maintenance performed on instrument (may cause bad data)

- Instructions for any software and advice on further processing of the data

- Details of any processing the instrument has already performed (reference the manual or other documentation)

- Any other information that may be relevant to future processing/analysis

- For processed data, the full details of the processing should be stored, including quality control processes.

v. Contributions

The contribution folder is designed to allow all users with authorisation to view a dataset, to enhance the dataset. This folder will be open to all users to upload any files to be included in the filing system. This would include processed data, presentations, media, papers and any scripts that have been written as well as any data in a further processed

form that may be useful. It is the responsibility of the data manager to maintain this folder, moving relevant contributions to the media folder and deleting old or obsolete contributions. Management decisions for files in this folder are taken by the data manager, and may be deleted as they decide. For users, it should not be considered as secure storage (copies should be retained), unless transferred to another folder by the data manager.

ii) Access and Permissions

a. Data Access

There are two methods of access for editing data on the NAS box:

Manual access will require a data manager to access the PRIMaRE server (or a separate serve) with login credentials using an SSH connection. The SSH connection can be accessed using third party software such as "Filezilla". Manual access allows a data manager to update the raw and processed data in their data source folder through a direct upload. It is important for the speed of the system that processing of raw data does not use the NAS box raw data folders as the source.

Automated access refers to the update of datasets automatically from a device which is sending information on regular basis to a device server. This device server could be the PRIMaRE server of a separate server. In both cases, the data must be sent to the PRIMaRE server before being automatically processed and then sent to the NAS box using scripts on the PRIMaRE server.

All access methods must be designed by the CDO to ensure that security restrictions meet the universities legal requirements for each dataset.

b. Data Manager

The data manager responsible for each data source folder has full read and write access for this folder and any sub folders within it. Full contact details of the data manager must be available to all users on the central log. It is be the responsibility of the data manager to maintain the contents of the data source folder and any sub folder.

c. User

A user is anyone who has access to view and use data on the NAS box, and is limited to internal staff. To control the read access to the data, groups will be created where the names will be the same as the data source folders. Each user will be made a member of the groups which allow access to the data they are authorised to view and copy from the server. The access of a user will be limited to read access for all folders within the groups they are members of except the contributions folders. The contributions folders are designed to allow users to upload files and data they feel would be useful to others when using this data. It is then the user's duty to contact the data manager of the relevant data source folder and have them file the information appropriately.

iii)   Data back-up

All data stored on the DMRL will be in RAID 5. Once a dataset is completed, it will be re-classified as either active (remains on DMRL), or archived.

Archived data will be transferred to the institutional repository, in a relevant collection (e.g. Renewable energy).

At present, hardware does not allow for continuous backup of the DMRL in addition to the mirror. This situation is under review.

3. Data review and deletion

i)   Data Review

- Data storage and access permissions will be reviewed annually by the Data Supervisor.

- During the review:

   a. The Data Supervisor will check that the central log is up to date.

   b. The Data Supervisor will ensure that the predicted data rate remains accurate, and highlight any projected difficulties with storage space.

   c. Data Managers will be asked the following questions for each dataset:

      i.   Is the processing software still valid?

      ii. Can we run it on current machines?

      iii. Is the data log entry up to date?

      iv. Do we need to keep all of the data? (This decision must refer to the funders/university policy on long-term archiving)

Is the data documented and named correctly as per section 2.i.b?Where issues arise with the projected capacity of the database, possible solutions will be discussed by a committee of all Data Managers, Data Supervisor, Group Leader and any other relevant staff.

ii)   Data Deletion

Where data are to be removed to the recycle bin or deleted, the Data Manager and Data Supervisor must be in agreement. Raw data can only be removed to the recycle bin or deleted with the agreement of the GroupL, and without contravention of the requirements for storage of completed datasets (see checklist for selecting data to keep, in appendix A). Removed data will be retained in a recycle bin until storage space is an issue. Subsequently, it will be deleted in preference for new data.

When data is deleted, this must be noted and the record retained on the central log.

4. Data sharing and archiving

### i)	Data Classification

At the outset of a project, data must be classified into one of the following three categories, defined in terms of their accessibility rights. This must reflect the University's legal requirements for the dataset, which are established at the project outset, and documented in the documentation folder. This process is the responsibility of the PI (for UoE), with support from the Data Supervisor.

i.	**Open access data (OAD)** – No limitations are placed on the use of this data by third parties. However, applications for these data during a live project, must be made using the open access data form (see appendix B), signed by the Group Leader and third party applicant. This form specifies the proposed use. Efforts should be made to ensure the third party user is able to use the data effectively for the proposed application. Data should not be released if the Data Manager, Data Supervisor, or Group Leader feel that the intended use, or outcome, may be misleading, inappropriate, or damaging to the University.

Once the project is closed, the full dataset will be reviewed, and transferred to the institutional repository, with any embargo set out by the funding body respected.

ii.	**Controlled data (CD)** – This includes data that are owned by UoE, and do not contain a requirement for open access. These data should not be released without a data agreement form (see appendix B), signed by a member of the UoE Legal Services Office, the Group Leader, and the relevant third party applicant.

Once the project is closed, the dataset will be uploaded to the insititutional repository with an indefinite embargo period (forever).

iii.	**Commercially sensitive data (CSD)** – Where data are collected with direct relevance to a commercial application, they must be considered commercially sensitive. Where this is the case, the Group Leader and Data Supervisor must be informed prior to collection. Access protocols will limit the viewing of these data to the relevant Data Manager and contractually associated staff. If a request is made to share these data, it must be taken forward to the commercial partner, and the release of data will be subject to their authorisation and/or the University's legal obligations. The contact details for the commercial partner must be kept up-to-date, and will be covered in the annual review.

Once the project is closed, the default actions will be to transfer the dataset to the institutional dark archive unless this contravenes the agreement with the commercial partner, which takes precedence.

Third party data can be stored, although they must be supported with clear documentation to identify source and ownership, as well as being classified as per the categories above.

No private data may be stored, all data must be (or have been) directly used for research within the Group.

Figure 3: The Institutional Repository
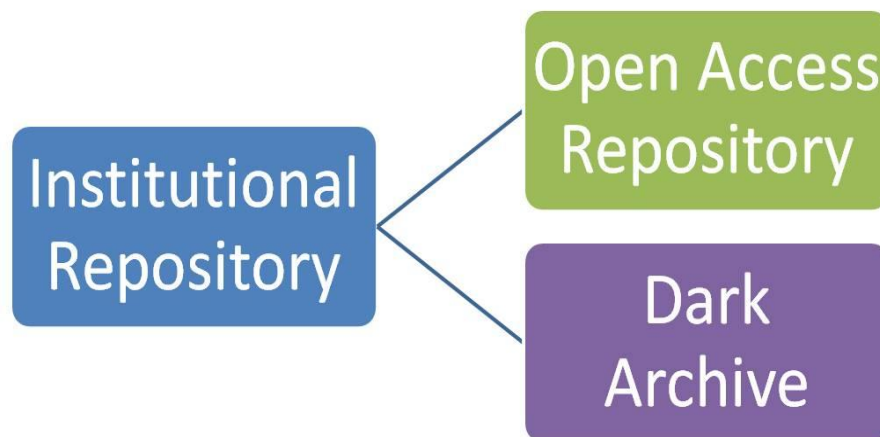
ii)   Data Sharing

Data will be shared according to the classification of the data and the University's legal obligations. Even in the case of open access data, it is important to keep careful control on live data that are shared to ensure that they are used in an appropriate manner. It is also imperative to ensure that commercial third party data are not released inappropriately.

- All live data are subject to authorisation for sharing, preservation and re-use (including open access data). Template documents for data sharing are available in a central documentation folder (see also appendix B).

- No live data will be shared without the data manager being aware.

- Where external sharing is authorised, access will be a remote link for partners, using a temporary access folder held on the server. **External users will not be granted access to the database.** Methods of access will be controlled by the CDO. Access

will be strictly limited to the data named in the data agreement. This will be setup and monitored in conjunction with the Data Supervisor.

- Data marked sensitive will not be made accessible when sharing. Rather, it will be sent, or transferred electronically directly to the recipient, and a receipt must be retained. The data must be encrypted when sending, with password or keys sent separately.

- Completed open access data will be uploaded to the University's open access repository.

- Completed controlled data will be uploaded to the University repository, with an indefinite embargo (forever). Access will continue to be controlled using the data access agreement.

Authorised internal users will have read access to non-sensitive live data. Therefore, these can be accessed directly via their remote link to the storage centre. Data use for an external publication must receive authorisation from the Group Leader. Furthermore, results, processing routines, or any relevant information found when using the data, should be put into the contributions folder of that dataset.

Where data are shared directly via e-mail, drop box (https://as.exeter.ac.uk/utils/dropbox/dropbox.php), or removable storage, a note must be made on the central log which includes information on:

- Which data were provided

- How they were sent

- Which version of processing routines were used (where appropriate)

- The filename of the relevant data access agreement(s)

5. Open Access to research papers

All research papers should be published on Open Access as soon as publisher restrictions will allow. Researchers should always comply with funder policy on Open Access publishing. Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed. The term "research papers" refers to articles published in peer-reviewed academic journals or conference proceedings. For more information on the University's repository for Open Access research papers and how to upload papers see http://as.exeter.ac.uk/library/resources/openaccess/ore/depositingguide/.

6. Policy revision and other relevant policies

The Marine Renewable Energy Group Policy on Research Data and Open Access and the procedures documented in this procedural document will be reviewed annually.

These documents may also be revised more frequently in response to issues that arise and in response to changes in funder and relevant University policies.

The University's relevant policies include:

- Open Access and Research Data Management Policy:
  http://hdl.handle.net/10036/4280

- Data Protection Policy:
  http://www.exeter.ac.uk/recordsmanagement/dataprotection/

- Code of Good Practice in the Conduct of Research:
  http://www.exeter.ac.uk/staff/development/research/support/codeofpractice/good
  practiceinresearch/

- Information Security Policy: http://as.exeter.ac.uk/it/regulations/infosec/policy/

- IT Regulations: http://as.exeter.ac.uk/it/regulations/

7. Specific procedures

a. Project setup

During the early stages of a project, prior to work commencing, the PI (for UoE) will nominate a Data Manager(s). The Data Manager will introduce the Data Supervisor to the proposed work package, whilst the Data Supervisor will brief them on the Group procedures and policy. At the earliest possible date, the following must be addressed:

- Review the University's legal obligations with regards data to be collected. This may require consultation with the Legal Services Office and/or the Open Access and Data Curation team. This process should classify the data types that will be collected. It should also produce a checkbox document for data sharing, to ensure policy is adhered to when sharing. A template is included as appendix C.

- Estimate the data accumulation rate for the project, make predictions of the total required storage during the lifetime of the project, and purchase hardware, where required. Where this is not possible, a suitable procedure for monitoring storage and purchasing increased hardware as required must be established. A minimum buffer of 1GB should be added for contributions, publications and other media. Note, where a large number of videos will be gathered as part of the project, these must be covered in the data accumulation rate.

- Setup the file structure for the new dataset

- Identify the user groups for the data to be collected, and assign access accordingly.

- Establish the end date of the project

- Setup a central log entry for this dataset

b. Closing a live project

Completion of a project occurs when the funding ceases, and the proposed work schedule comes to an end, and is defined in the project setup. At this stage, the data are reviewed with the Data Manager and the Data Supervisor. The PI (for UoE) and the Group Leader may also be included. The following points must be addressed

- Review the folder contents, ensure no duplication, and ensure the documentation and software are such that the data can be used in the future.

- If it is still in regular use for other live projects, or part of proposals, then it should remain in the DMRL and logged with the institutional repository. This status will be reviewed in the annual review.

- If the data are not in regular use, it should be transferred to the institutional repository in accordance with its classification (see section 4.i). If the data are removed from the DMRL, the central log must be updated.

- In the case of commercially sensitive data, any on-going storage must be in agreement with the commercial partner.

Appendix A

Selecting data: what to keep, what to delete?

Why select data?

If you are going to manage your research data effectively, a selection or appraisal process is essential. Deciding what should be kept and what can be disposed of saves on storage costs, enables you to find meaningful data more easily, and ensures you abide by any legal requirements for the data's retention and reuse.

How do I know what to keep?

Choosing what to keep will always involve subjective judgement, as no one can know exactly what information is going to be wanted in the future. The following checklist outlines some of the things to consider when assessing your data:

Are there any relevant institutional or legal requirements?
• Does the University/Research Group have a retention policy for this type of data?
• Does your funding body have a data policy that specifies a retention period for the project's data?
• Is the data affected by legislation such as Data Protection, Freedom of Information or copyright?

Does the data have scientific or historical value?
• Is the data vital to your project?
• Has it been used again in subsequent projects or research?

Is the dataset unique?
• Does it duplicate existing work or is it unique?
• Do other copies exist elsewhere, and if so will they be preserved?

Can the data be reused?
• Are there any intellectual property rights (IPR) issues relating to sharing or reuse of the data?
• Are human subjects involved and was consent given for archiving or reusing the dataset?
• Is the dataset in a format that allows others to reuse it without cost or other restrictions?

Can the data be easily recreated or replicated?
• Is the data recording one-off events that cannot be recreated?
• Can the data be replicated or re-measured without considerable cost or new external funding?

Is retaining the data going to be cost effective?
• Have you considered the costs of long-term preservation of this data?
• Do you have the funds available to do this?

Has the data been effectively documented?
• Is there a data dictionary explaining things such as field names and the context of the data?
• Is there sufficient documentation to allow the data to be found wherever it is stored?

Adapted from:
http://www.lib.cam.ac.uk/dataman/resources/Incremental_factsheet_Cambridge_selection.pdf

UNIVERSITY OF
EXETER

Appendix B

DATA ACCESS AGREEMENT

This agreement governs the terms on which access will be granted to the data generated by the University of Exeter ("the University") as part of the PRIMaRE project.

In signing this agreement, You are agreeing to be bound by the terms and conditions of access set out in this agreement.

For the sake of clarity, the terms of access set out in this agreement apply both to the User and the User's Institution (as defined below). User Institution and User are referred to within the agreement as "You" and "Your" shall be construed accordingly.

Definitions:

Data means the data obtained from the University listed in Schedule A.

User means a researcher whose User Institution has previously completed this Data Access Agreement and has received acknowledgement of its acceptance.

Publications means, without limitation, articles published in print journals, electronic journals, reviews, books, posters and other written and verbal presentations of research.

User Institution means the organisation at which the User is employed, affiliated or enrolled.

Terms and Conditions:

In signing this Agreement:

1. You agree to use the Data only for the [*insert purpose that* Data *is to be used for*].
2. You agree not to use the data or any part thereof for the creation of products for sale or for any commercial purpose.
3. Other than for the purpose set out in Clause1 above you agree to preserve, at all times, the confidentiality of information and Data.
4. Other than for the purpose set out in Clause1 above you agree not to transfer or disclose the Data, in whole or part, or any identifiable material derived from the Data, to others. Should You wish to share the Data with a collaborator outside your Institution, the third party must make a separate application for access to the Data.
5. You agree to acknowledge in any work based in whole or part on the Data, the published paper from which the Data derives, the version of the data, and the role of the University [and the relevant primary collectors and their funders].
6. You accept that the University, the original data creators, or copyright holders, or the funders of the Data or any part of the Data supplied:
    a)   bear no legal responsibility for the accuracy or comprehensiveness of the Data; and
    b)   accept no liability for indirect, consequential, or incidental, damages or losses arising from use of the Data, or from the unavailability of, or break in access to, the Data for whatever reason.
7. You understand and acknowledge that the Data is protected by copyright and other intellectual property rights, and that duplication, except as reasonably required to carry out Your research with the Data, or sale of all or part of the Data on any media is not permitted, although you may publish the data in accordance with Clause 5.

8. You recognise that nothing in this agreement shall operate to transfer to the User Institution any intellectual property rights relating to the Data. [The User Institution has the right to develop intellectual property based on comparisons with their own data].

9. You accept that this agreement will terminate immediately upon any breach of this agreement by You and You will be required to destroy any Data held.

10. You agree that you will submit a report to the University of Exeter, if requested, on completion of the agreed purpose. The University of Exeter agrees to treat the report and all information, data, results, and conclusions contained within such report as confidential information belonging to the User Institution.]

11. This agreement shall be construed, interpreted and governed by the laws of England and Wales and shall be subject to the non-exclusive jurisdiction of the English courts.

For and on behalf of User:

Name of

Applicant(s):   _____

                          _____

                          _____

                          _____

Signature of

Applicant(s):   _____

                          _____

                          _____

Date: _____

For and on behalf of User Institution:

Signature of
Institutional
or Administrative
Authority: _____

Print name: _____

User Institution: _____

Date: _____

For and on behalf of University of Exeter:

Signature of

Institutional or

Administrative

Authority: _____

Print name: _____

User Institution: _____

Date: _____

Schedule A

Appendix C

Application made by
Date (of completion of this form)
Data description
Data classification

Actions (delete those not required):	
Data agreement signed and authorised	☐
Commercial contact authorised	☐
Data Supervisor agreed transfer method	☐
Data encrypted and key prepared to send separately	☐

Any other obligations adhered to: