Check for updates

Implementation of a Data Management Quality Management Framework at the Marine Institute, Ireland

Adam Leadbetter[1] · Ramona Carr[1] · Sarah Flynn[1] · Will Meaney[1] · Siobhan Moran[1] · Yvonne Bogan[1] ·
Laura Brophy[1] · Kieran Lyons[1] · David Stokes[1] · Rob Thomas[1]

Abstract

The International Oceanographic Data and Information Exchange of UNESCO's Intergovernmental Oceanographic Commission (IOC-IODE) released a quality management framework for its National Oceanographic Data Centre (NODC) network in 2013. This document is intended, amongst other goals, to provide a means of assistance for NODCs to establish organisational data management quality management systems. The IOC-IODE's framework also promotes the accreditation of NODCs which have implemented a Data Management Quality Management Framework adhering to the guidelines laid out in the IOC-IODE's framework. In its submission for IOCE-IODE accreditation, Ireland's National Marine Data Centre (hosted by the Marine Institute) included a Data Management Quality Management model; a manual detailing this model and how it is implemented across the scientific and environmental data producing areas of the Marine Institute; and, at a more practical level, an implementation pack consisting of a number of templates to assist in the compilation of the documentation required by the model and the manual.

Introduction

The International Oceanographic Data and Information Exchange of UNESCO's Intergovernmental Oceanographic Commission (IOC-IODE) is the governing body of the global network of National Oceanographic Data Centres (NODCs). In 2013, IOC-IODE released guidance for NODCs to design and implement quality management systems for the successful delivery of oceanographic and related data, products and services (IOC-IODE 2013). IOC-IODE has been encouraging NODCs to implement a quality management system and to demonstrate they are in conformity with ISO 9001, the international standard for quality management. A stated goal of the IOC-IODE's guidance is to "promote accreditation of NODCs according to agreed criteria."

In parallel, funding bodies increasingly have policies on research data management. For example the Horizon 2020 guidance document on Findable, Accessible, Interoperable & Reusable data management from the European Commission (Directorate General for Research & Innovation 2016) describes good data management "...not as a goal in itself, but rather the key conduit leading to knowledge discovery & innovation and to subsequent data & knowledge integration & reuse".

In response to the IOC-IODE guidance and to the requirements of funding agencies, the Marine Institute included "Quality" as a goal in its Data Strategy (2017-2020, see Table 1), with a target of achieving the IOC-IODE accreditation as the NODC for Ireland. To facilitate the development of a Data Management Quality Management Framework (DM-QMF) for the Marine Institute a number of steps were taken. First, a working group was established to develop the DM-QMF model for the Marine Institute and to write the associated manual for submission to IOC-IODE for review prior to accreditation. Second, a number of quality objectives were identified, which the model and manual would need to support (see Table 2). Third, it was decided that an implementation pack would be created which would provide data stewards (see Section "Data steward") and data owners

---

✉ Adam Leadbetter
adam.leadbetter@marine.ie

[1] Marine Institute, Rinville, Oranmore, Galway, Ireland

Table 1 The focus areas of the Marine Institute's data strategy (2017–2020)

Policy	Making it clear how Marine Institute data should be managed
Governance	Ensuring Marine Institute data are categorised and handled appropriately
Quality	Defining processes to ensure the Marine Institute delivers high-quality reproducible data
Capability	Developing data expertise, processes and tools
Connectivity	Connecting Marine Institute data to national and international networks
Coordination	Facilitating national data contributions to global programmes

(see Section "Data owner") with a series of templates and guidelines to implement what is described by the manual. Finally, a series of workshops were held with data stewards and data owners in order to introduce them to the DM-QMF and to begin the process of documenting their datasets and data producing processes within the DM-QMF. Completion of DM-QMF implementation packs by the data stewards and data owners shows conformity of their data processes with the DM-QMF.

This paper will provide an overview of the roles required in the DM-QMF model; details the contents of the DM-QMF implementation pack and a short discussion of the lessons learned and the benefits of this approach will conclude the paper.

Definition of roles

In the DM-QMF manual, a number of roles are identified and also introduced above. Those relevant to this paper are detailed below.

Data owner

Following the definition in Gordon (2013), the role of Data Owner has the authority in the organisation to agree a dataset's classification and the retention schedule for a dataset. While the person tasked with this role may also be responsible for the stewardship or curation of the data values, the Data Owner is more likely to be in a team managerial role. A primary goal for this role is ensuring good data governance is achieved.

Data coordinator

The Data Coordinator role is closely aligned with the Data Administrator role as described in Gordon (2013). The Data Coordinator is responsible for the processes around data management within an organisational unit. Their responsibilities include oversight of the cataloguing of datasets whose Data Owner is a member of the business unit (in this model third-party datasets are catalogued by a centralised Data Management team); and facilitating the

Table 2 The quality objective's of the Marine Institute's Data Management Quality Management Framework

1	Deliver quality data products and services within the appropriate timeframes *(P/E/C/S)*
2	Maximise the Marine Institute's ability to collaborate with current and future customers and interested parties e.g. as data providers, data consumers, service providers & state bodies *(S)*
3	Support the Marine Institute in identifying the relevant competencies and dependencies required to deliver the Data Management Quality Management Framework *(E)*
4	Improve data dissemination facilities and services 'Infrastructure and Products', taking into account data, metadata, documentation, formats and protocol standards *(E/S)*
5	Be consistent with the Marine Institute policies, including the data policy *(C)*
6	Be measurable in order to facilitate reporting, evaluation and process improvement *(P/E)*
7	Take into account requirements of the data and of interested parties as per relevant service level agreements, legislative obligations, and national guidelines on open data *(E/S)*
8	Be relevant to the delivery of products and services optimised beyond the initial process requirements *(E)*
9	Improve the data management processes following the completion of performance evaluations *(E)*
10	Be communicated both internally to ensure implementation and externally to offer transparency *(C/S)*
11	Be consistent across the organisational areas of the Marine Institute *(C)*
12	Highlight where there may be security, access control or data protection concerns *(P/E/C)*

The objectives are grouped into four categories: Performance *(P)*, Effectiveness *(E)*, Conformity (C) & Satisfaction (S)

quality assurance of the data management processes in that unit. The Data Coordinator also acts as a liaison point between the central IT services in the organisation and their business unit which allows for data publication through centralised services. Through regular meetings between the Data Coordinators from the various business units, cross-organisational coordination of data management processes can be achieved.

Data steward

The Data Steward is involved with a dataset on a daily basis, and as such is responsible for many of the day-to-day activities around a dataset including the quality of the data; ensuring its safe archival and storage; and providing the required metadata and documentation around the dataset. Due to the technical scientific nature of their work within the organisational context, the Data Stewards will often blend aspects of the Business Data Steward and Technical Data Steward roles identified in Plotkin (2013). Therefore, as domain scientific experts they will understand the business needs fulfilled by the data they collect and curate but also will often have technical knowledge of database operations and numerical computing in scripting language environments. A Data Steward needs to operate within the bounds of organisational responsibilities and guidelines, and therefore needs to both be aware of these and other legislative requirements which may impact the datasets for which they are responsible; and to be supported by the organisation with appropriate training.

Data protection officer

A Data Protection Officer (DPO) is a leadership role in enterprise security (techniques and strategies for decreasing the risk of unauthorized access to data and IT systems and information) required by the European Commission's General Data Protection Regulation (GDPR). DPOs are responsible for overseeing data protection strategy and implementation to ensure compliance with GDPR requirements.

Data Management Quality Management Framework model

Initially, the DM-QMF model was developed as a visual overview of an early draft of the manual. However, it soon became apparent that this would be a powerful tool for framing the further development of the manual and for communicating the contents of the manual more broadly. The model has a focus on four key areas: inputs; planning; operations; and outputs. The model also has two supporting

areas of focus: performance evaluation and improvement; and input from stakeholders (see Fig. 1).

The inputs area includes direct requirements from customers and interested parties and as such describes the various considerations that should be made prior to moving into the planning phase. A number of other considerations are made. First is any standards that must be adhered to, both in terms of the data creation or collection process (for instance ISO17025 for laboratory based activities) or data reporting standards (such as ISO19115 and ISO19139 for datasets which will be reported to the European Commission's INSPIRE Spatial Data Infrastructure), the Marine Institute Strategy and Policies, any Legislative Drivers (e.g. Water Framework Directive) and document management systems. Depending on the individual process being considered, not all entities of the model may apply.

A Data Management Plan (see Section "Data management plan") should be developed at this point to manage the lifecycle of the process. The planning phase also includes inputs from stakeholders, and describes the process involved in taking all of the applicable inputs and generating an agreed set of requirements (see Section "Requirements document").

The operations phase describes taking the outputs from the planning phase and making them operational via a design and delivery stage as well as producing the various documents required for a data process; including process flows (see Section "Process flows") and standard operation procedure documents (see Section "Standard operating procedures"). Completion of the templates highlights any General Data Protection Regulation considerations which may exist with handling a dataset, and any issues with acceptance criteria and which are recorded in an issues log.

The outputs area describes the data product or service that is produced as a result of the data management process. It should include the data product, a data catalogue entry
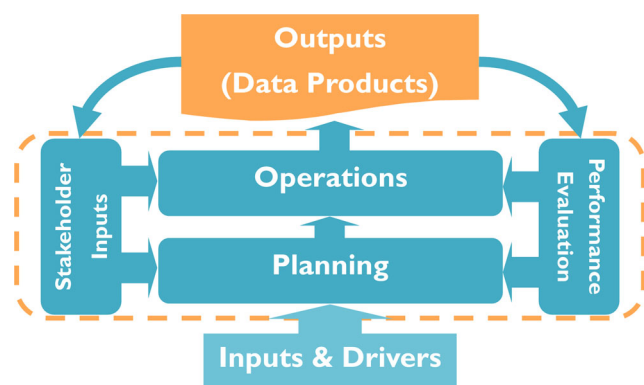

Fig. 1 An overview of the marine institute's Data Management Quality Management Framework model

(see Section "Data catalogue entries") containing all the relevant metadata, a statement of the quality measures applied to the process, as well as, where applicable, any interpretation applied to the data.

Performance evaluation (see Section "Performance evaluation") is an ongoing iterative phase that allows processes to be reviewed and evaluated periodically by Data Stewards and Data Owners with the support of Data Coordinators.

In addition to the model, an implementation pack consisting of a series of templates and guidelines has been developed to allow consistent documentation of the various sections of the model. This implementation pack is described in detail below.

Data Management Quality Management Framework implementation pack

The various elements of the implementation pack are outlined in Table 3 and are expanded in the sections below.

Data management plan

A Data Management Plan (DMP) describes what data will be created during a project, how they will be stored during the project, how they will be archived at the end of the project and how access will be granted (where appropriate). Although a DMP should be prepared before a project begins, it must be referred to and reviewed throughout, as well as after the project, so that it remains relevant.

In accordance with the Marine Institute's Data Policy (Marine Institute 2017) and in keeping with the Government's Open Data Policy (Government Reform Unit 2017) "...data will by default be made available for reuse unless restricted..." Most data generated during a project can be successfully archived and shared. However, some data are more sensitive than others. A DMP will help identify issues related to confidentiality, ethics, security and copyright

before initiating a project and it is important to consider these issues before initiating the project. Any challenges to data sharing (e.g. data confidentiality) should be critically considered in a plan, with solutions proposed to optimise data sharing.

Directorate General for Research & Innovation (2016) states that DMPs are a "key element of good data management".

Funding bodies do not usually ask for a lengthy plan; in fact in 2011 the US's National Science Foundation (NSF) policy stated all NSF proposals must have a data management plan of no more than two pages (National Science Foundation 2011). The UKs Natural Environment Research Council (NERC) proposed a short 'Outline Data Management Plan (ODMP)' ((Natural Environment Research Council 2019)) with the view that full data management is completed by the Principal Investigator (PI) within three months of the start date of a grant. The main purpose of an ODMP is to identify if a project will in fact produce data and the estimated quantity of said data.

Under H2020 the Commission provides a DMP template (Directorate General for Research & Innovation 2018), the use of which is voluntary; however the submission of a first version of a DMP is considered a deliverable within the first 6 months of the project. H2020 FAIR stipulates that a DMP should be submitted only as part of the ORD (Open Research Data) pilot; all other proposals are encouraged to submit a DMP but at the very least are expected to address good research data management under the impact criterion addressing specific issues. DMPs (under ORD pilot) should include information on:

- Data Management - during & after the project
- What data the plan covers
- Methodologies & Standards
- Data Accessibility – Sharing / Open Access
- Data Curation & Preservation

Good research data management criterion should address:

Table 3 The contents of the Marine Institute's Data Management Quality Management Framework implementation pack

	Implementation pack item	Format
1	Data Management Plan	Guidelines and template
2	Requirements document including acceptance criteria)	Guidelines and template
3	Process flow(s), including: process flow descriptions; General Data Protection Regulation concerns; and issues register	Examples, templates and links to registers
4	Standard operating procedure(s)	Template and examples
5	Data catalogue entry	Link to Marine Institute data catalogue and guidelines on creating metadata records
6	Performance evaluation, lessons learned and feedback	Templates

- Standards to be applied
- Data exploitation
- Data sharing
- Data accessibility for verification & reuse including reasons why the data cannot be made available, if applicable
- Data Curation & Preservation methods

In addition:

- Reflect current consortium agreements
- Intellectual Property Rights (IPR)

In general a DMP should contain the following elements to ensure the data will be managed to the highest standards throughout the project data lifecycle in keeping with the Marine Institute's Data Policy principles around the management of data. These elements include (but are not limited to):

- Project & Data Description
- Data Management
- Data Integrity
- Data Confidentiality
- Data Retention & Preservation
- Data Reuse (Sharing and Publication)

In order to prepare a DMP, there is evidence to suggest having a generic template available, with commentary, is useful in guiding a user in addressing the appropriate considerations. This can simply be in the format of a checklist of questions in a document or alternatively are electronic tools available, such as the Digital Curation Centre's (DCC) DMPOnline[1] to help navigate a user through the appropriate sections. As part of the Data Management Quality Management Framework (DM QMF) Implementation Pack a Word template has been created which utilises the DCC checklist. This has been piloted for several in-house Marine Institute Data Processes receiving very positive responses.

Data management costs relating to the preparation of data for deposit and ingestion, data storage, ongoing digital preservation and curation after the project, can be included in a data management plan. Good forward thinking can really help to illustrate, and achieve, time savings in accessing the data by avoiding the costly task of recreating data that has been lost or corrupt.

The UK Data Archive has developed a Costing Tool that can be used for costing data management in the social sciences. This is based on each activity (e.g. in the data management checklist) that is required to make research data shareable beyond the primary research team. It can be used to help prepare research grant applications.

---

[1] https://dmponline.dcc.ac.uk/

Requirements document

The requirements document should contain an agreed set of clear requirements for the data being produced. It is not intended to be an exhaustive list of requirements, rather a high level set of functional requirements that the process must achieve. These requirements maybe either prerequisites in order to commence a data process or requirements to be met in the design or output of a data process. For example, for the Marine Institute's process to publish data through an instance of an Erddap server (Simons 2017), prerequisites include: a dataset must have a public-facing record in the Marine Institute's data catalogue; and that the dataset must not contain personal, sensitive personal, confidential, or otherwise restricted data. In addition, the criteria for successfully meeting these requirements should be specified, ensuring the data produced meets the needs of consumers of the data.

Process flows

A Process Flow is a visual representation of an activity or series of activities, using standard business notation, illustrating the relationship between major components and demonstrating the logical sequence of events. A Process Flow describes 'the what' of an activity and a Procedure describes 'the how'. Together they form part of a Data Management Framework. The Process Flow may be split across multiple levels, but at the highest level should encompass the complete lifecycle of the data process (see Fig. 2). Process flow mapping involves gathering everyone involved in the process (administrators, contractor, scientists) together and determining what makes that process happen: inputs, outputs, steps and process time. A process map takes that information and represents it visually.

The visual aspect is key: but the benefits go beyond making it easier to understand or simple to grasp. Having every key team member aware and included improves morale by having a visual representation of what everyone is working towards. Where problems are obvious, team members have a part in creating the solution. In order to ensure consistency across a suite of process flows, they are drawn using the Business Process Model and Notation (Object Management Group 2011).

All parties can discover exactly how the process happens, not how it is supposed to happen. In creating the process flow, discrepancies can be clearly observed occurring between the ideal and the reality. Once a process is mapped, it can be examined for non-value-added steps. Unnecessary repetitions or time-wasting side-tracks, can be clearly identified and dealt with; being removed or altered as needed.
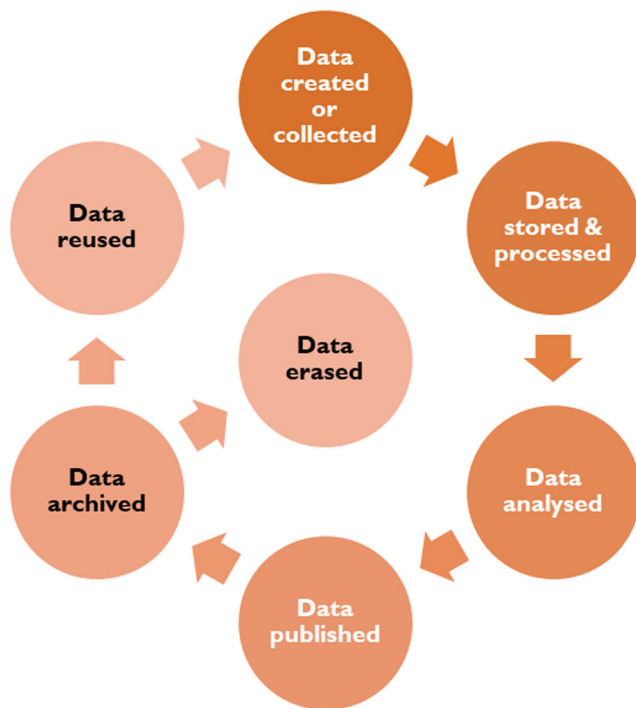
Fig. 2 An overview of the data lifecycle adopted in the marine institute's Data Management Quality Management Framework

A complete Process Flow can provide a clear vision of the future. After pinpointing problems and proposing solutions, there is an opportunity to re-map the process to what it should be. This shares the big picture with a team; each contributing member is then able to carry out improvements with a shared vision in mind. An example process flow for an ocean modelling dataset is shown in Fig. 4.

A process flow highlights duplicate processes across an organisation as well as variant practices, allowing an organisation to prune out the inefficient and propagate the most effective.

Each process flow is supplemented with an accompanying Process Flow Data Sheet (part of the Implementation Pack), which provides context to each process. Moreover, the Process Flow Data Sheet allows process owners and data stewards to record information, at an individual process flow level.

From a user perspective, the Process Flow Data Sheet is structured as a series of questions; mandatory information is clearly indicated, while optional information can be recorded as 'N/A' if deemed not applicable for a given process. Operational planning and control ensures that each process:

– Defines and responds to the requirements for the data product or service.

– Defines the acceptance criteria for the process output to ensure that requirements are met.
– Is fully documented through each step, providing traceability and confidence that each planned activity has been performed.
– Is modified only when changes are planned and reviewed to understand the impact when made operational.

Standard operating procedures

A Standard Operating Procedure is a set of step-by-step instructions compiled to perform the activities described by the Process Flow. Depending on complexity there can be multiple Standard Operating Procedures contained in a single Process Flow. Within the context of this implementation pack, the Process Flow and the Standard Operating Procedure are the main mechanisms used to capture and retain organisational knowledge. A template for Standard Operating Procedures has been developed as part of the Implementation Pack, and covers:

- Purpose and scope of the procedure
- Abbreviations and terminology used in the procedure
- Roles and responsibilities required to carry out the procedure
- Detailed description of the procedure

  – Data acquisition
  – Data processing
  – Data storage
  – Data access and security
  – Data quality control
  – Data backup and archive

- Reporting requirements (including legislative requirements on data delivery)
- Recommendations to improve the procedure

The documentation is then stored in a document management system, allowing for version control. Figure 3 shows an example of a Standard Operating Procedure written in Markdown and stored in a private GitHub repository. Where appropriate, the Standard Operating Procedures are being migrated from plain documentation to automated workflows, such as Jupyter notebooks, to demonstrate reproducibility in the data processing workflow (Fig. 4).

Data catalogue entries

The Marine Institute's data catalogue consists of an internal content management system and a public facing, standards compliant catalogue service. This decoupling of content and service is important as it allows a full data catalogue to be maintained inside the corporate firewall, with only those datasets which are deemed appropriate for public
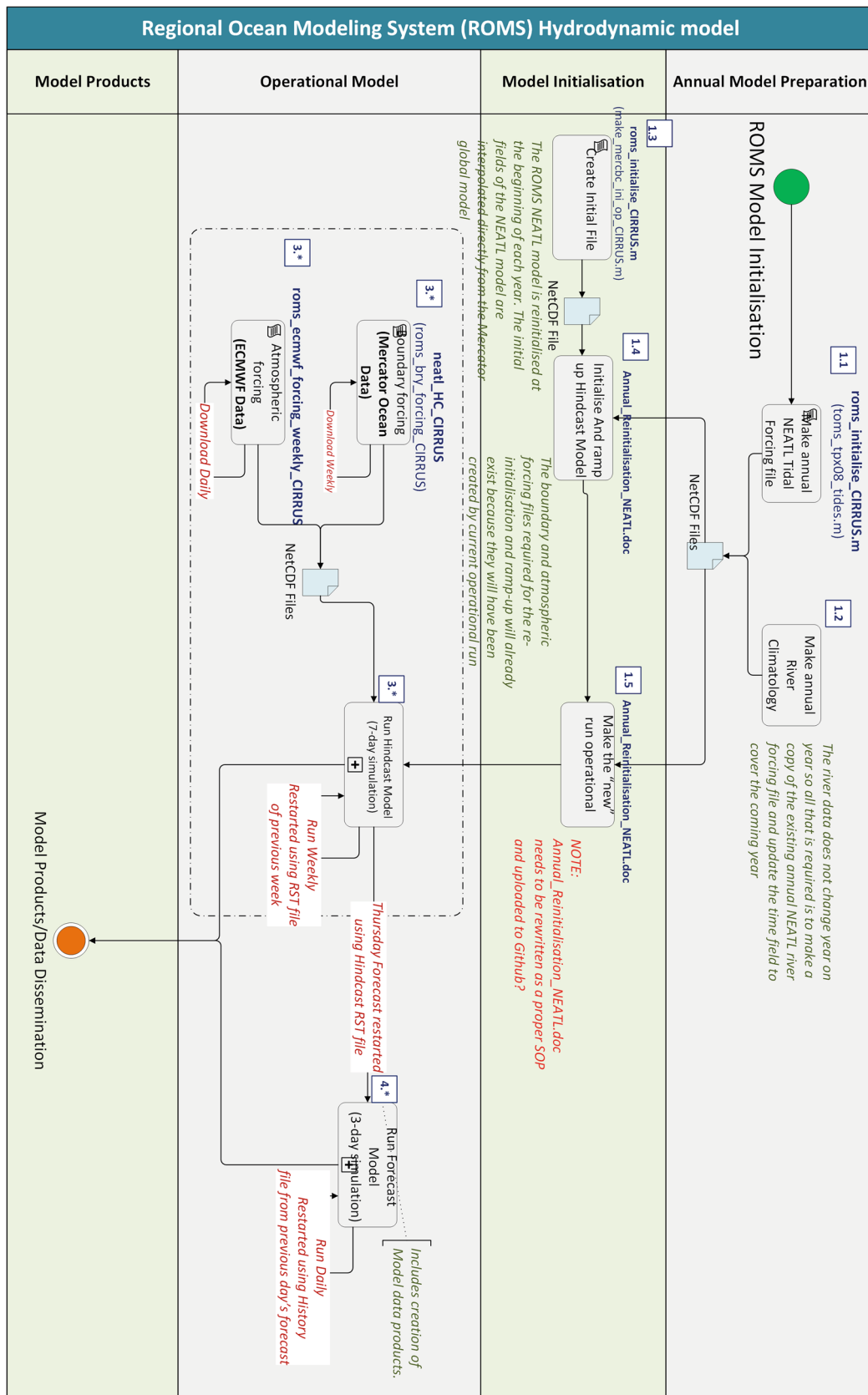
Fig. 3 An example process flow using the business process model and notation. this process flow details the regional ocean modelling system annual re-initialisation in the North East Atlantic domain and shows linkages to hindcast and forecast processes

Fig. 4 An example Standard Operating Procedure (SOP) showing the steps involved in a modelling sub-process for downloading forcing data from the European Centre for Medium Range Weather Forecasts (ECMWF). This SOP highlights details of scheduling; scripts in various programming languages; external prerequisites; and potential issues with the procedure

Task 3.1: Download ECMWF Atmospheric Forcing Data

Cron:

/home/matlab/scripts/cron.op/**roms_ecmwf_forcing**

Schedule:

Runs every night at 20:40

Description:

Task 1 consists of downloading the latest ECMWF atmospheric model forecast data which is published in GRIB file format to a location on our FTP site by Met Eireann. ECMWF runs two forecasts per day initialised at 0000 and 1200. Met Eireann is the agency responsible for the management of ECMWF data in Ireland so they publish a subset of the 0000 and 1200 forecast data for our region. The GRIB files are read and written to daily netcdf files using the ROMS parameter-naming structure. These files are then used by Task 2 to create the weekly atmospheric forcing files. This Task is also listed as Task 1 in the ROMS Forecast Model task list because the data downloaded by this task is also used for creating Forecast forcing files.

We use the first 12 hours of data (at 3-hourly interval) from each of the 2 daily ECMWF forecasts (at 0000 and 1200) to create the daily Hindcast netcdf forcing files.

Reference scripts:

Note: files are colour-coded as: ■ *'Shell Script'* ■ *'Python'* ■ *'MatLab'*

- /home/matlab/scripts/Operational/ROMS/**met_batch_forcing.m** ■
  - /home/matlab/scripts/Operational/ROMS/**ecmwf_download_FTP.m** ■
  - /home/matlab/scripts/Operational/ROMS/**ecmwf_daily_netcdf.m** ■
    - /home/matlab/scripts/Operational/ROMS/**create_met_nc.m** ■
    - /home/matlab/scripts/Operational/ROMS/**ecmwf_to_daily.m** ■
  - /home/matlab/scripts/Operational/ROMS/**ecmwf_daily_netcdf_fc.m** ■
    - /home/matlab/scripts/Operational/ROMS/**create_met_nc.m** ■
    - /home/matlab/scripts/Operational/ROMS/**ecmwf_to_daily_fc.m** ■

External Prerequisites:

- The successful completion of this task is dependent on the latest ECMWF data being made available by Met Eireann. The script, ecmwf_download_FTP.m, waits for up to 2 hours for files to become available.
- Every night (after roms_ecmwf_forcing has completed) a script is ran by IT Operations to download whatever data has been uploaed to Met Eireann on that day and then clear out the FTP folder. This data can be used for ROMS forcing in the event of roms_ecmwf_forcing failing.

Potential Issues:

Issue	Consequence	Alert
ECMWF data not available on FTP	met_batch_forcing will halt	Email to Admin indicating that data is missing
Corrupt ECMWF Grib file	Erroneous data written to daily forcing file which can carry all the way through to model run	NOTHING

Other Comments:

The issue/bug handling is not fully connected with the rest of the neatl_HC_CIRRUS flow. For example, an empty daily atmospheric forcing file may be created and subsequently used by the model to produce erroneous results!

---

consumption published to the wider community. Within the content management system, this differentiation of datasets is achieved through an actionable version of the Marine Institute data policy. The logic applied by this actionable data policy ensures that non-open categories of datasets remain in the internal catalogue only.

The internal content management system manages metadata related to datasets, dataset collection activities, organisations, platforms and geographic features. In this context a dataset may be comprised of the data from one or more collection activities, or may be a geospatial data layer, or may be non-spatial data that is logically grouped. A dataset optionally has a start and end time and an associated geographic feature. A dataset collection activity

is, for example, a research vessel cruise or survey; or the deployment of a mooring at a site. A dataset collection activity must have a start date, an end date, and must be associated with both a geographic feature and a platform (such as a research vessel). A dataset collection activity is also linked to an associated dataset. The concept of geographic feature here links a dataset or dataset collection activity to a representation of the spatial coverage of the dataset. At the coarsest level of detail this will be a bounding box of the extent of the dataset, but a finer level of detail is recommended such as a representation of the shape of a research vessel survey track.

In order to ensure that the metadata in the data catalogue has a level of consistency and interoperability, a number

Table 4 A summary of the performance evaluation questionnaire in the Data Management Quality Management Framework implementation pack

Performance Objective		Quality Objective
1		Quality Objective: Deliver quality data products and services within the appropriate time frames
	a.	What product / service is delivered by this flow?
	b.	Has this process flow been reviewed/updated/revised since the last evaluation?
	c.	Has the product / service been delivered to required deadlines since the last evaluation? If not please provide details.
	d.	Were any issues encountered since the last evaluation? If so are they on the Issues Register? If not please add them.
	e.	Is there a summary/checklist report published at the end of each process run? Who is responsible for running that report?
	f	Where is the summary report located & to whom is it available?
	g.	If the process involves more than one person (either working in parallel or sequence) is there a formal procedure for the transitions and handovers involved? Has this worked smoothly since the last evaluation?
	h	What checks are in place to ensure the integrity of the data has been maintained within the database/dataset/archived files?
2		Quality Objective: Maximise the MI's ability to collaborate with current and future customers and interested parties e.g. as data providers, data consumers, service providers & state bodies
	a.	Describe the mechanism for receiving feedback and if no formal mechanism is available what other channels is feedback received through?
	b.	What feedback has been received from customers and interested parties?
	c.	How has the feedback been considered and what action has it led to?
3		Quality Objective: Support the MI in identifying the relevant competencies and dependencies required to delivery this quality framework
	a.	Do the people involved in the process understand the process and the desired end point?
	b.	Do they have the necessary expertise and training to fulfil their responsibilities?
	c.	What additional training and/or refresher training may be required?
	d.	Are the people aware of the dependencies external to their processes with respect to handovers, expectations or dependencies? Briefly describe any dependencies in the work flow.
4		Quality Objective: Improve data dissemination facilities and services 'Infrastructure and Products', taking into account data, metadata, formats and protocols standards
	a.	Describe the infrastructure in place supporting the process and how this could be improved to enhance the process.
	b.	Where datasets are produced or used by the process are there entries in the Data Catalogue? Please identify them here.
	c.	How is the final/processed data or service product made available both internally and externally ?
5		Quality Objective: Be consistent with the Marine Institute Policies including Quality Policy, Data Policy...etc.
	a.	What categories of data, defined by the data policy, are handled by this process?
	b.	If personal or sensitive personal data are handled, does a General Data Protection Regulation assessment exist and is the Data Protection Officer aware?
6		Quality Objective: Be Measurable to facilitate reporting, evaluation and process improvement
	a.	How is the process is measured (e.g. outputs, products, quality, monetary value, time-frames)?
7		Quality Objective: Take into account the data and stakeholder requirements as per the Service Level Agreements, Legislative Obligations, and National Guidelines on Open Research Data...e.g. SeaDataNet, ICES, etc.
	a.	Illustrate how the deliverables match the requirements as stated in requirements document?
8		Quality Objective: Be relevant to the delivery of products and services optimised beyond the initial process requirements
	a.	How has the product or service delivered by the process been optimised? E.g. data are now available more widely to national/international programmes or networks.

Table 4 (continued)

Performance Objective		Quality objective
9		Quality Objective: Improve the 'Processes' following the completion of Performance Evaluations
	a.	Is the issues register up to date?
10		Quality Objective: Be communicated both internally to ensure implementation and externally to offer transparency
	a.	Is the Data Catalogue record populated with sufficient details of the description and lineage to provide transparency of the process externally?
11		Quality Objective: Be consistent across the Service Areas
	a.	Are all parts of the implementation pack in place?
12		Quality Objective: Highlight where there may be Security & Access Control concerns
	a.	How is the process in line with the organisational security policy?
	b.	Highlight any security concerns since the last evaluation

The performance objectives are introduced in Table 2

Table 5 A summary of the performance evaluation review checklist

Process Flow Summary Checklist

 1. Are all applicable phases of the data lifecycle represented by the process flow?

 2. Are all Standard Operating Procedures identified?

 3. Are all Standard Operating Procedures up-to-date?

 4. Are all Standard Operating Procedures logically represented in the process flow?

Quality Objective: Deliver quality data products and services within the appropriate time frames

 5. Is the product/service being delivered as expected?

 6. Is the process carried out within the allowed time frame agreed?

Quality Objective: Maximise the MI's ability to collaborate with current and future customers
and interested parties e.g. as data providers, data consumers, service providers & state bodies

 7. Is feedback on the process gathered?

 8. Is feedback on the process recorded?

Quality Objective: Support the MI in identifying the relevant competencies and dependencies
required to delivery this quality framework

 9. Does each person involved know where the work they are responsible for lies along the process flow ?

 10. Are they aware of the consequences if their task is not complete or run in a logical order?

Quality Objective: Improve data dissemination facilities and services 'Infrastructure and
Products', taking into account data, metadata, formats and protocols standards

 11. Are there any potential improvements to the process identified?

 12. Is the resulting product available internally?

 13. Is the resulting product available externally?

Quality Objective: Be consistent with the MI Policies including Quality Policy, Data Policy...etc.

 14. Is the process consistent with MI Policy, Data Quality Policy / Data Policy?

 15. Have the datasets been classified correctly?

 16. Have any Data Protection[a] considerations been highlighted/captured?

Quality Objective: Be measurable to facilitate reporting, evaluation and process improvement

 17. Is the output of the process clearly defined?

 18. Is the output of the process produced successfully and in line with the requirements document?

 19. Are the acceptance criteria for the output clearly stated?

Table 5 (continued)

Quality Objective: Take into account the data and stakeholder requirements as per the Service
Level Agreements, Legislative Obligations, and National Guidelines on Open Research Data

  20. Is there a Service Level Agreement or legal obligation associated with the process?

  21. If the answer to (20) is 'no', is such an arrangement needed? *Do not score*

  22. have the acceptance criteria been met?

Quality Objective: Be relevant to the delivery of products and services optimised beyond the
initial process requirements

  23. Has the process been optimised?

  24. If the answer to (23) is no, should the process be optimised? *Do not score*

  25. Have any improvements been identified and logged that could aid optimisation?

Quality Objective: Improve the 'Processes' following the completion of Performance
Evaluations

  26. Is the Issues Register up-to-date?

Quality Objective: Be communicated both internally to ensure implementation and externally to
offer transparency

  27. Is the Data Catalogue populated for the datasets from this process?

  28. Are the Data Catalogue records up-to-date?

Quality Objective: Be consistence across the organisation

  29. Have all parts of the Quality Management Framework pack been completed?

  30. If the answer to (29) is 'no', what percentage of the pack is complete? *Do not score*

Quality Objective: Highlight where there may be security and access control concerns

  31. Is the process owner familiar with the operational IT policies and procedures?

  32. Are there any concerns that should be captured?

[a]With particular reference to the European Commission's General Data Protection Regulation
Unless otherwise noted each question scores 1 for 'yes' and 0 for 'no', giving a possible total of 29

of controlled vocabularies are used and referenced. These may be domain specific, as in those used by the SeaDataNet community (Schaap and Lowry (2010), Leadbetter et al. (2014)) or more generalised, as in the ISO topic categories or as in the INSPIRE Spatial Data Infrastructure.

The internal content management system has functionality to export ISO 19115 metadata, encoded as ISO 19139 XML, to the public facing catalogue server software. In turn, and aligned with the requirements of the European Commission's INSPIRE Spatial Data Infrastructure, the catalogue server is compliant with the Open Geospatial Consortium's Catalog Service for the Web standard. The content management system also exposes INSPIRE compliant Atom feeds for data download services and Schema.org encoded datasets descriptions to enhance the findability of datasets.

Digital object identifiers for datasets

Following the guidance laid out in Leadbetter et al. (2013), digital object identifiers (dois) may be assigned to datasets in the Marine Institute data catalogue under certain circumstances. dois may only be applied to datasets which are in the public facing data catalogue, therefore in this system any non-open categories of datasets may not receive a doi. Further, for a dataset in the data catalogue to receive a doi it must have a publicly accessible download of the dataset associated with the data catalogue record. This is not the case for all datasets in the data catalogue as many do not have an associated data publication service, whereby the data catalogue record is used only as a discovery tool to highlight the existence of the dataset to potential users. The internal content management system allows the creation of DataCite metadata records for the minting of dois from the same database as is used to generate the ISO 19115 metadata records.

Performance evaluation

Within the Implementation Pack, the Performance Evaluation, Lessons Learned, and Feedback sections are designed to provide inputs to improve individual data processes and the quality management framework as a whole. It should be noted that the questions asked here are tightly coupled with the quality objectives laid out in Table 2 and therefore the

specifics may need some adjustment for use in other organisations. Bearing this in mind, the performance evaluation asks one or more questions against each of the performance objectives of the DM-QMF (see Table 4).

A reviewers' checklist template has also been developed, and is completed by reviewers during the review process (see Table 5). This allows for a common score to be applied to the process to assess the level of maturity of the process and to highlight areas for improvement.

Conclusions

In this paper we have shown a Quality Management Framework model and presented an implementation of that model (Fig. 5). It is important to bear in mind that this framework assures the quality of the data management process, and not the data values themselves. However, by bringing the quality assurance of the data management process to the fore it is hoped that the data quality will also be improved over time.

In implementing this framework at the Marine Institute, it has been shown that a high degree of coordination is required. The Data Coordinator roles described above have been key in the liaison between different business units, which in the past have operated as silos. The Data Coordinators meet each other and the central data management team regularly in order to prioritise and progress the implementation of the DM-QMF. The implementation pack arose in response to the need to have a consistent set of documentation for each data process under the DM-QMF and the Data Coordinators have been responsible for developing the content of these packs with the relevant Data Owners and Data Stewards. Awareness of the DM-QMF has been raised by issuing posters with overviews of the framework and more detail on the components of implementation packs throughout the

Marine Institute offices. Workshops on the DM-QMF have also been run with each team which is responsible for a data process, or data process, throughout the Marine Institute.

The main challenge to implementing the DM-QMF has been in resourcing the Data Owners and Data Stewards and freeing them from their normal day-to-day tasks in order to focus on the content of the DM-QMF packs. A related challenge has been in the reporting of progress to senior managers within the organisation in a manner which does not "point the finger" or could be seen as putting blame on to teams who have not been able to make progress due to these resourcing challenges or other, competing, priorities. This reporting is currently being achieved through a quarterly bulletin-style newsletter available throughout the organisation. Further, progress from individual teams has been reported through Data Stewards giving 30-minute lunchtime seminars on how the DM-QMF has been implemented within their teams. This approach has broadened the reach of the dissemination of information around the DM-QMF and has also reduced the perception that the DM-QMF is solely an exercise in data management and increased the perception that the DM-QMF has practical benefits throughout the Marine Institute.

The process flow has been shown to be an invaluable tool to illustrate to new starters their overall position and area of responsibility within an entire process. It clearly identifies all parties involved, ordering, handover points and dependencies. This information can be empowering to members enabling them to be conscious of the importance of their role in the delivery of the project or service as a whole. Over time, organisations become vulnerable to 'single points of failure' i.e. an element of a system that if it fails, will stop an entire process from working. This occurs when organisational knowledge is held by one individual and the process becomes a risk if they leave the organisation or are unable to attend work for some reason. In this Quality Framework, we have mitigated these risks by introducing a

Fig. 5 An overview of the marine institute's Data Management Quality Management Framework model with the various elements of the implementation pack superimposed
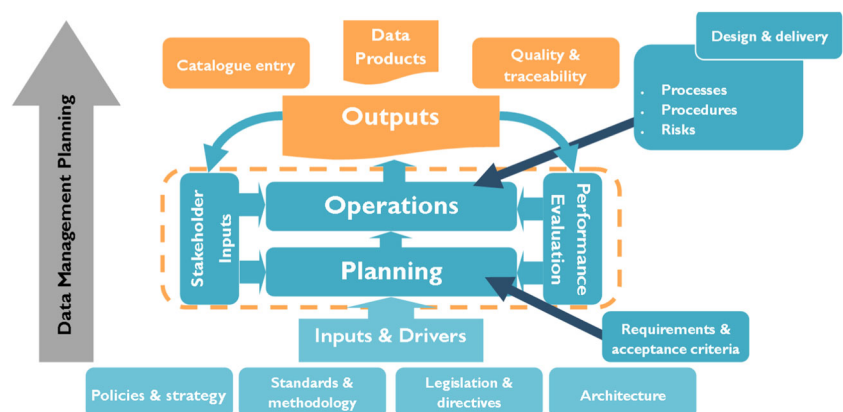
standardised framework for documenting processes. Once the risk is presented in a visual manner by the use of a process flow map it can be easily identified and other elements of the pack, such as the performance evaluation, can bring these risks and approaches to reducing them to the attention of the organisation.

Performance evaluation of the DM-QMF implementation packs through a peer-review process has also been an important part of the framework as it has enabled greater consistency in the content of the implementation packs across the Marine Institute. This consistency has also been improved through running "drop-in clinics" where all Data Coordinators are present for one hour, and any Data Stewards who have questions about completing an element of DM-QMF implementation packs may attend and bring their questions. This enables both Data Stewards and Data Coordinators to hear a range of opinions and to build consensus on the approach to take.

Future work will concentrate on evolution of this approach to close the gap between the approach taken here and the full ISO9001 model. This would require a gap analysis in order to identify areas of this approach which need to be strengthened with respect to that international standard. Secondly, while the content of the implementation packs described above contains much internal information for consumption only within the Marine Institute a further task could be the development of a Best Practices document or paper showing the common areas arising from the contents of multiple packs. Further, work is ongoing to provide an automated assessment of Data Stewardship Maturity (Peng 2018) based on the contents of the DM-QMF implementation packs in order to provide an objective assessment of the status of each data process.

References

Directorate General for Research & Innovation (2016) Guidelines on FAIR data management in horizon 2020. European commission

Directorate General for Research & Innovation (2018) Template horizon 2020 data management plan. European commission

Gordon K (2013) Principles of data management: facilitating information sharing second edition. BCS

Government Reform Unit (2017) Open data strategy (2017-2022). Department of public expenditure and reform, Dublin

IOC-IODE (2013) IODE Quality Management Framework For National Oceanographic Data Centres, IOC Manuals and Guides vol 67. Intergovernmental Oceanographic Commission of UNESCO, Paris

Leadbetter A, Raymond L, Chandler C, Pikula L, Pissiersens P, Urban E (2013) Ocean data publication cookbook, IOC manuals and guides, vol 64. Intergovernmental oceanographic commission of UNESCO, Paris

Leadbetter AM, Lowry RK, Clements DO (2014) Putting meaning into netmar–the open service network for marine environmental data. Int J Digital Earth 7(10):811–828

Marine Institute (2017) Marine institute data policy. https://www.marine.ie/Home/sites/default/files/MIFiles/Docs/DataServices/MarineInstituteDataPolicy2017.pdf, Galway

National Science Foundation (2011) Division of atmospheric and geospace sciences - advice to PIs on data management plans. National Science Foundation

Natural Environment Research Council (2019) Data management planning. UK Research and Innovation, Swindon

Object Management Group (2011) Business process model and notation (BPMN) 2.0. Object management group, Needham, Massachusetts

Peng G. (2018) The state of assessing data stewardship maturity – An overview. Data science journal, 17

Plotkin D (2013) Data stewardship: an actionable guide to effective data management and data governance. Newnes

Schaap DM, Lowry RK (2010) Seadatanet–pan-european infrastructure for marine and ocean data management: unified access to distributed data sets. Int J Digital Earth 3(S1):50–69

Simons RA (2017) ERDDAP. https://coastwatch.pfeg.noaa.gov/erddap. Monterey, CA: NOAA/NMFS/SWFSC/ERD