

## **Marine Environmental Data and Information Network (MEDIN): Accreditation Process for Data Archiving Centres**

### **Introduction**

A key objective of MEDIN is to establish an operational network of linked marine data archive centres (DACs) to provide secure long-term storage for marine data. This network will provide the capability to upload and retrieve data. Data contributors should have free access to their data managed within the DAC framework.

The required capabilities of DACs within the MEDIN framework are:

- To ensure the secure, long term, curation of key marine data sets, according to best practice and to relevant national and international standards.
- To make available clear, searchable information on their data holdings, by the generation and publication of metadata on the MEDIN portal.
- To form the first point of call of expertise for the management of marine data.

MEDIN has established an accreditation procedure to govern the process by which new Data Archive Centres are included into the network. Once accredited DACs must provide annual reports for the MEDIN Sponsors.

### **Accreditation Process**

There are six stages to the accreditation process, finishing with formal approval by the MEDIN Executive Team:

- *Initiation / Preparation*
- *Response to MEDIN DAC Requirements*
- *Review of DAC Response*
- *Updated Response to MEDIN DAC Requirements*
- *Recommendation from Expert Panel*
- *Accreditation by MEDIN Executive Team*

The first, preparation stage can take up to several years. Subsequent stages should take between 8-12 weeks before the final accreditation by the Executive Team.

Once accredited, the status and performance of DACs will be reviewed annually as part of the annual review process.

The expert panel who review the DAC response and provide the recommendations to the Executive team will only include members who are independent of the DAC being considered.

#### *Initiation / Preparation*

Involvement: MEDIN DAC working group, experts and organisation proposing to host a DAC.



Description: The MEDIN DAC Working Group identifies the need for a further Data Archive Centre within the MEDIN network. Working with interested parties the DAC Working Group proposes an outline scope, remit and *modus operandi* for the new DAC.

Duration: This part of the process can take between 6 months and 2 years, as it requires a consensus to be established between interested parties, and perhaps business plans to be developed.

#### *Response to MEDIN DAC Requirements*

Involvement: DAC host organisation

Description: The new DAC provides a detailed response to the list of MEDIN DAC requirements as detailed in the Appendix.

Duration: 2-4 Weeks

#### *Review of DAC Response*

Involvement: DAC Expert Panel

Description: An expert panel appointed by the DAC Working Group / DAC Executive Team reviews the DAC response and identifies where (1) further information is required, and (2) where the proposed arrangements do not meet MEDIN requirements.

Duration: 2 Weeks

#### *Updated Response to MEDIN DAC Requirements*

Involvement: DAC host organisation

Description: The DAC responds to the reviewers' comments and updates its arrangements as necessary (or proposes a work programme to do so).

Duration: 2-4 Weeks

#### *Recommendation from Expert Panel*

Involvement: DAC Expert Panel

Description: The Expert Panel provides recommendations to the MEDIN Executive Team on the DAC application

Duration: 2 Weeks

#### *Accreditation by MEDIN Executive Team*

Involvement: MEDIN Executive Team

Description: The Executive Team considers the Expert Group's recommendations and:

- (a) Confirms accreditation of the DAC.
- (b) Confirms accreditation of the DAC but recommends specific actions to be taken by the DAC to meet requirements.
- (c) Postpones accreditation of the DAC until specific actions are taken.
- (d) Recommend that an alternative solution be found to provide Data Archiving facilities for the data categories under consideration.

Duration: At MEDIN Executive Team Quarterly Meeting

## Marine Environmental Data and Information Network: Requirements for Data Archiving Centres

This document lists the requirements for an organisation to become a Data Archive Centre (DAC) under the Marine Environmental Data & Information Network (MEDIN). It also provides further explanatory information for each of these requirements to ensure that potential DACs are clear as to the evidence needed to be provided in order to be accredited as a DAC.

<b>Requirement</b>	<b>To be accredited DACs must provide</b>
<b>ORGANISATIONAL FRAMEWORK</b>	
<b>Generally exhibiting evidence of expertise and a track record in the scientific area of the data</b>	<p>DACs should describe the range and length of expertise of both the organisation and their staff.</p> <p>In addition, details of data sets or products available can also be provided</p> <p>Any appropriate affiliations (e.g. national or international bodies, etc.) should also be noted.</p>
<b>Committed to provide sufficient resources for defined period of time and plans for transition if and when it ends</b>	<p>In order to be accredited, a DAC must provide evidence that it is hosted by a recognised institution (ensuring long-term stability and sustainability) and that it has sufficient funding, including staff resources, IT resources and a budget for attending meetings, ideally for a 3 to 5 year period, and this information should be updated regularly.</p>
<b>Committed to return of data holdings to originators, or lodging with an alternative and suitable repository, if the DAC becomes unsustainable</b>	<p>A long-term stewardship plan should be available including:</p> <ul style="list-style-type: none"> <li>• A statement on how the DAC is financed and for how long.</li> <li>• Action that will be taken in the event that the DAC becomes unsustainable</li> </ul>
<b>Provide annual report as specified by MEDIN</b>	<p>Accredited DACs should provide an annual report to MEDIN according to the pro forma provided by MEDIN. The report comprises 4 sections as follows:</p> <ul style="list-style-type: none"> <li>• A short summary of the remit and status of the DAC</li> <li>• An overview of activities and developments in reporting year</li> <li>• Key Targets for the next reporting year</li> <li>• Report any changes against the specific MEDIN DAC requirements (in particular referring to any requirements placed as a condition of accreditation)</li> </ul> <p>Other suggestions for future reports might include:</p> <ul style="list-style-type: none"> <li>• Key Performance Indicators</li> <li>• Statement on readiness for INSPIRE compliance</li> </ul>
<b>QUALITY CONTROL AND MAINTENANCE</b>	
<b>Adherence to MEDIN Discovery Metadata Standard and appropriate international principles</b>	<p>MEDIN DACs need to provide evidence of adherence to these principles. Further information and links are given below.</p> <p>The MEDIN Metadata Discovery Standard must be used to record details of data sets. The fields used in the standard are compliant with other international conventions (INSPIRE, ISO19115), which means that the details can be transferred easily between organisations and queried by the MEDIN portal. The Metadata Discovery Standard also conforms to the <u>GEMINI2</u> profile. Publication of metadata in the MEDIN Metadata Discovery Standard and made available to the MEDIN Discovery Portal meets both INSPIRE compliance and</p>

	<p>UK Location Programme requirements for discovery services.</p> <p><u>ISO 19115</u> (Geographic Information - Metadata) is an international standard that sets out a number of metadata fields for describing spatial information datasets. <u>ISO 19139</u> (Geographic Information - Metadata - XML schema implementation) is the standard that aims to define an XML encoding for the metadata elements defined in <u>ISO 19115</u>.</p> <p>The <u>UK GEMINI Discovery Metadata Standard</u> is a defined element set for describing geo-spatial, discovery-level metadata within the United Kingdom. It is derived from, and therefore compliant with, <u>ISO 19115</u> Geographic Information – Metadata and the UK eGovernment Metadata Standard (<u>eGMS</u>). GEMINI was originated by the Association for Geographic Information and is currently being revised to produce GEMINI 2.</p> <p>A number of tools and documents to assist in creating MEDIN-compliant metadata are available from the Standards section of the MEDIN web-site.</p>
<p><b>Data collection according to defined quality principles and accepted procedures</b></p>	<p>MEDIN DACs need to provide evidence of defined quality principles and procedures.</p> <p>DACs may also be able to advise on data collection procedures and should be able to direct data collecting organisations to appropriate standards, where these exist.</p> <p>MEDIN is also in the process of deriving <u>data guidelines</u> comprising requirements as to what must be recorded when data of a certain theme is being collected. This allows easier reuse of the data in the future. For example, if benthic invertebrate samples are collected, the instrument used to sample, the sieve size and taxonomic list used to record species should also be stated and use common lists of terms. MEDIN approved data guidelines are available from the <u>standards pages</u> of the MEDIN web site. Where MEDIN data guidelines do not already exist, it is recommended that the resources available on the <u>other marine data standards</u> web pages should be used.</p> <p>Provision of advice and feedback to the original data collectors is valuable, covering information to be recorded alongside data, established quality assurance procedures to be used, etc.</p>
<p><b>Quality assurance of the collected data</b></p>	<p>MEDIN DACs should provide summaries of any quality assurance processes and algorithms that are in place. This should not be a detailed description of how the algorithms work but a broad summary of the checks that are run and, for example, whether data are visually inspected. The summary should include details of how any issues are resolved (e.g. are they returned to the data provider for rectification, fixed by the DAC, noted by quality flags in the data file and/or included in the accompanying metadata).</p> <p>In addition, details of any Quality Management System (QMS) or accreditation schemes implemented by the DAC should be provided. Where data have been collected in line with nationally or internationally agreed standards this should be indicated. For example:</p> <ul style="list-style-type: none"> <li>• Quality Assurance of Information for Marine Environmental Monitoring in Europe (<u>QUASIMEME</u>)</li> </ul>

	<ul style="list-style-type: none"> <li>• Biological Effects Quality Assurance in Monitoring Programmes (<u>BEQUALM</u>)</li> <li>• National Marine Biological Analytical Quality Control Scheme (<u>NMBAQC</u>)</li> <li>• ISO9000 accreditation</li> <li>• Data collected to internationally agreed standards within major scientific projects (e.g. <u>JGOFS</u>, <u>WOCE</u> protocols and standards)</li> </ul> <p>Where guidelines and standards are in use these should be mentioned. For example, the ICES Working Group on Marine Data Management has developed a series of “<u>Data Type</u>” <u>guidelines</u>, which have been designed to describe the elements of data and metadata important to the ocean research community. These guidelines are targeted toward physical-chemical-biological data types collected on oceanographic research vessel cruises.</p>
<p><b>Committed to advising third party organisations collecting similar types of data on procedures, and providing data-banking (warehousing) and curation facilities for such similar data from other sources</b></p>	<p>Short description of DAC</p> <ul style="list-style-type: none"> <li>• Short description of the remit of the DAC including the data types held and those accepted from external parties for archiving.</li> <li>• Licensing terms</li> <li>• Standard agreements covering: <ul style="list-style-type: none"> <li>• Transfer of a copy of data to a DAC</li> <li>• Transfer of ownership to DAC</li> <li>• Use of the data held by DAC by external users</li> </ul> </li> </ul> <p>Format requirements</p> <ul style="list-style-type: none"> <li>• Note that these are aspirational for new data being collected which needs to be submitted to a DAC. It is not intended that all historical data would need to be converted to these formats before acceptance by the DAC. Historical data needs to be addressed on a case by case basis.</li> <li>• At least one, but potentially more, format(s) that data can be submitted to the DAC.</li> <li>• Details of the process for establishing or agreeing alternative formats.</li> <li>• The format description would need to cover both format and syntax.</li> </ul> <p>It may be advantageous for the provider to submit data in their own format provided this is properly documented perhaps along with some sort of index of the data.</p>
<p><b>TECHNICAL INFRASTRUCTURE</b></p>	
<p><b>Databasing and banking with appropriate metadata standards</b></p>	<p>MEDIN DACs should provide documentation of their working practice and procedures. This should include:</p> <p>Information on the technical metadata for all holdings.</p> <ul style="list-style-type: none"> <li>• Descriptions of the data structures (both entities and attributes) within which the data are stored</li> <li>• Explanations of any lookups not obvious from the data holdings directly</li> <li>• Locations of data holdings on the network or other physical locations</li> <li>• Information on metadata schemes</li> <li>• Editorial advice on the content expected in each mandatory field of ISO xxx</li> <li>• List of any topic specific additional fields and accompanying editorial guidance</li> <li>• Information on georeferencing standards in use</li> </ul>
<p><b>Auditable process for long term custodianship and updating of data sets, with appropriate disaster planning</b></p>	<p>MEDIN DACs should have a security policy describing how the data holdings are protected from both malicious and accidental loss. Note that the security policy should exist but should not be made public as it potentially exposes vulnerabilities.</p>

	<p>A policy should include the following:</p> <ul style="list-style-type: none"> <li>• How the holdings are physically protected (e.g. how access to the building is controlled, how secure the building is, who has access)</li> <li>• Access to the network (if the holdings are accessible from the network) – what is the access policy, how is user access limited and by who, whether there is an internet link and details of how the firewall is configured and altered, how machines are patched, which users can log on to particular machines, policy on passwords (e.g. how often they are changed and how secure they need to be)</li> <li>• Policy when staff leave organisation</li> <li>• Database policy – how users are established, what rights they have, how often administrator passwords are changed, what control is there over allowable passwords</li> </ul> <p>How the data holdings are backed up – how often, where are the backups stored and how long for, how protected are the backups (e.g. fire proof safe, stored securely off site, who has access)</p>
<b>USER ACCESS AND COMMUNICATION</b>	
<p><b>Committed to, and focus on, customer service</b></p>	<p>DACs should provide information on:</p> <ul style="list-style-type: none"> <li>• Response times to enquiries for data and information</li> <li>• Description of aimed service level for responding to user requests (where these are cannot be met on-line).</li> <li>• Whether an Enquiries or Help Desk is available</li> </ul> <p>Details of surveys of customer satisfaction undertaken</p>
<p><b>Committed to raising awareness of the holdings and promoting the use of the data</b></p>	<p>Describe facilities available at the DAC to discovery data holdings:</p> <ul style="list-style-type: none"> <li>• Details of how the data can be searched or interrogated by interested users (e.g. On-line metadata search, physical access on site etc)</li> <li>• Short summary of any on-line search functionality</li> </ul> <p>Describe other search facilities used, e.g.</p> <ul style="list-style-type: none"> <li>• Discovery metadata available through the GI Gateway, National Biodiversity Network, UK MED Directory/EDMED, etc.</li> </ul> <p>The DAC should provide an indication of participation in conferences and exhibitions; production of promotional leaflets, flyers and articles</p> <p>In addition to the activities above the DAC should provide information on:</p> <ul style="list-style-type: none"> <li>• Data products available</li> <li>• Linkages with other organisations who use the data for generation of products</li> <li>• Current projects aiming to increase and promote data use</li> <li>• Statistics/metrics indicating data usage</li> </ul>
<p><b>Making datasets freely available wherever possible (not necessarily at zero cost)</b></p>	<p>MEDIN DACs should have a policy on data access. In general DACs should aim to make data sets freely available, although it is recognised there may be restrictions on access to data for a number of reasons including national security, commercial confidentiality, for scientific research to allow the principle investigators and their co-workers to exploit the data in the first instance. However, release of data to the wider community after a period of 1-3 years from data collection should be strongly encouraged. Metadata should be made available at zero cost and data should be made available at zero cost where</p>

	<p>ever possible.</p> <p>The data access policy should include the following:</p> <ul style="list-style-type: none"> <li>• Details of what can / cannot be obtained on-line (e.g. metadata only, full dataset download)</li> <li>• Licensing arrangements</li> <li>• The format(s) that data can be provided in</li> <li>• The media used for providing data (if data are not on-line)</li> <li>• Costs associated with data provision (or cost scales) – including cost of media as well as staff time</li> </ul> <p>Where ever possible, data policies should be in accordance with internationally agreed data policies (e.g. <a href="#">IOC Oceanographic Data Exchange Policy</a>, <a href="#">GOOS Data Policy</a>, <a href="#">WMO Resolution 40</a>, <a href="#">ICES Data Access Policy</a>, etc.)</p>
--	--

## Assumptions

1. It is accepted that there may be instances where there is more than one copy of a dataset within the MEDIN structures but that there will be one MASTER (original) version, held by the originator or transferred to a DAC
2. It is accepted that there may be instances where datasets of similar type are held in separate DACs
3. It is accepted that there will be a range of different levels of value added and commercial activity with the MEDIN DACs
4. There are Funders of Data Collection, Contributors of Data, Holders of Data and Users of Data in MEDIN (all subject to relevant sets of requirements) as well as DACs; these roles are not mutually exclusive.