



MEDIN: A Case Study on how GEMINI has been implemented for marine data

Sean Gaffney, MEDIN, 29 May 2020

Introduction

The Marine Environmental Data and Information Network (MEDIN) is a collaborative and open partnership, established in April 2008 to improve the management of and access to the UK's valuable marine data and information resources. MEDIN is a group under the UK Marine Science Coordination Committee (MSCC) and is supported and funded by a group of 16 public and private organisations¹. MEDIN holds and provides access to marine environmental data predominantly within UK territorial waters. However, MEDIN further provides archiving and access to some marine environmental data on a global scale.

MEDIN and its partners² share a commitment to open, accessible data, in line with government policy, whilst recognising that there are instances and sectors that are required to have legitimate restrictions on data publication, and to cover the production and, where required, maintenance of higher level datasets if not funded centrally. The MEDIN approach can be summarised as “Data should be as open as possible, and only closed if necessary”.

MEDIN has a web [portal](#), enabling users to find UK marine data. The portal is fundamental to the commitment MEDIN has to making data accessible. The portal is the most comprehensive of its type and currently provides access to over 220 TB of marine data, originating from more than 400 organisations and includes data collected for a range of scientific, policy, conservation and commercial uses.

MEDIN built the portal to operate against its own geospatial discovery metadata profile, which ensures all relevant information about a marine dataset is readily available to allow a potential user to make an informed decision about whether it is pertinent. MEDIN maintains this profile, titled the

¹ DEFRA: Department for Environment Food and Rural Affairs; NERC: Natural Environment Research Council; Scottish Government; BEIS: Department of Business, Energy and Industrial Strategy; Cyfoeth Naturiol Cymru / Natural Resources Wales; Met Office; The Crown Estate; Maritime and Coastguard Agency; UK Hydrographic Office; CEFAS: The Centre for Environment, Fisheries and Aquaculture Science; Joint Nature Conservation Committee; OceanWise; DAERA: Department of Agriculture, Environment and Rural Affairs, Northern Ireland; AFBI: Agri-Food and Biosciences Institute; National Oceanography Centre, Llywodraeth Cymru / Welsh Government

² <https://www.medin.org.uk/about/sponsors-and-partners>

“MEDIN Discovery Metadata Standard”, and provides tools to allow users to create the metadata that populates the MEDIN portal.

History of MEDIN Discovery Metadata Standard

The MEDIN Discovery Metadata Standard was devised back in 2009. From its inception, it was considered essential that the Standard not be developed in isolation, but that it be based on existing geospatial metadata standards. MEDIN investigated various different standards, but concluded that their requirements of signposting UK marine data meant that standards already in use throughout the UK should form the basis of the development.

The geospatial discovery metadata profile that was considered most suitable for MEDIN was UK GEMINI (hereafter referred to as GEMINI), owned by the Association for Geographic Information (AGI). The reasons for this were twofold. Firstly, development of the MEDIN profile against GEMINI would, through GEMINI’s conformance to the International Organisation for Standardisation (ISO) geospatial standard ISO 19115:2003, allow MEDIN metadata to be ISO compliant, which brings international interoperability, benefits and recognition. Secondly, by developing against GEMINI, MEDIN would become a valued route for UK organisations to become INSPIRE-compliant in how they signposted their data resources. This second consideration was very important for the public organisations involved in MEDIN as they had legal obligations under INSPIRE.

One important point to note is that while the MEDIN Discovery Metadata Standard was developed against GEMINI, it is not a duplicate of the GEMINI profile. MEDIN has adjusted the GEMINI profile to make it marine-specific, while continuing to adhere to the rules of GEMINI.

What makes the MEDIN profile different from GEMINI?

MEDIN, in keeping with best practice in marine data management, altered certain GEMINI elements, either making the population rules to be specific to data of a marine nature, or by altering the requirement level of some elements. In this latter case, the alteration was always done so that the requirement was tightened e.g. optional to mandatory, as that ensured the GEMINI rules were maintained.

One fundamental strategy MEDIN used in its discovery metadata profile has been to ensure that elements, as much as possible, are populated using terms from controlled vocabularies. MEDIN is fortunate in that the marine data science community lies at the forefront of ontological work and that the British Oceanographic Data Centre, who host MEDIN (and in turn is hosted by the National Oceanography Centre), also manages the NERC Vocabulary Service (NVS) and its associated services. Controlled lists served from the NVS are used throughout the MEDIN profile.

As an example of how MEDIN has adapted the GEMINI profile to match the needs of the marine community, the MEDIN profile requires that, unless an appropriate term is not available, at least one keyword to describe the dataset be taken from the SeaDataNet Parameter Discovery Vocabulary. This is a controlled vocabulary with worldwide marine exposure through the NVS service. By using this term list, MEDIN instantly ensures that the marine data in the metadata resource are described unambiguously using a community accepted practice. This use of the SeaDataNet list also has a follow-on benefit in that, through Linked Data, the terms have been mapped to the ISO Topic Category codes which are held in another NVS controlled vocabulary list. This enables marine data providers to build tools to populate metadata resources and have elements auto-populate, thus improving efficiencies.

Examples of other areas where controlled vocabularies are used by MEDIN include:

(i) Extent: MEDIN uses controlled vocabularies to ensure that any geographic area defined within the data resource is unambiguous. In this case, the vocabularies reference particular sea areas such as the North Sea, Irish Sea etc.

(ii) Responsible Party: MEDIN uses the European Directory of Marine Organisations (EDMO) vocabulary to provide governed content for responsible parties, while providing the capability in the element population for free text if no EDMO entry is applicable for the role being described.

(iii) Data format: a controlled vocabulary is used here to ensure that all understand precisely what the data resource holds. This is useful in a marine context due to the variety of data types being generated by originators.

(iv) Spatial reference system: MEDIN has chosen to restrict the element so that it is only populated by coordinate reference systems held within the European Petroleum Survey Group (EPSG) registry.

In earlier versions, the MEDIN profile included extra elements beyond those in GEMINI. This was done on the request of the marine community to maximise linkages between data resources, and to improve machine to machine interoperability. Examples are the addition of Parent Identifier, to allow MEDIN to link between series and dataset metadata, and Metadata Standard Name, which enables stakeholders to differentiate easily between a MEDIN profile and that of the wider GEMINI. These extra elements now form part of GEMINI 2.3, a decision partially influenced by MEDIN's presence on the GEMINI Working Group.

An example of an element which has had its requirement level tightened is Responsible Party. GEMINI mandates the role of distributor and metadata point of contact (defined as a separate element within GEMINI), but allows other roles as optional additional entries. MEDIN currently mandates five roles: distributor, owner, custodian, originator and metadata point of contact.

Element names can be slightly different as well, such as Dataset Language in GEMINI with the equivalent being Resource Language in MEDIN. This is a consequence of MEDIN's early work on its standard focussing on the ISO 19115 standard and on INSPIRE. While the MEDIN names may be different, the implementation will adhere to the rules of the corresponding GEMINI element.

Use of the MEDIN Standard

MEDIN work with academic, industry and government bodies to encourage and enable uptake and use of the MEDIN Standard. MEDIN promote a data clause to be used when commissioning new data collection activities. This clause stipulates that any new data should be accompanied by MEDIN compliant metadata. Members of the UK Marine Science Coordination Committee engage with their stakeholders, ensuring that marine metadata is described using the MEDIN standard.

The MEDIN website has a standards-focused page (<https://www.medin.org.uk/medin-discovery-metadata-standard>) which provides stakeholders with resources to generate MEDIN metadata.

These include the standard itself, a MEDIN-specific Schematron and two bespoke metadata generation tools. One of the tools is an online tool, hosted by DASSH (https://www.dassh.ac.uk/medin_metadata/login), and the other is an offline tool called Metadata Maestro, initially developed by SeaZone and now supported by HR Wallingford.

These tools ensure that metadata providers can meet their requirements to supply marine-specific GEMINI-compliant metadata, and the Schematron ensures that only valid records are presented on

the MEDIN Data Discovery portal (portal.medin.org.uk). Metadata Maestro provides users with an additional benefit, in that it allows users to extract and supply metadata in both MEDIN format or in standard GEMINI format only. This allows organisation who hold a mixture of marine and non-marine data to use one tool for their metadata needs and supply valid (non-marine records) to Portals such as FindOpenData or INSPIRE.

MEDIN's participation within GEMINI

MEDIN is an active member of the GEMINI Working Group, and regularly provides feedback on issues around interpretation and enhancement of the GEMINI profile itself. MEDIN has been heavily involved in the development of GEMINI from v2.2 to the latest version 2.3.

MEDIN is one of a number of groups within the UK actively working on implementing GEMINI 2.3. MEDIN has updated its Discovery Metadata Standard (now MEDIN Discovery Metadata Standard v3) and is engaged in active work updating a Schematron validation language that meets the both wider GEMINI requirements and those mentioned above as being specific to MEDIN. Work is also scheduled to update tools for metadata creators, so they can generate MEDIN metadata that adheres to GEMINI 2.3.

MEDIN considers the GEMINI profile to be an essential resource for geospatial metadata in the UK and will continue to both support it and develop the MEDIN standard from it, as this directly helps deliver MEDIN's vision of marine data being F.A.I.R (Findable, Accessible, Inter-operable, Re-usable) and facilitates interoperability across disciplines.