International Coastal Atlas Network Cookbook:

Understanding Metadata

Table of Contents

Introduction 3
What is metadata? 3
Why do you need metadata? 3
Metadata standards and profiles 4
ISO 19115 / 19119 / 19139 5
Dublin Core 5
INSPIRE 5
FGDC / NAP 5
Metadata hierarchy levels 6
Metadata editing tools 7
Metadata encoding 8
Metadata examples 9
Metadata and the ICWA prototype 11
Acknowledgements 11
Document Information 12

Introduction

This document provides a brief tutorial for those who wish to get an overview of metadata, with a focus on the ISO 19115 metadata standard. It is aimed specifically at members of the International Coastal Atlas Network (ICAN) community and more generally at scientists, data managers, and system developers. Aimed at scientists and data managers, this document includes a description of metadata and why we need it, metadata standards in use today, description of different metadata hierarchy levels, and a list of selected metadata editing tools available. Aimed at system developers, this document also includes information on metadata encoding using ISO 19139 XML and a reference to example metadata records located on the NETMAR WIKI. ISO 19115/19139 metadata is a requirement to connect to the International Coastal Web Atlas (ICWA) prototype.

What is metadata?

Geospatial metadata is "data about data". It contains information that documents the basic characteristics of a geospatial data resource. It can also document basic characteristics of geospatial applications or services. Metadata falls into broad categories where it answers the "what, why, when, who, where and how" questions about the resource. These questions include[1]:

- What: Title and description of the data.
- Why: A narrative summary detailing the reasons for data collection and its uses.
- When: When the data was created and the update cycles, if any.
- Who: Originator, data supplier, and possibly the intended audience.
- Where: The geographical extent based on latitude and longitude coordinates, geographical names or administrative areas.
- How: How the data was produced and how to access the data.

Why do you need metadata?

Metadata helps a user to find or discover the data that they need and, thereafter, evaluate whether this resource satisfies the user's requirements. Once a user has chosen the resource, usage metadata is then required to help fully understand and interpret the data. Metadata can be used in-house to help locate and use internal data resources. If a staff member leaves an organisation, important knowledge may also leave the organisation too. New staff members may have difficulty in taking up new responsibilities and fully understanding the organisation's

[1] Wilson, M., 2009, Chapter Three: Metadata -- Describing geospatial data, Spatial Data Infrastructure Cookbook.

data resources. Such undocumented data resources may lose value or cost time to relearn its value. Metadata can also be used to locate data resources published by other organisations, helping to minimise duplication of data collection and enabling more efficient and cost-effective use of this data. Specific examples of business cases for using metadata outlined by the Federal Geographic Data Committee (FGDC) include[2]:

Data Management:

- Preserve data history so that the data resource can be reused or adapted.
- Assess the age and character of data holdings to determine which data should be maintained, updated or deleted.
- Improve data accountability.
- Limit data liability by explicitly stating data limitations of use.

Project Management:

- Plan and document the data resources required for a project.
- Monitor data resources' development progress.
- Share data resources' development progress with project participants.
- Ability to access data characteristics for outsourced data production by ensuring metadata is a contract deliverable.

Metadata standards and profiles

In order for geospatial metadata to operate effectively between different organisations and data users, metadata must be compliant with international standards. Such standards provide a common structure and format to describe metadata. Standards enable improved metadata interoperability and integration, thus, facilitating more seamless sharing, searching, and discovery of metadata between organisations and users of geospatial data and services. Discovery metadata is the minimum amount of information that needs to be provided to help users find geospatial resources. Prominent metadata standards in use today include:

- ISO 19115 (Geographic information – Metadata)
- ISO 19119 (Geographic information – Services)
- ISO 19139 (Geographic information – Metadata – XML schema implementation)
- Dublin Core (ISO 15836)
- FGDC Content Standard for Digital Geospatial Metadata (CSDGM)

[2] Business Case for Metadata (http://www.fgdc.gov/metadata/metadata-business-case)

ISO 19115 / 19119 / 19139

The ISO 19115 standard provides a model or structure for describing geospatial data resources (e.g. digital datasets), while the ISO 19119 standard extends ISO 19115 to describe geospatial service resources (e.g. dataset view services). In turn, the ISO 19139 standard defines an XML schema for the physical implementation of these standards. As these are international standards, they are being widely adopted by governments and organisations around the world. While metadata entities and elements are well defined within the ISO standards, there is an extensive list of optional metadata elements on top of the mandatory metadata element set. It is left up to the system developer to define a specific information model or profile. There is no single metadata profile that fits all users' needs. Therefore, there is a need for metadata profiles to be adopted to support various regional, national, organisational, and communities' requirements.

The ISO 19115 standard is currently under formal review. The proposed ISO 19115-1 (Geographic information - Metadata - Part 1: Fundamentals) standard is expected to be approved in 2013, with the proposed XML schema implementation expected in 2015. The existing ISO standards are recommended until these revisions and associated implementations are formally approved and mature.

Dublin Core

The Dublin Core (ISO 15836) standard defines a cross-domain model or structure for describing web resources, typically general electronic documents. The Dublin Core Metadata Element Set contains fifteen properties. While Dublin Core can be successfully applied to describe geospatial resources, the ISO 19115 standard is more specialised in describing such geospatial resources. Since Dublin Core and 19115 are independent standards, therefore, a crosswalk is required to map from one standard to the other[3] [4]. Some metadata tools (e.g. GeoNetwork opensource), automatically provide a mapping from ISO 19115 to Dublin Core.

INSPIRE

Within Europe, the INSPIRE Directive[5] has defined a base metadata profile which is specified in European legislation. INSPIRE adopts the underlying ISO 19115 standard. However, full conformance to the ISO 19115 mandatory element set implies the provision of additional metadata elements which are not mandated by INSPIRE legislation. However, the INSPIRE metadata encoding guidelines respects these ISO 19115 mandatory elements by ensuring they are included as implementation recommendations. In addition, INSPIRE is defining thematic dataset specifications. Individually these dataset specifications have adapted some additional ISO 19115 elements on top of base INSPIRE metadata profile to help data evaluation.

FGDC / NAP

Within the U.S., the Content Standard for Digital Geospatial Metadata (CSDGM) standard is the U.S. federal standard, while the North American Profile (NAP) of ISO 19115 is the U.S. national standard. The Federal Geographic Data Committee (FGDC) developed CSDGM in the 1990s

[3] ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14857-00-2003-Nov.pdf
[4] OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile, Version 1.0.0, OGC 07-045
[5] http://inspire.jrc.ec.europa.eu

for federal agencies. CSDGM is often referred as the 'FGDC metadata standard'. This standard preceded the ISO 19115 standard. The American National Standards Institute (ANSI), the U.S. member body of the ISO, adopted ISO 19115 in December of 2003. The U.S. and Canada have aligned national profile development efforts with the cooperative development of the NAP. An initial profile was adopted in 2009. In September 2010, the FGDC formally endorsed the NAP. However, once the NAP profile is fully developed, the FGDC will process it as a federal standard. Therefore, transition from CSDGM to NAP is an on-going process[6].

Metadata hierarchy levels

Metadata may exist at different levels of granularity. The most common implemented levels are "dataset" and "series". A dataset is defined by ISO 19115 as an: "identifiable collection of data" (e.g. a raster map). A dataset series is defined by ISO 19115 as a: "collection of datasets sharing the same product specification" (e.g. a collection of raster maps captured from a common series of paper maps). Metadata for which no hierarchy is listed are interpreted to be "dataset" metadata by default. Finer levels of granularity include feature and attribute metadata. Examples include: feature type (e.g. a tunnel), feature instance (e.g. the Mont Blanc Tunnel), attribute type (e.g. overhead clearance associated with a tunnel), and attribute instance (e.g. overhead clearance associated with the Mont Blanc Tunnel). These data hierarchical relationships are illustrated in Figure 1.

In addition, metadata can also be used to describe geospatial services, which typically includes information on how to access and invoke such services. Service metadata is defined by ISO 19119 as: "a service metadata record describes a service instance, including a description of the services operations and an 'address' to access the specific service instance". An example of a service metadata is a description of a Web Map Service (WMS) that enables viewing of a raster map collection.

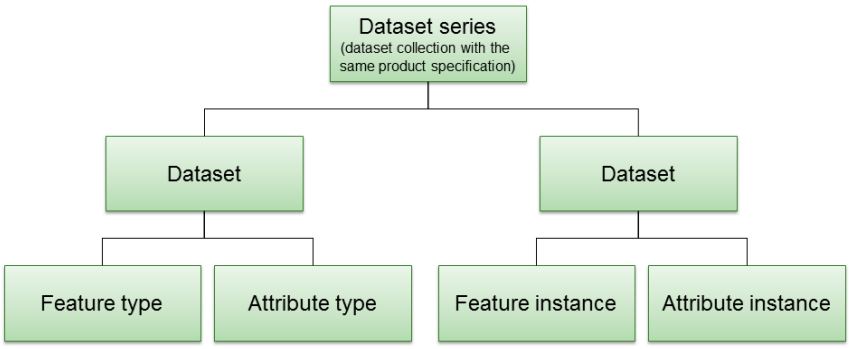

Figure 1: Metadata hierarchy[7]

[6] Preparing for International Metadata, Federal Geographic Data Committee, October 20, 2011
[7] ISO 19115:2003, Geographic information - Metadata

Metadata editing tools

A metadata editor is a program that is used for creating and editing metadata. It typically uses an intuitive graphical user interface which protects the user from the details of the underlying ISO 19139 XML document. A metadata editing tool may include functionalities such as:

- Creating, editing, deleting and viewing of metadata and metadata templates
- Metadata validation (support for XML schema and Schematron validation)
- Import and export of metadata
- Metadata search
- Automatic metadata generation
- Pre-processing and post-processing of metadata
- Extraction and transformation of metadata to different standards and formats
- Additional functionalities such as automatic selection of bounding box coordinates, thesaurus functions, etc.

Several metadata editing tools have been developed. Commonly used commercial metadata editing tools include:

1. ESRI ArcGIS Desktop (http://www.esri.com/)
2. Intergraph GeoMedia (http://www.intergraph.com/)
3. MapInfo Manager (http://www.pbinsight.com/)

Commonly used Opensource/freeware metadata editing tools include:

1. CatMDEdit (http://catmdedit.sourceforge.net/)
2. GeoNetwork opensource (http://geonetwork-opensource.org/)
3. M³Cat (http://www.intelec.ca/)

A screenshot of the GeoNetwork opensource metadata editor is shown in Figure 2. Example metadata fields illustrated include:

- title of the dataset (name in which the dataset is known)
- creation, publication and revision dates of the dataset
- identification code for the dataset (“SeaLevelRise” in the example)
- original owner of the dataset (orgainisation’s name and contact details)

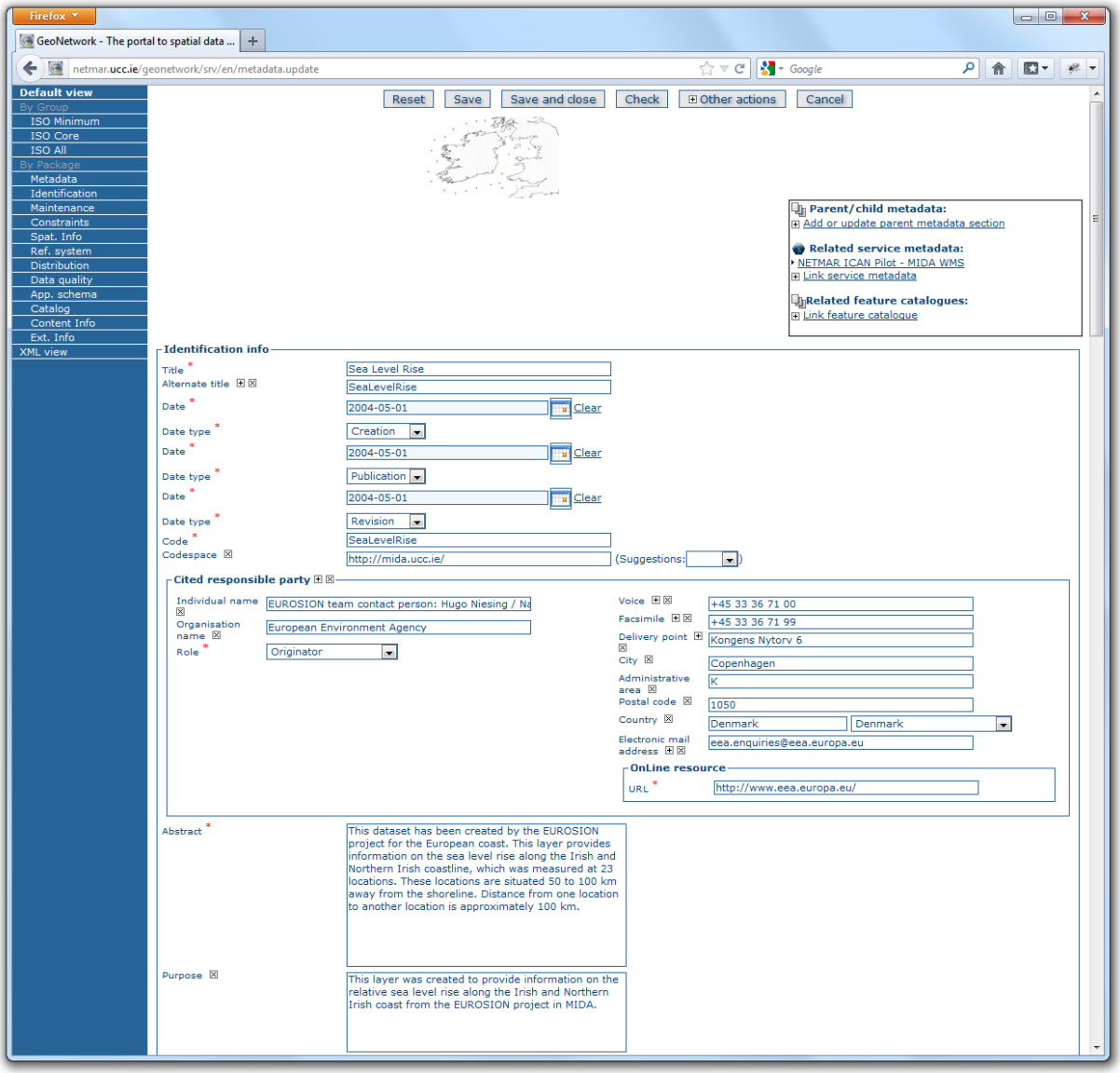

Figure 2: GeoNetwork opensource metadata editor

- dataset's abstract (narrative summary)
- purpose of dataset (summary of intentions)

An alternative to using geographic metadata editing tools is to directly edit the ISO 19139 XML document metadata using an XML editor (e.g. XMLSpy). However, this is only recommended for advanced users who have knowledge of ISO 19139 XML.

Metadata encoding

Metadata can be stored internally within a digital dataset or in a separate external file. Metadata can also be stored in a database to facilitate more efficient searching. The underlying metadata model (an ISO 19115 profile) is the first important aspect to metadata interoperability. At the

physical implementation level, ISO 19139 defines an XML implementation of the ISO 19115 metadata model using XSD (XML Schema Definition) schemas. The 19139 schemas define the structure of the XML metadata document. Therefore, an ISO 19139 XML implementation is the second important aspect to achieve interoperable metadata sharing and exchange between organisations and users of data. The details of the ISO 19139 encoding are not required to be understood by the typical metadata user. Most users typically use graphical based metadata search, view and editing tools. However, for system developers building spatial data infrastructures and who wish to understand more about the ISO 19139 encoding, some useful guidance resources include:

1. UK Gemini Encoding Guidance, version 1.2
   http://location.defra.gov.uk/wp-content/uploads/2010/04/UKGEMINI-Encoding-Guidance_20110505v1-2.pdf

2. Guidance notes for the production of discovery metadata for the Marine Environmental Data and Information Network (MEDIN), version 2.3.7
   http://www.oceannet.org/marine_data_standards/medin_approved_standards/documents/medin_schema_documentation_2_3_7_14mar12.pdf

ISO 19139 XSD schema sets for the full ISO 19115/19119 element set are available in the official ISO repository[8] and the OGC repository[9]. The unofficial EDEN repository[10] is a useful alternate as it has implemented patches to fix schema issues identified by users. It is recommended that metadata is validated against the ISO 19139 XML schema to ensure compliance with the standard. For metadata profiles which have specific requirements or constraints, then Schematron validation is also recommended in addition to XML schema validation. Schematron is a rule based validation language. Its strength is its ability to enforce additional constraints which XML schema cannot (e.g. specific attribute validation).

Metadata examples

For system developers who are familiar with XML, example metadata records can be downloaded from the NETMAR WIKI:

- http://eumis.nersc.no/en/wiki/-/wiki/Main/Metadata

Examples include 19115/19139 dataset metadata for vector and raster datasets, and ISO 19119/19139 service metadata examples for WMS and WPS. The details of this ISO 19139 XML are best described in the MEDIN and UK Gemini documents (referenced in the previous section).

Metadata keywords can be selected from a controlled vocabulary. A metadata user should decide if a controlled vocabulary should be used for their metadata keywords, and if so, whether an existing vocabulary can be reused or a new vocabulary created. Please refer to the

[8] http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/
[9] http://schemas.opengis.net/iso/19139/
[10] http://eden.ign.fr/xsd/isotc211/isofull/

"*Understanding Semantics*" cookbook for future details regarding vocabularies, thesauri and ontologies.

Figure 3 illustrates an example segment of a MIDA (Marine Irish Digital Atlas) metadata record which uses thesauri for keywords. These thesauri are registered in the NERC Vocabulary Server[11] (NVS). This metadata example can be downloaded from the NETMAR WIKI.

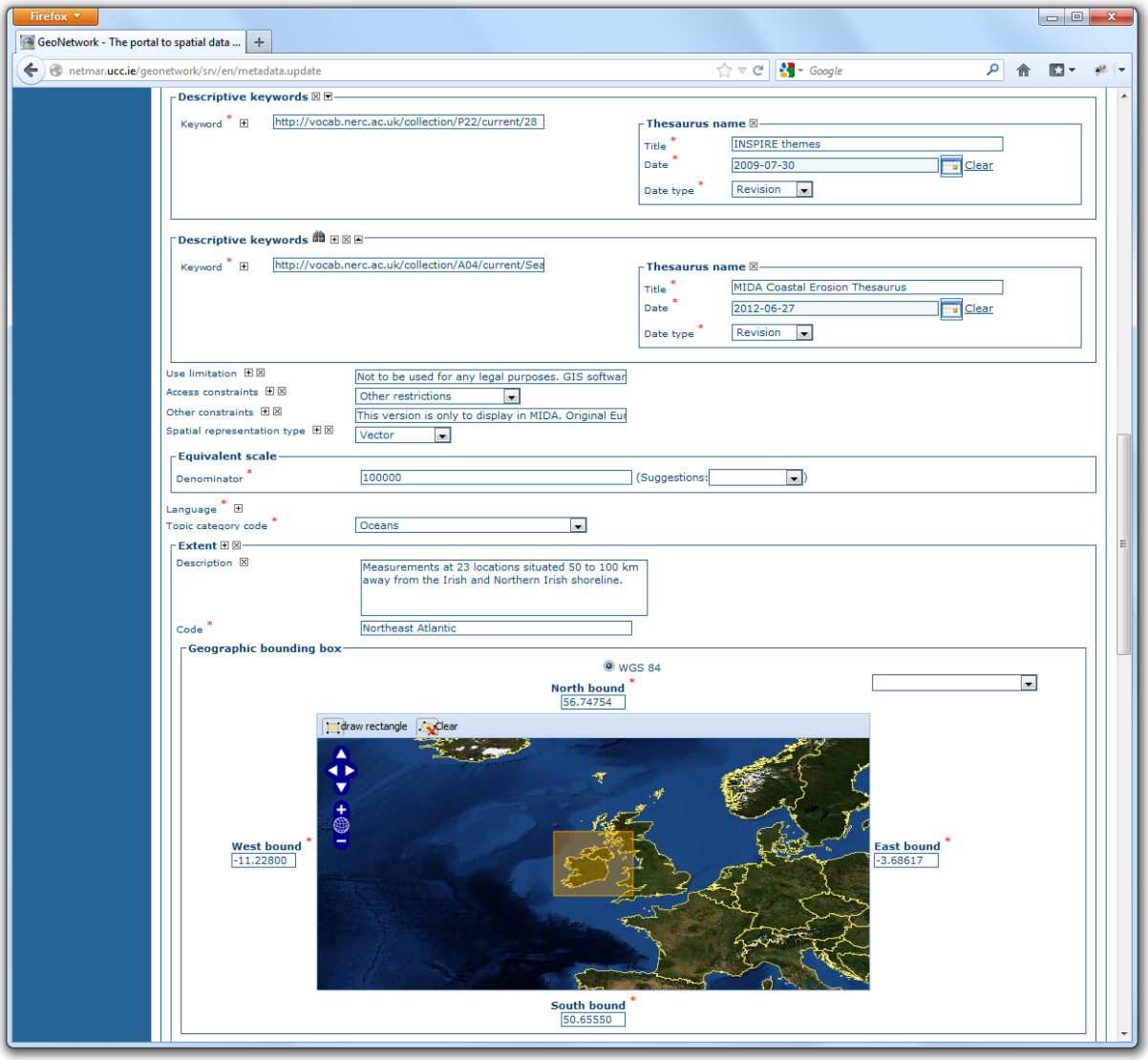

Figure 3: GeoNetwork opensource metadata editor (contains two "Descriptive keywords", each containing a keyword URI registered in NVS)

[11] http://vocab.nerc.ac.uk/

The two metadata keyword fields illustrated in Figure 3 are:

- **Keyword name:** Oceanographic geographical features
  NVS URI: http://vocab.nerc.ac.uk/collection/P22/current/28
  Thesaurus: INSPIRE themes

- **Keyword name:** Sea Level Change
  NVS URI: http://vocab.nerc.ac.uk/collection/A04/current/SeaLevelChange
  Thesaurus: MIDA Coastal Erosion Thesaurus

Other metadata fields illustrated in Figure 3 include:

- Dataset's use limitations (e.g. fitness for use) and access constraints (e.g. intellectual property rights)
- Dataset spatial type (i.e. "vector" dataset in the example)
- Dataset scale (i.e. "1:100000" in dataset in the example)
- High level thematic classification (i.e. "Oceans" in the example)
- Dataset's geographic area description, code identifier (i.e. "Northeast Atlantic" in the example) and geographic latitude/longitude coordinates

Metadata and the ICWA prototype

ISO 19115/19119/19139 metadata is required to connect to the International Coastal Web Atlas (ICWA) prototype. Profiles conforming to these standards should, in practice, connect to the ICWA. However, specific ICWA requirements regarding the encoding of semantic keywords are described in detail with snippet examples in the "*Connecting your Atlas to the ICWA prototype*" cookbook. In addition, a CSW (version 2.0.2) metadata catalogue server is required to publish and query metadata via the web. This is described in the "*Establishing a CSW metadata catalogue with GeoNetwork opensource*" cookbook.

Acknowledgements

This cookbook was written for the International Coastal Atlas community under the auspices of the NETMAR (Open Service Network for Marine Environmental Data) project. NETMAR is partially funded by the European Commission under Theme ICT-2009.6.4 ICT for environmental services and climate change adaptation of the Information & Communication Technologies FP7 Programme.

This document has been reviewed by, and incorporates comments from, Torill Hamre of the Nansen Environmental and Remote Sensing Center; Peter Walker of Plymouth Marine Laboratory; and John Helly of the San Diego Supercomputer Center. Thanks to the reviewers for their help in making the document clear and readable. Further feedback on this document is welcomed, and may be provided by contacting the author whose details are below.

Document Information

Author Declan Dunne, Coastal and Marine Research Centre

Contact d.dunne@ucc.ie

Version 2.1

Date 2012 July 30

Revisions

2.1	Responses to NETMAR internal review
2.0	Comments from ICAN community
1.0	2011 December 23