Report to the Marine Safety Umbrella Operation

Information Technology and Geographical Information Systems: Common Protocols & Procedures

Nick O'Neill, Ric Pasquali, Seamus Coveney

August 2006

DOCUMENT ISSUE SHEET

	Prepared By	**Prepared By**	**Approved By**	**Date**
REVISION				
DRAFT 1	R. Pasquali	N. O'Neill	D. Lewis	15th August 2006

TABLE OF CONTENTS:

1. INTRODUCTION:

The Marine Safety Umbrella Operation (MSUO) facilitates the cooperation between Interreg funded Marine Safety Projects and maritime stakeholders. The main aim of MSUO is to permit efficient operation of new projects through Project Cooperation Initiatives, these include the review of the common protocols and procedures for Information Technology (IT) and Geographical Information Systems (GIS).

This study carried out by CSA Group and the National Centre for Geocomputation (NCG) reviews current spatial information standards in Europe and the data management methodologies associated with different marine safety projects.

International best practice was reviewed based on the combined experience of spatial data research at NCG and initiatives in the US, Canada and the UK relating to marine security service information and acquisition and integration of large marine datasets for ocean management purposes.

This report identifies the most appropriate international data management practices that could be adopted for future MSUO projects.

2. MSUO PROJECT DATA INVENTORY

2.1 Marine Projects Identified:

A number of projects are funded under the INTERREG programme and under the MSUO remit. A web-search was carried out on each of these projects and contact was made with the relevant Data Managers. Table 1 shows the number of projects that were identified during the course of this process.

Marine projects are broadly subdivided into different areas around northern Europe, in turn these are incorporated into the MSUO brief under a number of different programmes: the North Sea programme, the North-West Europe programme, the Northern Periphery programme, the Baltic Sea programme and the Atlantic Area programme. A number of additional projects cover different maritime safety aspects in the areas mentioned above.

A metadata questionnaire (Appendix 1) was designed to assess data management in the different projects, but more importantly to characterise the data populating the database. A description of the type of data (vector or raster) was obtained along with different hardware and software used for data collection. In addition, data processing procedures, hardware calibration and data scope were obtained to assess the standards at the data gathering stage. An idea of data categories and classes were obtained based on format and brief descriptions to characterise the different datasets used in every project. This also provides an indication of the potential errors and limitations associated with the data gathering process.

The procedures associated with data updates and interoperability was obtained in order to assess the scope and frequency of data upgrade and associated digital transfer options and previous data inclusion.

MSUO Project Name	Project Acronym	Lead Partner	Other Partners	Start date	end date	website
North Sea Programme						www.interregnorthsea.org
Safety at Sea North Sea Region	S@S	Norwegian Coastal Administration	21	Sep-04	Sep-07	http://www.safetyatsea.se
Northern Maritime Corridor North Sea Region	NMC	Exec Comm of Northern Norway	8	2002	2005	http://www.northernmaritimecorridor.no
Motorway of the Northern Seas	NMC II	Rogaland County Council	8	2005	2008	http://www.northernmaritimecorridor.no
Save the North Sea		Keep Sweden Tidy Foundation	6	May-02	Dec-04	http://www.savethenorthsea.com
Pushing Offshore Wind Energy Regions	POWER	Bremerhavener Gesellschaft für Investitionsförderung und Stadtentwicklung mbH (BIS)	37	Jul-04	Jul-07	http://www.offshore-power.net/
Forum Skagerrak II		Sweden		Dec-03	Jul-07	http://www.forumskagerrak.com/
Scoping Study on Maritime Safety & Marine Pollution		umbrella for programme				http://www.interregnorthsea.org/
Combined functions in coastal defence zones	Comcoast	Rijkswaterstaat	9	Apr-04	Dec-07	http://www.comcoast.org/
North West Europe Programme						www.nweurope.org
Espace Manche Development Initiative	EMDI	Region Haute-Normandie	20+	Oct-04	Oct-07	http://www.emdi.certic.unicaen.fr/en
CYCLEAU	CYCLEAU	UK Environ Agency (SW England)	13	Oct-04	Oct-07	http://www.cycleau.com/
Creating a Sustainable Framework for ICZM	COREPOINT	UCC CMRC	11	Nov-04	Apr-08	http://corepoint.ucc.ie/
Schéma d'Aménagement Intégré du Littoral	SAIL II	Essex County Council	14	Jul-02	Jul-05	http://www.sailcoast.org/index.shtml
Freight Intermodiality & Exchange on Sea & Straits in Europe	FINESSE	SEEDA (SE England Development Authority)	9	Dec-03	Feb-06	http://www.finesse-project.net/
Marine & Yachting 2 in Lower North Sea & Irish Sea	MAYA II	Provincie Zeeland	13	Dec-03	Jul-06	http://www.maya-net.org/
Development of a Framework for Mapping European Seabed Habitats	MESH	JNCC	11	Apr-04	Apr-07	http://www.searchmesh.net/
Ports & Nature, Striking a New Balance	NEW! DELTA	Province of S. Holland	9			http://www.newdelta.org
Habitat Mapping	HABMAP	Countryside Council for Wales	5	Dec-03	Dec-06	http://www.habmap.org
Marine Senstive Areas in the Netherlands	SensMap	National Institute for Coastal & Marine Management, Netherlands	?	Jun-05	Jun-06	w.ecoserve.ie/index.php?option=com_content&task=view&id=98&Ite
Northern Periphery Programme						www.northernperiphery.net
Safety at Sea Northern Periphery	S@S	The Maritime and Coastguard Agency, UK	4			http://www.maritime-safety.org/S@S-NP-Contact-Information-g.asp
Northern Maritime Corridor Northern Periphery	NMC	Exec Comm of Northern Norway	8	2002	2005	http://www.northernmaritimecorridor.no
Baltic Sea Region Neighbourhood Programme						www.spatial.baltic.net
Baltic Master				Jul-05	Dec-07	http://www.balticmaster.org
OILECO	OILECO	Helsinki University, Finland	3	Apr-05	Dec-07	http://hykotka.helsinki.fi/oileco/
Programme for Civil Protection	EuroBaltic II					http://www.eurobaltic.srv.se/
Coastal Zone Management in the Baltic Region	Coastman	Royal Institute of Technology (KTH), Sweden	6	Mar-04	Mar-07	http://www.coastman.se
Intermodality and Interoperability in the Baltic Sea Area	InterBaltic	Klaipéda science and Technology Park (KSTP), Lithuania	42	Dec-05	Dec-07	http://pn.nfk.no/interbaltic/

Table 1: MSUO Other Maritime Projects Contacts

MSUO Project Name	Project Acronym	Lead Partner	Other Partners	Start date	end date	website
Baltic Master Maritime Saftey Across Borderd	BalticMaster	Region Blekinge, Sweden	41	Jul-05	Dec-07	http://www.balticmaster.org/index.aspx?page_id=1
Atlantic Area Programme						
Emergency Response to Coastal Oil, Chemical and Inert Pollution from Ships	EROCIPS	Devon County Council	14	Nov-05	Nov-08	http://www.erocips.org/
General Programme						
PPC 1a : Risk Assessment and Acceptance		BMT Ltd			May-06	
PPC 1b : The Cumulative Effects of Small Accidental and Operational Oil Spills from Shipping						
PPC 2a: Audit of Marine Data Sources						
PPC 2b: Information Technology and GIS: Common Protocols and Procedures		CSA				
PPC 2c : Passenger Vessel Safety		World Maritime University				
PPC 2d: Training and Maritime Safety Expertise						
PPC 3a : Maritime Safety Seminars						
PPC 3b : Creating Synergies: The development of European Advanced Information Systems (AIS)						
PPC 3c : Application of Protection Measures to Motorways of the Sea						
Other Contacts						
European Maritime Safety Agency	EMSA	EU				http://www.emsa.europa.eu/
British Oceanographic Data Centre	BODC	University of Liverpool				http://www.bodc.ac.uk/
SeaZone Solutions Limited	SeaZone					www.seazone.com
Nowcasting International	Nowcasting		5			www.nowcasting.ie
Oil Spill Resources Limited - Global Alliance	OSRL					www.oilspillresponse.com
Logica CMG	Logica					www.logicacmg.com
Techworks	Techworks					www.techworks.ie
Marine Environement and Security For the European A	MERSEA	EU	40			www.mersea.eu.org
Eurpean Global Ocean Observing System	EuroGOOS		3			
Helsinki Commission Baltic Marine Environment Protection	HELCOM		10			www.helcom.fi
OSPAR Commission	OSPAR					www.ospar.org
REMPECK (Data on Mediterranean Shipping)						
ITOF						
GMES			31	2004	2008	www.gmes.info
Joint Nature Conservation Committee	JNCC	UK Government Agency				www.jncc.co.uk
Other Projects						
Sea Level Change Effecting the Spatial Development in the Baltic Sea Region	SEAREG	Geological Survey of Finland				
Integrated Coastal Zone Development in the Baltic Sea Region	BALTCOAST	Ministry for Labour, Building and Regional Development of MV				
Integrating the Seaways of the Southern Baltic Sea into the PanEuropean Transport System	BALTIC GATEWAY	Region Blekinge, development partner				
BALTIC SEA INFORMATION MOTORWAYS	BaSIM	Technology Centre Lübeck				
	SUMMERI I and II					

Table 1: MSUO Other Maritime Projects Contacts

The confidentiality and constraints of the data, the data searching interface, the restrictions and availability of the data were investigated. The final section of the questionnaire outlines the dissemination and publication characteristics and the potential for data re-usability

2.2 Metadata Questionnaire Results:

Response to the questionnaire was poor primarily due to the absence of data managers on holidays. In some cases no data manager was appointed. Some of the projects co-ordinated by MSUO do not directly deal with GIS data and therefore didn't provide any information. Because there was no response from many of the projects their websites were interrogated to find out what data sets were being used and shared. In some cases where password access to data was required an application was made and access was granted. It was apparent that some of the MSUO projects have partners that actively gather data. Other projects access data from existing repositories without a requirement for data processing. Most projects described in this report are from the North West Europe programme and from the Baltic Sea programme. The "SAFESEANET " project was also investigated to identify progress on data integration between EU public bodies involved in maritime safety. SAFESEANET, a European Platform for Maritime Data Exchange between Member States' maritime authorities, is a network/internet solution based on the concept of a distributed database. The information on SAFESEANET is constantly available, reliable and confidential. Access to the Central Index is restricted and secure yet available 24/7 on TESTA. A summary of the current status of TESTA is provided in Appendix 2. TESTA is the European Community's own private, IP-based network. TESTA offers a telecommunications interconnection platform that responds to the growing need for secure information exchange between European public administrations. It is a European IP network, similar to the Internet in its universal reach, but dedicated to inter-administrative requirements and providing guaranteed performance levels. In the future sharing of data sets between MSUO projects may be through the medium of SAFESEANET and TESTA.

2.2.1 Data Characterisation:

Most of the data acquired by MSUO projects is obtained by differential GPS technology to produce point and vector data. Seventy five percent of the projects produce raster data as grids. The boundaries of these data are defined by digitising polygon coverage based on data attributes and geographic limitations. These are primarily derived from processing of attributes of the vector data to generate thematic maps in 50% of the cases, or represent a layer of data interpolated from *x,y* coordinates. The most popular method of expressing the data is through the use of Arc grids generated using ArcGIS. Some raster data is also output in GeoTIFF, DEM and jpeg formats. The vector data included in the respective databases generally share a common number of attributes (in most cases 5) such as FID (or the identifier for the each individual data entry), Shape (a text attribute to differentiate at the data entry stage point, polygons and lines), POLYGON (an independent polygon identifier using for the merging of datasets), GUI (a 2 letter country code and 6 digit code based on ISO 3166-1 known as the Globally Unique Identifier) and additional fields that reflects the data type. There is an overall tendency to try and reduce the amount of data attributes in order to make the data more manageable and easily referenced. The temporal aspect of the data is generally recorded by introducing the date and time expressed in 2 separate fields.

The standards relating to raster data also constitutes a large part of the metadata included in the MSUO projects. A range of different sets have been report as included in the different project databases. The more common ones and those observed in 75% of the projects relate to layers of data based on the attributes of the measurements included in the vector dataset. These are commonly generated to make statistical analysis of the attribute occurrences with other parameters monitored. These data are mostly displayed as Arc Grids and as DEMs if topological information is used. At least 50% of projects deal with older raster image data. This is generally included as georeferenced jpeg or GeoTIFF images digitized into the database.

Datums and projections used in each project are dependant on the area covered by the project. Eighty per cent of the projects appear to be using datum WGS84 and the relevant projection for the geographical spread of the metadata.

Polygon shapefiles are generated to determine the outlines of either interpreted metadata attribute maps, or images from remote mapping surveys such as satellite/airborne images, air photos, multibeam, side scan sonar or 3D exploration seismic. A global project bounding polygon is then generated to outline the geographical limitations of the project. This is used as a reference in the metadata attribute tables as a geographical reference.

All of the projects who responded reported using previous or existing data for incorporation in their datasets. This was achieved by either digitising existing map data (60%) from raster images or recording through an interview process of other researchers and by incorporating these verbal responses into a standard GIS format.

The potential errors associated with the inclusion of data from different datasets can be hard to quantify during the data acquisition process. In the case of raw data this may be associated with hardware calibration errors. However limitations related with dataset inclusion are minimised through the raw data processing procedures specific to each project. On the other hand, in over 85% of the projects, the major errors occur with the inclusion of older data. Conversion errors in MS Excel and MS Access are common in 75% of the procedures presently adopted. This is also true of the inclusion of metadata in XML and ArcCatalog when different standards from different project regions and projects completed prior to the INSPIRE initiative are applied. Errors associated with digitising existing or older raster based data generate inaccuracies associated with point, polygon or vector locations. These errors can sometimes be hard to quantify.

Fifty per cent of the projects show that there are time limitations to the datasets. This is true particularly of habitat monitoring programs. Changes in biological nomenclature structure and their associated attributes within the dataset introduce errors. These also include spatial changes over time. Modification of data inclusion procedures varies throughout the course of a project.

2.2.2 Data Processing Procedures and Software:

Few of the respondents provided detailed information concerning data processing procedure. The respondents treated data processing as the basic technique of incorporating raw collected data to be included in the database. The MESH project produces a number of habitat map shapefiles based on existing data and on additional data gathered in the field. A *MESH Data Exchange Format* has been put in place to standardise raw data before its inclusion. This determines the common attributes of different datasets and standardises data projection prior to inclusion in the final dataset. Topology testing and cleaning are also performed prior to incorporation. The remaining projects gather data as *x,y* coverages in

Microsoft Excel or Microsoft Access. These are subsequently imported into the database using the correct projections (e.g. Lat/Long, WGS 84 etc).

A brief description of the software used in the compilation and presentation of metadata throughout the different MSUO projects was provided by the different database managers. One hundred percent of respondents are using ESRI ArcGIS (v 9.1). Seventy five percent of the raw data gathered by the projects is obtained using Excel or Access. These are imported into ArcCatalog and subsequently included in the main database. All the data managers contacted confirmed that staff responsible for the construction, maintenance and update of the project database were given appropriate training.

2.2.3 Update and Interoperability:

Seventy five percent of the projects contacted have shown that data update frequency occurs on a yearly basis. The remaining 25% indicate that this takes place at six monthly intervals.

The changes relate to the increase in geographic coverage of the primary metadata entries or updates to the existing data.

At present interoperability between projects is very basic. Vector or raster data is only available in two forms, a visual publicly available web based GIS database (25% of the respondents) or by individual request for data to the relevant database manager. In this case the data is transferred by CD, ftp, e-mail or secure website log in. However individual project partners are granted access to all datasets from a secure part of the project website.

2.2.4 Dissemination, Publication and Data Transfer Options:

Most of the MSUO projects have a large number of partners. These have different roles in the projects and are often located in different countries. Several partners may be gathering metadata for inclusion in the final GIS database.

Ninety percent of the projects who responded to the questionnaire indicate that data is transferred from one partner to another by means of a local network. Access to data is achieved through an ftp site that permits the fast transfer of large data files.

All respondents confirmed that the information gathered and stored in the datasets of each project is available in the public domain for dissemination and publication. This is achieved through the use of a WebGIS interface on the project website through which the different data layers are published. When these data are specific to the project the resulting digital vector files are made accessible for viewing to the public but the actual raw metadata remains the property of the project partners. In 75% of cases this is made accessible to them through the use of username and password protected access of the project website, but in 25% of cases this is only made available through direct request to the data manager.

Where previous or external data are included in the GIS database the original maps are made available to the public through the WebGIS interface with the permission of the original data owner. Sensitive data specific to the project is often listed in the WebGIS interface of the project websites but not specifically made available in the public domain. There are different dissemination procedures based on the data included in the database and unique agreements between partners. As these procedures are project and case specific, they cannot be easily summarised. However, the majority of the dataset is shown on the WebGIS interface based on individual thematic layers. This is true of the majority of the vector data gathered. In 80% of cases the raw data is never made available on the internet.

The inclusion of raster data to the WebGIS interface is related primarily to thematic layers of dataset attributes. Raster imagery, which is commonly used in the various datasets, is often not included in the web based interface. Raster imagery files are too large and cumbersome to make available on an interactive map server. However the ArcCatalog reference for the relevant images is often included in the public interface.

The less sensitive data is targeted to a wider audience and made available generally as a list of raster images derived from the relevant dataset layers. This is true of approximately 50% of the projects and varies based on the project specific confidentiality agreements.

In all of the projects specific procedures are in place for the publication and reproduction of data. In most cases project partners are able to access and use the data based on the confidentiality and data use agreements specific to the project. This is facilitated by the metadata managers. Protocols for the release of data to third parties for the purpose of publication are provided by the various agreements. However applications for data release are made to the data managers and are assessed on an individual basis.

Secure (password protected) parts of the project websites are used for frequent upload of new or revised data by project partners to the database. A secure *ftp* site connection for the transfer of large files is used by 50% of the respondents which constitute the more data intensive projects contacted. The projects with lower data handling requirements allow partners to submit data on CDs and provide additional backup options. All of the projects use metadata submissions forms to characterise the type of data to be included in the database.

2.2.5 Data Storage and Confidentiality:

The range of MSUO marine data projects, the large number of partners, and the variety of project scopes and topics means that there are different levels of confidentiality and data storage options used by MSUO projects. Confidentiality agreements are applicable to all partners and relate specifically to data disclosure and generally follow the guidelines of directive 2003/4/EC. These guidelines clarify the basis upon which data is submitted to the project, the circumstances under which it can be held and also disseminated.

2.2.6 Data Re-usability:

A broad spectrum of thematic marine metadata is now available for the regions covered under the MSUO remit. This data has been gathered between 2001 and 2007. There are frequent data overlaps between different projects. This leads to datasets from different projects being re-used.

Nearly all projects are making provisions for simplified updating of the datasets after the completion of the project. Provisions for the maintenance and upkeep of the data subsequent to the project end are common practice and the data managers are responsible for database maintenance. The metadata from each of the projects can be made available by request to the metadata managers.

3. DATA AMANGEMENT METHODOLOGIES:

3.1 Common issues of potential difficulty with GIS data.

The advent of relatively user-friendly desktop GIS applications has made it possible for professionals from a wide range of diverse backgrounds to access, manipulate and utilise GIS technologies. GIS means different things to different users. For a large number of users it simply represents a convenient way of customising maps, while others may regard it primarily as a means of geocoding data, or even as a tool for modelling earth surface processes. While different applications require different approaches, a number of core data management issues are common to all applications of GIS.

These core issues apply equally across the diversity of domains within which GIS is normally utilised. The GIS needs of civil engineers, environmental managers, landscape planners, governmental agencies, private developers, academic research professionals and the general public may differ markedly, but common difficulties arise for all where GIS data management is not given adequate consideration. While these difficulties may often be overcome by the originator of GIS outputs, difficulties often become insurmountable when GIS outputs are passed on to a second-generation user.

The commitment of MSUO to “co-operating to create, maintain and implement a safer maritime environment (MSUO, 2006)” can be facilitated by the implementation of standards for sharing GIS data between MSUO projects. GIS outputs from MSUO projects are now required to conform to standards that ensure their utility for second-generation GIS users.

GIS can be considered as having five principal integrated components. These include: people, software, data, applications, and hardware. Table 2 below outlines common problems that apply to GIS outputs across a wide range of application areas. Solutions to these problems will be addressed later (Section 5) in this document.

	GI Data Issue	Problem
People	Training of staff	Little or inadequate training causing problems to data originators and second hand users
	Personnel issue	Staff not directly assigned to GIS data management
Software	Adequate Software	Graphic manipulation programs do not have the flexibility to handle GIS data
Data	Interoperability	Non-interoperable vector, raster, ASCII, RDMS, and graphics formats cause problems
	Attributes	Inappropriate data attributes can make GIS data unusable for second-generation users
	Editing Rights	Data reliability issues arise from multiple users have editing rights to GIS data
	Digitising	Poor digitising of data leads to serious reworking of GIS datasets for second-generation GIS data users
	Raster	Poor quality image data makes analysis difficult and maps unprintable
	Portability	Large files are difficult to use and cause problems in data dissemination, security and transfer
	Metadata	None or poor data can render data useless or cause confidence problems to second generation users
	File naming	Non-logical file names causes difficulties for second-generation users
	Vector formats	Software-native formats often problematic
	Raster formats	Software-native formats often problematic
	Spatial Database formats	Software-native formats often problematic
	Database formats	Software-native formats often problematic
	ASCII formats	Non-delimited files
	Large file size	Large files and email reliability
	Web transfer	Large files and email reliability / security
	Hardware	Data quality reflects inadequate system
	Data quality - vector	More files than needed is a problem
	Data quality - raster	Poor quality image data
	Data quality - attributes	None or poor attributes
	Data filing	Poor data housekeeping can make data retrieval an unnecessarily involved process
Applications	Projections	No projection / incorrect projections used causes problems to second generation data users
	Georeferencing	Incorrectly georeferenced raster data can make data very difficult to use or even unusable
Hardware	Centralised Storage	Lack of centralised storage for GIS data causes problems for data retrieval, security and quality
	Processing Hardware	Retrieval, security and quality issues arise if hardware is shared for GIS data processing and storage
	Access	GIS system administration by untrained staff leads to data quality, cataloguing and security problems
	GPS survey data	Device-native formats often problematic
	Remotely sensed data	Software-native formats often problematic

Table 2: *Common problems in GIS outputs*

3.2 GIS data standard Issues

GIS interoperability

Interoperability refers to the capability of software and hardware to share data with different machines from different vendors. Interoperability is crucial to ensure that GIS data from all MSUO projects can by used by GIS data users on subsequent MSUO projects.

A number of interoperability standards are in existence, all providing detailed instructions for developers and users of GIS data. A review of GIS data standards with the addresses of relevant websites is given in Appendix 3. The Open Geospatial Consortium (OGC) is the most developed of the interoperability standards. OGC is an international industry consortium of private companies, government agencies and universities participating in a consensus process to develop publicly available interface specifications. OGC specifications support interoperable solutions that "geo-enable" the Web, wireless and location-based services, and mainstream Information Technology. As a consequence, OGC specifications focus primarily on the concerns of technology developers of all kinds of GIS applications. The significance of OGC standards for the average GIS user is that its goals and objectives outline the manner in which interoperability has become a central to GIS applications that involve data dissemination.

OCG strategic goals

- To lead in the establishment of standards that allow seamless integration of GIS content into a range of areas not traditionally considered the domain of GIS data.
- The provision of free and openly available standards.
- To accelerate market assimilation of interoperability research.

While the majority of the aims of the OGC are pitched somewhat above the concerns of the general GIS user, their emphasis on seamless integration between software and hardware platforms is of relevance to any GIS output that will be disseminated by any means. Therefore, the principle of interoperability is not limited to high-end users alone. Data interoperability is now recognised as an essential component of GIS data management in the environmental and marine fields. The concept of interoperability is therefore of central importance for MSUO projects, and will be applied to the commonly occurring problems identified in Table 2, to create a set of GIS instructions for MSUO data managers.

GIS software Interoperability

At a global level, the most commonly used desktop dedicated GIS software applications come from ESRI (Environmental Systems Research Institute) and MapInfo. Each of these vendors provides a suite of software applications intended to meet the needs of GIS users for desktop mapping, database management, data visualisation, process modelling and web delivery. Worldwide, ESRI and MapInfo account for approximately 97% of all GIS users. The percentages using various GIS software are outlined overleaf (Table 3).

Software vendor	Products	% users
ESRI	ArcView, ArcInfo, ArcSDE, ArcIMS, ArcPad	77%
MapInfo	MI Professional, Vertical Mapper, MapX, MapExtreme, MapBasic	20%
Intergraph	MGE, GeoMedia, FRAMME	16%
Bentley	Microstation	12%

***Table 3:** Percentages of GIS professionals using different GIS software products. (source: gisjobs.com, 2006)*

As the dominant desktop GIS worldwide, the majority of GIS outputs from MSUO projects originate in ESRI formats. Therefore, the formats for vector, raster, image, attribute, text and other GIS-related data that issue from all MSUO projects in the future will have to be ESRI compatible. The dominant position held by ESRI has led to all other vendors adopting ESRI import and export routines to handle ESRI data. MapInfo, Intergraph, Bentley, AutoCAD and a host of smaller vendors all supply export, import and/or direct-read filters for ESRI data. The range of native ESRI formats is quite numerous however, and the range of filters, importers and format converters that are used by all the other vendors to handle ESRI data are more numerous still.

In order for interoperability to be workable, it must be easy to understand and easy to implement. A single format that can be read by all GIS software applications is the ideal. Such standards do exist, but are often lost in the array of options and sub-options available to the user. Non-specialist users are even more likely to be confused. Interoperable formats for vector, raster, image, attribute, text and other GIS-related data will be outlined in section 5.

For details of the main software vendors' positions on interoperability refer to the following URL's:

ESRI:
http://www.esri.com/software/standards/index.html

MapInfo:
http://www.mapinfo.com/location/integration?txtTopNav=686a2545d8a37f00dev-vcm100001a031dc7____&txtLeftNav=308ef23c222d1010vcmprd00000034021dc7____

Intergraph:
http://www.intergraph.com/interoperability/default.asp

Bentley Systems:
http://www.bentley.com/en-US/Markets/Geospatial/Products/Geospatial+Products.htm

3.3 Metadata

Metadata is structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities (Durrell, 1985). Put more simply, the term metadata refers to data about data, or more straightforwardly, information about data. The benefits of metadata are:

- It makes the data useable for second-generation users.
- It provides an inventory of the dataset(s).
- It helps determine data reliability and data currency.
- It facilitates accuracy verification.
- It saves time and financial resources.

Metadata is crucial wherever GIS data is intended for dissemination to second-generation users. Without metadata, second-generation users cannot verify the content, relevance and quality of the data. Metadata is essential wherever data is made available over the web (via web portals for downloading GIS data from data repositories). Without it, diverse datasets could not even be identified, let alone used. Data from all MSUO projects will be available for dissemination to all subsequent MSUO projects. This means that metadata will have to be supplied for all GSI datasets, for all projects.

Two main metadata standards exist for GIS data. ISO 19115 is the *de facto* standard for GIS metadata. The other main metadata standard; namely the U.S. FGDC metadata standard (which is currently implemented by ESRI in their ArcGIS product) is currently in the process of harmonising with ISO 19115.

ISO 19115 defines a comprehensive schema for describing geographic information, specifying information about the identity, geographical extent, quality, spatial and temporal characteristics, spatial reference characteristics and distribution of digital geographic data. It is applied to the cataloguing of GIS datasets and online GIS data repositories, and for the description of GIS datasets and the individual geographic features and properties within these datasets. It achieves this by providing specification for defining:

- Mandatory and conditional requirements for the adequate description of geographic datasets.
- The minimum set of metadata required to serve the full range of metadata applications, which include data discovery, determining data fitness for use, data access, data transfer, and use of digital data.
- Optional metadata elements designed to facilitate a more extensive description of geographic data, if required.
- A method for extending metadata to fit specialized needs.

Though ISO 19115 is applicable primarily to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data (source: http://www.iso.org).

The mandatory and conditionally mandatory requirements of ISO 19115 form the basis of the MSUO metadata form outlined in section 5.

4. **INTERNATIONAL BEST PRACTICE*:***

As part of the Project Cooperation Initiatives of MSUO, this study aims at developing suitable protocols to be adopted in the future for the development of GIS databases. As part of this a review of the common best practice procedures is included below. These are aimed at providing the best international guidelines for the development of a data protocol handbook for MSUO.

The commitment of GIS vendors to interoperability has already been discussed. These commitments arise out of recognition of a demand from GIS users. Data is the raw material of GIS, and the data needs of users are accommodated by national mapping agencies, State bodies, semi-state organisations, Non-Governmental Organisations (NGO's) and private sources. Most of these agencies, bodies and organisations have recognised the need to formalise their GIS data management protocols.

4.1 INSPIRE Methodologies and Recommendations:

The INSPIRE directive by the European Parliament and Council (2004/0175 COD) is aimed at providing a common infrastructure to provide better spatial information for policy making and implementation within the Member States. The directive focuses primarily on the improvement of monitoring information of the environment, air, water, soil and natural landscape. This proposal has been aimed at reinforcing existing directives such as GMES and GALILEO aimed at promoting the availability of public sector information. The INSPIRE directive is primarily aimed at reviewing metadata, spatial data sets, spatial data services, networking services and technologies, agreements on sharing, access and use, coordination and monitoring mechanisms and processes and procedures.

The general situation on spatial information in Europe is one of fragmentation of datasets and sources, gaps in availability, lack of harmonisation between datasets at different geographical scales, and duplication of information collection. These problems make it difficult to identify suitable data, and to access and use data that is available. Awareness is growing at national and at EU level about the need for quality geo-referenced information to support environmental policy and environmental management. The INSPIRE (INfrastructure for SPatial InfoRmation in Europe) initiative is a proposal for a new EU Directive, to establish an infrastructure for spatial information in the EU. The proposal aims to provide a framework for the provision of relevant, harmonised and quality geographic information for the purpose of formulation, implementation, monitoring and evaluation of Community policy-making (source: http://www.ec-gis.org/inspire/home.html)

The central principles of INSPIRE are:

- Data should be collected once and should be maintained at the level where this can be done most effectively.
- It should be possible to seamlessly combine spatial data from a range of different sources and to share it between many users and applications.
- Spatial data should be collected at one level of government and be shared between all levels of government.
- The spatial data that is required as a support for governance should be available on conditions that do not restrict its extensive use.
- It should be easy to discover which spatial data is available, to evaluate its fitness for purpose, and to know which conditions apply for its use.

The INSPIRE proposals are in the process of being refined, and will form the basis of a Directive for interoperability of GIS data at some point in the near future. Examples of implementations of interoperability plans for Environmental and Marine organisations and projects are already available. Of the numerous examples that exist, relevant marine examples include:

- The framework for Mapping European Seabed Habitats (MESH) Review of standards and Protocols for Seabed Habitat Mapping.
- Marine Electronic Highway (MEH).
- The Australian Ocean Data Centre (AODC) Marine Community Profile of ISO 19115
- The Marine Metadata Interoperability Project (MMI).

Examples one and two above relate primarily to data interoperability, while items three and four focus more specifically on the issues of metadata as a support for data interoperability.

4.2 Large Global GIS Data Projects:

A selection of large ongoing projects that use GIS marine datasets were examined to identify best practice for data management procedures.

4.2.1 Global Ocean Observing System (GOOS):

GOOS is a permanent global system for observations, modelling and analysis of marine and ocean variables to support operational ocean services worldwide. GOOS provides accurate descriptions of the present state of the oceans, including living resources; continuous forecasts of the future conditions of the sea for as far ahead as possible, and the basis for forecasts of climate change. In the remit of the GOOS project, Initial Operational Systems (IOS) for observations and data includes a comprehensive data system, specifying procedures for collection, quality control, comparison of observations from different sources, dissemination and utilisation of metadata.

The GOOS project is a highly distributed system based on the contribution of numerous operational agencies, data centres, and research organisations from both the oceanographic and meteorological communities. The information management is accomplished in an iterative fashion connecting the data collection stations that contribute to the programme. This is achieved through the Initial Observing System (IOS) which has the role to review, rationalise and modify the data being collected in the different parts of the project prior to establishing additional, data specific analyses. This technique has proven invaluable for dealing with the diversity and wide global distribution of data allowing the system to be highly distributed and evolving.

Data management characteristics for the GOOS project varies due to the nature of ten data categories outlined below:

- Operational Marine Coastal and Ocean Short Range Forecasting and Analyses
- Seasonal to Interannual Climate Prediction
- Numerical Weather Predictions
- High-Quality Procedures for Climate products
- Biodiversity and Habitat
- Natural and Man Made Hazards
- Environmental Indices

- Fisheries Productivity
- Satellite Data Processing Systems
- Regional, Integrated Systems

The broad diversity of the data types gathered under the scope of the GOOS project range requires a variety of different data management methodologies adapted to each category.

The more time dependant data gathered such as marine, coastal and ocean forecast data from ships and buoys or numerical modelling data is time dependant and has to be included in to the database prior to its cut off. Hence the real and near real data gathering and processing must be highly automated and include a second integration process prior to its distribution. This is generally achieved by submitting simple text or ASCII format data.

Standard physical variable data such as seasonal or interannual climate data, or biodiversity or habitat data is partly time dependant but not as critical due to its in situ collection. This however requires having specific standards and protocols to respect identification and description parameters. More significant quality control steps need to be taken when including these into a dataset. Manual intervention is frequently required to remove duplicates, pre-process the inputs and provide internal consistency prior to inclusion in the database. Due to the large number of organisations and country sources for the GOOS project much of this data is significantly standardised prior to it inclusion in the database.

Standard data processing procedures and data management programmes used on the GOOS project were based on the standards and methodologies of the International Oceanographic Commission (IOC) and World Meteorological Organisation (WMO). A detailed data flow charted for all the GOOS data and its partners is provided in figure 1 below.

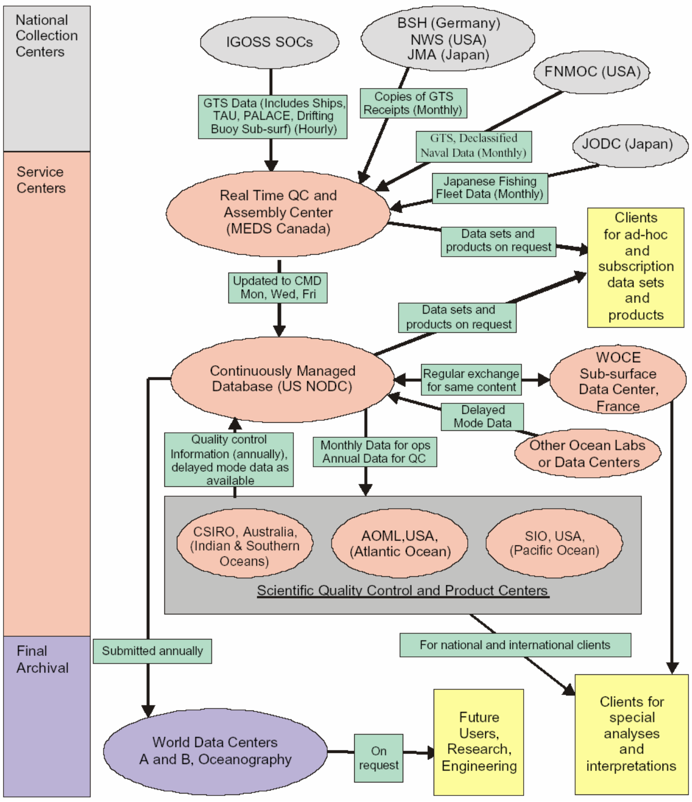

Figure 1: *Data Flow Chart for the GOOS project*

4.2.2 Global Monitoring for Environment and Security (GMES):

GMES is an EU based initiative for the implementation of information services dealing with environment and security with the aim to facilitate the availability of reliable data to end-user, national policy and decision makers as well potential investors. This initiative has occurred on the basis that information in these sectors has often been inadequate, unreliable or out of date in the majority of the member states the lack of continuous monitoring has lead to gaps and incompatibilities of data. The main objective is to secure sustainable and coherent information for future environmental and security policies. The primary metadata management methodologies for the GMES project follow the rules and guidelines set out by the INSPIRE initiative.

GMES have encountered a variety of data integration problem that have been addressed through a series of specific procedures. Specifically problems of a technical nature (e.g. functionally different retrieval systems, lack of technical standards, language differences and uncertain data quality), of an institutional nature (e.g. differences between conditions of access and security and use and a lack of incentives to maintain databases) and a legal and commercial nature (e.g. issues of ownership, privacy and confidentiality, intellectual property rights, pricing and licensing) were identified as the main barriers to the integration of metadata. Speed of access of data and seamless real time integration in the relevant databases required standardisation to allow successful delivery to the end users.

The capability to transfer data freely and seamlessly between different information systems is observed to be low and prevents the harmonisation of different data in a consistent format in order to derive and deliver information. The main obstacles to interoperability were found to be:

- insufficient data and exchange standards

- the lack of consensus on sound methods and arrangements for linkage between different data sources (e.g. Earth Observation, ground-based and management-derived data) and systems for generating information

- missing arrangements for integration across different sectors and policy areas, for example, the linkage of environmental and non-environmental data.

Data- and computationally intensive areas of GMES, such as real-time modelling based on earth observation data or climate modelling, high-performance networks and GRID-based computing were identified as being essential for mining, sharing and analysing data and visualising results. A structured framework for data integration and information management, for European shared information was focused on key architectural and user-oriented requirements implemented by GMES:

- **Openness**, based on agreed open standards, facilitating **seamless communication** and **Interoperability**, i.e. the ability of different devices or systems (usually from different vendors) to work together, as well as enabling **user service autonomy**;
- **Federated architecture**, enabling systems to grow and evolve;
- **Simplicity** of architecture (e.g. modularity of components), to break the complexity barrier, systems must be made easier to design, administer and use;
- **Self-configuration, programmability**, **scalability** (e.g. to handle various levels of operational load and external conditions);
- **Dependability**, i.e. the system's resilience to security threats or breakdown;
- **User-friendliness** of services and interfaces, e.g. in the handling of user request services, access control, workflow management, delivery management, visualisation, data extraction (e.g. “multilinguality”), multiuser sessions, administration;
- **Data security**, protection of provider and user data against alteration, theft and misuse;
- **Quality of service;**
- **Ubiquity of access**, including global reach**.**

The GEANT (EU based initiative) along with GRID provides the necessary infrastructure to access all main public data sources and to support not-for-profit activities. Thematic sub-networks were subsequently developed, based on the INSPIRE recommendations, to encompass existing networks, as well as a range of new scientific networks bringing together scientists working in specific domains (atmosphere, ocean, in-land water, coastal zones etc.). GMES impacts directly on a variety of end users but also relies heavily on the involvement of different types of organizations in the public and private sector to successfully deliver the metadata infrastructure networks. These range from European Commission services, agencies of the European Union, inter-governmental organisations; non-governmental organisations as well as small and medium service companies (SMEs) involved at the core of the GMES implementation. In the light of this, identifying and addressing the needs for information, technologies and effective data policies has been crucial for the implementation of GMES and has been achieved through a variety of contractual agreements.

An example of the data flow chart for the GMES projects is summarized in figure 2 below. The diagram shows the processes involved with the inclusion of Marine Core Data within the remit of the GMES projects.

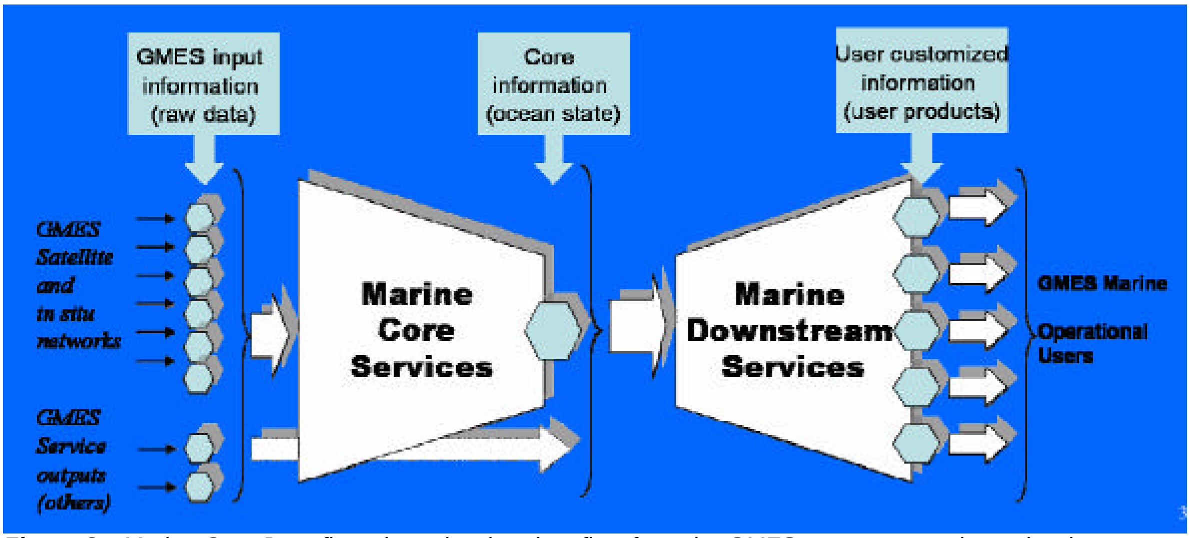

Figure 2: *Marine Core Data flow chart showing data flow from the GMES partner networks to the downstream end users.*

4.2.3 International Marine Organisation (IMO):

IMO is the United Nations' specialized agency responsible for improving maritime safety and preventing pollution from ships and is structured similarly to the UN with a general assembly and more sector specific committees and sub-committees.

The scope of the IMO is to provide machinery for cooperation among Governments in the field of governmental regulation and practices relating to technical matters of all kinds affecting shipping engaged in international trade; to encourage and facilitate the general adoption of the highest practicable standards in matters concerning maritime safety, efficiency of navigation and prevention and control of marine pollution from ships.

The organisation provides a means of formulating legislation relating to different aspects of safety at sea. This process is facilitated through the emplacement of the Global Integrated Shipping Information System (GISIS) which provides real time safety information for the member countries. A number of affiliate bodies and programmes to IMO provide the technical data background for the implementation of legislation and the GISIS. These are:

- GESAMP (http://gesamp.imo.org/) - The Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP) was established in 1967 by a number of United Nations Agencies. GESAMP deals with all scientific aspects on the prevention, reduction and control of the degradation of the marine environment to sustain life support systems, resources and amenities.

- **Global Ballast Water Programme** (http://globallast.imo.org/) - The Global Ballast Water Management Programme (GloBallast) is a three year, US$10.2 million initiative under the International Waters portfolio of the Global Environment Facility (GEF).

- **Global Marine Litter Information Gateway** (http://marine-litter.gpa.unep.org/)- Joint UNEP GPA Coordination Office/International Maritime Organization marine litter (marine debris) node of the GPA Clearing-House Mechanism.

- **London Convention 1972** (http://www.londonconvention.org/)- Website for the Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter, 1972

- **REMPEITC-Carib** (http://www.rac-rempeitc.org/)- The Regional Marine Pollution Emergency, Information and Training Center for the Wider Caribbean Region (REMPEITC-Carib is an International Maritime Organization (IMO) office which assists the countries in the region in preventing, preparing for and responding to major pollution incidents.

- **REMPEC** – (http://www.rempec.org/) The Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC). A databases and tools have been developed by REMPEC with a view to assisting in the activities related to prevention of, preparedness for and response to marine pollution emergencies. It contains the TROCS database developed to assist in taking decisions related to marine pollution emergencies caused by hazardous and noxious substances (HNS) and by certain crude and refined oils, and MIDSIS-TROCS, a computerised decision-support system based on the TROCS database.

- **PEMSEA** – (http://www.pemsea.org/index.htm) The Regional Programme for Marine Pollution Prevention and Management in the East Asian Seas region.

IMO does not deal directly with metadata of any description. However, their role is quite important as they represent one of the end-users of maritime databases gathered in the scope of all the other projects and organisations described in section 4 of this report. The importance of the harmonization of data standards of different projects and countries is highlighted by IMO through the implementation of maritime safety policies and legislation.

4.2.4 Other Projects:

Marine Electronic Highway (MEH) projects

A MEH system is a system of technology, people and processes that enables third party access to marine environmental and marine operational data. The MEH concept provides a framework for the implementation of tools to record, store, manage, model, analyse and access oceanographic and other data, and to present the results to a broad base of specialist and non-specialist users. As such, it has many of the attributes of a spatial data infrastructure.

An effective MEH system requires a transferable, distributed, standards-based approach that can handle real-time inputs, multiple data formats, and diverse reporting requirements. It must be independent of proprietary systems and formats, and it must provide system-independent access to all data and information. Since an MEH involves real-time dissemination and data updating, it is uniquely suited to operation over the web, and typically involves the use of web-based spatial data warehouse technologies and web mapping services. (source: http://www.acops.org/Gillespie.pdf).

The concept of a global MEH is receiving some attention (source: http://www.acops.org/Gillespie.pdf), but to date most MEH assessments have been conducted as proof of concept. The largest application of MEH to date is the East Asia Marine Electronic Highway. The World Bank agreed funding in early 2006 for the International Maritime Organization (IMO) to implement the first phase of the development of a regional Marine Electronic Highway in the Straits of Malacca and Singapore. The project aims to create an integrated system of physical infrastructure, hardware and software, processes and resources to improve the safety of navigation and the prevention of marine pollution in the busy seas of the area. (source: http://www.imo.org/Newsroom).

Its implementation signals increasing recognition of the importance of marine data interoperability for the management of Marine safety worldwide. MSUO's commitment to framing interoperability standards for marine GIS data within MSUO projects represents the expression of this recognition at EU level.

The Marine Metadata Interoperability Project

The goal of the Marine Metadata Interoperability (MMI) project is to promote the exchange, integration and use of marine data through enhanced data publishing, discovery, documentation and accessibility, developing web applications and stand-alone tools to enable sophisticated interactions across marine data systems. It aims to simplify some of the complexity of metadata by providing direct straightforward guidance for scientists and data managers on how metadata can be used to find, access, and use suitable data, and how metadata can be applied to their own datasets to make it easier for others to use their data. MMI achieves this by fostering communication and collaboration among its members, and by providing forums for discussion of diverse topics related to marine data management (source: http://marinemetadata.org). MMI's recommendations are based on the International Standard for Geographic information (ISO 19115).

The metadata requirements for marine GIS data for MSUO projects will be based on the core metadata elements of ISO 19115. These are outlined below in Table 4.

Framework for Mapping European Seabed Habitats (MESH) (http://www.searchmesh.net/)

MESH is an international marine habitat mapping programme that started in spring 2004 and will continue until 2007. It consists of a consortium of 12 partners across the UK, Ireland, the Netherlands, Belgium and France, funded under EU INTERREG IIIB. The project will result in common protocols and information systems to be used at a transnational level.

The main output of MESH will be a seabed habitat map for the MESH region, compiled using the European Environment Agency's classification system (EUNIS). The project will include habitat modelling for those areas where information is incomplete or inconsistent so as to predict habitat distribution for unsampled areas. International standards based on the best available expertise will help ensure the quality of future mapping programmes.

A metadata catalogue of mapping studies will be developed to store information defining who undertook each study, when it was undertaken, its purpose and geographical area. This "discovery metadata catalogue" will be implemented along rules recommended by the EU Inspire proposal and ISO 19115. The catalogue will also hold information on the techniques and standards used, in addition to the types of data collected and where these are archived. It will help partners identify sources of data to develop and test protocols and encourage improved meta-data standards for mapping studies. The end products of the MESH project will be:

1. A database of mapping studies (including metadata on all GIS datasets).
2. A web-delivered GIS showing the completed habitat maps.
3. Guidance documentation for marine habitat mapping (including a review of explanatory protocols and standards), for the application of Remote Sensing, Acoustic Sounding, video and imagery capture and in-situ sampling to marine habitat mapping.
4. A report describing case histories of habitat mapping.
5. A stakeholder database and an international conference with published proceedings (source: http://www.searchmesh.net/)

Items 1 & 2 above are components that are requirements of all MSUO projects. The MESH Data Exchange Format (DEF) defines the required format of the GIS vector files that will be transferred between MESH partners. The document is available from the MESH website at http://www.searchmesh.net/PDF/MESHDataExchangeFormat_WEB_v6.pdf. The WebGIS designed for the MESH project is in still under development. It has been reviewed by CSA and provides a good model for use by future MSUO projects.

AODC Marine Community Profile of ISO 19115

The Australian Ocean Data Centre Joint Facility (AODCJF) provides a cross-departmental government approach to ocean data management. It will develop a national multi-agency data management system for Australia to manage the ocean data resources of the partner agencies through a distributed network. The AODC has defined a Marine Community Profile (MCP) which supports the documentation and discovery of marine spatial datasets and forms the foundation of their Marine Catalogue. The MCP has been developed in accordance with the rules established by the International Standard ISO 19115 Geographic information – Metadata, and is comprised of a subset of the ISO standard, including all core metadata elements of ISO 19115 (Source: http://www.aodc.gov.au/index.php?id=19).

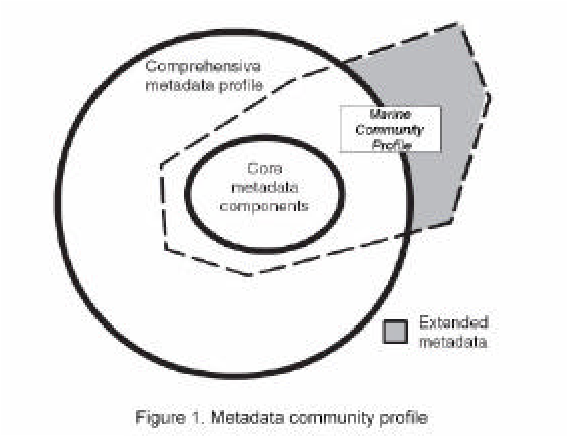

Figure 3: *Metadata element structure for the AODC project*

Figure 3 above shows how additional metadata elements under the Marine Community Profile are not in the International Standard. For this reason ISO 19115:2003 has been extended.

The standard is used to identify the metadata required to describe digital geographic data based on aggregations of datasets, individual geographic features and the various classes of objects that compose a feature. The metadata can be expressed as containing one or more Metadata Sections (UML Packages) and one or more Metadata Entities (UML classes).

The Marine Community Profile uses only a subset of the full number of these elements and has defined the minimum number of metadata elements required to describe a marine dataset. These fields are outlined in Table 4 below and are coded using “M” to indicate a mandatory element; “O” indicating an optional element and “C” a mandatory element under certain conditions.

Core metadata for marine geographic datasets

Dataset title (M) (MD_Metadata > MD_Identification.citation > CI_Citation.title)	**Spatial representation type** (O) (MD_Metadata > MD_DataIdentification.spatialRepresentationType)
Dataset reference date (M) (MD_Metadata > MD_Identification.citation > CI_Citation.date)	**Reference system** (O) (MD_Metadata > MD_ReferenceSystem)
Dataset responsible party (O) (MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty)	**Lineage** (O) (MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
Geographic location of the dataset (by four coordinates) (M) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > MP_Extent > EX_GeographicExtent > EX_GeographicBoundingBox	**Additional extent information for the dataset** (**temporal**)(C) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent)
Spatial resolution (O) (MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale)	**On-line resource** (O) (MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Dataset language (M) (MD_Metadata > MD_DataIdentification.language)	**Metadata file identifier** (M) (MP_Metadata.fileIdentifier)
Dataset character set (C) (MD_Metadata > MD_DataIdentification.characterSet)	**Metadata standard name** (O) (MD_Metadata.metadataStandardName)
Dataset topic category (C) (MD_Metadata > MD_DataIdentification.topicCategory)	**Metadata standard version** (O) (MD_Metadata.metadataStandardVersion)
Metadata language (C) (MD_Metadata.language)	**Metadata character set** (C) (MD_Metadata.characterSet)
Metadata date stamp (M) (MD_Metadata.dateStamp)	**Abstract describing the dataset** (M) (MD_Metadata > MD_Identification.abstract)
Distribution format (M) (MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)	**Metadata point of contact** (M) (MD_Metadata.contact > CI_ResponsibleParty)

Table 4.: *Core metadata elements for the description of Marine Datasets*

A data dictionary describes the characteristics of the metadata defined by the UML models specific to the Marine Community Profile and is accompanied by a dictionary categorising the various UML model package described.

5. **COMMON DATA MANAGEMENT PROTOCOLS***:*

This report has examined common issues of potential difficulty with GIS data. The results of a data management questionnaire were reviewed. The websites of MSUO sponsored projects were investigated to assess ease of access to data and interoperability. Examples of international best practice data management of marine data were examined.

Recommendations for data management of future MSUO projects are proposed in the table below.

GIS DATA MANAGEMENT METHODOLOGIES for MSUO Projects

1 GIS personnel	*1.1*	*GIS management* Individual personnel must be assigned to GIS management tasks. One member of the project team should be responsible for data management.
	1.2	*Training* GIS data management personnel must be trained in the GIS package used by the project.
	1.3	*Data access rights* Access to GIS data should be limited to trained GIS staff
	1.4	*Data editing rights* Editing rights to GIS data should be limited to trained GIS staff
2 Hardware	*2.1*	*Hardware* Dedicated networked storage device or dedicated workstation must be used. This avoids distributed locations for project data on individual PC hard drives or personal folders on servers.
	2.2	*Hardware access* Hardware administration should be carried out by trained IT staff
3 Software	*3.1*	*Software* Dedicated GIS software must be used for GIS work (ESRI, MapInfo is the most commonly used and strongly recommended)
4 Georeferencing	*4.1*	*Projections* All GIS datasets should be made available to subsequent MSUO projects in WGS84 / ETRF89 projection
	4.2	*Georeferencing* All raster files used should be georeferenced to WGS84 / ETRF89 geographic coordinates
5 GIS data formats	*5.1*	*Vector formats* All vector datasets should be made available to subsequent MSUO projects in .e00 format

5.2 Raster formats

All raster image datasets must be made available to subsequent MSUO projects in Georeferenced .tiff (geotiff) or Georeferenced jpg (jgw) format with accompanying world files

Non-image Raster gridded data must be made available in simple ASCII format.

5.3 Relational Database formats

All non-spatial database datasets (excluding those that accompany shapefiles or MapInfo tab files) must be made available to subsequent MSUO projects in Microsoft Access format

5.4 Spatial Database formats

Where spatial database datasets are used, these must be made available to subsequent MSUO projects in ESRI geodatabase format or in Oracle Spatial format

5.5 ASCII formats

Simple text file GIS data should be supplied in tab delimited text format

5.6 GPS survey data

Survey data should be supplied as vector .e00 format only

5.7 Remotely Sensed data

Remotely Sensed image data should be supplied in Band Interleaved Line (.bil), Band Interleaved Pixel (.bip) or Band Sequential (.bsq) formats

6 GIS Attributes

6.1 GIS Data Attributes

GIS attributes should be added to all vector point, polyline and polygon data before exporting to .e00 format

7 Data portability

7.1 File size

Files larger than 10MB should be provided as contiguous mappable tiles

8 Data Administration

8.1 File naming

A consistent logical schema for filenames should be used throughout the MSUO project and cross reference these in metadata form

8.2 Data filing

Files for transfer to MSUO should be organised in logical sub-folders

9 Data Quality

9.1 Data quality - vector

GIS data should be digitized by theme (i.e. multiple thematically related objects in one file)

9.2 Digitising quality

Data managers should ensure that edge-matching in digitised datasets using routines provided by software vendor

9.3 Data quality - Raster

Raster maps scanned from paper sources should be at image quality of 300dpi

10 Metadata

10.1 Metadata

Metadata should be recorded according to the requirements of the MSUO metadata form provided below.

MSUO - Core metadata for geographic datasets

Dataset title (Mandatory) (MD_Metadata > MD_Identification.citation > I_Citation.title)	**Spatial representation type (Optional)** (MD_Metadata > D_DataIdentification.spatialRepresentationType)
Dataset reference date (Mandatory) (MD_Metadata > MD_Identification.citation > CI_Citation > CI_Date.date and CI_dateType)	**Reference system (Optional)** (MD_Metadata > MD_ReferenceSystem)
Dataset responsible party (Optional) (MD_Metadata > MD_Identification.pointOfContact > CI_ResponsibleParty)	**Lineage statement (Optional)** (MD_Metadata > DQ_DataQuality > LI_Lineage.statement)
Geographic location of the dataset (by four coordinates or by geographic identifier) (Conditonally mandatory) (MD_Metadata > MD_DataIdentification.geographicBox or MD_DataIdentification.geogrphicIdentifier)	**On-line resource (Optional)** (MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Dataset language (Mandatory) (MD_Metadata > MD_DataIdentification.lauguage)	**Metadata file identifier (Optional)** (MD_Metadata.fileIdentifier)
Dataset character set (Conditonally mandatory) (MD_Metadata > MD_DataIdentification.characterSet)	**Metadata standard name (Optional)** (MD_Metadata.metadataStandardName)
Dataset topic category (Mandatory) (MD_Metadata > MD_DataIdentification.topicCategory)	**Metadata standard version (Optional)** (MD_Metadata.metadataStandardVersion)
Spatial resolution of the dataset (Optional) (MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)	**Metadata language (Conditonally mandatory)** (MD_Metadata.language)
Abstract describing the dataset (Mandatory) (MD_Metadata > MD_Identification.abstract)	**Metadata character set (Conditonally mandatory)** (MD_Metadata.characterSet)
Distribution Format (Optional) (MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)	**Metadata point of contact (Mandatory)** (MD_Metadata.contact > CI_ResponsibleParty)
Additional extent information for the dataset (vertical and temporal) (Optional) (MD_Metadata > MD_DataIdentification.extent > EX_Extent)	**Metadata date stamp (Mandatory)** (MD_Metadata.dateStamp)

Reproduced from
The OpenGIS™ Abstract Specification
Topic 11: OpenGIS(tm) Metadata
(ISO/TC 211 DIS 19115)
Version 5

APPENDIX 1

GIS DATA MANAGEMENT METHODOLOGIES QUESTIONNAIRE

A. GIS personnel

1 GIS management

Q Are personnel assigned to GIS management in your organisation?

Y / N

2 Training

Q Have GIS management staff been trained in the use of GIS?

Y / N

3 Data access rights

Q Is access to GIS data limited to trained GIS staff?

Y / N

4 Data editing rights

Q Are editing rights to GIS data limited to trained personnel?

Y / N

B. Hardware

5 Hardware

Q What kind of hardware is used for GIS data

Network Server	Y / N
Dedicated Workstation	Y / N
Shared access PC	Y / N

6 Hardware access

Q Is access to GIS data limited to trained personnel?

Y / N

C. Software

7 Software

Q Is dedicated GIS software used for GIS work?

Y / N

D. Georeferencing

8 Projections

Q Are geographic coordinates applied to all GIS datasets created?

Y / N

9 Georeferencing

Q Are all raster files used georeferenced to geographic coordinates?

Y / N

E. GIS data formats

10 Vector formats

Q What GIS software format is used for vector data?

11 Raster formats

Q What GIS software format is used for raster data?

12 Relational Database formats

Q What software format is used for GIS compatible relational database data?

13 Spatial Database formats

Q What software format is used for spatial database (e.g. Oracle spatial) data?

14 ASCII formats

Q Do you use text files to import or export simple geographic features?

Y / N

15 GPS survey data

Q Do you import GPS data to GIS?

16 Remotely Sensed data

Q Do you use Remotely Sensed image data in GIS?

F. GIS Attributes

17 GIS Data Attributes

Q Do you use GIS database (e.g. ESRI .dbf) or map labels for feature attributes?

G. Data portability

18 File size

Q How do you handle large GIS files that need to be emailed?

19 Web transfer

Q Do you ever use web FTP (File Transfer Protocol) to transfer large files?

Y / N

20 Web dissemination

Q Do you use web GIS services to download or upload GIS data?

Y / N

H. Data Administration

21 File naming

Q Does your organisation use logical or personalised file names for GIS data?

Y / N

22 Data filing

Q Do you have procedures in place for logical fling of digital GIS data?

Y / N

I. Data Quality

23 Data quality - vector

Q Do you prefer to digitise GIS data by object, or by multiple objects in one file?

24 Digitising quality

Q Do you ever have problems with non edge-matching thematic GIS data?

Y / N

25 Data quality - Raster

Q Do you ever have to use poor quality base map raster images in your GIS?

Y / N

26 Data Quality - Attributes

Q Have you ever encountered poor attribute data in the GIS data you use?

Y / N

F. Metadata

27 Metadata

Q Do you record metadata (data about data) about the GIS data you use?

Y / N

APPENDIX 2

APPENDIX 3

APPENDIX 3

GIS Data Standards

A wide array of GIS data standards have been emerging over recent years. The standards outlined by in the ISO 191100 series are regarded as the best global standards. Interoperability is one of the key motivating factors driving the establishment of international standards, so metadata standards for GIS data have received much attention. ISO 19115 appears to be the *de facto* standard for GIS metadata, and its core components offer a solid template on which we can base our recommendations. In fact, many other existing metadata standards are beginning to harmonise with IOS 19115. The U.S. FGDC metadata standard is a case in point. Standards for a wider range of GIS data issues are provided in the rest of the ISO 191100 series. For more information on each standard, click the links provided in the list following.

ISO 19100 suite of standards

The 19100 suite covers:

- 19101: Reference model http://eden.ign.fr/std/iso_19101?set_language=en&cl=en
- 19102: Overview http://eden.ign.fr/std/iso_19102?set_language=en&cl=en
- 19103: Conceptual schema language http://eden.ign.fr/std/iso_19102?set_language=en&cl=en
- 19104: Terminology http://eden.ign.fr/std/iso_19104?set_language=en&cl=en
- 19105: Conformance and testing http://eden.ign.fr/std/iso_19105?set_language=en&cl=en
- 19106: Profiles http://eden.ign.fr/std/iso_19101?set_language=en&cl=en
- 19107: Spatial schema http://eden.ign.fr/std/iso_19107?set_language=en&cl=en
- 19108: Temporal schema http://eden.ign.fr/std/iso_19108?set_language=en&cl=en
- 19109: Rules for application schema http://eden.ign.fr/std/iso_19109?set_language=en&cl=en
- 19110: Feature cataloguing methodology http://eden.ign.fr/std/iso_19110?set_language=en&cl=en
- 19111: Spatial referencing by co-ordinates http://eden.ign.fr/std/iso_19111?set_language=en&cl=en
- 19112: Spatial referencing by geographic identifiers http://eden.ign.fr/std/iso_19112?set_language=en&cl=en
- 19113: Quality principles http://eden.ign.fr/std/iso_19113?set_language=en&cl=en
- 19114: Quality evaluation procedures http://eden.ign.fr/std/iso_19114?set_language=en&cl=en
- 19115: Metadata http://eden.ign.fr/std/iso_19115?set_language=en&cl=en

- 19116: Positioning services http://eden.ign.fr/std/iso_19116?set_language=en&cl=en
- 19117: Portrayal http://eden.ign.fr/std/iso_19117?set_language=en&cl=en
- 19118: Encoding http://eden.ign.fr/std/iso_19118?set_language=en&cl=en
- 19119: Services http://eden.ign.fr/std/iso_19119?set_language=en&cl=en
- 19120: Functional standards http://eden.ign.fr/std/iso_19120?set_language=en&cl=en
- 19120 / Amendment 1: Functional standards - Amendment 1 http://eden.ign.fr/std/iso_19120?set_language=en&cl=en
- 19121: Imagery and gridded data http://eden.ign.fr/std/iso_19121?set_language=en&cl=en
- 19122: Geographic information/Geomatics - Qualifications and Certification of Personnel http://eden.ign.fr/std/iso_19122?set_language=en&cl=en
- 19123: Schema for coverage geometry and functions http://eden.ign.fr/std/iso_19123?set_language=en&cl=en
- 19124: Imagery and gridded data components http://eden.ign.fr/std/iso_19124?set_language=en&cl=en
- 19125-1: Simple feature access - Part 1: Common architecture http://eden.ign.fr/std/iso_19125_1?set_language=en&cl=en
- 19125-2: Simple feature access - Part 2: SQL option http://eden.ign.fr/std/iso_19125_2?set_language=en&cl=en
- 19125-3: Simple feature access - Part 3:COM/OLE option http://eden.ign.fr/std/iso_19125_3?set_language=en&cl=en
- 19126: Profile - FACC Data Dictionary http://eden.ign.fr/std/iso_19126?set_language=en&cl=en
- 19127: Geodetic codes and parameters http://eden.ign.fr/std/iso_19127?set_language=en&cl=en
- 19128: Web Map Server interface http://eden.ign.fr/std/iso_19128?set_language=en&cl=en
- 19129: Imagery, gridded and coverage data framework http://eden.ign.fr/std/iso_19129?set_language=en&cl=en
- 19130: Sensor and data models for imagery and gridded data http://eden.ign.fr/std/iso_19130?set_language=en&cl=en
- 19131: Data Product Specification http://eden.ign.fr/std/iso_19131?set_language=en&cl=en
- 19132: Location based services possible standards http://eden.ign.fr/std/iso_19132?set_language=en&cl=en
- 19133: Location based services tracking and navigation http://eden.ign.fr/std/iso_19133?set_language=en&cl=en
- 19134: Multimodal location based services for routing and navigation http://eden.ign.fr/std/iso_19134?set_language=en&cl=en
- 19135: Procedures for registration of geographical information items http://eden.ign.fr/std/iso_19135?set_language=en&cl=en

ISO 19115

http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70

Comment

ISO 19115 is the International Metadata Standard for Geographic Information.

Background

ISO 19115:2003 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

ISO 19115:2003 is applicable to:

- The cataloguing of datasets, clearinghouse activities, and the full description of datasets.
- Geographic datasets, dataset series, and individual geographic features and feature properties.

ISO 19115:2003 defines:

- Mandatory and conditional metadata sections, metadata entities, and metadata elements.
- The minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data).
- Optional metadata elements - to allow for a more extensive standard description of geographic data, if required.
- A method for extending metadata to fit specialized needs.

Though ISO 19115:2003 is applicable to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.
NOTE Certain mandatory metadata elements may not apply to these other forms of data.

Source:
http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70

FGDC-STD-001-1998

Comment

A U.S. based standard. FGDC-001-1998 is older than the ISO 19115 standard, is compliant with ISO 19115, and is moving to increased harmonisation with it.

ISO 19115 therefore represents the best template for best practice.

Background

U.S. Federal Geographic Data Committee (FGDC) promotes the coordinated development, use, sharing, and dissemination of geographic data in the U.S.

The FGDC is composed of representatives from the United States Departments of Agriculture, Commerce, Defence, Energy, Housing and Urban Development, the Interior, State, and Transportation; the Environmental Protection Agency; the Federal Emergency Management Agency; the Library of Congress; the National Aeronautics and Space Administration; the National Archives and Records Administration; and the Tennessee Valley Authority. Additional Federal agencies participate on FGDC subcommittees and working groups. The Department of the Interior chairs the committee.

FGDC subcommittees work on issues related to data categories coordinated under the circular. Subcommittees establish and implement standards for data content, quality, and transfer; encourage the exchange of information and the transfer of data; and organize the collection of geographic data to reduce duplication of effort.

The Open Geospatial Consortium and GML

OGC provide a huge array of data standards that overlap substantially with the ISO 19100 series. OGC focuses particularly on the area of web interoperability of GIS data. Discussions about OGC standards often focus on the issue of Geography Markup Language (GML). The Geography Markup Language (GML) is the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. GML offers huge scope for interoperability, because similar to XML, its logic and structure is independent of software platform, and can be read or edited in a simple ASCII text editor. This is an obvious advantage when disseminating GIS data over the web. Implementation of GML in mainstream GIS products is still at an early stage however (http://en.wikipedia.org/wiki/Geography_Markup_Language)

OGC is an international industry consortium of over 300 companies, government agencies and universities participating in a consensus process to develop publicly available interface specifications. The OGC is necessary because cooperation is necessary to solve the difficult interoperability issues in the geospatial marketplace.

OpenGIS® produce specifications for interoperable solutions that "geo-enable" the Web, wireless and location-based services, and mainstream IT. OpenGIS® is designed to serve as a global forum for the collaboration of developers and users of spatial data products and services, and to advance the development of international standards for geospatial interoperability.

Aims

1. Provide free and openly available standards to the market, tangible value to Members, and measurable benefits to users.
2. Lead worldwide in the creation and establishment of standards that allow geospatial content and services to be seamlessly integrated into business and civic processes, the spatial web and enterprise computing.
3. Facilitate the adoption of open, spatially enabled reference architectures in enterprise environments worldwide.
4. Advance standards in support of the formation of new and innovative markets and applications for geospatial technologies.
5. Accelerate market assimilation of interoperability research through collaborative consortium processes.

What do they do?

1. Specification Program - Work to arrive at approved (or "adopted") OpenGIS® Specifications.
2. Interoperability Program - Hands-on engineering initiatives to accelerate the development and acceptance of OpenGIS® Specifications.
3. Outreach and Adoption - OGC and its members offer resources to help technology developers and users take advantage of OGC's open standards.

OGC Europe (OGCE)

OGCE is actively involved in a number of European Union Projects:

- ETEMII - European Territorial Management Information Infrastructure (http://www.ec-gis.org/etemii/)
- GETIS - Geoprocessing networks in a European Territorial Interoperability Study (http://www.pcigeomatics.com/getis/)
- GINIE - Geographic Information Network In Europe (http://www.ec-gis.org/ginie/)
- INSPIRE - INfrastructure for SPatial InfoRmation in Europe (http://www.ec-gis.org/inspire/home.html).

OCG documents

For a list and detailed descriptions OGC specification documents please follow the following link. http://www.opengeospatial.org/specs/?page=specs