MEDIN Resources Linking to the UK Location Programme / data.gov and INSPIRE

1. Introduction

Under the INSPIRE directive all organisations holding Public Geospatial Data are required to publish metadata, and provide data view and download services for their data. The UK Location Programme (UKLP) is working in cooperation with data.gov to establish a central national framework to support UK organisations in meeting these obligations. This framework includes a national geo-portal and an architecture for publishing data view and download services. Data.gov and UKLP together have a wider remit beyond INSPIRE to improve access to public data. Thus UKLP / data.gov and INSPIRE are essentially data and metadata publication mechanisms. They will allow users to search, view and download public spatial data, but they do not support basic data management capability.

MEDIN is a national initiative to improve the management of, and access to, UK marine data and information. MEDIN has a remit to address data management as well as accessibility. Through its network of accredited Data Archive Centres, MEDIN provides a secure data archiving and management capability which helps to ensure the quality, validity and authority of data made available to users. This aspect is important for most uses, but especially important to users who are compiling data for reporting against legislative obligations, for scientific evidence, and for planning, all of which are key drivers for MEDIN.

There is clearly much scope for MEDIN and UKLP/data.gov to work together and share some resources. This note describes how MEDIN will link metadata and data services to the UKLP / data.gov system, and how to use the MEDIN resources through these links to meet INSPIRE metadata and data publishing obligations.

2. Overview of UKLP / data.gov system

The UK Location Programme refers to *"Data Providers"* – organisations that create the data and supply data for web publications along with its metadata; and *"Data Publishers"* - organisations that publish the data on the web and supply data services to users.

MEDIN has established a network of *Data Archive Centres (DACs)*, with which all partners are encouraged to archive their data, and also a central *Discovery Metadata Portal and Catalogue* to which all partners are required to publish their discovery metadata.

MEDIN is ready to act as the *Metadata and Data Publisher* on behalf of the Marine Community, through the onward publication of its Discovery Metadata Catalogue to data.gov, and the provision of data view and download services by the DAC network. Figure 1 provides a schematic from the UKLP on how data publishers can act on behalf of data providers and link to the central UKLP / data.gov service.
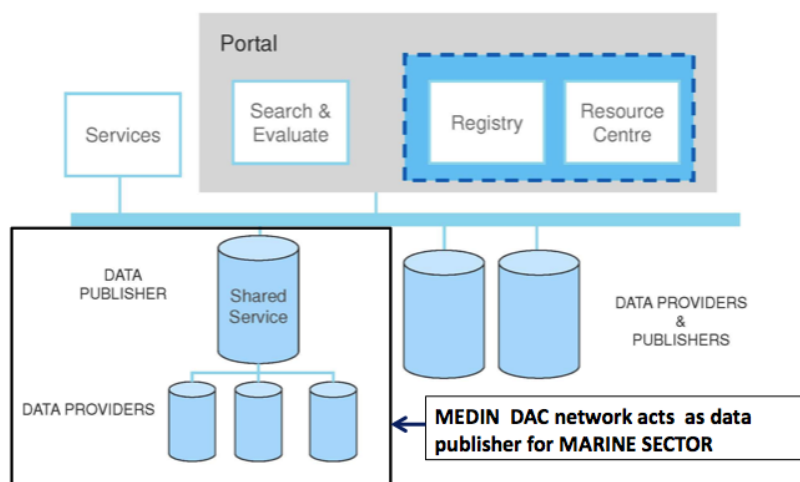

Figure 1. MEDIN as a thematic metadata /data assembly and publication centre for the marine community, into the UKLP framework

This note covers four metadata /data services:

<u>Discovery Metadata Service</u>
Discovery metadata provide a description of the data held within a data set. Each data set must have a Discovery Metadata record to describe it. These descriptions are the basis of most initial searches for data, through a Discovery Portal.

<u>Data View Service</u>
A Data View Service will provide over the web a visual representation of the data within a data set. The discovery metadata record will provide a link to the view service of the data described.

<u>Data Download Service</u>
A Data Download Service provides over the web a direct download of the data identified in the discovery metadata record.

<u>Data Transformation Service</u>
A Data Transformation Service will allow the user, over the web, to transform data from one format to another. For instance, a transformation service could be used to transfer a data set from a format that does not meet an INSPIRE specification to a format that does.

3. Generating and Publishing Discovery Metadata

Overview – The MEDIN Discovery Portal and Metadata Catalogue

MEDIN has developed a marine discovery portal to provide a single central access point through which all major UK holdings of marine data can be searched. The UKLP /data.gov have developed the data.gov portal for all UK public geospatial data. MEDIN is maintaining a separate Discovery Portal to data.gov, as the data.gov search portal will not include all MEDIN data and the MEDIN portal provides marine specific search capabilities supported by an enhanced portal functionality and a higher level of metadata content (see later discussion).

Thus all MEDIN partners are encouraged to generate MEDIN format discovery metadata records, and to publish these metadata records to the MEDIN Discovery Portal, rather than generate data in the generic UK Gemini 2.1 format and publish direct to the UKLP portal. MEDIN will publish metadata to data.gov as requested by the data owner.

Once metadata records are generated and published by MEDIN Data Archive Centres and MEDIN partners, they are harvested onto a central MEDIN metadata catalogue. This catalogue is managed and hosted for MEDIN by the Science and Technology Facilities Council (STFC), and searched through the MEDIN portal on the oceannet web page[1].

MEDIN has put in place and tested the mechanism to publish metadata records on to data.gov, through a CSW (Catalogue Service for Web) operated by STFC, and so on to INSPIRE. The metadata set to be re-published in this way will be a subset of the full MEDIN metadata catalogue as not all data within MEDIN will be relevant to UKLP/data.gov or INSPIRE (e.g. non-public data, non geospatial format data, data not covered by INSPIRE themes) .

Discovery Metadata Standard

MEDIN has developed a specially adapted version of the Gemini 2.1 discovery metadata standard to suit the specific requirements of the marine community. This is a higher-level standard than the generic version, but remains entirely compliant with Gemini 2.1, INSPIRE requirements and with the relevant ISO standards. It includes the use of particular vocabularies and code lists within certain fields. These vocabularies are widely accepted across the UK and European marine communities, are subject to a managed update process, and ensure that the data content of marine data sets is precisely specified according to recognised international standards.

An up to date specification is maintained on the MEDIN web pages at
http://www.oceannet.org/marine_data_standards/medin_disc_stnd.html

---

[1] http://portal.oceannet.org/search/full

How Can I Create Discovery Metadata? Guidelines and Tools

On its website MEDIN provides comprehensive guidelines on how to generate MEDIN Standard Discovery Metadata, an online metadata creation tool, a downloadable stand alone metadata tool and a tool to generate MEDIN format metadata from ESRI created GIS layers. A further metadata tool to link to large databases will be available shortly. MEDIN also runs a metadata helpline (phone 01752 633291, email medin.metadata@mba.ac.uk).

How Do I Publish my Metadata?

It is recommended that all UK marine data is archived within a MEDIN accredited DAC, and that the DAC publish metadata records on behalf of the data owner. If you prefer to manage data within your own organisation, and publish your metadata yourself, you should ask MEDIN for guidance. You can either publish through an OGC CSW service, or using a server configured with OAI-PMH (Open Archive Initiative - Protocol for Metadata Harvesting) software (downloadable from http://www.dlese.org/Metadata/tool/index.php). In either case you must register your metadata publishing service with MEDIN. Again contact the MEDIN metadata helpline with any queries.

Do I want MEDIN to publish my Metadata records to UKLP and INSPIRE?

All metadata published to MEDIN will be searchable through the MEDIN Discovery Portal. MEDIN will also publish metadata records, *on request*, to data.gov and thence onward to INSPIRE. You should decide if you wish all or some of your metadata to be accessible through data.gov and the INSPIRE Portal, and advise MEDIN accordingly.

Can I use the UKLP metadata tool to produce metadata?

No – the UKLP metadata tool will not produce MEDIN compliant discovery metadata and so these data would not be searchable within the MEDIN portal. Please use one of the MEDIN metadata tools available on the MEDIN website.

4. Data View Services

Overview

UKLP / data.gov have defined a technical architecture to support the publication of data view services. MEDIN is planning to act as a data publisher on behalf of the marine community through its network of DACs. The intention is that the MEDIN DACS will publish INSPIRE compliant data view (and download) services which are accessible to the UKLP/data.gov framework.

The INSPIRE Timetable requires that View services for Annex I and Annex II data sets are available in an initial operating capability by May 2011, with full view services available by November 2011. The equivalent dates for Annex III data, which includes most marine data, is December 2013[2]. An overview of the coverage of the INSPIRE "annexes" is given at the end of this note.

Although the intention is that the MEDIN DACs will publish view services on behalf of those whose data are held within the DAC, details of the standards for View Services are included below for information.

An overview of the standards and specifications for view services are given below.

View Service Standards

The UKLP (and INSPIRE) data view services will be based on the use of the OGC standard for Web Mapping Services (WMS). Web Mapping Services present location information as an image (GIF, JPEG or PNG).

This can be displayed in a map viewer, either singly, or in combination, against a base map layer.

Some further guidance on INSPIRE requirements for the view service is given below:

Styles: There are no specific requirements relating to Styles, other than a list of rendering styles for a given layer should be provided.

Coordinate Reference System: The WMS must support at least the Coordinate reference system defined in the INSPIRE Annex I Theme.

Image Format: The view service must support at least one of: Portable Network Graphics (PNG); Graphics Interchange Format (GIF) without compression.

---

[2] This date may have already slipped. To be confirmed

Feature Information: There are no INSPIRE requirements for the support of Feature Information, but this can be optionally provided

Quality of Service: There are some requirements in terms of Performance (maximum response time to a Get Map request of 5 s); Capacity (must be able to serve at least 20 simultaneous service requests); and Availability (the View service must be available 99% of the time)

Note that once a view service is published the metadata record should be updated to suit (in the "Resource Locator" and "Coupled Resource" Elements.

What Should I Do?
MEDIN will publish further guidance shortly, when it has agreed with the DACs how to implement view services. In the meantime please contact us with any queries.

5. Data Download and Transformation Services

The INSPIRE timetable requires that initial capability for Data Transformation and Data Download services for Annex I and Annex II data shall be available by 28 June 2012, with services fully conforming to the regulation by 28 December 2012. The equivalent deadline for Annex III data (most of the marine data) is December 2013.[3]

Whilst the full technical specification, and hence the UKLP technical architecture to support the publication of these services, is not yet available, a regulation passed in November last year has provided some initial requirements and quality of service specifications, as follows:

Download Operations
The download service shall at least provide the following:
- Get Download Service Metadata (response < 10s)
- Get Spatial Data Set (response < 30s, download rate > 0.5 Mbytes s^{-1}/ > 500 spatial objects s^{-1})
- Describe Spatial Data Set (response < 10s, download rate > 0.5 Mbytes s^{-1}/ > 500 spatial objects s^{-1})
- Link Download Service

The minimum number of simultaneous service requests served shall be 10 per second, the number of requests processed in parallel may be limited to 50. The minimum availability of a network services shall be 99%.

Transformation Operations
The transformation service shall at least provide the following:
- Get Transformation Service Metadata
- Transform
- Link Transformation Service

The minimum number of simultaneous transformation service requests served shall be 5 per second. The minimum availability of a network services shall be 99%.

Whilst MEDIN DACs may be expected to support Download Services, they may not be required to support Data Transformation Services.

Further details are provided in Commission Regulation No 1088/2010, see http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:EN:PDF

Data Specifications
INSPIRE is in the processing of defining specifications for the format and content of data sets as they should be made available for download through the INSPIRE framework. MEDIN is helping to collate responses to the draft specifications that are relevant to marine data, and will continue to provide support to the marine community as these standards are finalised and implemented

---

[3] This date may have slipped. To be confirmed.

6. INSPIRE, UKLP / data.gov and MEDIN Coverage and Functionality

In this section we discuss the coverage of INSPIRE, data.gov / UKLP and MEDIN, and the functionality offered by the different initiatives.

Data Types and Themes

INSPIRE is specifically directed at public geospatial data, with a particular focus on data for environmental policy making. UKLP / data.gov has a wider coverage than INSPIRE, and includes data needed for sustainable development, socio-economic analysis and other purposes. MEDIN covers all types of marine data, including public and private data holdings, and non-spatial data.

For reference, the definition of Geospatial Data from INSPIRE documentation includes data that are:

> *"...usually stored as coordinates and topology, and comes in three forms:*
> *Vector data – points, lines and polygons.*
> *Raster data – digital satellite imagery, or air photos, and elevation shading or digital elevation models.*
> *Text data – alphanumeric data that includes coordinate referencing (e.g. latitude, longitude, and height), also referred to as 'geo-referenced', or "geo-encoded".*
> *...excludes other forms of Raster Maps, e.g. PDF, and data that might be associated with a location, but otherwise is not geospatial in form."*

The table below compares the coverage in terms of data types and themes.

INSPIRE	UKLP / data.gov	MEDIN
Focus is environmental policy making All public geospatial data, falling within data types specified within INSPIRE themes. Publishers must provide data view and download services	Broader than INSPIRE, information for sustainable development, socio-economic analysis and many other purposes. Includes all public UK spatial data lying within INSPIRE themes. Public data are "government held non-personal data collected or generated in the course of public service delivery" Legal obligation to publish only applies if the data comes under an INSPIRE data theme, but data owners are encouraged to publish any location information they believe will be of benefit for sharing and re-use. Private organisations are also encouraged to publish the location information they create, particularly if associated with the delivery of a public task, or social benefit. Publishers must provide data view and download services?	Thematically narrower than data.gov - marine data and information only – but includes all types of data and information. Intention is to provide single central source for searching all marine data and information held in the UK. Includes: - Data not in "geospatial" form, e.g. reports, sample analyses, . - public and private data, - "raw" data (large proportion – which is validated and calibrated but not worked up into higher level analyses such as maps, gridded data sets, statistical representations) which may be the data of widest public /policy interest?. - High volume data such as video and multi-beam survey data – not easily available as download - Other data outside strict INSPIRE specifications for which view or download services are not currently available and are not planned to be available in short – middle term.

Thus we can identify a <u>core</u> set of data for which there is a complete overlap between INSPIRE and MEDIN coverage, an <u>additional</u> set which includes the wider remit of UKLP /data.gov, and a further <u>MEDIN specific</u> set, which does not fall within the current UKLP/remit, as described below:

<u>Core</u>
Marine data of types specified within the INSPIRE Annexes, held by public bodies or on behalf of public bodies by publishers (MEDIN DACs).

<u>Additional level</u>
Public environmental data not specifically identified within INSPIRE themes, non-public data whose owners are happy to publish

<u>MEDIN Specific Data</u>

- Non geospatial format data (pdf format, reports, ...)
- Non-public data the IPR owner does not want to publish to data.gov / INSPIRE
- Data for which view or download services are not available

Metadata Standards

Gemini 2 is the discovery metadata standard adopted by the UKLP / data.gov. It has been slightly adapted in order to meet the INSPIRE specifications.

MEDIN has defined a marine profile of the Gemini2 discovery metadata standard for use by the marine community. This standard entirely compliant with Gemini2 but has more tightly specified content in some fields through mandatory use of controlled vocabularies, and some additional fields. This is necessary to support a more powerful search capability (including pull down lists)which is available on the MEDIN portal and was specified following a user consultation exercise.

UKLP / data.gov and MEDIN have both developed metadata tools. The MEDIN tool will produce Gemini2 compliant metadata which can be published to, and searched on, the data.gov portal, but the converse is not true. The UKLP/data.gov tool can NOT be used to produce metadata for publication to the MEDIN portal.

Discovery Portals

The table below provides a summary comparison of the functionality offered by the data.gov.uk and MEDIN discovery portals:

Data.gov: http://data.gov.uk/data	MEDIN Discovery Portal: http://portal.oceannet.org/search/full

Free text based on search of all metadata content
 	
Not yet publically available: Map based, can delineate bounding box with map tool, gazetteer search.
	Free text based on metadata content
	
	Advanced structured search. Includes capability to compile combined field specific search (including author, parameter, resource type, topic category, lineage, public access limitations, data originator and data format) with pull down lists available
	
	Map based,
	    o can delineate bounding box with map tool,
	    o can select regions from different lists (countries, gazetteer, charting progress regions, ICES rectangles
	    o option to view and download retrieved metadata in range of formats

The MEDIN portal functionality is based on a user requirements survey. A questionnaire to users in Autumn 2010 has identified some further functionality that will be built into an enhancement to the portal in 2011-12.

Glossary

CSW – Cataloguing Services for the Web
DAC – Data Archive Centre
MEDIN – Marine Environmental Data and Information Network
OAI-PMH - Open Archive Initiative - Protocol for Metadata Harvesting
OGC – Open Geospatial Consortium
STFC – Science and Technology Facilities Council
UKLP – UK Location Programme
WMS – Web Mapping Service

References

UKLP Getting Started Notes – all available at location.defra.gov.uk/resources/getting-started
Guide 1: UK Location, an introduction
Guide 2: Organisation and Data Coverage
Guide 3: What Needs to Happen and When
Guide 4: Publishing Discovery and View Services
Guide 5: UK Implementation of INSPIRE
UKLP and AGI "System Supplier Workshop Presentations
   http://location.defra.gov.uk/resources/media/presentations-agi-system-supplier-sig-25-jan-2011/
WMS "primer": http://www.gommap.org/gommap/docs/ogc-wms-primer.pdf
Guide to Distributing your data products by WMS 1.1.1 : http://oceanesip.jpl.nasa.gov/esipde/guide.html
INSPIRE View Service Technical Guide:
http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_View_Services_v2.12.pdf

Links

MEDIN: http://www.oceannet.org
UKLP: http://location.defra.gov.uk
Data.gov: http://data.gov.uk/
INSPIRE: http://inspire.jrc.ec.europa.eu/
DLESE OAI software: http://www.dlese.org/oai/
OGC WMS information: http://www.opengeospatial.org/standards/wms

INSPIRE Themes (see http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/

Annex I	*Annex II*	*Annex III*
1 Coordinate reference systems	*1 Elevation*	1 Statistical units
2 Geographical grid systems	2 Land cover	2 Buildings
3 Geographical names	3 Orthoimagery	3 Soil
4 Administrative units	*4 Geology*	4 Land use
5 Addresses		5 Human health and safety
6 Cadastral parcels		6 Utility and governmental services
7 Transport networks		*7 Environmental monitoring Facilities*
8 Hydrography		8 Production and industrial facilities
9 Protected sites		9 Agricultural and aquaculture facilities
		10 Population distribution and demography
		11 Area management/restriction/regulation zones & reporting units
		12 Natural risk zones
		13 Atmospheric conditions
		14 Meteorological geographical features
		15 Oceanographic geographical features
		16 Sea regions
		17 Bio-geographical regions
		18 Habitats and biotopes
		19 Species distribution
		20 Energy Resources
		21 Mineral Resources

Latest INSPIRE Key Dates

03 Dec 2010	Metadata available for spatial data sets and services corresponding to Annex I and II
09 Nov 2011	Discovery and view services operational
28 Jun 2012	Members states shall provide Transformation Service with initial operating capability
28 Jun 2012	Members states shall provide Download Service with initial operating capability
28 Dec 2012	Transformation services operational
28 Dec 2012	Download services operational
03 Dec 2013	Metadata available for spatial data sets and services corresponding to Annex III
October 2015	Newly collected and extensively restructured Annex II and III spatial data sets available
October 2020	Other Annex II and III spatial data sets available in accordance with IRs for Annex II and III