SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA MANAGEMENT

A Guide to creating EDMED entries using MIKADO

Project Acronym : SeaDataNet II
Project Full Title : SeaDataNet II: Pan-European infrastructure for ocean and marine data management
Grant Agreement Number : 283607

A Guide to creating EDMED entries using MIKADO - 16/02/2011
Sdn-userdesk@seadatanet.org – www.seadatanet.org

A Guide to creating EDMED entries using MIKADO

EDMED is a searchable directory of datasets relating to the marine environment. It covers a wide range of disciplines including marine meteorology; physical, chemical and biological oceanography; sedimentology; marine biology and fisheries; environmental quality monitoring; coastal and estuarine studies; marine geology and geophysics etc. Datasets are catalogued in EDMED irrespective of their format (e.g. digital databases or files, analogue records, paper charts, hard-copy tabulations, geological samples, biological specimens etc).

Originally established in 1991 within the EU Marine Science and Technology (MAST) framework, the EDMED format was revised and upgraded as part of the SEA-SEARCH (2003 - 2005) initiative and a brand new search interface has been implemented. EDMED, along with other marine databases has been developed further by the EU SeaDataNet (2006 - 2011) and SeaDataNet II (2011 – 2015) projects.

The purpose of an EDMED record is to allow users to discover marine data they are interested in. A discovery metadata record should allow a user to make a decision as to whether or not they are interested in accessing the data it describes. You should try to bear in mind that the resulting dataset should be easily extractable from a database to deliver to a user. The dataset should not be too small or too large that it becomes unusable.

Although the directory is targeted primarily at data sets that can be made accessible to other users, encouragement is also given to holders of working data sets, or data of a confidential or restricted availability, to make their data known through EDMED. You should aim to minimise the number of similar data set descriptions and group like data within single datasets, whether by data type, project or other criteria.

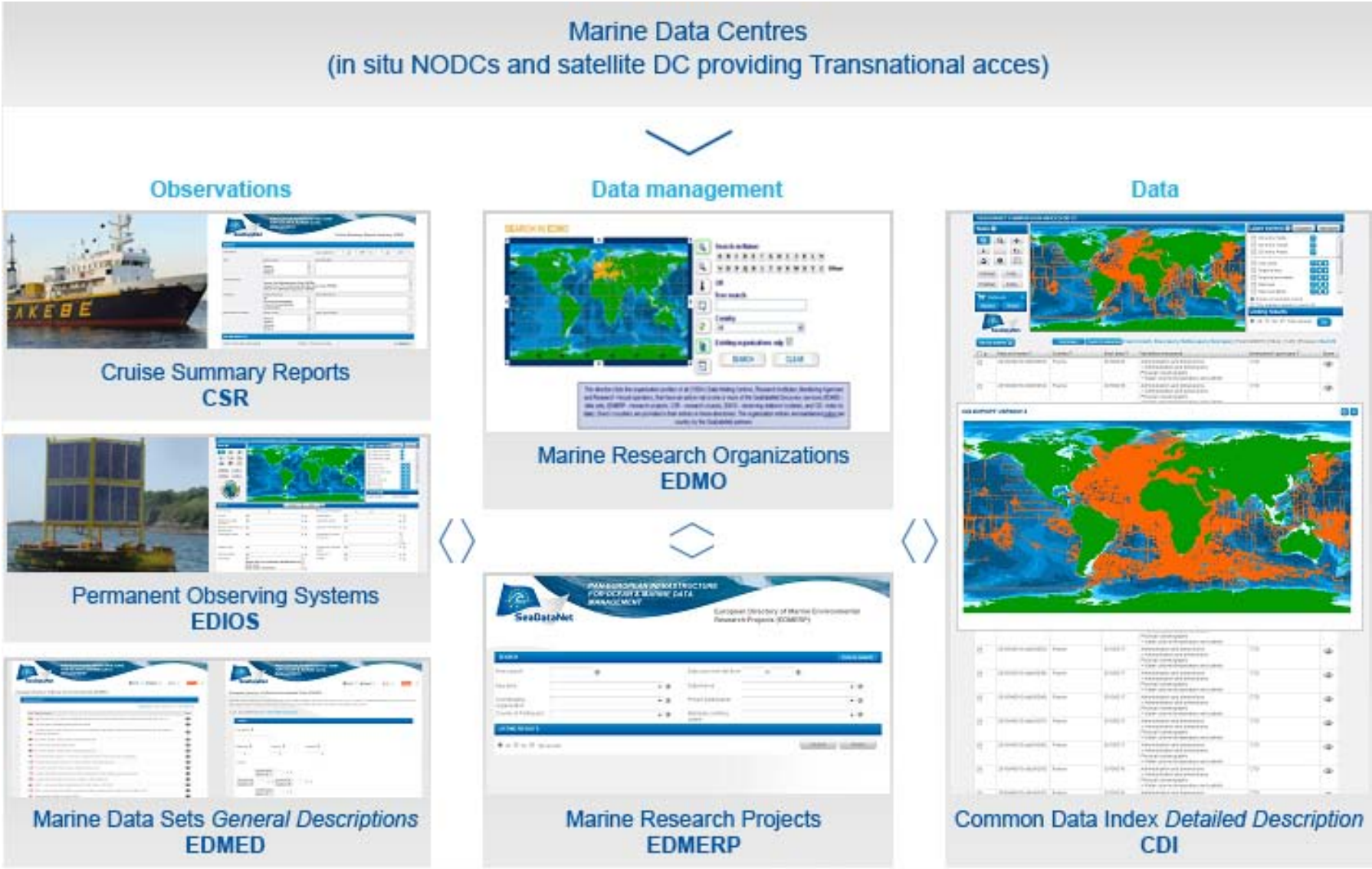
An EDMED record should describe when data were collected and where, what sort of data were collected and what instruments were used and how you can get hold of the dataset.

Directory entries are prepared by institutes and collated by the national centres that are responsible for populating and maintaining their national directories. These national directories are combined to provide a single centralised system managed by BODC.

This document is intended to assist in completing EDMED entries. Its focus is on the content of EDMED entries and complements the ‘User manual and instructions for updating EDMED, EDMERP, EDIOS, EDMO and CSR’, which explains in some detail how to produce XML files in manual or automatic mode.

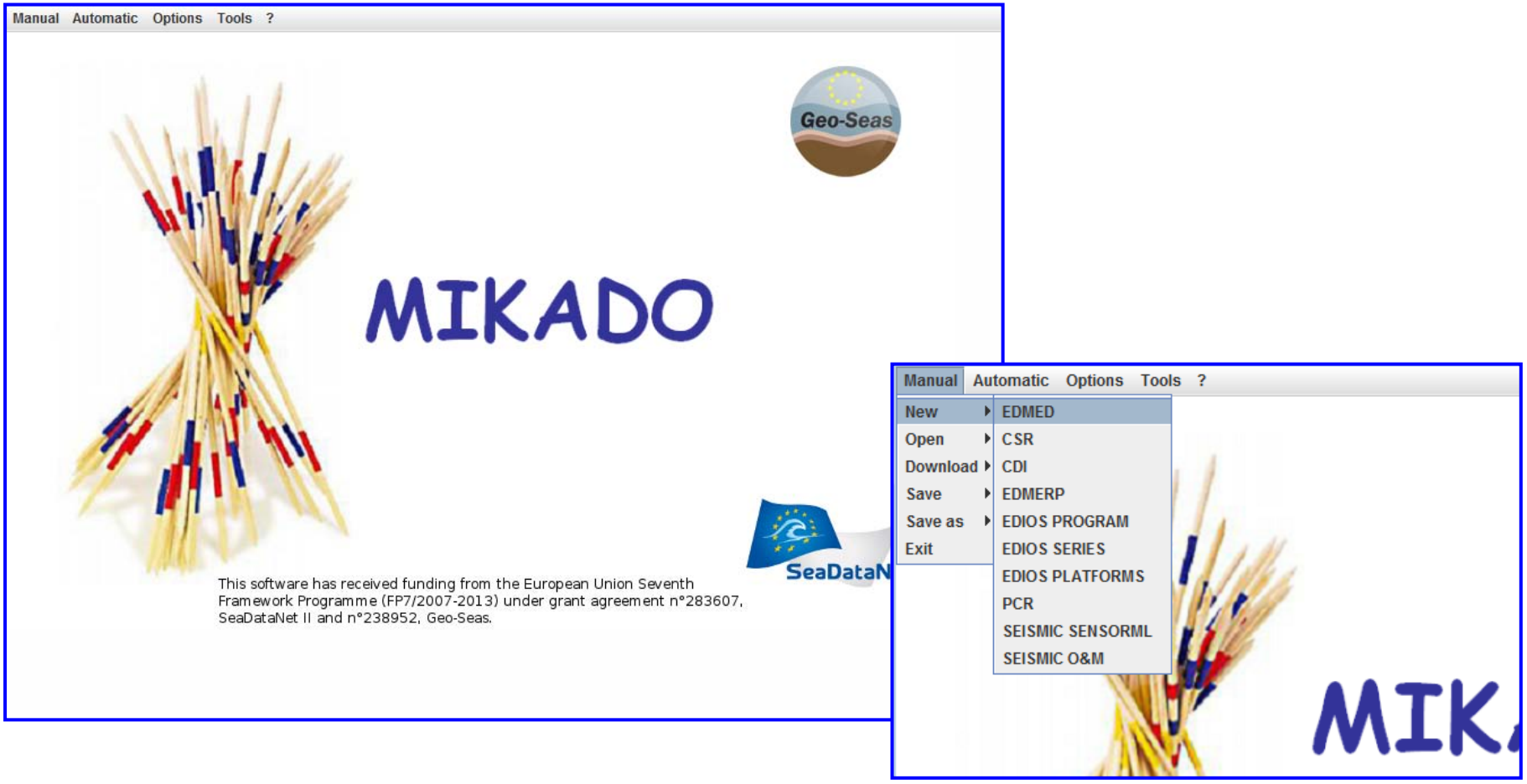
Examples of EDMED entries:

- **Bathymetry data (conventional and multi-beam) of the East Mediterranean Sea (1986-)**
- **Sardines from the Portuguese continental coast (1930-)**
- **Hydro-chemistry station data from ICES International Bottom Trawl Surveys in the North Sea**
- **Dataset of the trace metals, nutrients and chlorophyll a in the nepheloid layer of the Gulf of Riga**
- **Isle of Man Government Laboratory (GAL) Coastal Monitoring Sites network data sets**
- **Zooplankton of Kamyshovaya bay, Sevastopol, Ukraine**
- **Bristol Channel Suspended Sediments Data Bank (collected by the Institute of Oceanographic Sciences (IOS), Taunton) (1974-1978)**
- **IMARES, Monitoring sedimentation and vegetation on saltmarshes in the Dutch Wadden Sea (1993 -)**


Marine Data Centres
(in situ NODCs and satellite DC providing Transnational acces)
Observations
Data management
Data
Cruise Summary Reports
CSR
Marine Research Organizations
EDMO
Permanent Observing Systems
EDIOS
Marine Data Sets General Descriptions
EDMED
Marine Research Projects
EDMERP
Common Data Index Detailed Description
CDI

Be warned! You cannot save until you have filled in all your mandatory values. This means that you cannot save a partial document but is to maintain the integrity of the output xml.

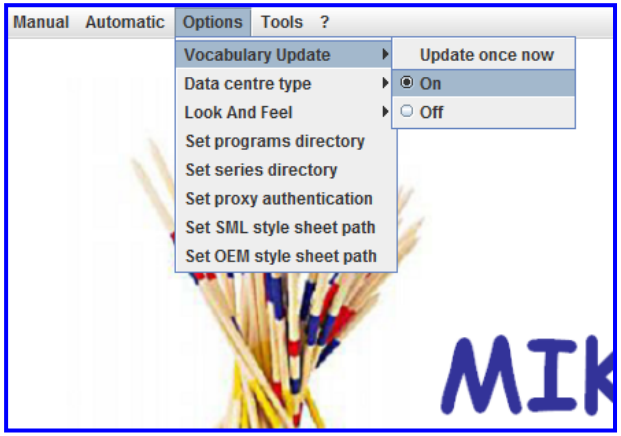
Startup page



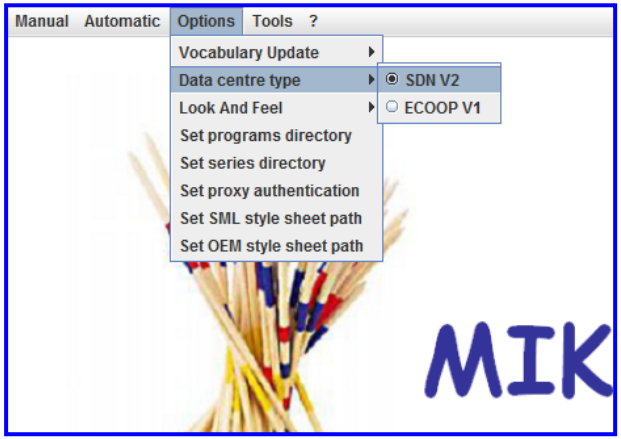
Automatically updating vocabulary lists

In order to make sure MIKADO automatically gets the latest codes every time it is opened, users need to set the vocabulary update to '**On'**.

When '**On'** and when MIKADO is launched, you should see in the bottom left hand corner of the launch screen the vocabulary names appear one by one. If not, it most likely is not updating.



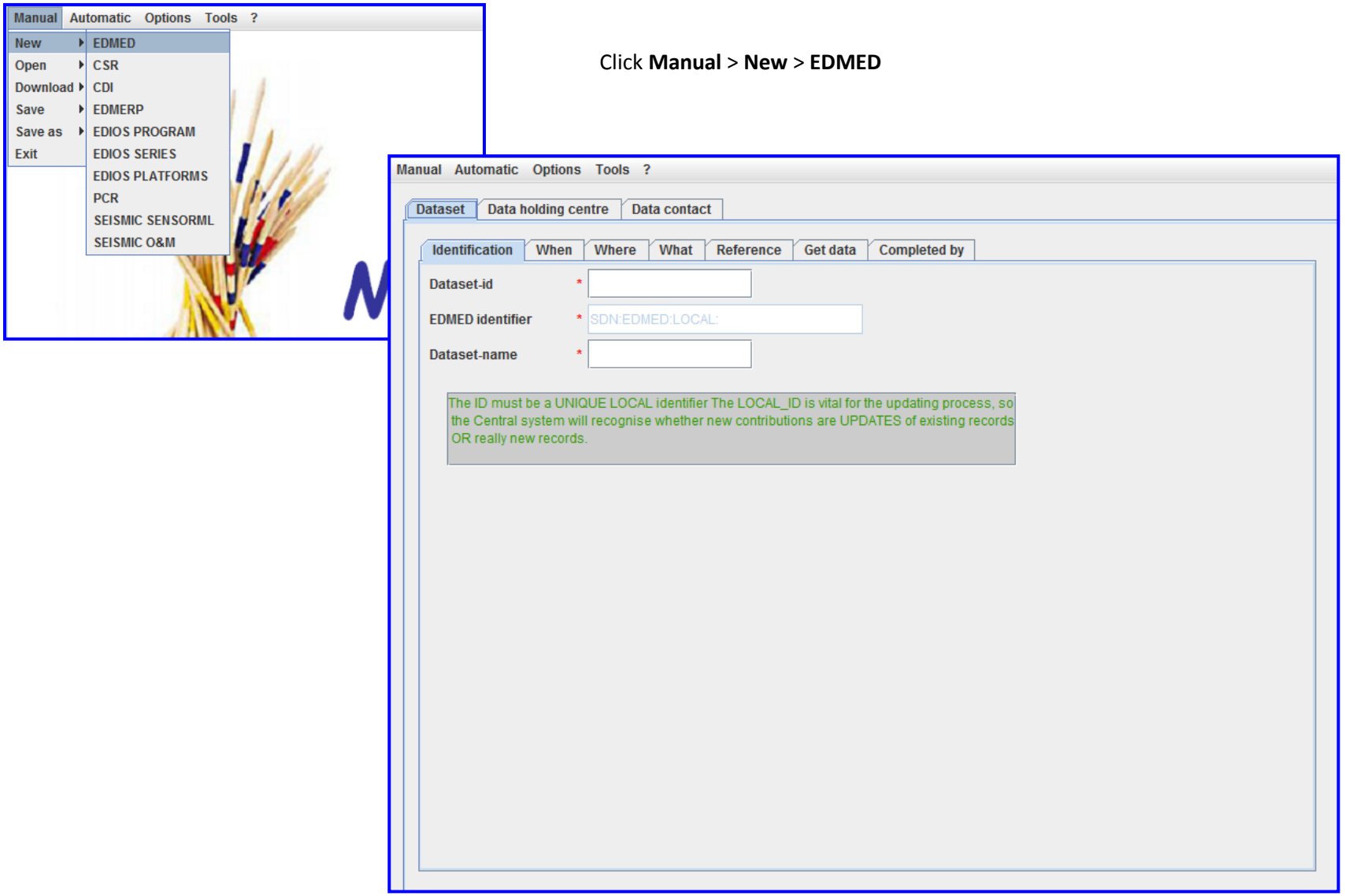
Click **Options** > **Vocabulary Update**> **On**



Also, users need to make sure that the SeaDataNet vocabularies are chosen.

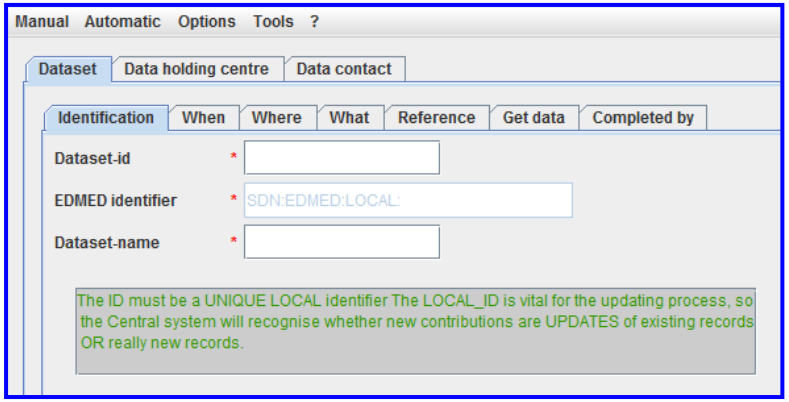
Click **Options** > **Data centre type** > **SDN V2**

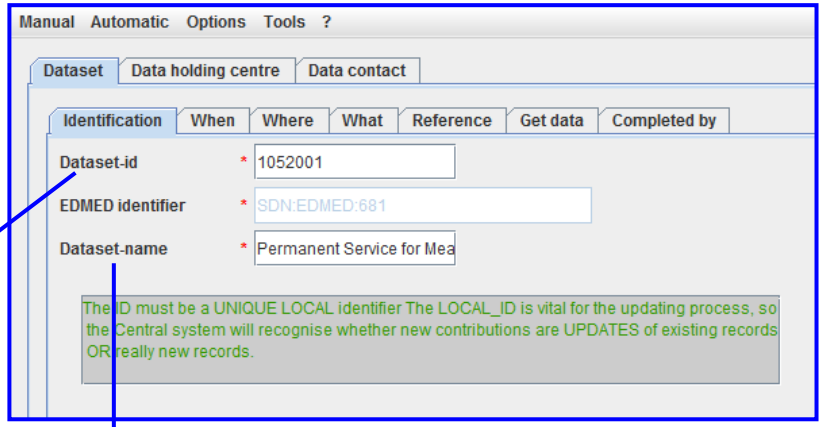
Creating a new EDMED



Click **Manual** > **New** > **EDMED**

Page 1.a – Dataset: Identification

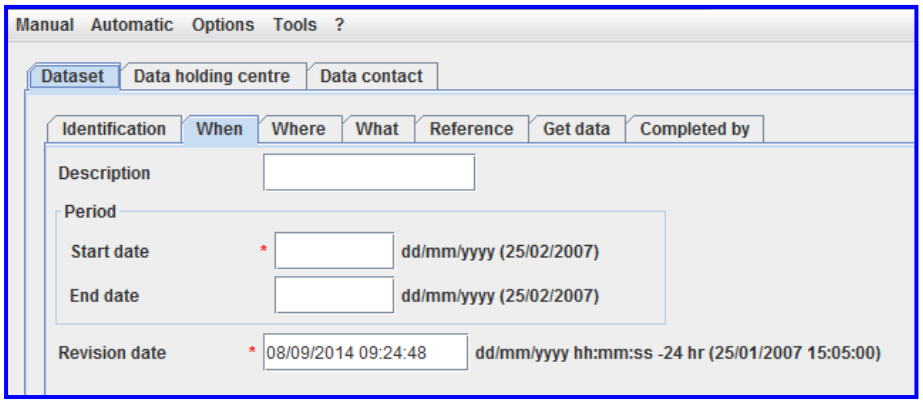




Dataset-id: This should be a unique local identifier, possibly containing unique numeric values, so that each dataset is easily identifiable. It could include name or collate-id of your centre, the year and a sequential number. **Mandatory**. Maximum 80 characters.

Dataset-name: The name of the data set. It should be similar to a journal article title and give some indication of the content of the data set and its temporal and spatial coverage. **Mandatory**. Maximum 160 characters.

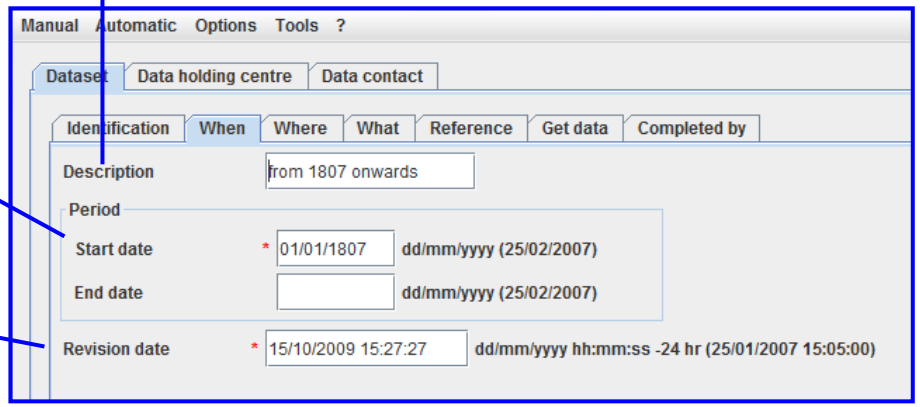
Page 1.b – Dataset: When



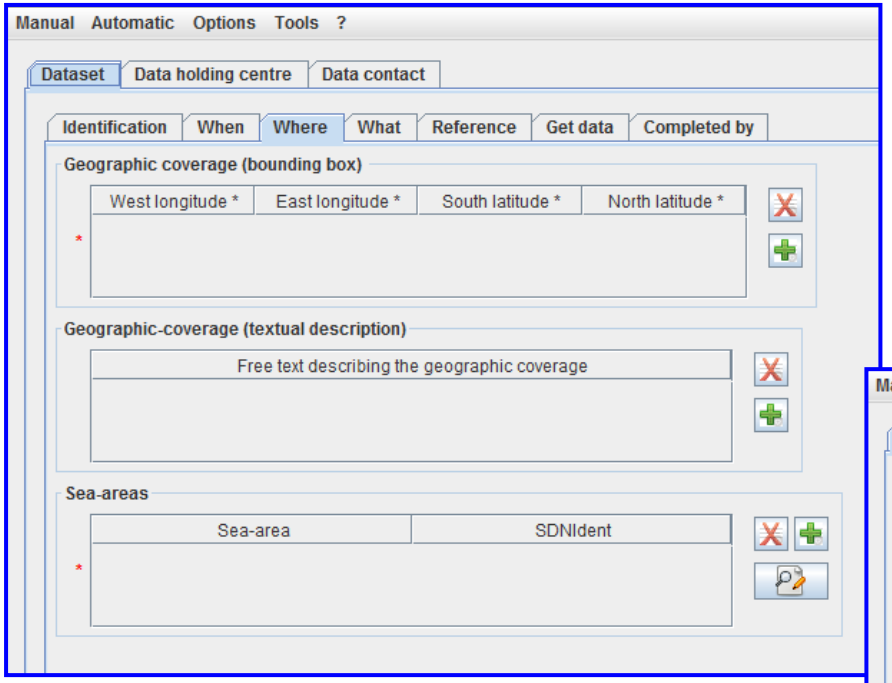
Description: The range of dates covered within the data set. Any major gaps should be noted. If the data set is to be updated, enter the most recent date and write 'ongoing'. Maximum 1000 characters.

Start date and End date: (dd/mm/yyyy) Earliest data and most recent data within the data set. Start date is **Mandatory**. If dataset is ongoing, leave End date blank. End date can't be in the future.

Revision date: (dd/mm/yyyy hh24:mm:ss) Date of last revision of the dataset. This is filled in automatically. **Mandatory**.

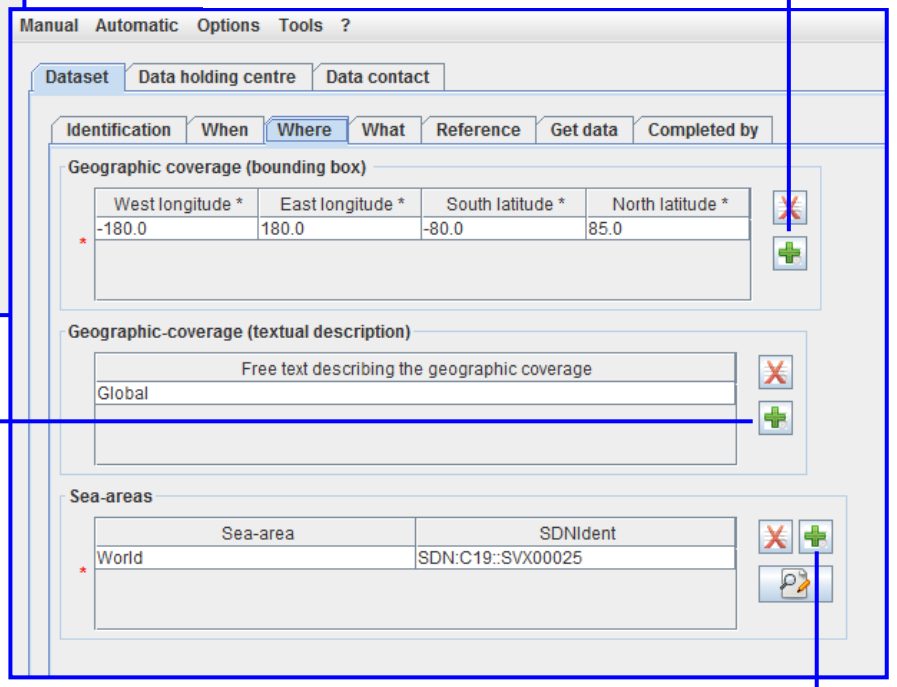


Page 1.c – Dataset: Where



Geographic coverage (bounding box): limits of dataset in degrees latitude and longitude. For a point measurement, fill in West and East identically, and South and North identically.

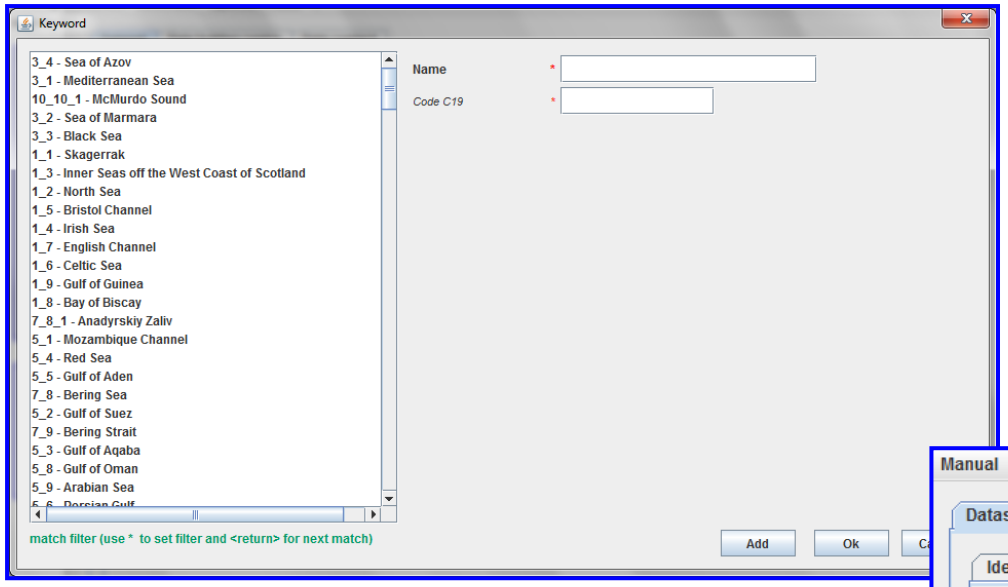
Several Bounding Boxes may be provided for schematising tracks. Northern latitudes and eastern longitudes are entered as positive and southern latitudes and western longitude are entered as negative. Values can be entered to 4 decimal places. Mandatory.

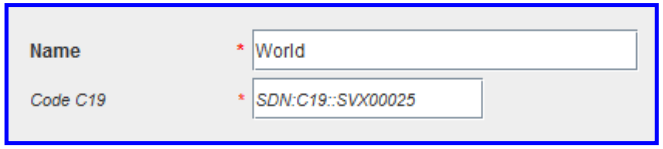


Geographic-coverage (textual description): General description of the geographic distribution of the data, using geographic names and/or latitude and longitude as appropriate. For offshore data, the ocean/sea areas should be clearly identified, while for coastal and estuarine data the description should include the name of the region/country. Maximum 2000 characters.

Select the appropriate **Sea-areas** from a list. Only include an ocean if the data set has ocean wide coverage. **Mandatory**.

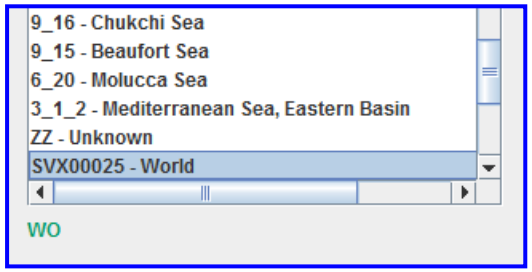
Clicking on the  symbol will open up the Keyword window. You can search for a particular sea area by holding down the shift key and typing a *. You can then type in a text string.

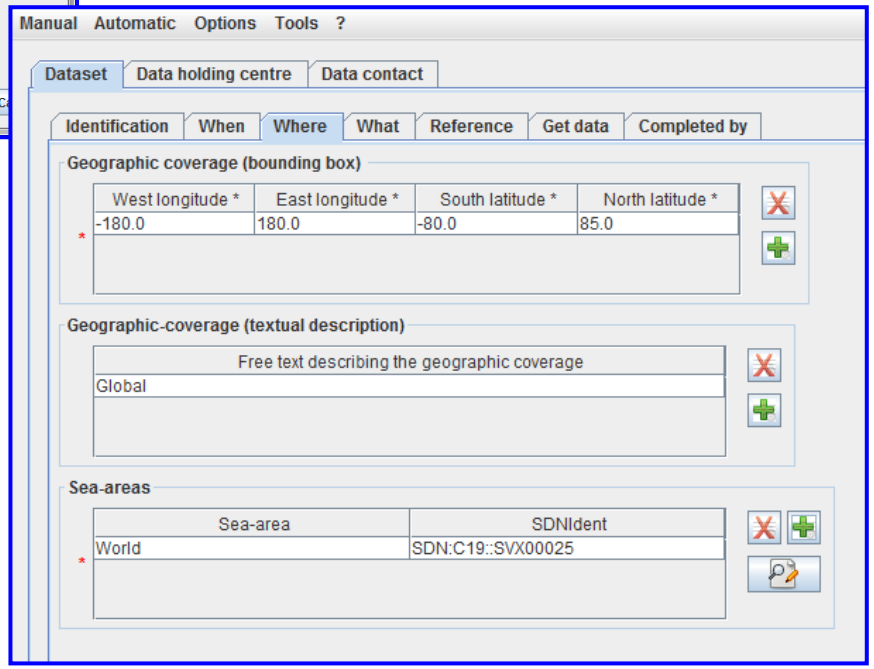




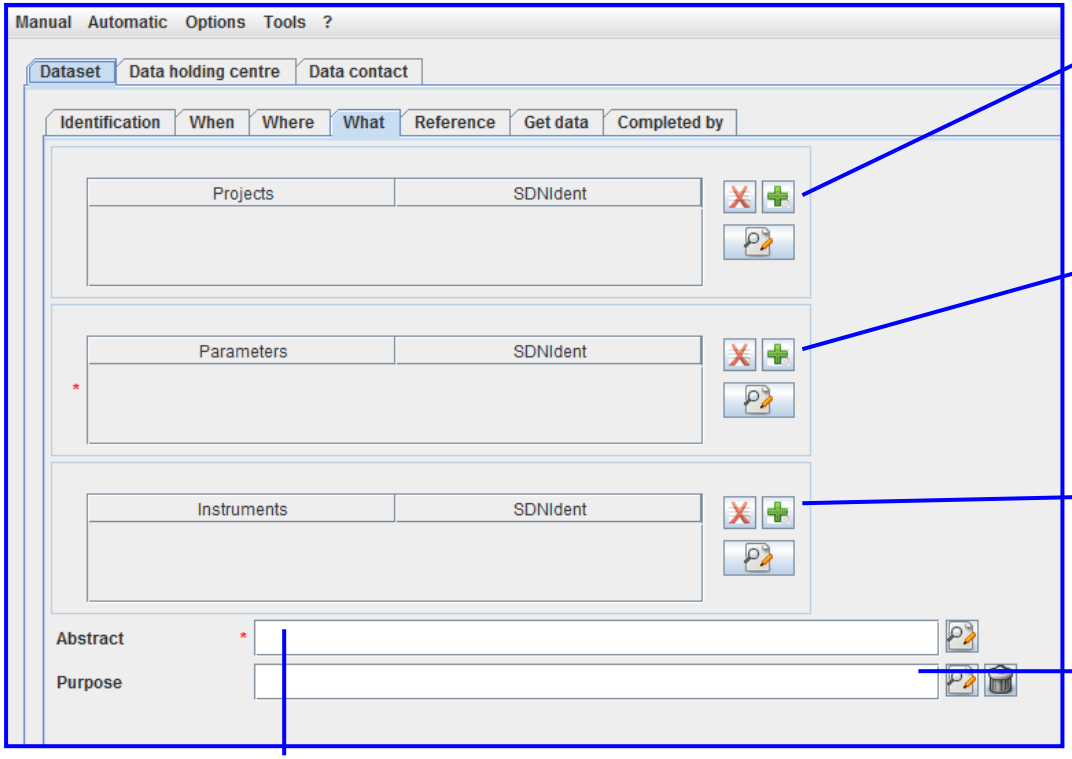
Typing * then **WO** will bring up the **Keyword** and code for Worldwide.

Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more sea areas to add.





Page 1.d – Dataset: What



Projects: Links to the European Directory of Marine Environmental Research Projects (EDMERP). Users can select relevant projects, providing they already exist in EDMERP, by clicking on the .

Parameters: Links to the SeaDataNet parameter vocabulary. Users can select multiple relevant parameters by clicking on the . **Mandatory**.

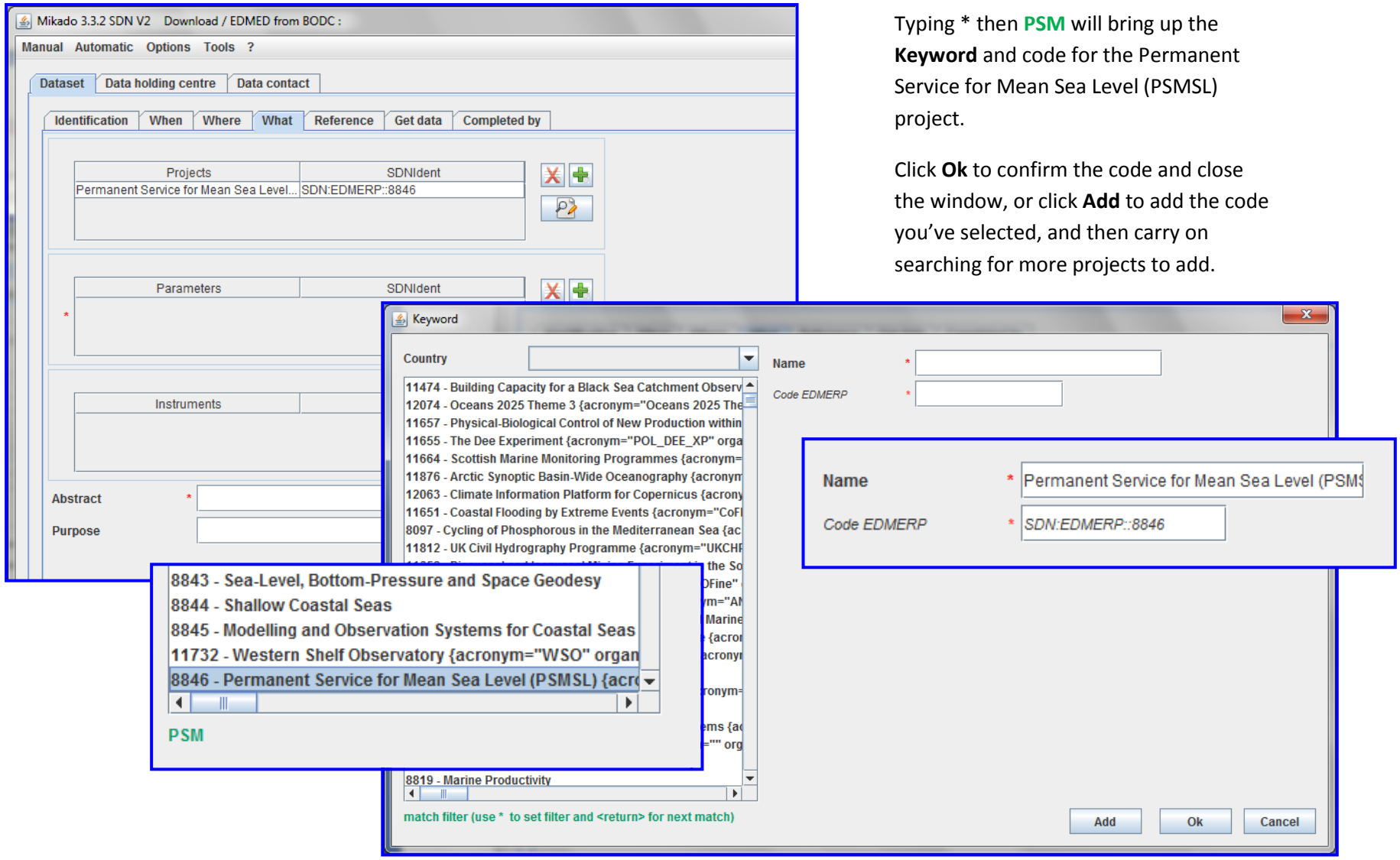
Instruments: Links to the SeaDataNet device categories. Users should select the instrument/gear/methodology used to collect the data by clicking on the .

Purpose: Summary of the intentions with which the resource was developed. Maximum 4000 characters.

Abstract: **Mandatory**. Maximum 4000 characters. This is a concise abstract and should contain brief statements for describing the data set. The information should include

- a description of the measurements/samples, the purpose for which they were collected, and the platforms, instrumentation and methods of sampling used in their collection.
- a statement on the level to which the data are processed and quality controlled, and any known limitations on their reliability.
- arrangement of data e.g. time series/depth series per station, underway tracks arranged by cruise, synoptically arranged data, gridded data, contoured maps etc.
- an estimate of the amount of data expressed in terms of the number of stations, sites, observations, cores, months of recording, miles of track, net hauls, or other units as appropriate.
- a statement of data sources i.e. which organizations contributed data to the data set.

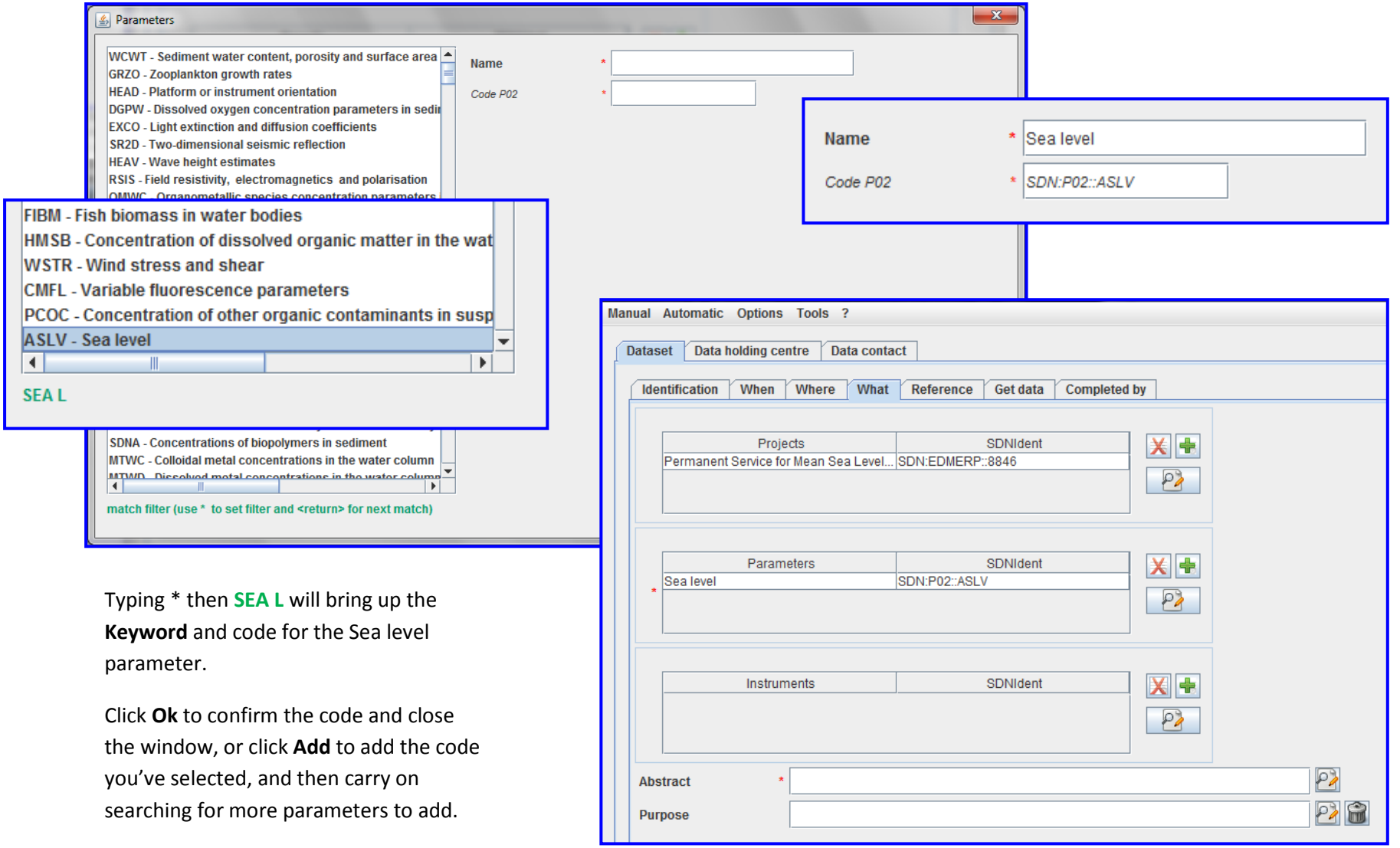
Clicking on the  symbol will open up the Keyword window. You can search for a particular project by holding down the shift key and typing a *. You can then type in a text string.



Typing * then **PSM** will bring up the **Keyword** and code for the Permanent Service for Mean Sea Level (PSMSL) project.

Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more projects to add.

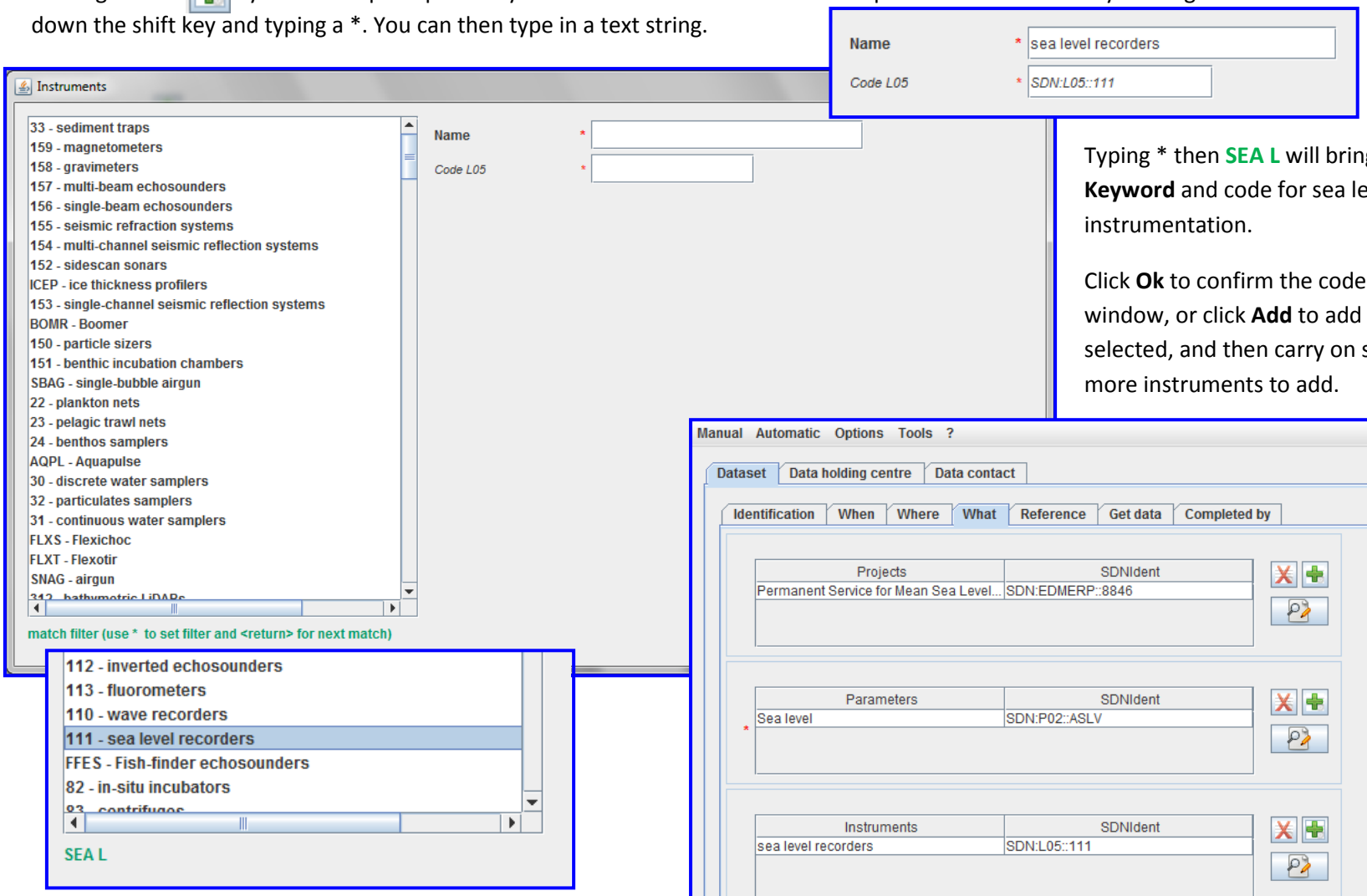
Clicking on the  symbol will open up the Keyword window. You can search for a particular parameter by holding down the shift key and typing a *. You can then type in a text string.



Typing * then **SEA L** will bring up the **Keyword** and code for the Sea level parameter.

Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more parameters to add.

Clicking on the  symbol will open up the Keyword window. You can search for a particular instrument by holding down the shift key and typing a *. You can then type in a text string.



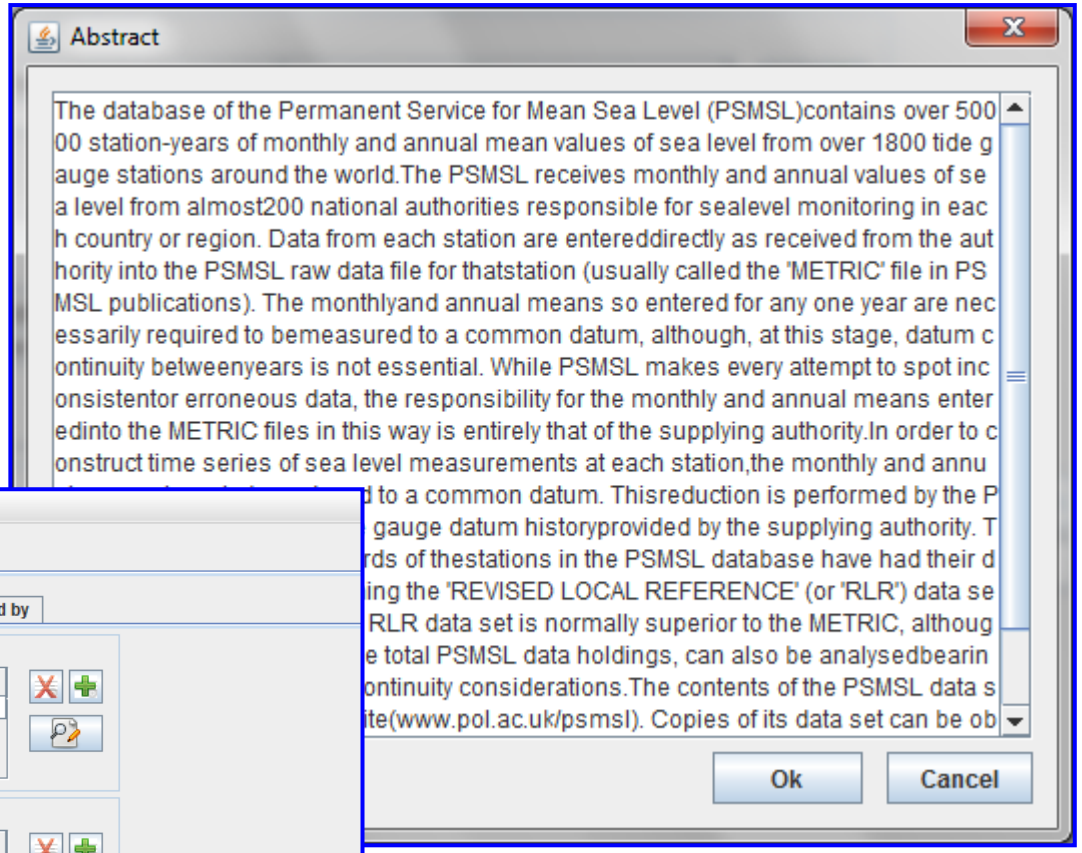
Typing * then **SEA L** will bring up the **Keyword** and code for sea level recorder instrumentation.

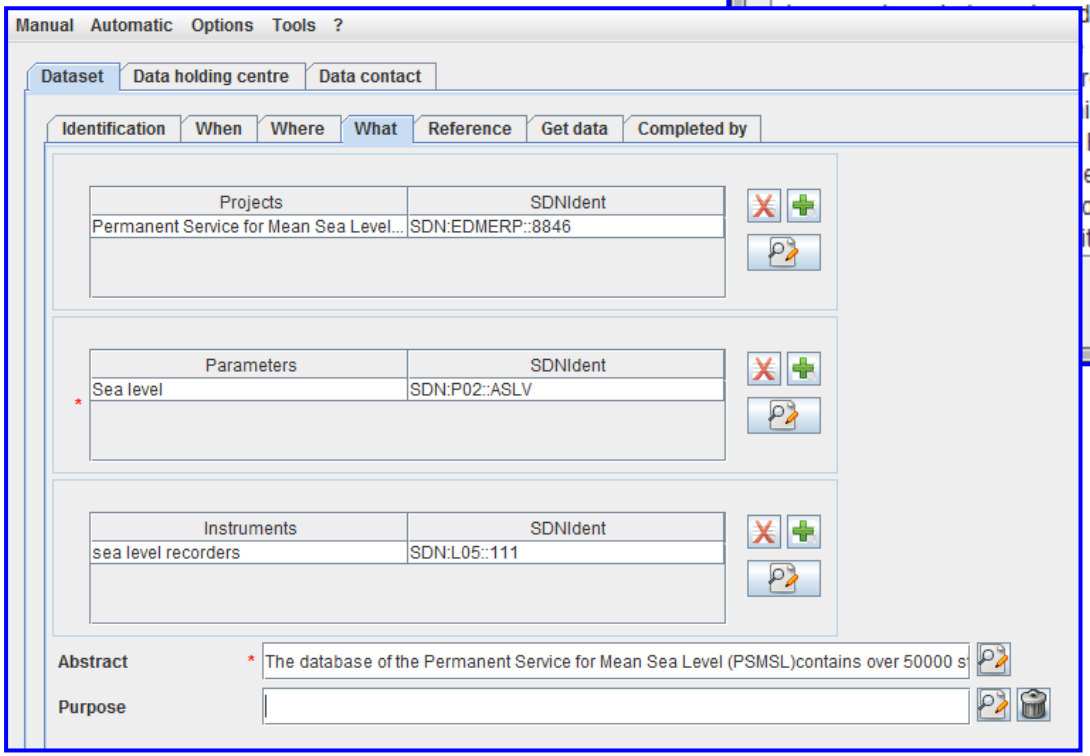
Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more instruments to add.

Click on the  symbol to open up the Abstract text box.

This is a concise abstract and should contain brief statements for describing the data set.

Maximum 4000 characters.





Instructions for completing the Abstract (taken from NERC Guidance for Authors of Discovery Metadata)

Requirements

- The abstract should describe the contents of the resource in plain language for a non-expert user (first year undergraduate level).
- Write the abstract in sentences.
- The abstract should describe the resource in question, NOT the project/activity which produced it. Some details of the activities which produced the resource may be pertinent, in which case they should be included.
- The first few one or two sentences should summarise the contents of the resource.
- Where they are used, specialist terms should be explained in full.
- Where acronyms and abbreviations are used, they should be reproduced in full. They should take the format - Full Name (ACRONYM). For example “Land Ocean Interaction Study (LOIS)”

Recommendation

- It is recommended that the abstract is organised using the "What, Where, When, How, Why, Who" structure.

The abstract is an ‘executive summary’ that allows the reader to determine the relevance and usefulness of the resource. The text should be concise but should contain sufficient detail to allow the reader to ascertain rapidly the scope and limitations of the resource.

Write for readers, *not* robots and write complete sentences rather than fragments. For example:

Regional Geochemical data from drainage basin reconnaissance survey carried out as part of a bilateral aid project between the UK Department for International development (DFID) (formerly ODA) and Indonesian Government. Some 23,000 stream sediment samples collected and analysed.	✘
This dataset consists of regional geochemical data from a drainage basin reconnaissance survey. The survey was carried out as part of a bilateral aid project between the UK Department for International development (DFID) (formerly ODA) and the Indonesian Government. Some 23,000 stream sediment samples were collected and analysed.	✓

The maximum length of an abstract is 4000 characters, but it can be much shorter.

Keep sentences short. The average length of a sentence should be about 15-20 words. Very long sentences can almost always be avoided by breaking them up in some way. To help organise thinking, the author may like to use the following structure:

1. **What**
   A description of what has been recorded and what form the data takes. This should immediately convey to the reader precisely what the resource is.
2. **Where**
   A description of the spatial coverage. This should include, where relevant, whether the coverage is gridded or scattered data; whether the coverage is even or very variable
3. **When**
   A description of the temporal coverage (e.g. the period over which data were collected)
4. **How**
   A brief description of methods and instrumentation used.
5. **Why**
   For what purpose was the data collected?
6. **Who**
   The party/parties responsible for the collection and interpretation of data.
7. **Completeness**
   Are any data absent from the dataset? Explain which data are included or excluded and why.

One or more of these elements may not always be applicable; where they are not applicable they may be omitted. If you can't easily summarise or describe the resource it could be a sign that it isn’t fully understood.

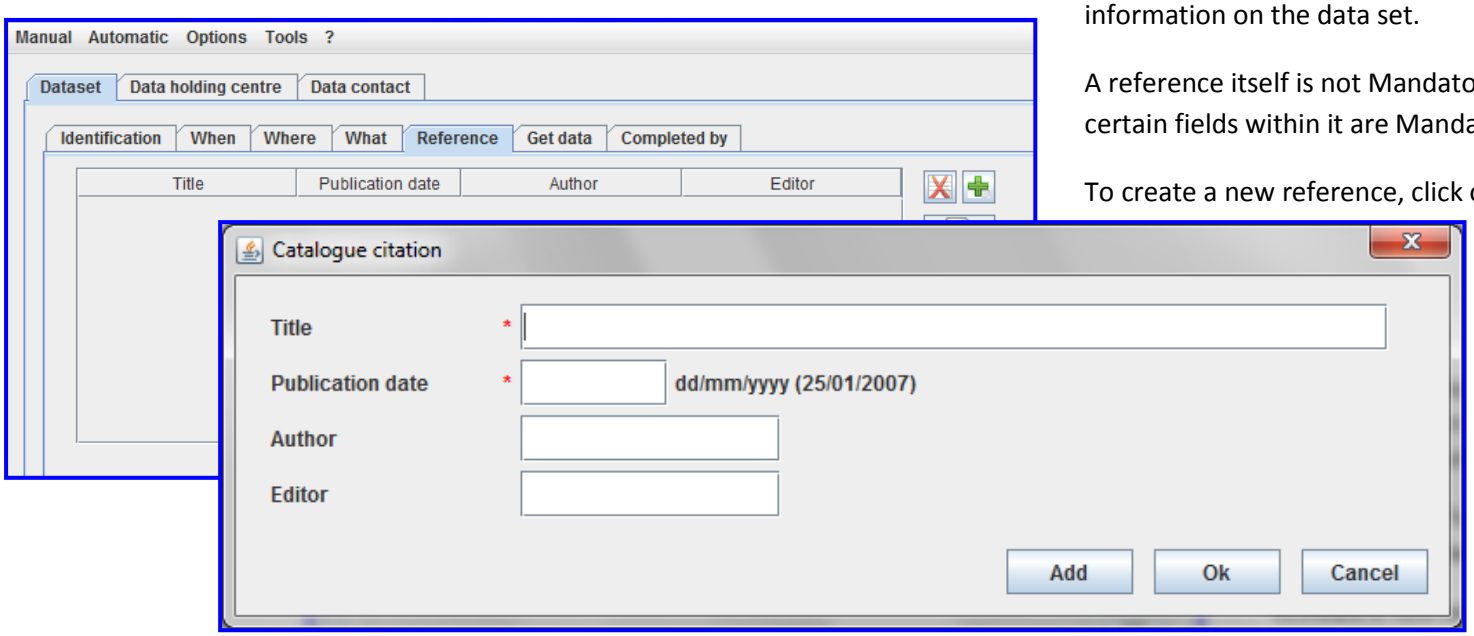
This dataset contains a variety of atmospheric measurements including time series of air temperature, wind speed and direction, precipitation, irradiance and humidity.	
A comprehensive atmospheric sampling programme provided measurements of atmospheric particulates, aerosols and gases, including hydrocarbons, nitrogen, oxygen, ozone and sulphur species, carbon monoxide, carbon dioxide, and nitrous and hydrochloric acids. Additional measurements of photolysis rates and ion and radical concentrations were also collected.	**What**
The data were collected from the vicinity of the north Norfolk coast between 1994 and 1997.	**Where**
The bulk of the data were collected during two field campaigns in the winter (October/November) of 1994 and the summer (May/June) of 1995.	**When**
During these campaigns data were collected continuously from the University of East Anglia (UEA) Atmospheric Observatory at Weybourne on the north Norfolk coast. The widest range of parameters is available for this station. An instrumented vessel (MV Guardian) was stationed offshore to provide a second sampling site to allow changes in a given air mass to be monitored.	**How**
The Imperial College London Jetstream Research aircraft made one flight during each campaign to provide a link between the two surface stations and four additional flights in 1996 and 1997.	
The River-Atmosphere-Coast Study (RACS) was the component of the LOIS programme looking at processes from the river catchment into the coastal sea.	**Why**
Professor John Plane from the Environmental Sciences Department at UEA was the scientific co-ordinator of this sub-project of LOIS.	**Who**
The data are held by BODC as a series of ASCII data files conforming to the NASA AMES 1001 format together with a PDF document that describes the data set	**Completeness**

Example abstract 1.

The Northern Seas Programme dataset comprises hydrographic, biogeochemical, biological and meteorological data. Hydrographic profiles provided measurements of parameters such as temperature, salinity, fluorescence and dissolved oxygen, while current velocities and acoustic backscatter were also measured. A comprehensive water sampling programme permitted the collection of biogeochemical data including concentrations of various organic compounds, dissolved gas concentrations and radioactivity. Water samples were also analysed for phytoplankton, zooplankton and viruses. Larger biological samples were obtained from the water column using trawl nets and cetacean distributions were monitored using hydrophone arrays. Sediment samples were collected at various locations and analysed for biogeochemical parameters and zoobenthos. Sample data were supplemented by those derived from experiments, while bathymetry and meteorological parameters were measured across the study area.	**What**
Data collection was undertaken in the Irish and northern North Seas, across the NE Atlantic and up to the marginal Arctic pack ice zone. This includes the territorial waters of the UK, Norway and the Russian Arctic, and extends from coastal fjords to the ocean margins.	**Where**
The data were collected during the period 2001-2007 over a number of cruises: RRS Discovery cruise D257, RRS James Clark Ross cruises JR75 and JR127, RRS Charles Darwin cruise CD176 and FS Poseidon cruise PO300/2.	**When**
Measurements were taken using a variety of instrumentation, including conductivity-temperature-depth (CTD) profilers with attached auxiliary sensors, bathymetric echosounders, sediment samplers, trawl nets and acoustic Doppler current profilers (ADCPs), while incubation chambers were used for shipboard experiments	**How**
The programme was designed to advance the understanding of how marine systems in Northern Seas respond to environmental and anthropogenic change and was developed in three themes: Theme A - Understanding fjordic systems insights for coastal and oceanic processes; Theme B - Ocean Margins: the interface between the coastal zone and oceanic realm; Theme C - Measuring and modelling change: sea sensors and bioinformatics. Theme B included the Ellett Line Time Series	**Why**
The Northern Seas Programme was co-ordinated by the Scottish Association for Marine Science (SAMS). Data from the programme are held at the British Oceanographic Data Centre.	**Who**

Example abstract 2.

Page 1.e – Dataset: Reference



The **Reference** page should be used to list any information sheets or published literature that includes further information on the data set.

A reference itself is not Mandatory, but if one is created, certain fields within it are Mandatory.

To create a new reference, click on  .

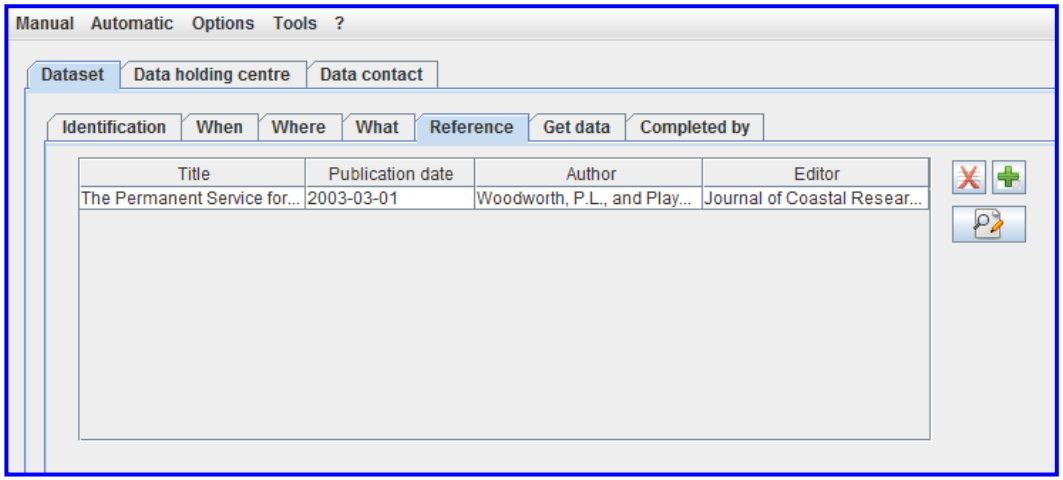
Title: The title of the journal article, book, etc. **Mandatory**. Maximum 500 characters.

Publication date: The date the article was published. dd/mm/yyyy. **Mandatory**.

Author: Names of the responsible people. Maximum 200 characters.

Editor: Name of the Publisher of the journal, Editor of a book, etc. Maximum 100 characters.

Click **Ok** to confirm your entry and close the window, or click **Add** to add the entry you've created, and then carry on creating more references.



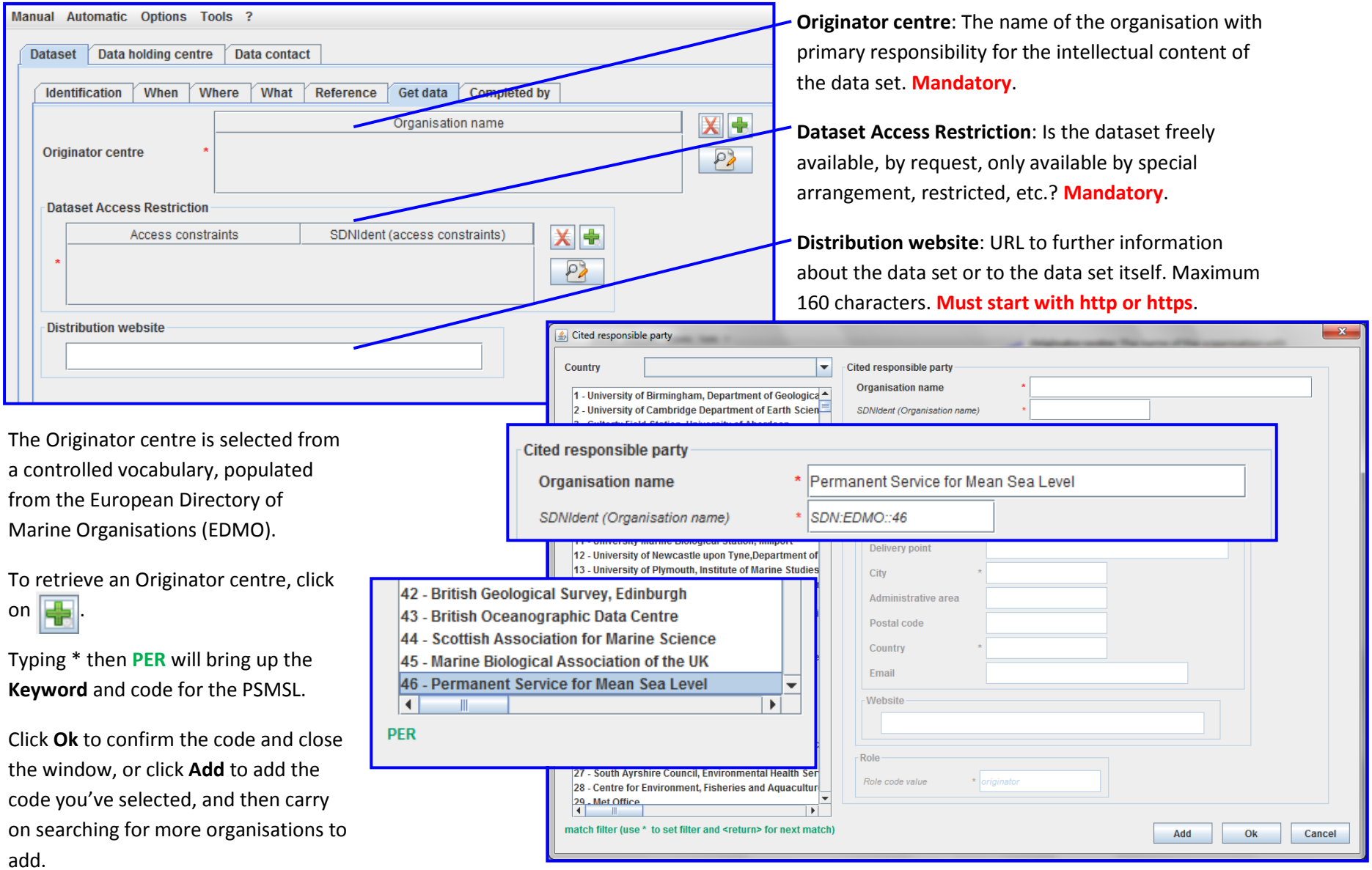
Page 1.f – Dataset: Get data

The Get data page provides information on how to get hold of the dataset.

Originator centre: The name of the organisation with primary responsibility for the intellectual content of the data set. **Mandatory**.

Dataset Access Restriction: Is the dataset freely available, by request, only available by special arrangement, restricted, etc.? **Mandatory**.

Distribution website: URL to further information about the data set or to the data set itself. Maximum 160 characters. **Must start with http or https**.

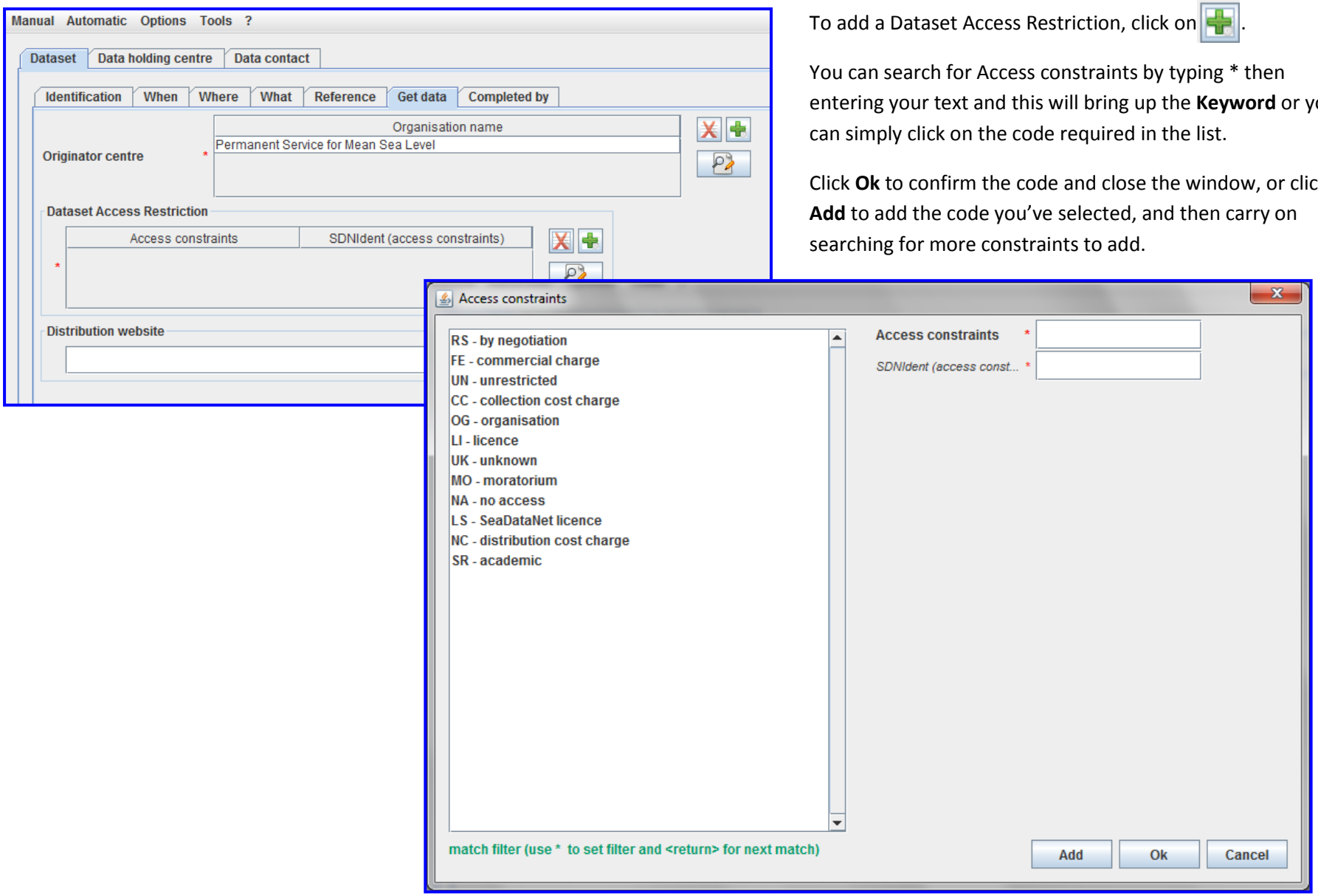


The Originator centre is selected from a controlled vocabulary, populated from the European Directory of Marine Organisations (EDMO).

To retrieve an Originator centre, click on .

Typing * then **PER** will bring up the **Keyword** and code for the PSMSL.

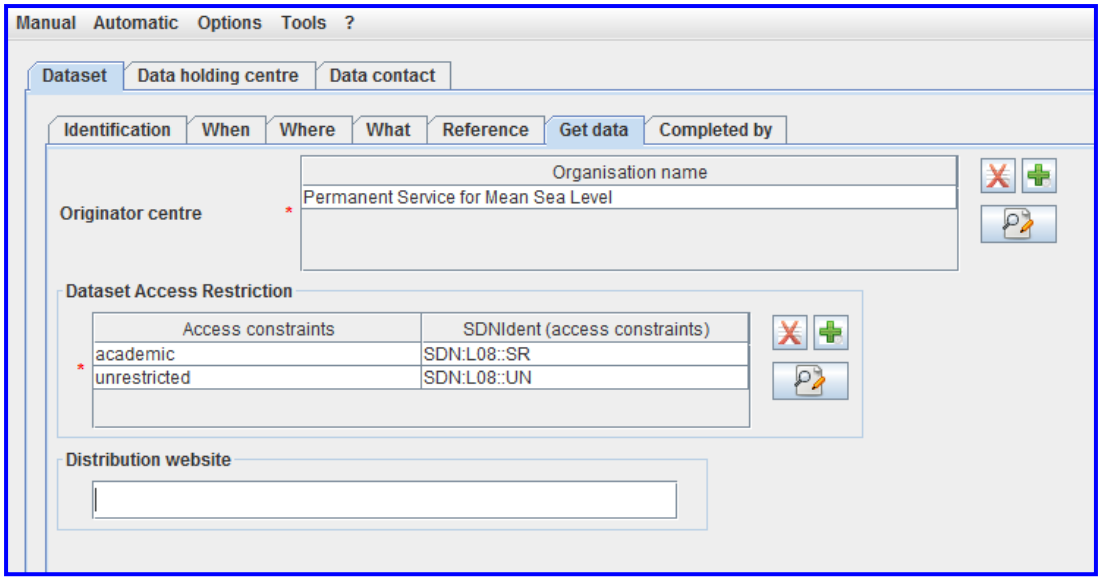
Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more organisations to add.



To add a Dataset Access Restriction, click on .

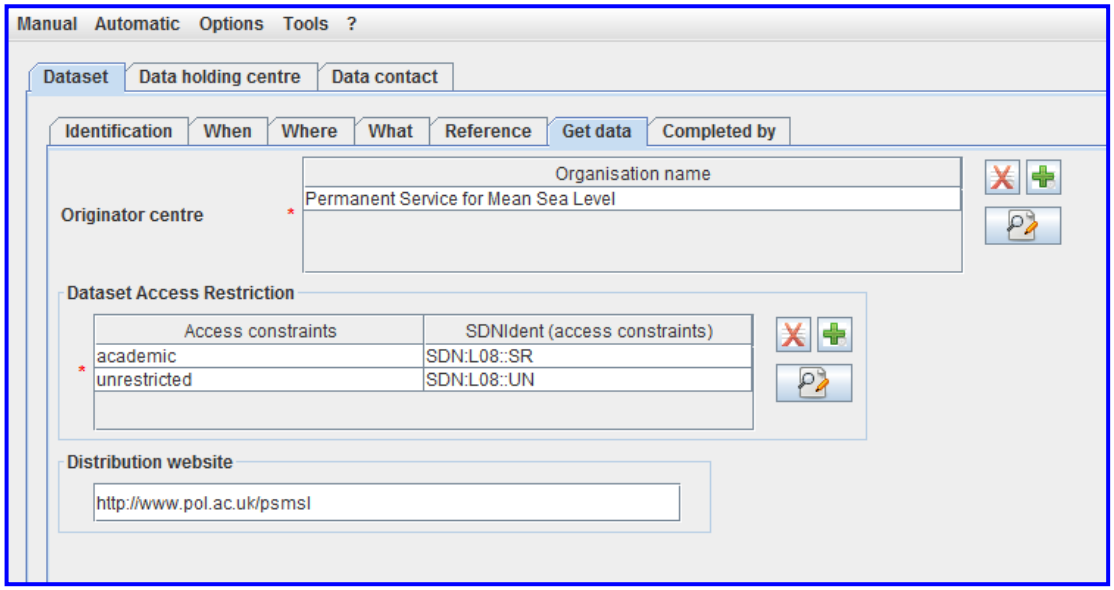
You can search for Access constraints by typing * then entering your text and this will bring up the **Keyword** or you can simply click on the code required in the list.

Click **Ok** to confirm the code and close the window, or click **Add** to add the code you've selected, and then carry on searching for more constraints to add.



The Distribution website field is a free text box and should contain the URL to further information about the data set or to the data set itself.

Maximum 160 characters. **Must start with http or https**.



Page 1.g – Dataset: Completed by

The Completed by page contains information about the organisation that prepared the EDMED description.

To retrieve a Collate-centre, click on  .

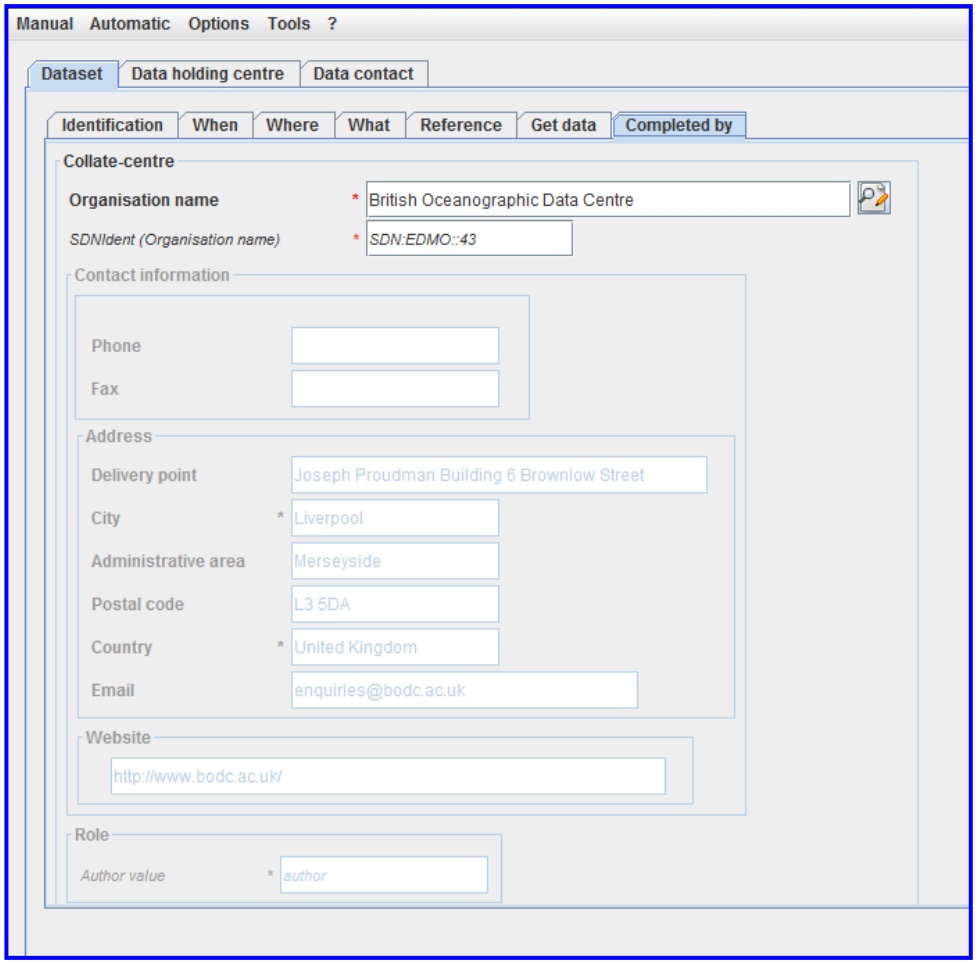
Typing * then **BRITISH O** will bring up the **Keyword** and code for the British Oceanographic Data Centre.

Click **Ok** to confirm the code and close the window.

Selecting an EDMO code will then automatically fill in the:

- Contact information
- Address
- Website
- Role

with the information from EDMO.



Page 2 – Data holding centre

The Data holding centre page contains information about the organisation holding and distributing the dataset.

To retrieve an Organisation name, click on  .

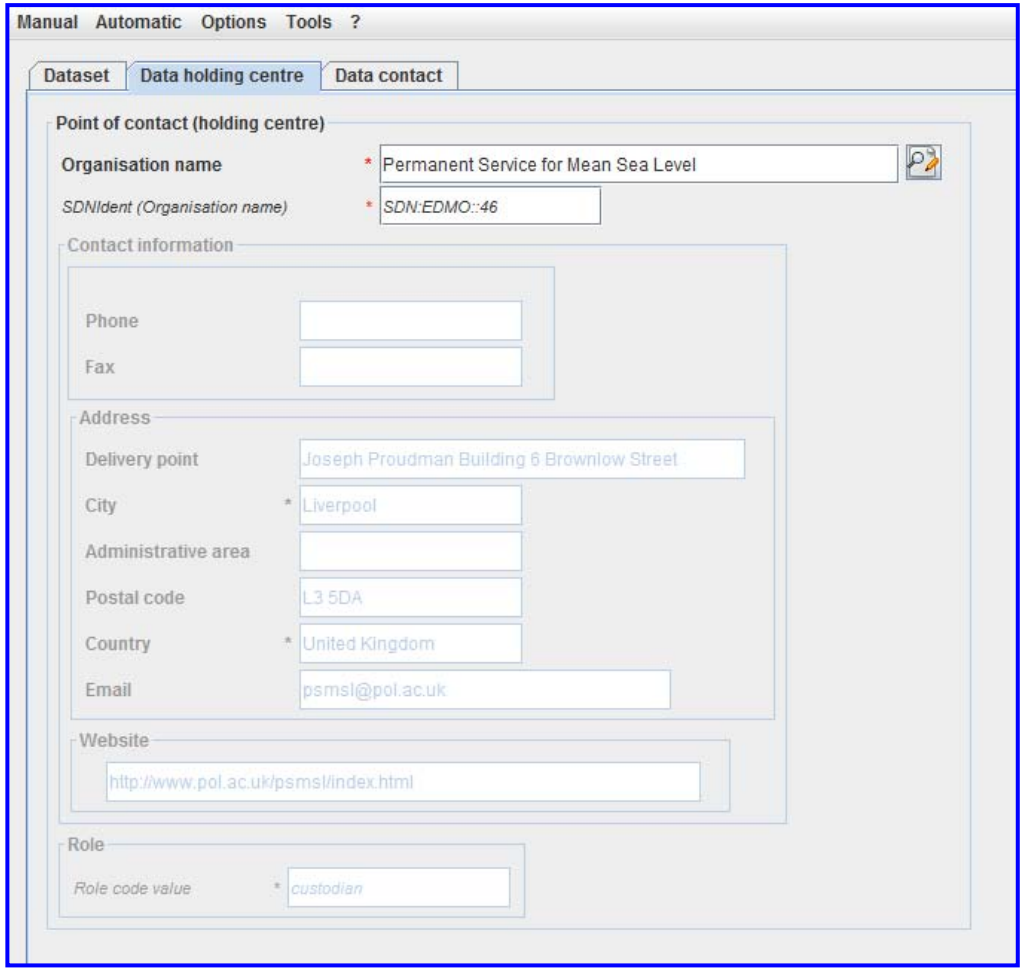
Typing * then **PER** will bring up the **Keyword** and code for the Permanent Service for Mean Sea Level.

Click **Ok** to confirm the code and close the window.

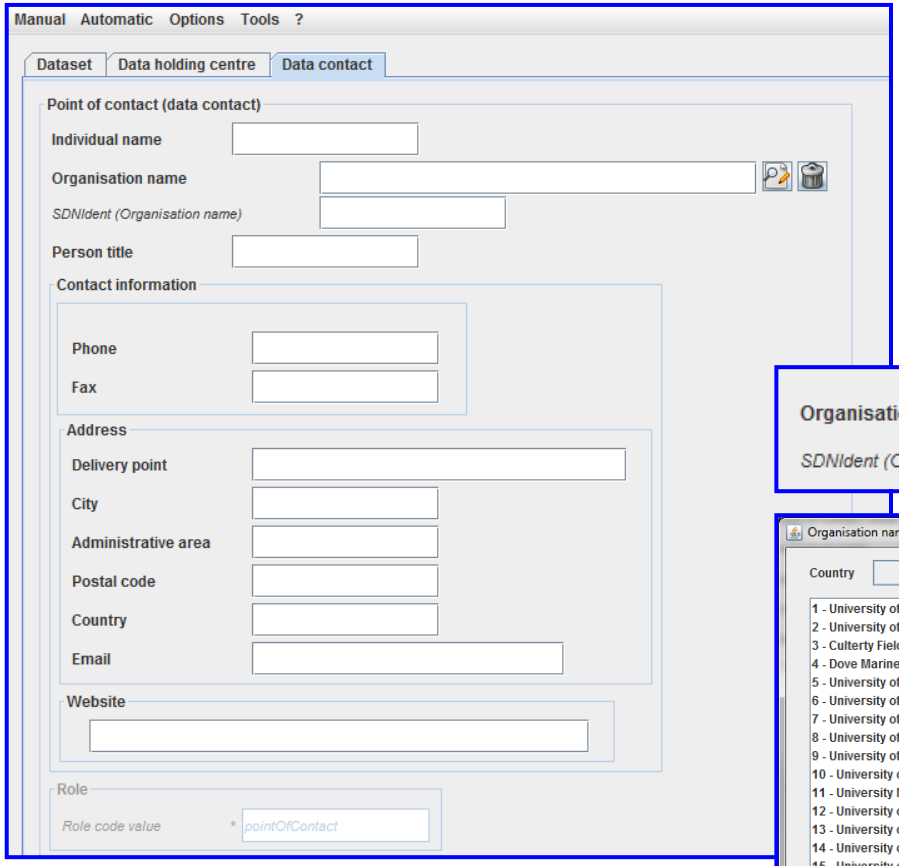
Selecting an EDMO code will then automatically fill in the:

- Contact information
- Address
- Website
- Role

with the information from EDMO.



Page 3 – Data contact

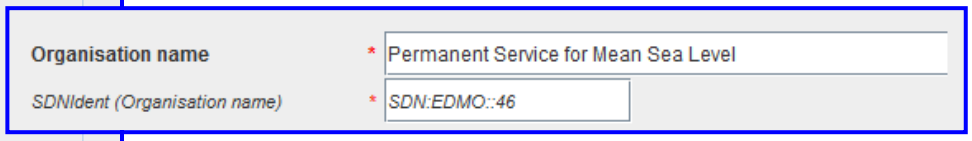


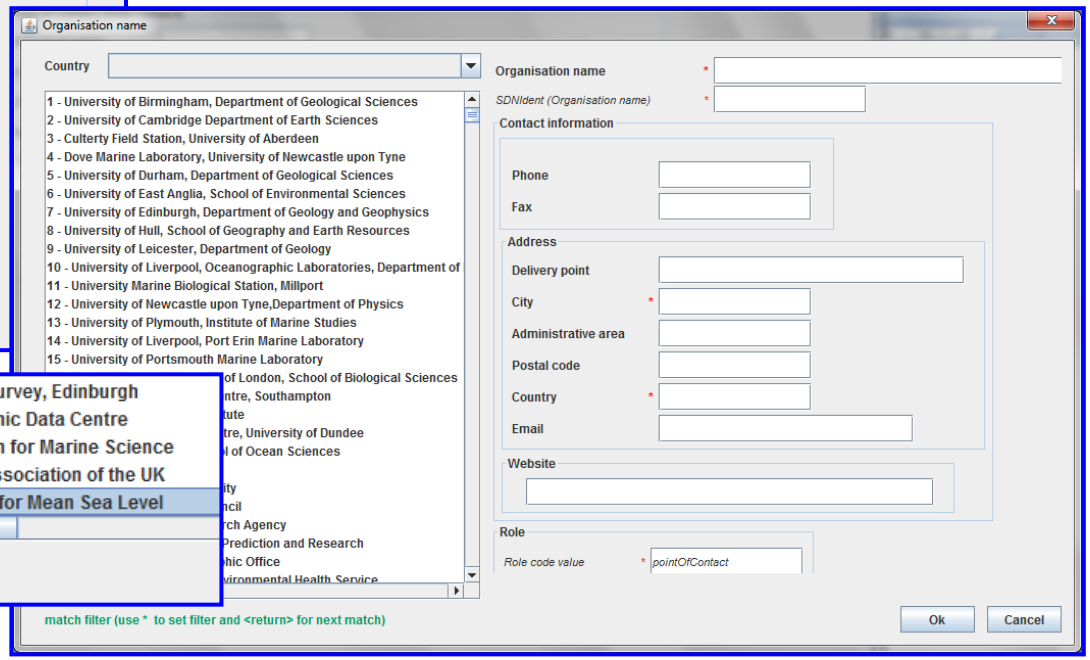
The Data contact page contains information about the organisation managing the dataset.

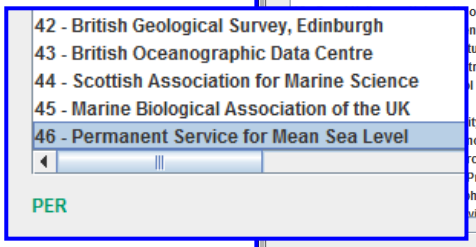
Individual name: Name of the person to contact for data information, working at the Data Centre, managing the data set. Maximum 80 characters.

Organistation name: Selected from the EDMO list.

Person title: Role or job title of person to contact for data information e.g. 'Enquires Officer', 'Database manager'. **THIS SHOULD NOT BE DR. or PROF.**. **Mandatory**. Maximum 80 characters.







To retrieve an Organisation name, click on  .

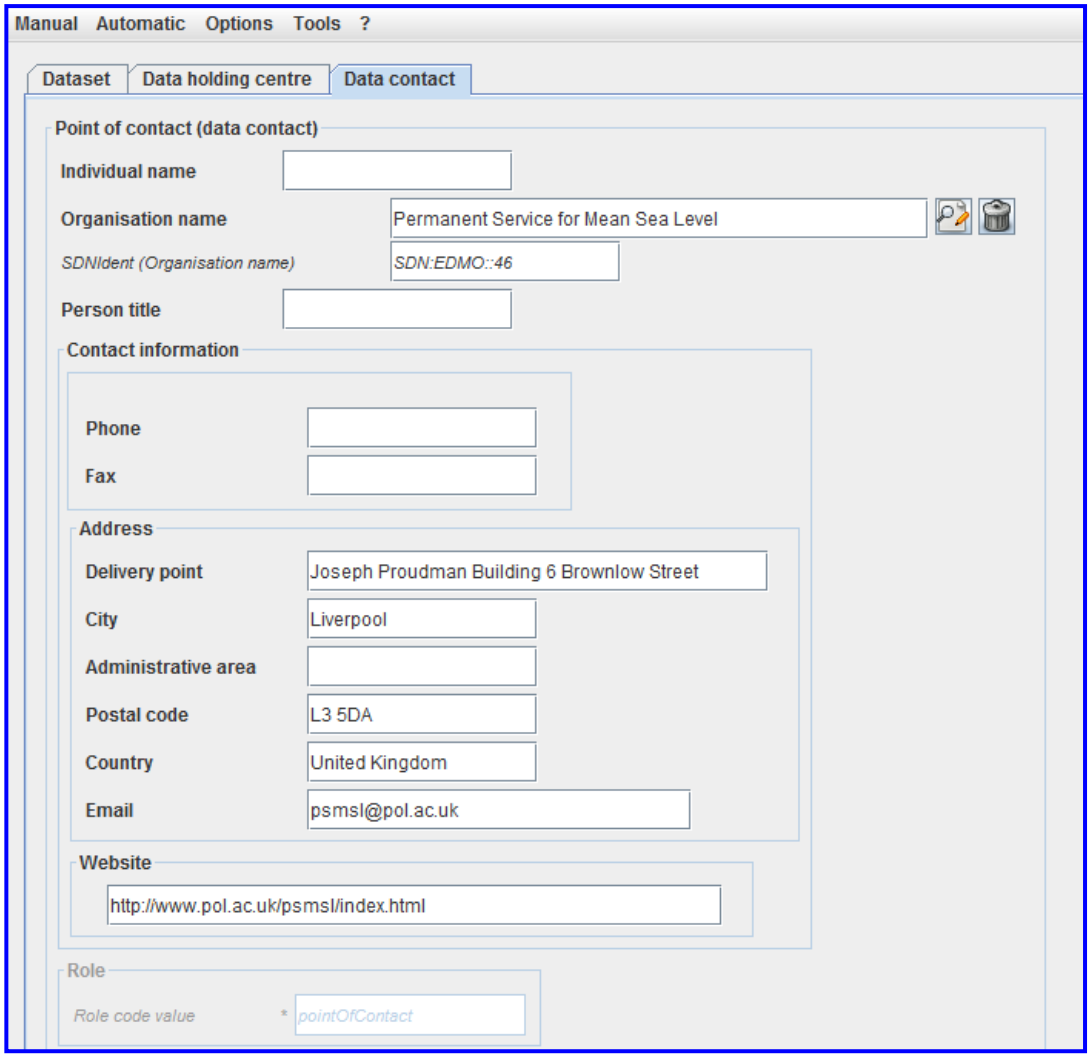
Typing * then **PER** will bring up the **Keyword** and code for the Permanent Service for Mean Sea Level.

Click **Ok** to confirm the code and close the window.

Selecting an EDMO code will then automatically fill in the:

- Contact information
- Address
- Website
- Role

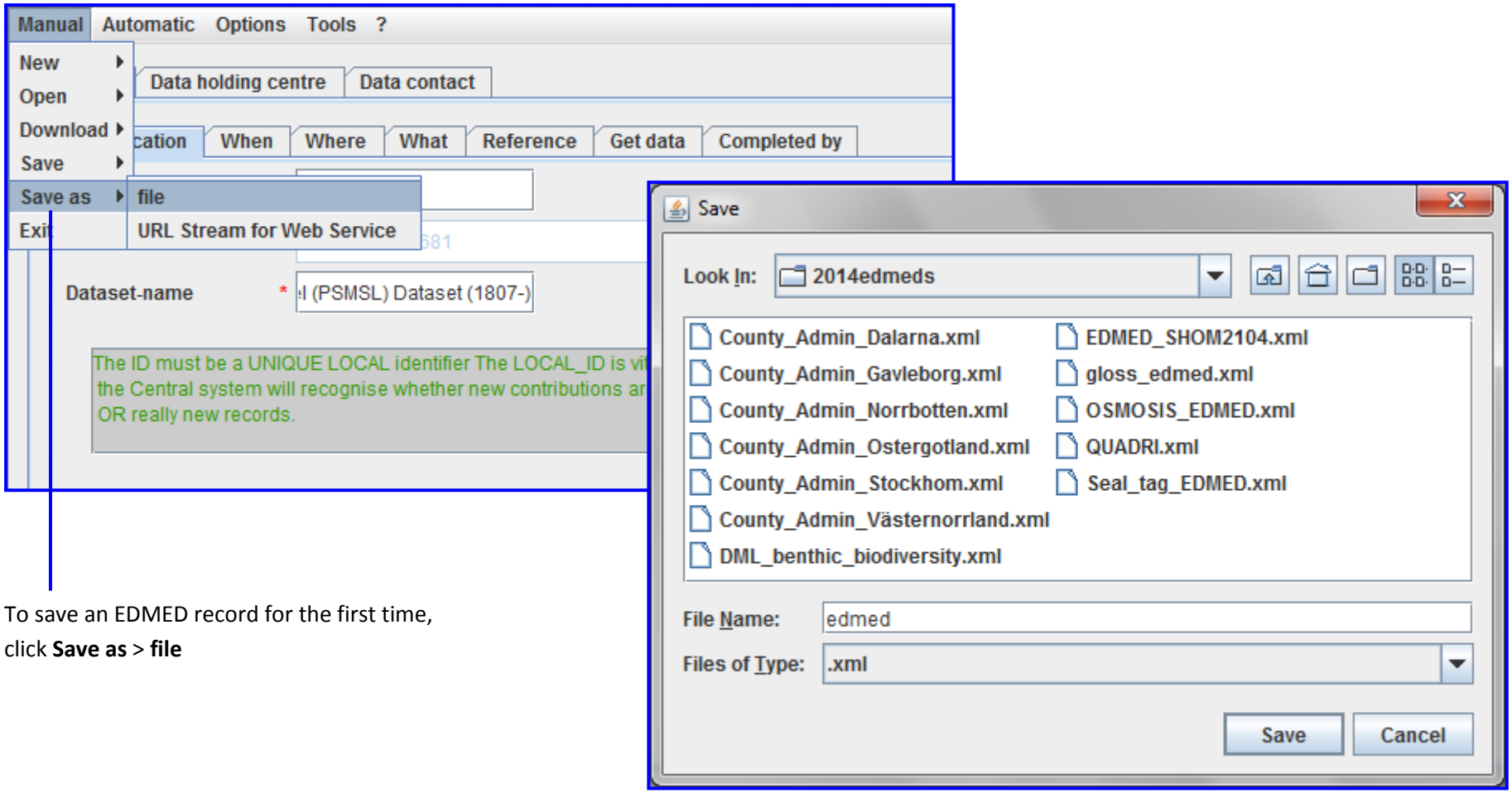
with the information from EDMO.



Saving your EDMED record for the first time

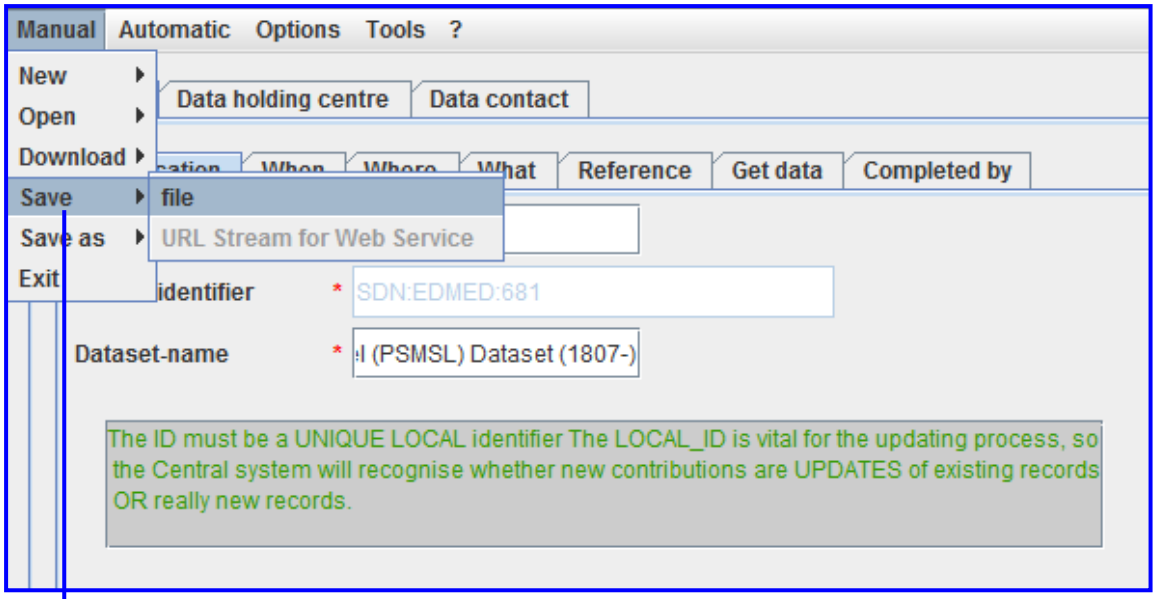
You cannot save until you have filled in all your mandatory values! This means that you cannot save a partial document and is to maintain the integrity of the output xml.

Be warned again! When you have filled in a 'plain text' box, make sure you click on another box; this effectively "exits" the box. If you don't click elsewhere and try to save, it does not clue in that you typed something in the box because there was no "exit" of the box.



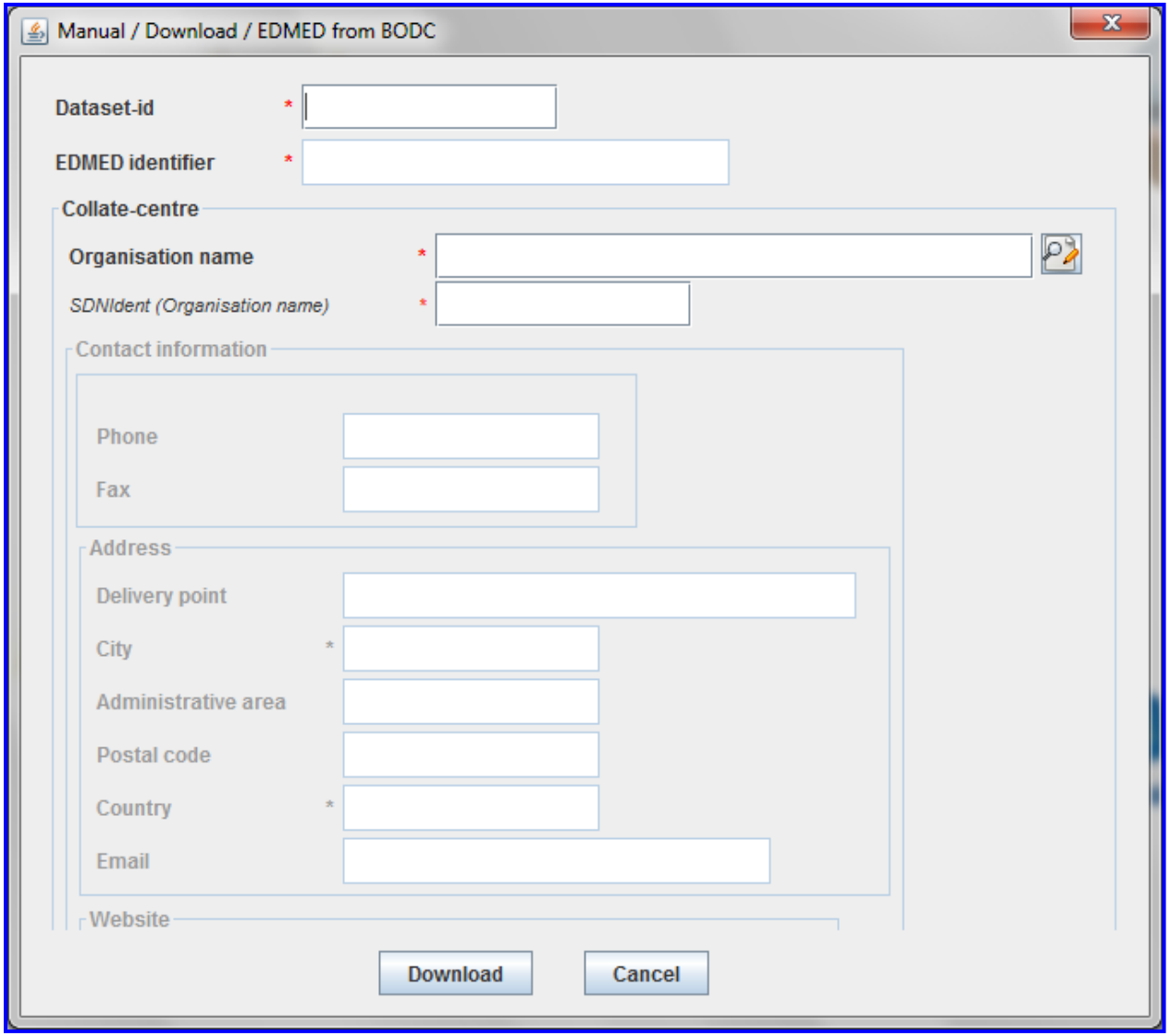
To save an EDMED record for the first time, click **Save as** > **file**

Saving an EDMED record after the first time



To save an EDMED record after the first time, click **Save** > **file**

Updating an existing EDMED entry



Dataset-id: Fill in the text box with your dataset's id.

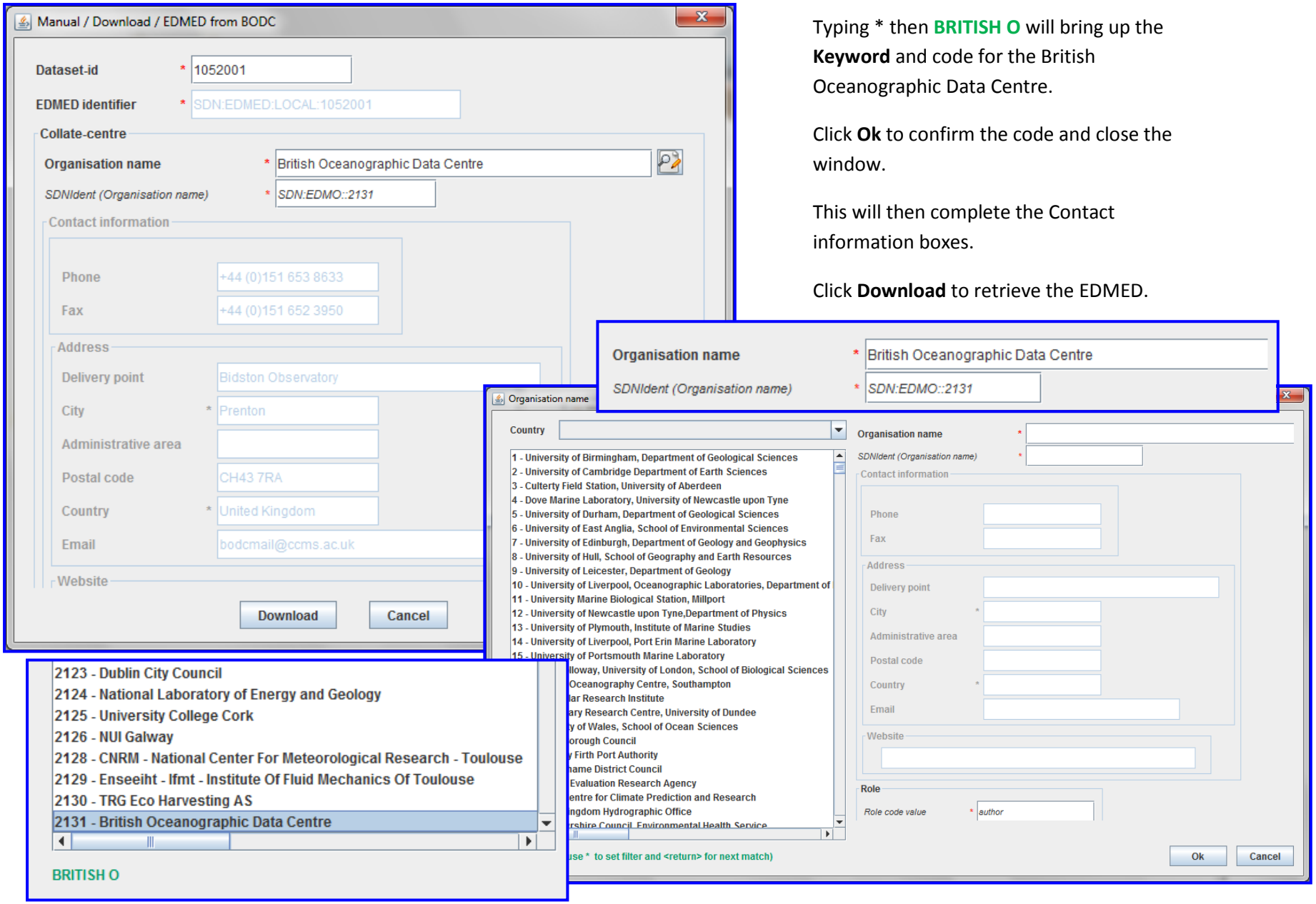
This will then automatically complete the **EDMED identifier**.

To retrieve a Collate-centre, click on  .

When an Organisation name has been selected the fields:

- Contact information
- Address
- Website
- Role

will automatically be filled in with the information from EDMO.



Typing * then **BRITISH O** will bring up the **Keyword** and code for the British Oceanographic Data Centre.

Click **Ok** to confirm the code and close the window.

This will then complete the Contact information boxes.

Click **Download** to retrieve the EDMED.

Report any bugs in Mikado to: sdn-userdesk@seadatanet.org