

Metadata Information Requirements

Introduction To This Document

Management of environmental research data is a core activity for the ERC with particular emphasis on the application of appropriate data management techniques to ensure their long-term availability and accessibility. Environmental research data are often irreplacable; they are always unique, if only in the spatial location and temporal characteristics of their collection. They can also be extremely expensive and difficult to collect or generate. For these reasons the EPA and the ERC attach great importance to the ongoing development of systems that will ensure maximum benefits are derived from research data once acquired.

This document describes the metadata that researchers must submit to the EPA using the SAFER-Data system. SAFER-Data is an acronym for *Secure Archive For Environmental Research Data*.

The SAFER-Data system is available at http://erc.epa.ie/safer

The term **Dataset** is defined as follows:

- a collection of one or more digital computer files:
- a database which is accessed using database system software;
- part or one of the above as specified by some filtering, querying, or extraction.

The term *Metadata* can be described as follows:

- text-based information that provides information about a dataset or collection of datasets;
- capturing the basic characteristics of a dataset resource. It represents the who, what, when, where, why and how of the dataset resource. Geospatial metadata are used to document geographic digital resources such as Geographic Information System (GIS) files, geospatial databases, and earth imagery.
- Allows persons not familiar with the dataset to find (a) who created/owns the
 dataset, (b) what the dataset represents, (c) when it was created or edited, (d) why
 it was created, (e) how it is used or accessed, and (f) any other information that will
 assist in evaluating the fitness for purpose of the dataset.

Several acronyms are used in the remainder of the document. Unless otherwise stated ERC-DM indicates the ERC Data Manager, ERC-C indicates the ERC Co-ordinator, and EPA-M indicators the EPA Research Manager.

This document is part of a series of ERC Data Management documents. The other documents in this series are "Providing Large Volumes of Data To The ERC", "SAFER-Data User Guide", and "SAFER-Data Usage and Usage Policy". While these documents have been developed as documents for usage by users of data services provided by the ERC the wider distribution of these documents is not restricted. They may be circulated "as is" to other organisations or individuals who may have an interest in data management or whom are developing a data management system or data management policy of their own.

Contact Person: Peter Mooney peter.mooney@nuim.ie

The requirement that all research projects funded by the EPA under ERTDI or STRIVE provide metadata and all generated datasets is outlined in the Terms and Conditions "Guide For Grantees" (Section 8 Final Reporting Guidelines).

http://www.epa.ie/downloads/pubs/other/corporate/oea/research/

Table of Contents

Introduction To This Document	1
The ERC's ISO 19115 Metadata Subset	4
Explaination of Metadata Terminology	4
Identification Information Metadata Fields	5
Resource Title:	
Resource Keywords:	
Resource Language:	
Resource Scope:	
Metadata Progress:	6
Project Code/Reference Number:	6
Internal Thematic Area:	
ISO 19115 Topic Category:	7
Frequency of Changes To Metadata Record:	7
Temporal Information Metadata Fields	8
Commencement of Research Work Date:	8
Earliest Recorded Date/Time value:	
Most Recent Recorded Date/Time value	8
Geographical Information Metadata Fields	10
Geographical Description:	10
West Bounding Coordinate:	
East Bounding Coordinate:	
South Bounding Coordinate:	
North Bounding Coordinate:	
Access Restrictions and Constraints Metadata Fields	
Resource Access Limitations:	
Resource Usage Limitations:	12
Resource Availability From SAFER-Data:	
Additional Information Metadata Fields	
Resource Lineage:	
Links To Internet Resources:	
Resource Abstract	
Supplementary Information:	
Dataset Supply and Presentation Modes Metadata Fields	
Digital File Types:	
Resource Supply Mode:	
Resource Supply Types:	16

Responsible Parties Contact Information	.18
Responsible Party ROLE CODES:	.19

The ERC's ISO 19115 Metadata Subset

The ISO 19115 metadata standard for geospatial metadata defines an extensive set of metadata elements. The Environmental Research Centre Research Data metadata profile uses only a subset of the full number of these elements. This subset is the minimum number of elements required to properly describe data resources originating from the environmental research community funded by the EPA. These core metadata elements are required to properly identify the dataset or data resource. The more comprehensive the metadata description is the better third party users can understand the data resource without ambiguity. The SAFER-Data system collects ISO 19115 compliant metadata. Authorised users (data providers) provide metadata through a series of web-based forms. This removes the burden of having to learn the complexities of the ISO 19115 metadata standard.

The following is a brief overview of the the key elements in the metadata subset:

- Who: Who created the dataset? Who owns the dataset? Who is responsible for the maintenance and future updates to the datasets?
- Where: Where can one obtain access to the dataset? Where is the geographical areas or regions represented within the dataset resource?
- When: When is the dataset available for access? What are the date and time ranges (timeseries) respresented within the dataset resource?
- What: What topic categories does the dataset resource exist in? What parameters are measured in the dataset? What instrumentation was used? What data quality procedures were applied to the dataset?
- Why: Why was the dataset resource created? Why was a particular geographical study region chosen?

Explaination of Metadata Terminology

Each metadata field will be described using the following headings:

- Name of Field: This is the name or label this metadata field is referred to on SAFER-Data and within accompanying documentation.
- **Description**: This is a description of the meaning or purpose of this metadata field.
- Required: Some metadata fields are not required while others are optional.
 - Mandatory: This field must be filled in by the data provider
 - Optional: This field is optional and is not required
 - Conditional: This field is mandatory under certain conditions
- Input Type: Metadata fields which allow "Free Text" input allow data providers to type descriptions into the table. Fields which are "Code Lists" allow data providers to choose attribute values from a list of allowable values. Fields which are "Automatic" are inputted automatically by SAFER-Data.
- Data Type: The information in a metadata field can be of the following types:

textual (mixture of text characters and numerical characters); numeric (numerical values only); or date (an expression of date and/or time)

- Default Values: In the case of some metadata fields a default value exists. This
 default value is used rather than leaving the field empty
- Maximum Occurances: This applies to "Code List" and selected "Free Text" fields. It outlines the maximum number of values the metadata field can be assigned. In the case of "Code List" a value of 1..N indicates that one may choose between 1 and N values from the list, where N is the total number of items in the list. In the case of a "Free Text" field N indicates that one my assign N text items to this metadata field.
- Allowable Values: This describes the valid range of values that the metadata field
 can be assigned. In the case of "Code Lists" this corresponds to the set of values in
 the list. In the case of "Free Text" fields it outlines any restrictions on the type of text
 that can be assigned to this metadata field.
- INSPIRE: This indicates if this metadata field is specified in the INSPIRE Directive Metadata Specifications. In the case of some fields the information is collected for EPA purposes only and is not a subset of ISO 19115.

This document purposely omits ISO 19115 technical jargon and notation. The ERC feel that this will make this document more accessible to a wider audience most of whom are not familiar with the complexities of ISO 19115.

Identification Information Metadata Fields

Resource Title:

Metadata Characteristic	Values and Description
Name of Field	Resource Title
Description	The title by which this resource is known by
Required	MANDATORY
Input Type	Free Text
Data Type	Text
Default Values	No Default Value
Maximum Occurances	1
INSPIRE	YES

Resource Keywords:

Metadata Characteristic	Values and Description
Name of Field	Resource Keywords
Description	A list of relevant keywords describing the content
	and nature of this data resource
Required	MANDATORY
Input Type	Free Text
Data Type	Text
Default Values	No Default Value
Maximum Occurances	1N
INSPIRE	YES

Resource Language:

Metadata Characteristic	Values and Description
Name of Field	Resource Language
Description	The natural language which is used to describe
	this data resource and used in any supporting
	documentation. In some cases several languages
	may be represented
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"ENG" - for English language
Maximum Occurances	1N: Number of Official European Languages
INSPIRE	YES

Resource Scope:

Metadata Characteristic	Values and Description
Name of Field	Resource Scope
Description	The class or type of information resource which
	the metadata is describing
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"dataset"
Maximum Occurances	1N
INSPIRE	YES

Metadata Progress:

Metadata Characteristic	Values and Description
Name of Field	Metadata Progress
Description	The current status of the data resources or
	datasets which this metadata is describing. You
	must specify the stage at which the project is
	currently at
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"ongoing"
Maximum Occurances	1
INSPIRE	YES

Project Code/Reference Number:

Metadata Characteristic	Values and Description
Name of Field	Project Code/Reference Number
Description	For EPA funded projects this is an identification
	code or reference number unique to the project
	which created the datasets or data resources
Required	OPTIONAL
Input Type	Free Text
Data Type	Text

Default Values	Supply the value "Not Applicable"
Maximum Occurances	1
INSPIRE	NO (EPA Internal Use Only)

Internal Thematic Area:

Aetadata Characteristic	Values and Description
Name of Field	Internal Thematic Area
Description	For EPA funded projects this is the thematic area under which the project was initially funded.
Required	OPTIONAL
Input Type	CODE_LIST
Data Type	Text
Default Values	"Not Applicable"
Maximum Occurances	1
INSPIRE	NO (EPA Internal Use Only)

ISO 19115 Topic Category:

Metadata Characteristic	Values and Description
Name of Field	ISO 19115 Topic Category
Description	This is a high-level geographic data thematic
	classification to assist in the grouping and
	searching of available geospatial datasets. It is
	understood that there are overlaps between
	general categories and users are encouraged to
	select the most appropriate
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	Must select a default value (See below for an
	explaination of the values in CODE_LIST)
Maximum Occurances	1N
INSPIRE	YES

Frequency of Changes To Metadata Record:

Metadata Characteristic	Values and Description
Name of Field	Frequency of Changes To Metadata Record
Description	This describes how often the owner of this
	metadata resource will update the information in
	the metadata fields.
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"continual"
Maximum Occurances	1

INSPIRE YES

Temporal Information Metadata Fields

Commencement of Research Work Date:

Metadata Characteristic	Values and Description
Name of Field	Project Start Date
Description	This is the approximate date (accuracy to one month) when the project commenced. More specifically when work on creating the dataset resources described by this metadata began
Required	MANDATORY
Input Type	Date Format
Data Type	DATE
Default Values	None – a date value must be supplied
Maximum Occurances	1
INSPIRE	YES

Earliest Recorded Date/Time value:

Metadata Characteristic	Values and Description
Name of Field	Resource Start Date
Description	This is the earliest recorded date, date time, or
	time within the dataset resources. If the dataset is
	not timeseries based the date of first
	work/analysis on the dataset(s) is provided
Required	MANDATORY
Input Type	Date Format
Data Type	DATE
Default Values	None – a date value must be supplied
Maximum Occurances	1
INSPIRE	YES

Most Recent Recorded Date/Time value

Metadata Characteristic	Values and Description
Name of Field	Resource End Date
Description	This is the most recently recorded date, date time, or time within the dataset resources. If the dataset
	is not timeseries based the date of most recent work/analysis on the dataset(s) is provided
Required	MANDATORY
Input Type	Date Format
Data Type	DATE
Default Values	None – a date value must be supplied
Maximum Occurances	1
INSPIRE	YES

Geographical Information Metadata Fields

Geographical information is a requirement of all projects providing metadata to the ERC. This information is also very important for building mapping and data services for environmental research datasets.

A minimum bounding rectangle (bounding box) must be specified for all projects. The default bounding box for EPA funded research projects is the island of Ireland.

Bounding box coordinates (See below) must be specified in Decimal Degrees. This is equivalent to the INSPIRE Specification for geographical bounding boxes.

If your project studies many geographical locations you can specify bounding boxes for each location. Using SAFER-Data you can specify as many bounding boxes as required.

Geographical Description:

Metadata Characteristic	Values and Description
Name of Field	Geographical Description
Description	This field allows the data owner to describe in
_	free text the geographical area(s) studied or
	analysed in the project. Data owners are
	encouraged to be as verbose as necessary in this
	field. One can include details of placenames,
	local geographical features, travel directions,
	environmental conditions, etc.
Required	OPTIONAL
Input Type	Free Text
Data Type	Text
Default Values	No Default Value
Maximum Occurances	N/A
INSPIRE	YES

West Bounding Coordinate:

Metadata Characteristic	Values and Description
Name of Field	West Bound Longitude
Description	The Western-most coordinate of the limit of the
	dataset geographical extent. This is expressed in
	longitude in decimal degrees.
Required	MANDATORY
Input Type	Free Text (numerical)
Data Type	NUMERIC
Default Values	-10.7857421875
Maximum Occurances	1
INSPIRE	YES

East Bounding Coordinate:

Metadata Characteristic	Values and Description
Name of Field	East Bound Longitude
Description	The Eastern-most coordinate of the limit of the
	dataset geographical extent. This is expressed in

	longitude in decimal degrees.	
Required	MANDATORY	
Input Type	Free Text (numerical)	
Data Type	NUMERIC	
Default Values	-5.55908203125	
Maximum Occurances	1	
INSPIRE	YES	

South Bounding Coordinate:

Metadata Characteristic	Values and Description
Name of Field	South Bound Latitude
Description	The Southern-most coordinate of the limit of the
	dataset geographical extent. This is expressed in
	lattitude in decimal degrees.
Required	MANDATORY
Input Type	Free Text (numerical)
Data Type	NUMERIC
Default Values	51.3580615731909
Maximum Occurances	1
INSPIRE	YES

North Bounding Coordinate:

Metadata Characteristic	Values and Description
Name of Field	North Bound Lattitude
Description	The Northern-most coordinate of the limit of the
	dataset geographical extent. This is expressed in
	lattitude in decimal degrees.
Required	MANDATORY
Input Type	Free Text (numerical)
Data Type	NUMERIC
Default Values	55.3291444084507
Maximum Occurances	1
INSPIRE	YES

Access Restrictions and Constraints Metadata Fields

Resource Access Limitations:

Metadata Characteristic	Values and Description
Name of Field	0
Description	This is a specification of any limitations placed upon the access (and subsequent usage) of the
	dataset resources associated with this metadata
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"noRestrictions"
Maximum Occurances	1N
INSPIRE	YES

Resource Usage Limitations:

Metadata Characteristic	Values and Description
Name of Field	Usage Limitations
Description	This field allows the data owner to specify in exact detail the nature of any limitation imposed upon third parties who access and use the datasets provided. This should include any special access requirements, licenses, costs, etc that should be considered before third party access and usage. One can also specify applications for which this data is not suitable. It is recommended that data owners provide details to this optional field
Required	OPTIONAL
Input Type	Free Text
Data Type	Text
Default Values	None
Maximum Occurances	N/A
INSPIRE	NO

Resource Availability From SAFER-Data:

Metadata Characteristic	Values and Description
Name of Field	Public Access Type
Description	This describes the type of access third parties will
	have to the metadata and associated datasets. This
	includes the rights to download the datasets from
	SAFER-Data. This field is required by SAFER-
	Data.
Required	MANDATORY
Input Type	CODE_LIST (Values Explained Below)
Data Type	Text
Default Values	"Fully Publicly Available"
Maximum Occurances	1
INSPIRE	NO

Using this field data owners can express the level of public availability that their datasets are given by SAFER-Data. This is a very important consideration. There are three availability levels and are organised as follows:

- Public Availability (PUBL): This is the most open level of availability. This means
 that metadata and the dataset files corresponding to the metadata are available to
 any user. Users agree to the terms of a data download disclaimer (see below)
 before downloading any of these data files to their local computer. This is the
 availability level recommended by ERC.
- 2. Semi Public Availability (SEMI): This is a more restricted level of availability. This means that the metadata is available to any user to browse, view, or print. Any data files corresponding to the metadata remain hidden. The data files can only be accessed by the data owner and the System Administrators. Only the data owner can choose to change the level of availability from SEMI to PUBL. Users are encouraged to use SEMI if their data is not yet ready to be made publicly available. Such resources are frequently monitored and will be automatically changed to PUBL 12 months after after publication date of the project report.
- 3. Private (PRIV): This is the most restricted level of availability. Only the dataset owner and the System Administrators can view or access the metadata or datasets. This means that neither the metadata nor any associated datasets are available to any users. This level of availability is not recommended. It essentially hides metadata and dataset resources from any type of access. It should only be used in circumstances where users are creating metadata for the first time or where users must wait for a 3rd party to check their metadata before making it publicly available. Resources marked as PRIV cannot remain so indefinitely. Such resources are frequently monitored and will be automatically changed to PUBL 12 months after publication date of the project report.

Additional Information Metadata Fields

Resource Lineage:

Metadata Characteristic	Values and Description
Name of Field	Resource Lineage
Description	Essentially this field specifies "why" this data resource was created. What were the conditions that caused this project to be funded? What environmental events, gaps in scientific knowledge, etc are addressed by the existance of this data resource. This field should contain a general explaination of the data owners knowledge of the lineage of the data resource
Required	OPTIONAL
Input Type	Free Text
Data Type	Text
Default Values	None
Maximum Occurances	1
INSPIRE	YES

Links To Internet Resources:

Metadata Characteristic	Values and Description
Name of Field	Internet-based Links
Description	This field allows data owners to supply a set of
	hyperlinks to related internet services or websites.
	These services or websites may offer additional
	information on the data resources.
Required	OPTIONAL
Input Type	Free Text (must supply full URL)
Data Type	Text
Default Values	NON
Maximum Occurances	N
INSPIRE	YES

Resource Abstract

Metadata Characteristic	Values and Description
Name of Field	Resource Abstract
Description	This field closely resembles the abstract field in a
	journal paper or report. Data owners are
	encouraged to be as verbose as necessary. Ideally
	the abstract should provide an overview of the
	data resources and the scientific work that created
	them.
Required	OPTIONAL
Input Type	Free Text

Data Type	Text	
Default Values	None	
Maximum Occurances	N/A	
INSPIRE	YES	

Supplementary Information:

Metadata Characteristic	Values and Description
Name of Field	Supplementary Information
Description	This field allows data owners to include any
	information they feel is necessary for third parties
	to fully evaluate the data resource. This field
	offers the opportunity to provide information that
	cannot be provided using any of the other
	metadata fields. Examples of information types
	are as follows: bibliography references,
	specification of software requirements,
	acknowledgements of funding, library
	information, hardware or instrumentation details,
	acknowledgements of collaborations, etc.
Required	OPTIONAL
Input Type	Free Text
Data Type	Text
Default Values	None
Maximum Occurances	N/A
INSPIRE	NO

Dataset Supply and Presentation Modes Metadata Fields

The metadata fields below provide information on the types of dataset files contained in this resource (*digital file types*), how the dataset resource is actually delivered to third parties (*resource supply mode*), and the types of information media on which the resource is supplied (*resource supply type*).

Digital File Types:

Metadata Characteristic	Values and Description
Name of Field	Digital File Type
Description	This field allows data owners to specify the types
	of software file formats the dataset resources are
	available in. All major scientific software file
	types are represented.
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"No Files Attached" (if no datasets are available)
Maximum Occurances	1N
INSPIRE	YES

If a software file type is not presented in the CODE_List by SAFER-Data data owners are encouraged to give details of these software file types in "supplementary information" field.

Resource Supply Mode:

Metadata Characteristic	Values and Description
Name of Field	Resource Supply Mode
Description	This field represents the types of physical media or physical media infrastructure used to allow third parties to access your dataset resources. Datasets available using SAFER-Data are "online" supply mode. If data owners will copy datasets to media such as CD/DVD these supply modes should be specified.
Required	MANDATORY
Input Type	CODE_LIST
Data Type	Text
Default Values	"online"
Maximum Occurances	1N
INSPIRE	YES

Resource Supply Types:

Metadata Characteristic	Values and Description
Name of Field	Resource Supply Types
Description	This field specifies if soft copy or hard copy
	versions of the dataset resources are available.
	Electronic datasets will principally be a
	combination of "tableDigital", "imageDigital",

"mapDigital", "documentDigital". If hardcopy

versions of these are only available the data

owner should specify these

Required MANDATORY **Input Type** CODE_LIST

Data Type Text

Default Values "documentDigital" - as all resources on SAFER-

Data will correspond to a completed report

Maximum Occurances1..NINSPIREYES

Responsible Parties Contact Information

Responsible Party information includes the name and position for individuals or organisations that are responsible for the resouce. This responsibility covers the metadata and the dataset files. It is a separte table of information to the dataset metadata.

When a metadata resource record is created on SAFER-Data the contact information corresponding to the person logged on as the data owner is used automatically as a Responsible Party Contact. After this other Responsible Party Contacts can be added automatically.

All fields below are specified in the INSPIRE metadata requirements. All of the fields are FREE TEXT with the exception of the Country and Role fields which are CODE_LIST. It is important that the role of each additional Responsible Party is outlined (see explaination of roles below). A Responsible Party can have many roles.

ResponsibleParty Field Name	Description and Required Values	
Individual Name	The full [Title] [First Name] [Middle Names if Any] [Surname	
	this party. Please ensure correct punctuation of name information.	
Position	A statement of the current employment position held by the	
	individual named. For example: Senior Lecturer, or Technician.	
Primary Organisation	The principal organisation to which this individual is affliated. The	
	full un-abbreviated name of the organisation should be stated.	
Telephone	The telephone number of the individual named. This should	
	include relevant country dialing codes and instructions regarding	
	telephone extension numbers. This number should be located	
	within the principal organisation and must not be a private	
	telephone such as a cellular phone for private use.	
Fax Number	The facsimile number of the individual named. This should	
	include relevant country dialing codes and instructions. If the	
	individual does not have a Fax machine it would be useful to	
	include the number of a machine to which they have access to	
	send and recieve fax messages. This number should be located	
	within the principal organisation and must not be a private	
	telephone such as a cellular phone for private use.	
Email	The electronic mail address of the individual named. This email	
	should belong to the the principal organisation or other	
	organisation to which the individual is affiliated. It should not be a	
	private email address.	
Internet Address	An Internet address containing information on how the individual	
	may be contacted. This can be a staff or personell page on the	
	principal organisation's website. Alternatively if the individual has	
	a public website or has contact information on the website of	
	another project this can be included. The Internet address should	
	not direct users to social networking websites such as Bebo,	
	Facebook, MySpace, etc. Such links will be removed.	
Office, Building, Street Address	This is the physical office location of the individual. The office	
	number, building name or number, and street address should be	
	provided. This address should not reveal the private home address	

ResponsibleParty Field Name	Description and Required Values	
	of the individual.	
Postal Code	The postal code for the address of the individual.	
Administrative Code	The administrative region or entity within which the address of the individual is located.	
City	The city for the address of the individual	
Country	The country for the address of the individual. You should select ONE country from the CODE_LIST provided.	
Role	The function(s) performed by the responsible party in respect to project that generated the dataset resources documented in the metadata. (See table below)	

Responsible Party ROLE CODES:

The role names are specified with separate words appended and following lowercase Uppercase styling. This is inline with INSPIRE metadata guidelines.

RoleName	Role Code	Role Description
resourceProvider	001	The party that supplies the resource. Default: The EPA
Custodian	002	Party that accepts accountability and responsibility for the data and ensures appropriate care and maintenance of the resouce
Owner	003	The party that owns the resouce
User	004	The party that users the resouce
Distributor	005	The party that distributes the resource to other third parties users. In the case of SAFER-Data this is the EPA.
Originator	006	This party is the original creator of the dataset resources corresponding to this resource. They had a principal function in creating the initial data capture, data storage, or data generation.
PointOfContact	007	The party who can be contacted for acquiring knowledge about or aquisition of the dataset resources corresponding to this resource. This party should understand that third parties will be provided with their contact details (as supplied in the above format).
PrincipalInvestigator	008	The Principal Investigator (PI) is the lead scientist on this project. The PI is the person who takes direct responsibility for completion of a funded project, directing the research and reporting directly to the funding agency.
Processor	009	A party who has processed the data in a manner such that the dataset resources were modified. This processing can include data capture, data reformatting, reprojection, QA/QC, etc.
Publisher	010	A party who publishes the resource. This is not relevant for most EPA funded projects.

RoleName	Role Code	Role Description
Author	011	A party who authored the resource. This is not
		relevant for most EPA funded projects.