Requests for changes to metadata standards

Rachel Heaven, British Geological Survey on behalf of the Natural Environment Research Council (NERC), SIS Discovery Metadata Project.
17/02/2011

Contents

Glossary:

ISO 19115 / 19119 and 19139: International standards for geographic metadata and the xml schema specification for its serialisation and encoding.
INSPIRE: Geographic metadata for digital datasets required by EC directive; based ib core ISO metadata elements and it is implied that it should be encoded as ISO 19139 xml (though not specified).
GEMINI2: UK specific implementation of the INSPIRE standard, encoded in ISO 19139 xml. The GEMINI2 standard is stricter than INSPIRE; all GEMINI2 compliant metadata will also be INSPIRE compliant but not necessarily vice versa.
MEDIN: Marine Environment Data Information Network's profile of the ISO standard. It is stricter than GEMINI2 because of the use of specialised codelists and vocabularies. All MEDIN compliant metadata will also be GEMINI2 and INSPIRE compliant but not necessarily vice versa.
NERC Discovery Metadata standard: based upon the MEDIN standard, and is defined by specifying a small number of exceptions or "relaxations" where constraints specific to the marine community are not relevant in the wider context of NERC. All

NERC compliant metadata will also be GEMINI2 and INSPIRE compliant but not necessarily vice versa.
NERC DDS: the NERC Data Discovery Service, ingests metadata and provides downloads of metadata both in NERC standard format

1. Unique Resource Identifier

Summary:

The INSPIRE metadata standard has matched this to the ISO element CI_Citation.identifier.

Our opinion is that this attribute needs to be matched to ISO element MD_Metadata.datasetURI instead of (or in addition to) CI_Citation.identifier

Reasoning:

- URI values are one of the most important piece of information to identify the resource, and certainly are for resources that are exposed as RDF resources for use in the semantic web as linked data (http://data.gov.uk/linked-data).

- Multiple occurrences of CI_Citation.identifier are allowed in the ISO standard, each with a code and codespace or authority, so out of all the various identifiers that may be found there is no mechanism for an application to recognise which is the URI.

- The ISO element MD_Metadata.datasetURI is optional, has maximum occurrence of 1, and a description "Uniformed Resource Identifier (URI) of the dataset to which the metadata applies"). This seems the obvious place for the URI information to be encoded.

- ISO19139 xml metadata files should be able to be consumed by anyone around the world, not just those in the INSPIRE community. Applications from outside the INSPIRE community will probably look for URI in MD_Metadata.datasetURI element.

Severity: SEVERE : important piece of information is encoded in the wrong xml element and therefore invisible to applications

Changes required to: INSPIRE , GEMINI2 , MEDIN , NERC

Immediate solution (NERC):

1. Update the NERC Metadata Standard to mandate encoding of URI to MD_Metadata.datasetURI in addition to CI_Citation.identifier (anticipating any change in GEMINI2 and INSPIRE and MEDIN)
2. Update the NERC DDS to include the MD_Metadata.datasetURI element in the xml download

Longer term solution:

1. Lobby the INSPIRE metadata working group to change their regulations
2. Lobby the GEMINI metadata working group to update their guidance to mandate encoding of URI to MD_Metadata.datasetURI in addition to CI_Citation.identifier (anticipating any change in INSPIRE)
3. Notify the MEDIN Metadata Standards group – they may also want to anticipate the changes in INSPIRE and GEMINI

2. Spatial Resolution

Summary:

INSPIRE metadata regulations state that "a resolution distance shall be expressed as a numerical value associated with a unit of length"
GEMINI2 metadata standard states this is "measure of the granularity of the data (in metres)"
Many datasets cannot have their resolution described in metres or in length, but can be described in other units or properties. We would like the standards to be changed to allow this.

Reasoning:

For global datasets (e.g. marine and atmospheric) the resolution could well be in angular units for which the corresponding ground distance varies greatly. Additionally GEMINI2 v2.1 part 2 guidance notes rules for how to fill offer 'For image data, it is the resolution of the image' – this is not possible in metres. It would be useful to describe resolution in units of other properties also e.g. time, pressure.

Severity: MEDIUM : attribute is optional and will be left empty even though suitable information for ISO 19139 is available

Change level required: INSPIRE, GEMINI

Immediate solution (NERC):

Supply a null value if length units in metres cannot be provided.

Longer term solution:

Lobby GEMINI and INSPIRE to make this change.

3. Geographic Bounding Box

Summary:

It is not possible to describe a bounding box that crosses the 180/-180 degree meridian in GEMINI compliant metadata: only one box is allowed and numerical checks ensure that the eastern coordinate cannot have a higher value than the western. Some datasets will have this location and this

Severity: HIGH : mandatory element that cannot be provided for some datasets

Change level required: GEMINI

Immediate solution (NERC):

Split the dataset into two metadata records, one either side of the 180/-180 degree meridian, then a single bounding box can be provided for each.

Longer term solution:

Lobby GEMINI to allow multiple bounding boxes (as in INSPIRE and ISO 19115).