

## Infrastructure - Decision #8616

### Consider expanding isotc211's indexing component's keyword XPath to cover topicCategories

2018-06-15 00:13 - Bryce Mecum

<b>Status:</b>	New	<b>Start date:</b>	2018-06-15
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>		<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Milestone:</b>	None		

#### Description

From

[https://repository.dataone.org/software/cicore/trunk/cn/d1\\_cn\\_index\\_processor/src/main/resources/application-context-isotc211-base.xml](https://repository.dataone.org/software/cicore/trunk/cn/d1_cn_index_processor/src/main/resources/application-context-isotc211-base.xml)

The current XPath for the keyword field pulls out:

```
//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyword/gmx:Anchor/text() |
//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyword/gco:CharacterString/text()
```

ISO also defines MD\_DataIdentification/gmd:topicCategory which is defined as "The main theme(s) of the dataset." and is required (recommended) when describing a dataset. It's conditional, and repeatable. An example from a PANGAEA doc is

```
...
    <ns0:topicCategory>
      <ns0:MD_TopicCategoryCode>geoscientificInformation</ns0:MD_TopicCategoryCode>
    </ns0:topicCategory>
  </ns0:MD_DataIdentification>
```

I think it's improve recall to include in our keywords list. It appears to be a controlled vocabulary so we could even make more direct use of it. The controlled vocabulary appears to be (From the MI\_Metadata workbook):

Domain:

- farming
- biota
- boundaries
- climatologyMeteorolgyAtmosphere
- economy
- elevation
- environement
- geoscientificInformation
- health
- imageryBaseMapsEarchCover
- intelligenceMilitary
- inlandWaters
- location
- oceans
- planningCadastre
- society
- structure
- transportation
- utilitiesCommunicationgeoscientificInformation, health, imageryBaseMapsEarchCover, intelligenceMilitary, inlandWaters, location,

oceans, planning Cadastre, society, structure, transportation, utilities Communication

Both NCEI and PANGAEA make use of this field in their ISO docs.