



Métadonnées de Collection
et Jeux de Données

V0.7

Document de travail

Le 27 novembre 2019

epointal	Métadonnées Documents de travail Version 0.7	ForM@Ter 
----------	--	--

Objet

Ce document précise les métadonnées et paramètres retenues par ForM@Ter pour la création de son MétaCatalogue.

V0.1	Premières spécifications	epointal	Septembre 2019
V0.2	Précisions et premiers code listes	epointal	Novembre 2019
V0.7	Les contacts...	epointal	27 novembre 2019

Introduction

Pour la création du catalogue de métadonnées, nous avons les objectifs suivants:

- pouvoir exporter dans différents formats les métadonnées.

- ISO19139 valide INSPIREⁱ
- DataCiteⁱⁱ
- Format Pivot DataTerra
- UMM NASAⁱⁱⁱ
- compatible avec OGC:O&M

Pour répondre à ces impératifs, nous utilisons l'implémentation ISO19139 à laquelle nous ajoutons 3 thésaurus (variables, plateformes et instruments).

Définitions

Collection (ou series)

Une collection est un regroupement de jeux de données réalisés dans les mêmes conditions:

- même plateforme,
- même variable
- et si possible même instrument

Jeu de données ou granule (ou dataset)

En simplifiant, il s'agit du fichier de données.

Un jeu de données ne peut appartenir qu'à une unique collection.

Les champs essentiels

Titre	Type	Card	Description
hierarchyLevel	series ou dataset	1	Type de fiche
dateStamp	Date iso8601	1	Date de maj des métadonnées (automatique)
metadataContact	#Contact	1..n	Contact pour les métadonnées
dataIdentification			
identifier	URI, URN, DOI, ...	1..n	Identifiant des données
title	Texte	1	Titre du jeux de données ou de la collection
abstract	Texte ou html	1	Description du jeu de données ou de la collection
graphicOverview	url + titre + description	1..n	capture d'écran, image d'illustration
dataContact	#Contact	1..n	Contact pour les données (organisation + email)
date	Date iso8601	1..3	Date de création ou publication ou de maj des données
Variable/product	#descriptiveKeyword	1..n	Mot clés d'un thésaurus à réaliser
platform	#descriptiveKeyword	1	Thésaurus à réaliser
spatialExtent	Bounding Box (WGS84)	1..n	Etendue(s) spatiale(s).
temporalExtent		1..n	Etendue(s) temporelle(s)
resourceConstraints	#resourceConstraints	1..n	Contraintes d'accès et d'utilisation, license... et les restrictions d'accès
onlineResource	#onLineResource	0..n	lien divers (téléchargement, visualisation, répertoire ftp, page d'information, service WMS, service WFS)
dataQuality			
lineage	texte	0..1	description de la généalogie des données

Liste détaillée des champs

M pour Mandatory
 C pour Conditional
 R pour Recommended
 O pour Optional
 N/A pour Not applicable

La liste suivante récapitule les champs d'une fiche de métadonnées, ainsi que son exigibilité pour ForM@Ter, la norme ISO19139+INSPIRE et DataCite.

Les dénominations sont le plus possible reprise des dénominations ISO19139.

Les champs plus complexes sont précisés ensuite.

Il s'agit des métadonnées minimales à fournir, elles peuvent être étendues.

Dénomination	Type	Défaut	ForM@Ter	Card	ISO19139+INSPIRE	DATA CITE	Description
Metadata							
fileIdentifier	URI, URN ...		M	1	M		Identifiant unique et permanent de la fiche de métadonnée
parentIdentifier	identifiant de fiche de métadonnée parent		C	0..1	C	N/A	l'identifiant de la fiche parent si elle existe. Un seul parent possible.
metadataLanguageCode	fre et/ou eng	eng	M	1..2	M		Langue de la fiche de métadonnées (peut être bilingue fre et eng)
dateStamp	Date ou DateTime ISO8601		M	1	M		Date de mise à jour de la fiche de métadonnées
contact	#Contact		M	1..n	M		Contact avec au moins un mail pour les métadonnées
metadataCharacterSet	MD CharacterSet Code	utf-8	M	1	M		Encodage de la fiche de métadonnées
hierarchyLevel	series dataset	dataset	M	1	M	M	Niveau de la fiche
referenceSystemInfo	Liste à établir Specification INSPIRE	WGS 84 (EPSG:4326) ou ITRS69 (EPSG:4258) ?	M	1..n	M		Le(s) système(s) de coordonnées de référence utilisé(s) pour les données
temporalCRS	Liste à établir		C	0..n	C	N/A	Référentiel temporel des données si on utilise un référentiel particulier autre que UTC (datetime) ou calendrier grégorien (date)
verticalCRS	Liste à établir		C	0..n	C	N/A	Si a du sens pour les données

identificationInfo - Identification de la collection ou du jeu de données							
identifier	URI, URN, DOI, ...		M	1..n	M	M	Identifiant unique pour la collection de jeux de données. Peut avoir plusieurs identifiants La réf dans la DB + un DOI
status	MD_ProgressCode	completed	O	0..1			
title	Texte		M	1	M	M	Titre
abstract	Texte		M	1	M	M	Description de la donnée
date	maj	Date ISO8601	M	0..1	M au moins 1 des 3		Date de maj du jeu de données ou de création ou de la collection
	création			0..1			
	publication			0..1		M	
pointOfContact	#Contact		M	1..n	M	M	Contact pour les données avec son rôle
descriptiveKeyword	#Keyword		R	0..n	R	R	Mots clés divers libres ou provenant de thésaurus
GEMET Inspire Keyword	GEMET INSPIRE		M	1	M	O	Catégorie INSPIRE GEMET1.0
topicCategory	TopicCategories	geoscientific information	M	1..n	M	O	Catégorie très générale
spatialRepresentationInfo	MD_SpatialRepresentationTypeCode	raster	M	1..n			Type de représentation spatiale
variables productType	Liste à établir Thésaurus ForM@Ter		M	1..n	O		grandeur mesurée ou calculée
platform	Liste à établir Thésaurus ForM@Ter		R	1..n	O	N/A	nom de la station, du puit, du satellite....
instrument	Liste à établir Thésaurus ForM@Ter		O	0..n	O	N/A	nom de l'instrument
dataCenter	Liste à établir Thésaurus ForM@Ter		M	1	O	N/A	nom du centre de données
spatialResolution	Nombre (voir #spatialResolution)		O	0..n		N/A	Résolution spatiale
timeResolution	liste à établir Thésaurus GCMD		C	0..n		N/A	pour les series temporelles uniquement Mot clé
dataLanguageCode	eng ou fre	eng	O		M	O	Langue des données
dataCharacterSet	MD_CharacterSet	utf-8	C	0..1	C		Encodage des données (si

		Code						différent de utf-8)
spatialExtent		bbox (WGS84)		M	1..n	M	M une seule	La bbox est exprimée dans un référentiel WGS84 (ESPG:4326)
temporal Extent	begin	Date ISO8601		M			N/A	
	end			O (rien ou now si on going)				
dataMaintenance Frequency		MD Maintenance FrequencyCode	asNeeded	O	0..1	O		Fréquence de la maintenance des données
resourceConstraints								
limitationOnPublic Access		#limitationOnPublicAccess	No limitations on public access	M	1			Limitation d'accès public juridique
accessAndUseConditions		license, citation, tarif, authentication #ConditionsApplyingToAccessAndUse		M	1..n			Conditions applicables à l'accès et à l'utilisation
distributionInfo								
format				M	1..n	M		Format des données
onlineResource		#onlineResource		R	0..n			Lien de téléchargement ou d'accès
distributor		#contact		R	1	O	M	responsable de la diffusion des données (publisher pour datacite dans geonetwork)
qualityInfo								
conformity				M	0..1		N/A	Conformité des données ou des métadonnées? aux normes ?
lineage	source				0..n			Description de la généalogie des données
	statement	texte		R				

Détails des champs complexes

Les identifiants

La résolution spatiale

2 possibilités pour la résolution spatiale: une distance en mètre ou un dénominateur.

Exemples ISO19139

Xpath:

/gmd:MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:spatialResolution

Les mots clés

Les mots clés sont triés par [MD_KeywordTypeCode](#).

Cette liste a été étendue, un grand nombre de ces valeurs n'apparaît pas dans la liste des codes de 2005: <https://www.isotc211.org/2005/resources/Codelist/gmxCodellists.xml>

Il faut donc utiliser la liste de la NOAA

https://data.noaa.gov/resources/iso19139/schema/resources/Codelist/gmxCodellists.xml#MD_KeywordTypeCode

MD_KeywordTypeCode	
discipline	
place	
stratum	
temporal	
theme	
dataCentre	« nouveau »
featureType	« nouveau »
instrument	« nouveau »
platform	« nouveau »
process	« nouveau »
project	« nouveau »
service	« nouveau »
product	« nouveau »
subTopicCategory	« nouveau »
taxon	« nouveau »

Mot clé libre

Un mot clé libre est défini par le mot clé + le type de mot clé ([MD KeywordTypeCode](#))

Mot clé de thésaurus

Les thésaurus sont des vocabulaires contrôlés, c'est à dire une liste de termes si possible avec leur définition. Ils peuvent être hiérarchiques ou non.

Ce vocabulaire sera disponible sous forme de fichier au format [skos-rdf](#) .

Les thésaurus ForM@Ter, en cours d'élaboration, sont:

- dataCenter
- variable/product
- plateforme
- instrument

ainsi que:

- timeResolution

De nombreux autres thésaurus peuvent être utilisés:

- Geonetwork par exemple, utilise le thésaurus [natureearth-and-seavox](#) comme thésaurus de localisation.
- La NASA a créé plusieurs thésaurus: [GCMD keywords](#) , aussi bien de localisation, que de mots scientifiques ou d'instruments
- Nous pouvons aussi citer les «Climat and Forecast Standard Names » dont les différentes versions se trouvent sur la page <http://cfconventions.org/standard-names.html>

		description	exemple
keyword	keyword	dénomination du concept	age_of_sea_ice
	keywordURI	uri du concept	http://mmisw.org/ont/cf/parameter/age_of_sea_ice
thesaurus	thesaurusName	dénomination du thésaurus (y compris version)	Climate and Forecast (CF) Standard Names (v.69)
	thesaurusURI	uri du thésaurus	http://mmisw.org/ont/cf/parameter
	thesaurusDateRevision	date de mise à jour du thésaurus	(si existe: généralement change de version à chaque révision)
	thesaurusDatePublication	date de publication du thésaurus	2019-10-17

Cas particulier du GEMET1.0

La directive INSPIRE demande d'ajouter un mot clé [GEMET1.0](#)

Exemples ISO19139

Xpath:

/gmd:MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords

Mots clés libres

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Spain</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>Andalusia</gco:CharacterString>
    </gmd:keyword>
    <gmd:type>
      <gmd:MD_KeywordTypeCode codeListValue="place"
codeList="http://standards.iso.org/iso/19139/resources/gmxCodelists.xml#MD_KeywordTypeCode">Place</
gmd:MD_KeywordTypeCode>
    </gmd:type>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

Ici, les deux mots clés sont du même type, ils sont alors dans la même balise

gmd:descriptiveKeywords

Mot clé d'un thésaurus

```
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gmx:Anchor xlink:href="http://inspire.ec.europa.eu/theme/lu">Land use</gmx:Anchor>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gmx:Anchor xlink:href="http://www.eionet.europa.eu/gemet/inspire_themes">GEMET -
INSPIRE themes, version 1.0</gmx:Anchor>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2008-06-01</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode
codeList="http://standards.iso.org/iso/19139/resources/gmxCodelists.xml#CI_DateTypeCode"
codeListValue="publication">Publication</gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

Les contacts

Désignation	Type	ForM@Ter + ISO19915 + INSPIRE	DataCite	Remarque
organisationName	string	M		
individualName	string	O		
orcid	string	N/A	O	
positionName	string	O		
contactInfo				
address				
deliveryPoint	string	O		
city	string	O		
administrativeArea	string	O		
postalCode	string	O		
country	string	O		
electronicMailAddress	email	M		
phone		O		
onlineRessource	#onlineRessource	O		
role	#CI_RoleCode	M		

Les rôles

La liste des rôles ISO19115 (2003) a été étendue dans la norme ISO19115-1 (2007)

ISO19115: https://www.isotc211.org/2005/resources/Codelist/gmxCodellists.xml#CI_RoleCode

ISO19115-1:

https://data.noaa.gov/resources/iso19139/schema/resources/Codelist/gmxCodellists.xml#CI_RoleCode

identifiant	description	norme	Rôle DataCite
resourceProvider	party that supplies the resource	ISO19115	
custodian	party that accepts accountability and responsibility for the data and ensures appropriate care and maintenance of the resource	ISO19115	
owner	party that owns the resource	ISO19115	
user	party who uses the resource	ISO19115	
distributor	party who distributes the resource	ISO19115	Publisher
originator	party who created the resource	ISO19115	Creator

pointOfContact	party who can be contacted for acquiring knowledge about or acquisition of the resource	ISO19115	
principalInvestigator	key party responsible for gathering information and conducting research	ISO19115	Creator
processor	party wha has processed the data in a manner such that the resource has been modified	ISO19115	
publisher	party who published the resource	ISO19115	Publisher
author	party who authored the resource	ISO19115	Creator
coAuthor	the individual(s) or organization(s) who name(s) should appear after the first name in a citation for the resource (use author to denote the first name in the citation). while it is possible to have a co-author and principal investigator/collaborator be the same individual or organization, co-author is no the same as nor synonymous with principle investigator or collaborator	ISO19115-1	Creator
collaborator	party who assists with the generation of the resource other than the principal investigator	ISO19115-1	Creator
contributor	the individuals or organizations whose contributions deserve recognition in the citation. contributor is mutually exclusive from author, co-author, principal investigator, and collaborator. use ISO MD_Identification credit field to identify individual or organizations that should be given acknowledgement only.	ISO19115-1	
editor	the individual who has made a corrective or editorial change to the resource as part of a systematic revision process.	ISO19115-1	
funder	the individual or organization which has provided all or part of the finances associated with the resource.	ISO19115-1	
mediator	a class of entity that mediates access to the resource and for whom the resource is intended or useful	ISO19115-1	
rightsHolder	the individual or organization who has ownership of the legal right to the resource.	ISO19115-1	
stakeHolder	an individual or organization who has an interest in the resource and/or is affected by or affects the actions of the resource	ISO19115-1	

Exemple ISO19139

Xpath:

- Contact pour les métadonnées : `gmd:MD_Metadata/gmd:contact/CI_ResponsibleParty`
- Contact pour les données:
`gmd:MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:PointOfContact/CI_ResponsibleParty`
- Contact pour la maintenance des métadonnées
- Contact pour la maintenance des données
- Contact pour la distribution
- ...

Exemple simple

```
<gmd:CI_ResponsibleParty>
  <gmd:organisationName>
    <gco:CharacterString>My Organisation</gco:CharacterString>
  </gmd:organisationName>
  <gmd:positionName>
    <gco:CharacterString>Searcher</gco:CharacterString>
  </gmd:positionName>
  <gmd:contactInfo>
    <gmd:CI_Contact>
      <gmd:address>
        <gmd:CI_Address>
          <gmd:electronicMailAddress>
            <gco:CharacterString>catalogue@myorganisation.fr</gco:CharacterString>
          </gmd:electronicMailAddress>
        </gmd:CI_Address>
      </gmd:address>
    </gmd:CI_Contact>
  </gmd:contactInfo>
  <gmd:role>
    <gmd:CI_RoleCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_RoleCode"
      codeListValue="owner"/>
    </gmd:role>
  </gmd:CI_ResponsibleParty>
```

Exemple complet

```

<gmd:CI_ResponsibleParty>
  <gmd:individualName>
    <gco:CharacterString>John Doe</gco:CharacterString>
  </gmd:individualName>
  <gmd:organisationName>
    <gco:CharacterString>My Organisation</gco:CharacterString>
  </gmd:organisationName>
  <gmd:positionName>
    <gco:CharacterString>Searcher</gco:CharacterString>
  </gmd:positionName>
  <gmd:contactInfo>
    <gmd:CI_Contact>
      <gmd:phone>
        <gmd:CI_Telephone>
          <gmd:voice>
            <gco:CharacterString>+33 (0)9 88 83 00 07</gco:CharacterString>
          </gmd:voice>
          <gmd:facsimile>
            <gco:CharacterString>+33 (0)9 88 83 00 10</gco:CharacterString>
          </gmd:facsimile>
        </gmd:CI_Telephone>
      </gmd:phone>
      <gmd:address>
        <gmd:CI_Address>
          <gmd:deliveryPoint>
            <gco:CharacterString>13 rue de Rivoli</gco:CharacterString>
          </gmd:deliveryPoint>
          <gmd:city>
            <gco:CharacterString>PARIS</gco:CharacterString>
          </gmd:city>
          <gmd:postalCode>
            <gco:CharacterString>13001</gco:CharacterString>
          </gmd:postalCode>
          <gmd:country>
            <gco:CharacterString>France</gco:CharacterString>
          </gmd:country>
          <gmd:electronicMailAddress>
            <gco:CharacterString>catalogue@myorganisation.fr</gco:CharacterString>
          </gmd:electronicMailAddress>
        </gmd:CI_Address>
      </gmd:address>
    </gmd:CI_Contact>
  </gmd:contactInfo>
  <gmd:role>
    <gmd:CI_RoleCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\_19139\_Schemas/resources/codelist/ML\_gmxCodeLists.xml#CI\_RoleCode"
      codeListValue="owner"/>
    </gmd:role>
  </gmd:CI_ResponsibleParty>

```

Les ressources en ligne

Les contraintes d'accès et d'utilisation

Xpath ISO19139:

MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:resourceConstraints

La directive INSPIRE est loin d'être claire concernant les contraintes d'accès et d'utilisation.

On trouve sur le site wiki de georezo.net, la description de différents cas:

https://georezo.net/wiki/main/donnees/inspire/aide_a_la_saisie_des_metadonnees_inspire

Les restrictions d'accès public

Il s'agit des restrictions légales d'accès et non des restrictions pratiques comme l'authentification.

La liste des valeurs est imposée par la directive INSPIRE.

Code	Description
noLimitations	No limitations on public access
INSPIRE_Directive_Article13_1a	Public access to spatial data sets and services would adversely affect the confidentiality of the proceedings of public authorities, where such confidentiality is provided for by law.
INSPIRE_Directive_Article13_1b	Public access to spatial data sets and services would adversely affect international relations, public security or national defence.
INSPIRE_Directive_Article13_1c	Public access to spatial data sets and services would adversely affect the course of justice, the ability of any person to receive a fair trial or the ability of a public authority to conduct an enquiry of a criminal or disciplinary nature.
INSPIRE_Directive_Article13_1d	Public access to spatial data sets and services would adversely affect the confidentiality of commercial or industrial information, where such confidentiality is provided for by national or Community law to protect a legitimate economic interest, including the public interest in maintaining statistical confidentiality and tax secrecy.
INSPIRE_Directive_Article13_1e	Public access to spatial data sets and services would adversely affect intellectual property rights.
INSPIRE_Directive_Article13_1f	Public access to spatial data sets and services would adversely affect the confidentiality of personal data and/or files relating to a natural person where that person has not consented to the disclosure of the information to the public, where such confidentiality is provided for by national or Community law.
INSPIRE_Directive_Article13_1g	Public access to spatial data sets and services would adversely affect the interests or protection of any person who supplied the information requested on a voluntary basis without being under, or capable of being put under, a legal obligation to do so, unless that person has consented to the release of the information concerned.
INSPIRE_Directive_Article13_1h	Public access to spatial data sets and services would adversely affect the protection of the environment to which such information relates, such as the location of rare species.

Exemples ISO19139

```
<gmd:resourceConstraints>
  <gmd:MD_LegalConstraints>
    <gmd:accessConstraints>
      <gmd:MD_RestrictionCode
codeList="http://standards.iso.org/iso/19139/resources/gmxCodeLists.xml#MD_RestrictionCode"
codeListValue="otherRestrictions"/>
    </gmd:accessConstraints>
    <gmd:otherConstraints>
      <gmx:Anchor xlink:href="http://inspire.ec.europa.eu/metadata-
codeList/LimitationsOnPublicAccess/INSPIRE_Directive_Article13_1d">Public access to spatial data
sets and services would adversely affect the confidentiality of commercial or industrial
information, where such confidentiality is provided for by national or Community law to protect a
legitimate economic interest, including the public interest in maintaining statistical
confidentiality and tax secrecy.</gmx:Anchor>
    </gmd:otherConstraints>
  </gmd:MD_LegalConstraints>
</gmd:resourceConstraints>
```

Dans la plupart des cas, il n'y a aucune limitation d'accès public:

```
<gmd:resourceConstraints>
  <gmd:MD_LegalConstraints>
    <gmd:accessConstraints>
      <gmd:MD_RestrictionCode
codeList="http://standards.iso.org/iso/19139/resources/gmxCodeLists.xml#MD_RestrictionCode"
codeListValue="otherRestrictions"/>
    </gmd:accessConstraints>
    <gmd:otherConstraints>
      <gmx:Anchor xlink:href="http://inspire.ec.europa.eu/metadata-
codeList/LimitationsOnPublicAccess/noLimitations">No limitation to public access</gmx:Anchor>
    </gmd:otherConstraints>
  </gmd:MD_LegalConstraints>
</gmd:resourceConstraints>
```

Les conditions d'accès et d'utilisation

Code	Value
conditionsUnknown	Conditions to access and use unknown
noConditionsApply	No conditions apply

Code listes

Voir la page: [ensemble des code listes ISO TC211](#)

Status

Rôles

CRS

Exemples de fiches

GeoJSON

ISO19139

DataCite

Points à préciser

Référentiel spatial

Thesaurus à établir

- Les variables
- Les plateformes
- Les instruments

Code liste à établir

CRS

Documentations:

Exemple d'implémentation INSPIRE pour les métadonnées d'un jeu de données

https://inspire.ec.europa.eu/sites/default/files/md_2.0_datasets_example.xml

Guide d'implémentation ISO 19139 INSPIRE

<https://inspire.ec.europa.eu/sites/default/files/documents/metadatas/inspire-tg-metadatas-iso19139-2.0.1.pdf>

DataCite métadonnées:

https://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadataKernel_v4.2.pdf

Code listes et thésaurus GCMD - NASA

<https://gcmdservices.gsfc.nasa.gov/static/kms/>

Pour les codes listes ISO19115

https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19115_and_19115-2_CodeList_Dictionaries

epointal	Métadonnées Documents de travail Version 0.7	ForM@Ter 
----------	--	--

A propos de sos: <https://github.com/52North/SOS>

Spécification CRS

[INSPIRE specification CRS v3.0.pdf](#)

- i La directive INSPIRE est une directive européenne sur le partage des données géographiques et comprend un ensemble de recommandations pour l'implémentation des métadonnées.
- ii DataCite est un organisme international de bibliothèques et autres services des sciences de l'information dont l'objectif est de faciliter l'accès aux données numériques. DataCite possède son propre [Metadata Schema](#)
- iii Le [modèle de métadonnées de la NASA](#)