



Métadonnées de Collection
et Jeux de Données

V0.5

Document de travail

Le 20 novembre 2019

epointal	Métadonnées Documents de travail	
----------	-------------------------------------	---

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
V0.3	Détails des champs complexes	epointal	18 Novembre 2019
V0.4	Détails des champs complexes	epointal	19 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...
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							
identifiant	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

epointal	Métadonnées Documents de travail	
----------	-------------------------------------	---

dataCharacterSet		MD_CharacterSet Code	utf-8	C	0..1	C		Encodage des données (si différent de utf-8)
spatialExtent		bbox (WGS84)		M	1..n	M	M une seule	
temporalExtent	begin	Date ISO8601		M			N/A	
	end			O (rien ou now si on going)				
dataMaintenanceFrequency		MD_MaintenanceFrequencyCode	asNeeded	O	0..1	O		Fréquence de la maintenance des données
resourceConstraints - legalConstraints								
useLimitation			No conditons apply	R	1..n			Texte décrivant comment citer la ressource par exemple nom de la license, url de la license
useConstraints		MD_RestrictionCode		C	0..n			Mot clés: License copyright ...
accessConstraints		Liste Code INSPIRE	No limitations on public access	M	1..n			restriction d'accès suivant la liste INSPIRE
distributionInfo								
format				M	1..n	M		Format des données
onlineResource		voir onlineResource		R	0..n			Lien de téléchargement ou d'accès (surtout pour les dataset!)
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/gmxCodelists.xml>

Il faut donc utiliser la liste de la NOAA

https://data.noaa.gov/resources/iso19139/schema/resources/Codelist/gmxCodelists.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 »

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

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
- platform
- instrument

ainsi que:

- timeResolution

De nombreux autres thésaurus peuvent être utilisés:

- Geonetwork par exemple, utilise le thésaurus [naturalearth-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)

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

	thesaurusDate publication	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

Les ressources en ligne

Les contraintes d'accès et d'utilisation

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 des différents cas possibles:

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 et non des restrictions pratiques comme l'authentification.

Les conditions d'accès et d'utilisation

Code listes

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

epointal	Métadonnées Documents de travail	ForM  Ter
----------	-------------------------------------	--

Status

Rôles

CRS

topicCategory

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/metadata/inspire-tg-metadata-iso19139-2.0.1.pdf>

DataCite metadonnées:

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

https://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadatasKernel_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

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](#)