

Améliorer la FAIRisation des données météorologiques à l'aide de la ressource lexicale INMEVO



IIRIT



Mouna Kamel, **Nathalie Aussenac-Gilles**, Cassia Trojahn

mouna.kamel@irit.fr



SEMANTICS 4 FAIR





Contexte : rendre plus visible des jeux de données météo

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

MÉTÉO FRANCE

Rechercher une ville, un pays...

f t in

Aujourd'hui Demain Vigilance météo

PREVISIONS METEO DES PLAGES METEO MARINE MÉTÉO DES FORÊTS CLIMAT ACTUS & DOSSIERS NOS SERVICES

Ajouter une ville +

MÉTÉO FRANCE

AUJOURD'HUI DEMAIN WEEK-END 7 JOURS 15 JOURS TENDANCES

Mardi 04
Soirée
Nuit
Mercredi 05

PRÉVISIONS VENT

15° 18° 20° 18° 18° 22°
15° 18° 20° 21° 18° 21°
19° 21° 22° 23° 20°
20° 22° 20° 26° 23°
21° 21° 24° 26° 25°

A LA UNE

Comment Météo-France vous accompagne pendant l'été

La saison estivale est plus que jamais une période de mobilisation générale pour les équipes de Météo-France



Contexte: Accessibilité des données météorologiques

Des données météo

- Cruciales pour diverses applications et domaines (météorologie, climatologie, transports, agriculture, tourisme, médecine, etc.).
- Issues de différentes sources (stations météorologiques, satellites, radars, etc.).
- Ouvertes, accessibles sur des portails (Meteo-France, worldweather, infoclimat ou meteociel).

The screenshot shows the World Weather Information Service (WWIS) website. At the top, there is the WMO logo and the text 'WORLD METEOROLOGICAL ORGANIZATION Weather - Climate - Water'. The main heading is 'World Weather Information Service Official Forecasts'. Below this, there is a navigation bar with links: Home, Personalize, What's New, Links, Severe Weather, Cloud Atlas, About, Apps, Support. A search bar is present with the text 'Please enter city / country / territory name'. Below the search bar, there is a section with text: 'This global website presents OFFICIAL weather observations, weather forecasts and climatological information for selected cities supplied by National Meteorological & Hydrological Services (NMHSs) worldwide. The NMHSs make official weather observations in their respective countries. Links to their official weather service websites and tourism board/organization are also provided whenever available. Weather icons are shown alongside worded forecasts in this version to facilitate visual inspection. By June 2023, WWIS provided official weather information for 3451 cities in which 3307 cities are available with forecast from 139 Members while 2209 cities are available with climatological information from 171 Members. Suggestions to enrich the contents of this website are welcome.' Below this text, there are three radio buttons: 'Present Weather' (checked), 'Satellite Image', and 'Forecast / Climatology'. A world map is displayed with various weather icons (sun, clouds, rain, snow, etc.) placed over different regions. On the right side of the map, there is a dropdown menu with the text 'Please select' and a list of countries: Afghanistan, Albania, Algeria, Andorra, Angola, and Antigua and Barbuda.



Contexte: Accessibilité des données météorologiques

Des données météo

- Cruciales pour diverses applications et domaines (météorologie, climatologie, transports, agriculture, tourisme, médecine, etc.).
- Issues de différentes sources (stations météorologiques, satellites, radars, etc.).
- Ouvertes, accessibles sur des portails (Meteo-France, worldweather, infoclimat ou meteociel).

Meteociel.fr Observations, Prévisions, Modèles En temps réel

Accueil | 9140 visiteurs

• Modèles - AROME [PE-AROME 2.5km] - [Carte zoom dynamique AROME 1.3km] - [WRF-NMM 2km] - [Liste des modèles]

• Menu Utilisateur

- › Inscription
- › Se connecter
- › Meteociel Android
- › Meteociel iOS
- › Héberger image
- › Forums Meteociel
- › Chat météo (5)

• Temps réel

- › Observations
- › Poster vos obs.
- › Détails des obs.
- › Galerie des obs.
- › Carte des photos
- › Partager une photo
- › Température
- › Température min
- › Température max
- › Température mer
- › Point de rosée
- › Windchill - Ressenti
- › Humidex
- › Isobares
- › Pression
- › Vent moyen

8 Domaines : France Espagne Italie Roy.-Uni Allemagne Belgique Suisse DOMTOM

5 Zones : Pays Nord-Ouest Nord-Est Sud-Ouest Sud-Est Archives Zoom dynamique Ensembliste

Cartes du modèle AROME 1.3km

AROME HD 1.3km (souvent plus rapide à sortir que le 2.5km) :

Températures :	Précipitations :	Cumuls précipitations :	Vent :	Autres :
Temp. 2m / (Avec iso) Temp. Max 2m / Min 2m Windchill - Humidex Point rosée 2m	Précipitations Réfl. Radar Max Type précip.	Accumulations préc. Cumul précip. neige depuis 0h Cumul grapel depuis 0h Cumul neige + grapel depuis 0h	Vent moy. 10m Rafales 10m Rafales Max.Ech	MUCAPE Couches nébu. Humidité 2m Simu Sat IR - Col.

AROME 2.5km (Plus long à sortir mais avec plus de paramètres) :

Températures :	Précipitations :	Cumuls précipitation :	Vent :	Autres :
Temp.2m Temp.2m (iso) Temp. Max 2m / Min 2m Windchill - Humidex Point rosée 2m Temp 850hPa Temp 500hPa	Précipitations Réfl. Radar 925hPa / Max. Type précipitations Eau précipitable Dynamique d'alti. : Alt 1.5 PVU TA 500hPa ThetaE 850hPa Theta'W 850hPa	Accumulations préc. Cumul précip. neige depuis 0h	Vent moyen 10m Rafales 10m Rafales Max.Ech Vent altitude : Vent 925hPa / 700hPa V. Vert 700hPa VV2 700hPa / 600hPa Jetstream 300hPa	MUCAPE Résumé Humidité 2m Humidité 700hPa Couches nébu. Nébulosité Pression/Geop 500hPa

Derniers : Arome 0Z Arome 3Z Arome 6Z Arome 9Z Arome 12Z Arome 15Z Arome 18Z Arome 21Z

Précharger < Anim. >

Mardi 4 juillet 2023 11:00 locale (été)

Run AROME 1.3km 6 Z du Mardi 4 juillet 2023

Précipitation en 1h (mm), risque de neige ou pluie verglaçante (hachures blanches) risque de neige rouillée/arçel (hachures orange)



Contexte: Accessibilité des données météorologiques

Des données météo

- Cruciales pour diverses applications et domaines (météorologie, climatologie, transports, agriculture, tourisme, médecine, etc.).
- Issues de différentes sources (stations météorologiques, satellites, radars, etc.).
- Ouvertes, accessibles sur des portails (Meteo-France, worldweather, infoclimat ou meteociel).

Données publiques

Accueil | Textes officiels et conditions d'accès | **Catalogue** | Données libres d'accès | Informations sur les stations | Référence tarifaire | Géoservices | Votre avis | F.A.Q.

Rechercher []

Déjà inscrit ?
nathalie.aussenac-gilles@u [OK] | Créer un compte
S'inscrire

Catalogue >

Guide de démarrage PDF

Catalogue PDF

- Observations In situ
- Observations Radar
- Climatologie
- Modèles et données de prévisions
- Prévisions climatiques

Filtre avancé

Conditions d'accès :
Aucun filtre

Mots-clés :

- Climatologie
- Etat de mer
- Evapotranspiration

CATALOGUE

1 | 2 | 3 | 4 | 5 | 6 | Page suivante

ASCII

Données SYNOP essentielles OMM

Données d'observations tri-horaires issues des messages d'observation en surface (SYNOP).

Conditions d'accès :

- Sans redevance

Moyens d'accès :

- Téléchargement direct

En savoir +

Données infra-horaires

Précipitations et autres paramètres au pas de temps 6 minutes.

Conditions d'accès :

- Avec redevance

Moyens d'accès :

- Publiothèque
- Hors ligne

En savoir +

Animation mosaïque radar France métropole

Précipitations instantanées en coordonnées cartésiennes estimées à partir de réflectivité radar.

Conditions d'accès :

- Sans redevance

Moyens d'accès :

- Visualisation

Bulletins climatiques de France métropolitaine et outre-mer

Bulletins climatiques mensuels ou quotidiens sur la France, un département ou une région.

Conditions d'accès :

- Sans redevance

Moyens d'accès :

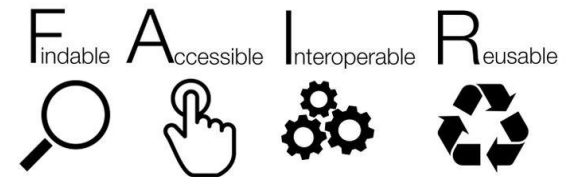
- Téléchargement direct



Contexte: Accessibilité des données météorologiques

Points faibles des données de MétéoFrance

- peu ou mal exploitées
- décrites par les producteurs de données pour les scientifiques du domaine



Objectifs

- Rendre les jeux de données plus faciles à trouver, Accessibles, Interopérables et Réutilisables
- Garantir la conformité de ces données aux principes FAIR
- les rendre accessibles aux non-experts du domaine



Contexte : vers un modèle sémantique de métadonnées

Recommandations FAIR

- représentation formelle et sémantique des métadonnées
- à l'aide de vocabulaires ou d'ontologies standards

Schémas de métadonnées pour les données météo

- RDF data cube + SOSA pour décrire 100 années de données relatives aux températures [L. Lefort et al., 2012]
- RDF data cube + SOSA/SSN + Time + QUDT, GeoSPARQL pour décrire le jeu de données SYNOP [N. Yacoubi et al., 2022]

Notre proposition : **dmo-core** [C. Trojahn et al., 2022]

- Description de schémas de jeux de données,
- notamment les jeux de données tabulaires aux formats CSV ou JSON

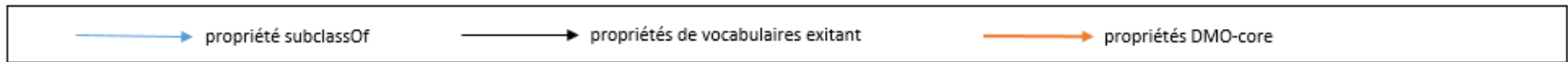
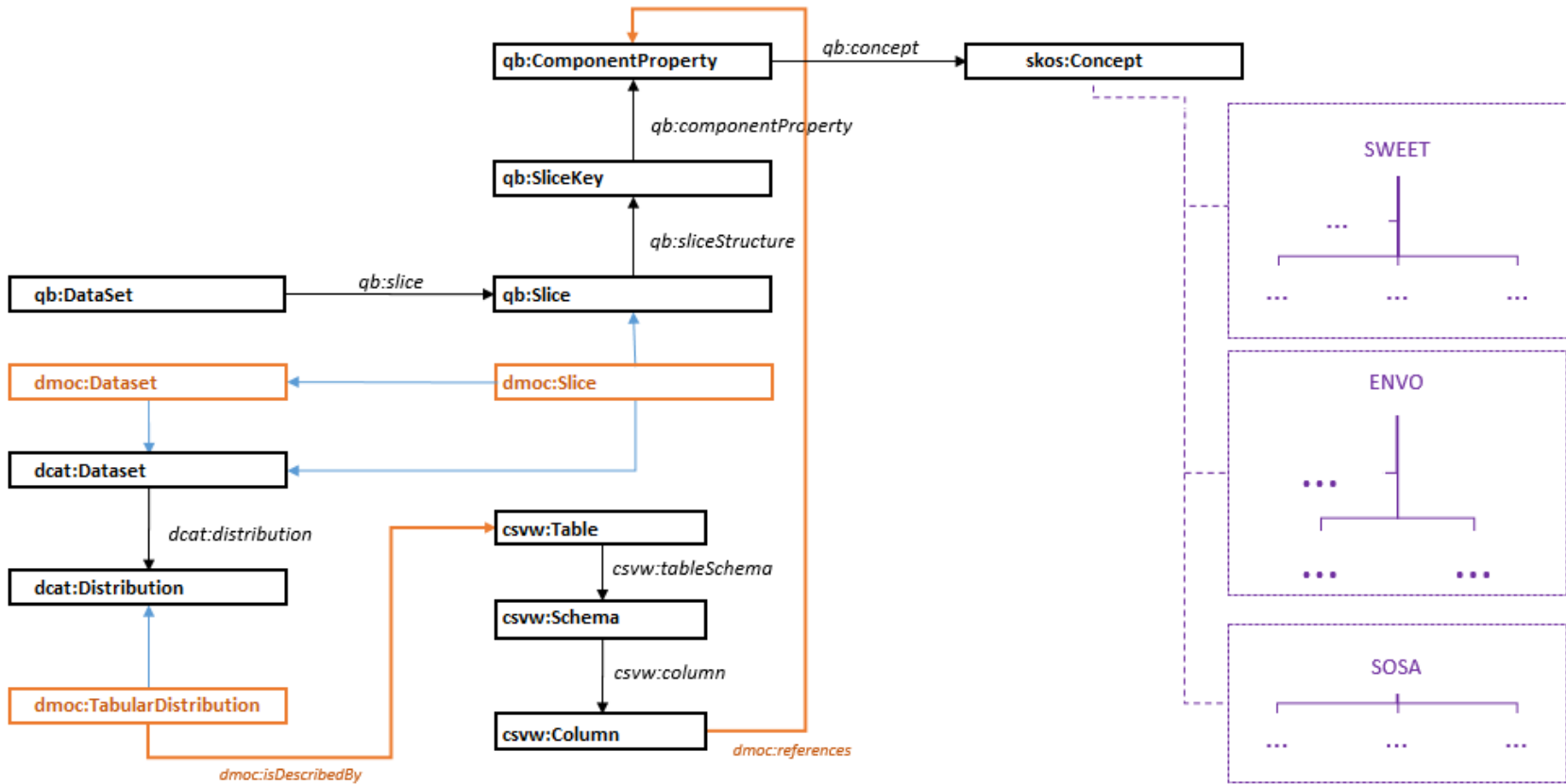


Contexte : Le modèle sémantique dmo-core

- Représenter des métadonnées pour décrire des jeux de données tabulaires et leur structure
- Intègre des vocabulaires et ontologies FAIR
 - DCAT** : <https://www.w3.org/TR/vocab-dcat-2/>
 - RDF Data Cube** : <https://www.w3.org/TR/eo-qb/>
 - CSVW** : <https://www.w3.org/ns/csvw>
- Instancier ce modèle = créer des métadonnées décrivant un jeu de données
- Permet de sémantiser les colonnes en les reliant à des concepts d'ontologies de domaine



Le modèle sémantique dmo-core : description partielle





Utilisation du modèle sémantique dmo-core : Limites des ontologies de domaine

Les concepts des ontologies de domaine ne sont pas toujours bien documentés.

Ex **SWEET** : peu de définitions, peu de termes en français

Pression de la mer
Sea level Pressure :
définition en anglais
Unité : Pa

Température :
Pas de définition
Unité : K

numer_sta	date	pmer	tend	cod_tend	dd	ff	t	td	...
7005	2,02E+13	103180	-80	8	120	1.800000	274.350000	272.750000	...
7015	2,02E+13	103320	0	5	80	4.700000	275.250000	275.150000	...
7020	2,02E+13	102870	-70	8	80	1.300000	280.550000	279.450000	...
7027	2,02E+13	103080	0	0	100	4.200000	275.750000	275.750000	...
7037	2,02E+13	103190	-30	8	130	2.200000	272.250000	272.250000	...
7072	2,02E+13	103320	-20	8	60	1.100000	270.650000	269.550000	...
7110	2,02E+13	102740	10	0	180	0.600000	282.750000	282.650000	...
7117	2,02E+13	102760	-20	8	130	0.500000	281.550000	280.950000	...
7130	2,02E+13	102940	-90	8	110	3.100000	278.350000	278.050000	...
...

Extrait du jeu de données SYNOP de météo-France



Proposition : ajouter des définitions

The screenshot shows the Météo France website interface. At the top left, there are logos for the République Française and Météo France. A search bar is present with the placeholder text "Rechercher une ville, un pays...". To the right of the search bar are social media icons for Facebook, Twitter, and LinkedIn, along with a notification bell icon. Further right are two maps of France labeled "Aujourd'hui" and "Demain", and a "Vigilance météo" icon. Below these are navigation tabs: "PREVISIONS", "METEO DES PLAGES", "METEO MARINE", "MÉTÉO DES FORÊTS", "CLIMAT", "ACTUS & DOSSIERS", and "NOS SERVICES". A blue bar below the navigation contains the text "Ajouter une ville" with a plus icon.

The main content area is titled "MÉTÉO FRANCE" and features a navigation bar with tabs: "AUJOURD'HUI" (highlighted in pink), "DEMAIN", "WEEK-END", "7 JOURS", "15 JOURS", and "TENDANCES". Below the tabs is a weather map of France showing temperature forecasts for "Mardi 04" and "Mercredi 05". The map includes icons for "PRÉVISIONS" and "VENT". A sidebar on the left shows the current date and time: "Mardi 04", "Soirée", "Nuit", and "Mercredi 05".

On the right side of the main content area, there is a featured article titled "A LA UNE" with a sub-headline "Comment Météo-France vous accompagne pendant l'été". The article text reads: "La saison estivale est plus que jamais une période de mobilisation générale pour les équipes de Météo-France".



Proposition : ajouter des définitions

Pourquoi ajouter des termes et définitions ?

- améliorer l'interopérabilité des jeux de données grâce à des métadonnées enrichies de termes et de définitions
- relier les colonnes des jeux de données tabulaires aux concepts définis dans cette ressource.

Où trouver des définitions ?

- Vocabulaire Météorologique International produit par l'Organisation Météorologique Mondiale (**WMO**)



Proposition : ajouter des définitions

Sous quelle forme les ajouter ?

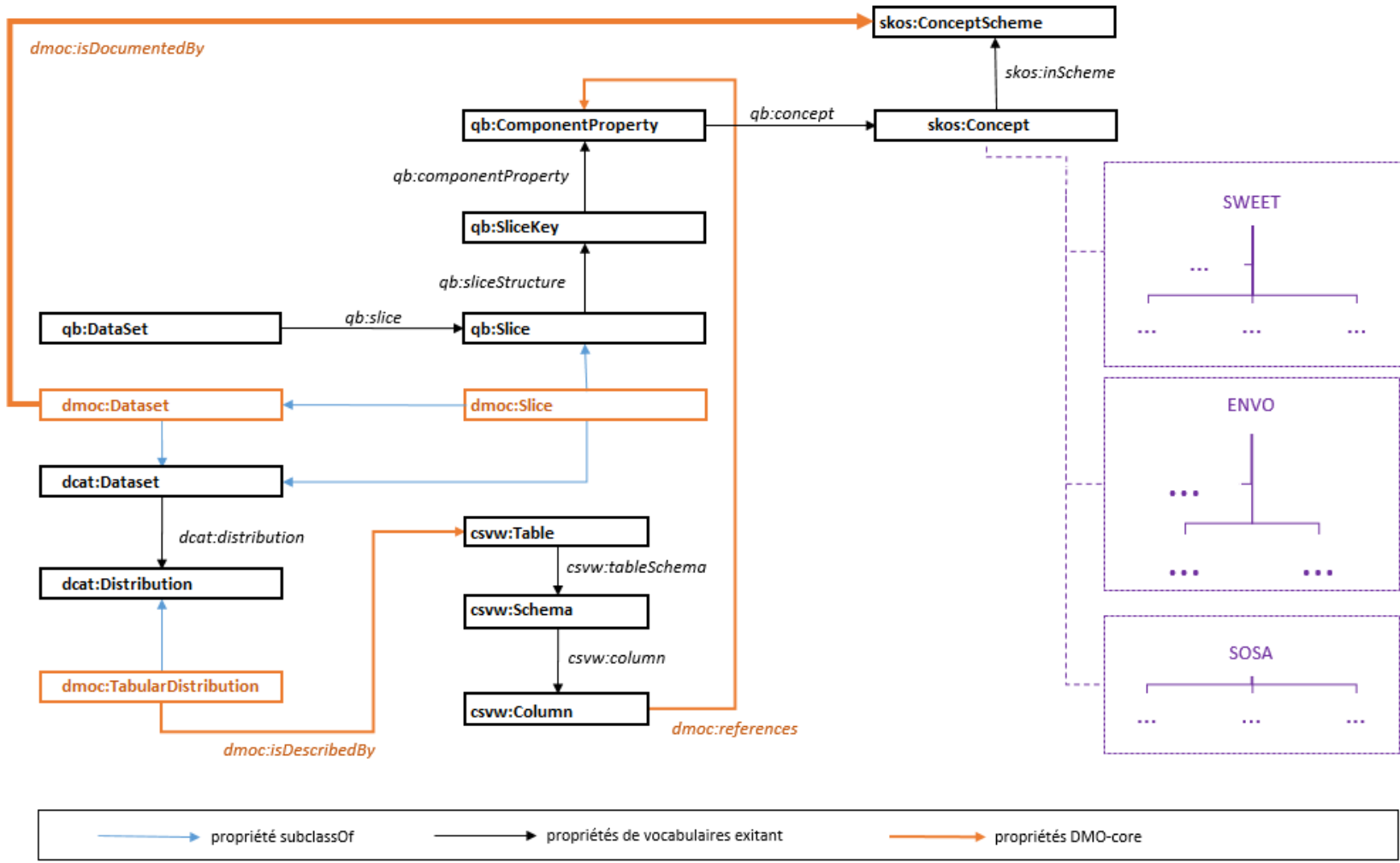
- Représentation formelle de lexiques en SKOS [N. Calzolari, 2008]
- Intégration à un graphe de connaissances [A. Toral & M. Monachini, 2007], [J. McCrae et al., 2011]
- **Créer la ressource lexicale INMEVO**

Où ajouter une ressource lexicale et des définitions ?

- **Option 1 : Enrichir dmo-core**
offrir la possibilité d'intégrer une ressource lexicale
- **Option 2 : enrichir SWEET en liant SWEET et INMEVO**



Enrichissement de dmo-core





Construction de la ressource INMEVO

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

MÉTÉO FRANCE

Rechercher une ville, un pays...

f t in

Aujourd'hui Demain

Vigilance météo

PREVISIONS METEO DES PLAGES METEO MARINE MÉTÉO DES FORÊTS CLIMAT ACTUS & DOSSIERS NOS SERVICES

Ajouter une ville +

MÉTÉO FRANCE

AUJOURD'HUI DEMAIN WEEK-END 7 JOURS 15 JOURS TENDANCES

Mardi 04
Soirée
Nuit
Mercredi 05

PRÉVISIONS VENT

15° 18° 20° 18° 22°
18° 20° 21° 18° 21°
19° 21° 22° 23°
20° 22° 25° 20°
22° 20° 26° 23°
21° 21° 24° 26° 25°

A LA UNE

Comment Météo-France vous accompagne pendant l'été

La saison estivale est plus que jamais une période de mobilisation générale pour les équipes de Météo-France



Construction de la ressource INMEVO

Lexique source WMO

- format pdf (de mauvaise qualité)
vocabulaire intégré dans la base de données UNTERM¹ (ONU), mais non accessible
- autour de 2800 entrées lexicales
- chaque entrée lexicale définie en 4 langues (EN, FR, ES, et RU)

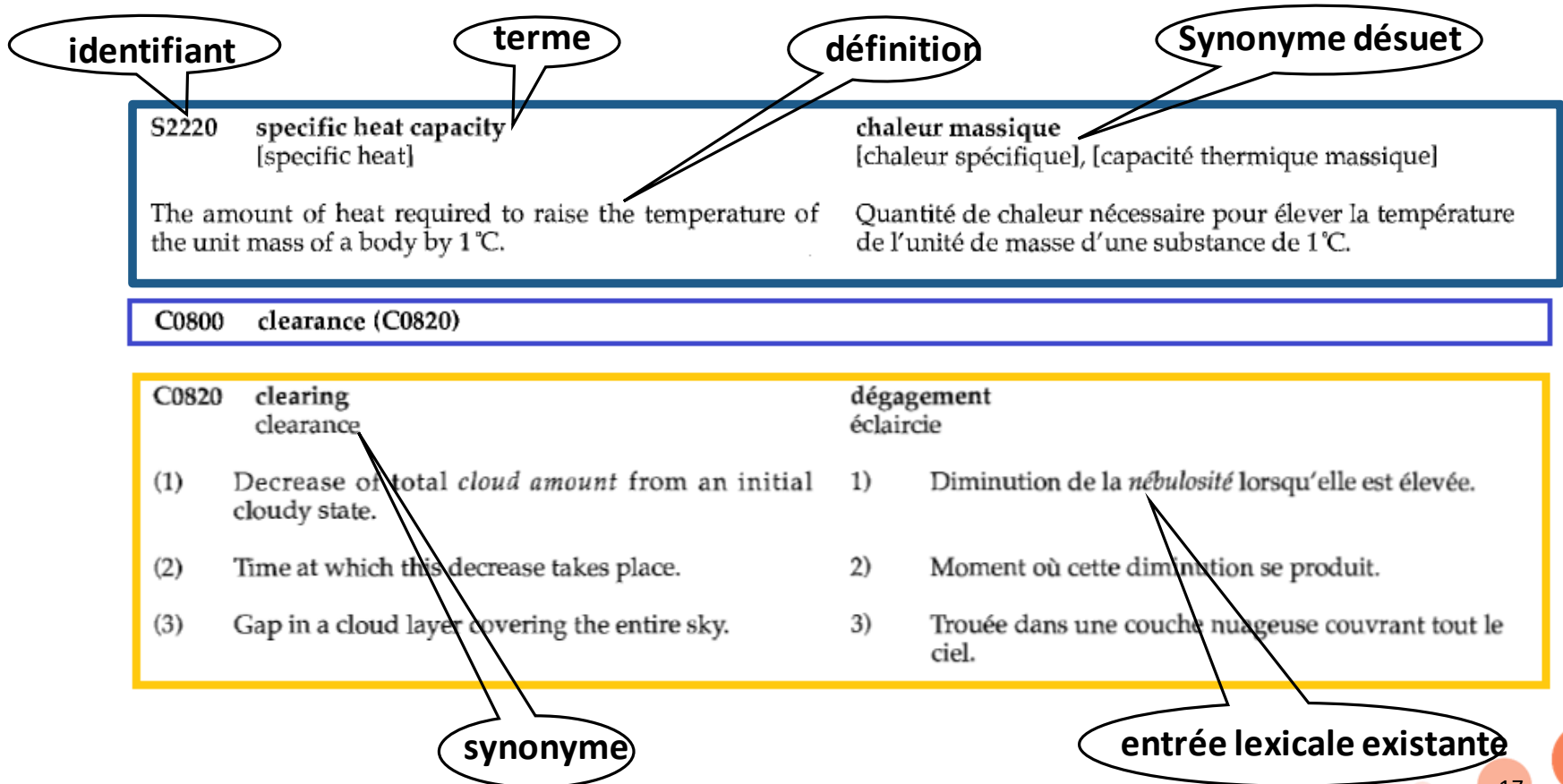
- Mise en forme matérielle du texte régulière
- Indices typographiques qui vont permettre d'exploiter la sémantique

(1) <https://unterm.un.org/unterm2/fr/>

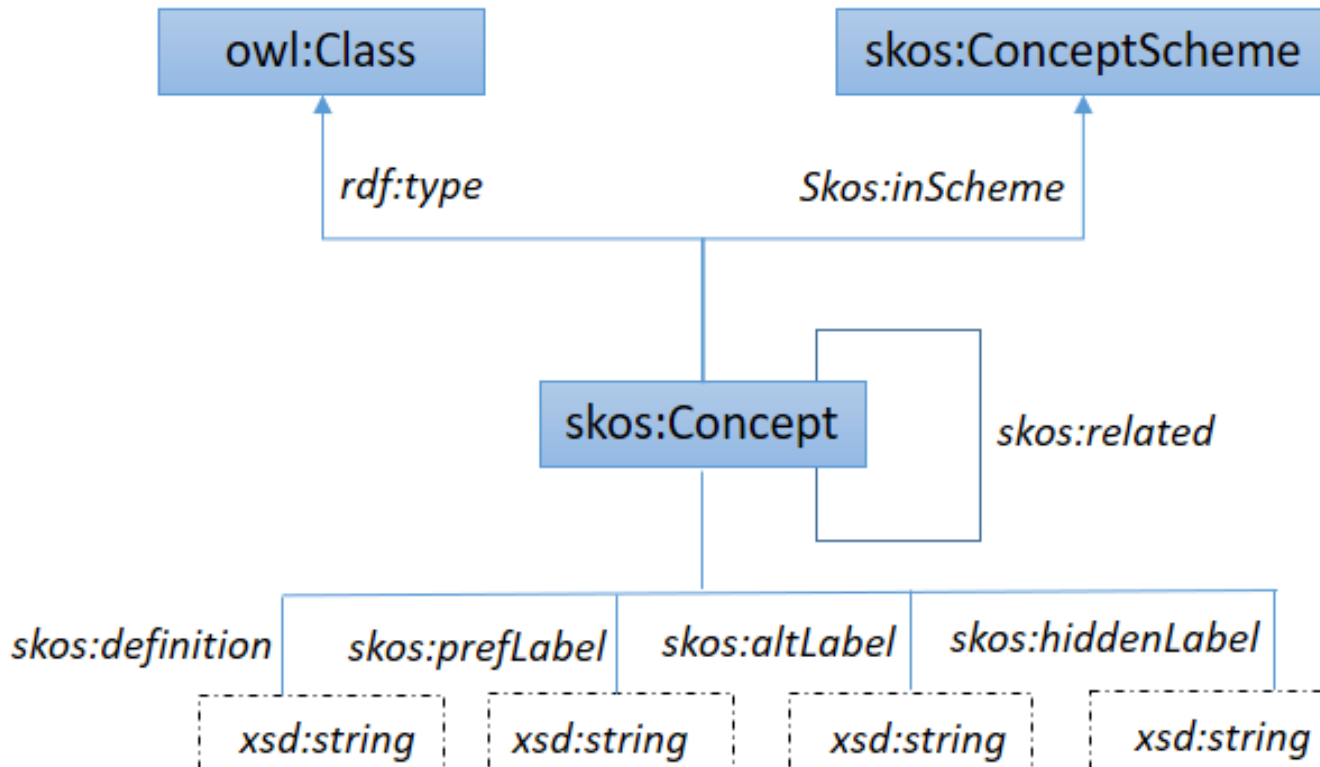


Construction de la ressource INMEVO

Description du lexique



Représentation formelle de WMO





Construction de la ressource INMEVO

Extraction des connaissances à partir des pages pdf

Difficultés

- Document pdf de mauvaise qualité :
 - Problèmes d'OCRisation
 - Pertes de la mise en forme (gras, italique, tabulation, etc.)
 - Perte du parallélisme
- Langues exploitées : anglais et français

Trois étapes principales

- Extraction et représentation des connaissances en EN
- Alignement (par traduction DeepL) des connaissances en EN avec les connaissances en FR
- Représentation des connaissances extraites en FR



Construction de la ressource INMEVO

Exemple de représentation en SKOS d'une entrée lexicale

E1070	evaporation evaporation of water	évaporation évaporation de l'eau
(1)	Emission of water vapour by the free surface of liquid water at a <i>temperaure</i> below its <i>boiling point</i> .	1) Émission de vapeur d'eau par une surface libre d'eau liquide à une température inférieure à son <i>point d'ébullition</i> .
(2)	Amount of water evaporated.	2) Quantité d'eau évaporée.

inmevo:E1070	rdf:type	skos:Concept ;
	skos:prefLabel	"evaporation"@en ;
	skos:prefLabel	"évaporation"@fr ;
	skos:altLabel	"evaporation of water"@en ;
	skos:definition	"Emission of water vapour by the free surface of liquid water at a temperaure below its boiling point."@en ;
	skos:definition	"Amount of water evaporated."@en ;
	skos:definition	"Emission de vapeur d'eau par une surface libre d'eau liquide a une température inférieure a son point d'ébullition."@fr ;
	skos:definition	"Quantité d'eau évaporée."@fr ;
	skos:related	inmevo:B1040 ;
	skos:inScheme	inmevo:INMEVO .



Construction de la ressource INMEVO

Exemples de problèmes rencontrés

C0820	clearing clearance	dégagement éclaircie
(1)	Decrease of total <i>cloud amount</i> from an initial cloudy state.	1) Diminution de la <i>nébulosité</i> lorsqu'elle est élevée.
(2)	Time at which this decrease takes place.	2) Moment où cette diminution se produit.
(3)	Gap in a cloud layer covering the entire sky.	3) Trouée dans une couche nuageuse couvrant tout le ciel.

inmevo:C0820rdf:type	skos:Concept ; skos:prefLabel skos:prefLabel skos:altLabel skos:altLabel skos:definition skos:definition skos:definition skos:definition skos:definition skos:related skos:inScheme	"clearing"@en ; "éclaircie"@fr ; "clearance"@en ; ??? @fr ; "Decrease of total cloud amonnt from an initial cloudy state."@en ; "Time at which this decrease takes place."@en ; "Gap in a cloud layer covering the entire sky."@en ; "Diminution de la nébulosité lorsqu'elle est élevée."@fr ; "Moment oti cette diminution se produit."@fr ; "Trouée dans une couche nuageuse couvrant tout le ciel."@fr ; inmevo:C1460 ; inmevo:INMEVO .
----------------------	--	---



Evaluation

- 30 entrées lexicales choisies aléatoirement
- comparaison manuelle entre document et connaissances extraites

Tableau 1. Pourcentage de connaissances extraites correctement

label@en	96%
label@fr	70%
def@en	96%
def@fr	86.6%
Termes reliés	84.2%
ocerisation@en	63%
ocerisation@fr	61%



Construction de la ressource INMEVO

Evaluation : erreurs les plus fréquentes pour le français

- Traduction des termes spécifiques au domaine n'est pas toujours pertinente
sonde de battage (R0970) → *threshing selon DeepL*
→ *ramsonde selon WMO*
- Correspondance incorrecte entre paragraphes lorsque la page contient des termes de la même famille

C3470 cyanometer

Instrument for determining the blueness of the sky.

C3480 cyanometry

Measurement of the shade of blue of the sky.

cyanomètre

Instrument servant à déterminer la teinte du bleu du ciel.

Instrument for determining the shade of blue in the sky (DeepL)

cyanométrie

Détermination de la teinte du bleu du ciel.



Conclusion

- Construction de la ressource INMEVO en cours
 - version de bien meilleure qualité bientôt en ligne
 - nouvelle méthode pour mieux conserver le parallélisme
 - abandon de DeepL
 - Exploitation des pages comme des tableaux
- Enrichissement de dmo-core pour associer un lexique
- Description du jeu de données SYNOP avec des termes INMEVO



- **Evaluer le caractère FAIR de la description de SYNOP**
 - Utilisation de FOOPs!
- **Améliorer la qualité de la ressource INMEVO**
 - Meilleure extraction pour le français
 - Différencier les termes représentant des phénomènes (éclaircie), des termes représentant des propriétés de phénomène (pression de la mer)
- **Aligner INMEVO avec les ontologies de domaine (SWEET)**