



## Spécialisation d'un logiciel de gestion de métadonnées sémantiques pour la description des jeux de données

Présenté par :  
**Louis Mendy**

Tutrice universitaire :  
**Farah Benamara**

Supervisé par :  
**Nathalie Aussenac-Gilles**  
**Amina Annane**  
**Cassia Trojahn**

2020/2021



## Utilisation d'ontologies pour la recherche de jeux de données météorologiques

Présenté par: Alexandre Champagne

Spécialité: M2 Données et Connaissances

Durée du stage: 29 Mars 2021 au 1er Septembre 2021

Supervisé par: Nathalie Aussenac-Gilles,  
Cassia Trojahn, Amina Annane

Laboratoire: MELODI - IRIT

Tutrice universitaire: Lynda Tamine

# INTRODUCTION

- Plus d'un million de jeux de données disponibles sur le web
  - Jeux de données inexploitablees ⇒ Absence de métadonnées
- Proposition : les principes FAIR
- Nos stages s'inscrivent dans le cadre du projet Semantics4FAIR qui a pour but de faciliter la recherche et l'accès aux données scientifiques en adhérant aux principes FAIR
  - Application aux jeux de données météorologiques de Météo France

# Exemple de Jeux de données et Métadonnées - SYNOP

A	B	C
numer_sta	date	pmer
7005	2,021E+13	102230
7015	2,021E+13	102230
7020	2,021E+13	102250
7027	2,021E+13	102320
7037	2,021E+13	102250
7072	2,021E+13	102330
7110	2,021E+13	102320
7117	2,021E+13	102270
7130	2,021E+13	102420
7149	2,021E+13	102320

```
:station_dataset a dcat:Dataset ;
geodcatap:custodian :MeteoFrance ;
dct:conformsTo <http://www.opengis.net/def/crs/EPSG/0/4326> ;
dct:created "2020-02-29"^^xsd:date ;
dct:description " la liste des stations météo-France pour les données synop... "@fr ;
dct:identifiant <https://donneespubliques.meteofrance.fr/?fond=produit&id\_produit=90&id\_rubrique=32> ;
dct:language <http://publications.europa.eu/resource/authority/language/ENG> ;
dct:provenance :synop_provenance ;
dct:spatial <http://publications.europa.eu/resource/authority/country/FRA>,
:spatial_coverage ;
dct:subject <http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/climatologyMeteorologyAtmosphere>,
<http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/environment>,
<http://inspire.ec.europa.eu/metadata-codelist/TopicCategory/geoscientificInformation> ;
dct:title "LA liste des stations météorologiques pour les DONNÉES SYNOP ESSENTIELLES OMM"@fr ;
```

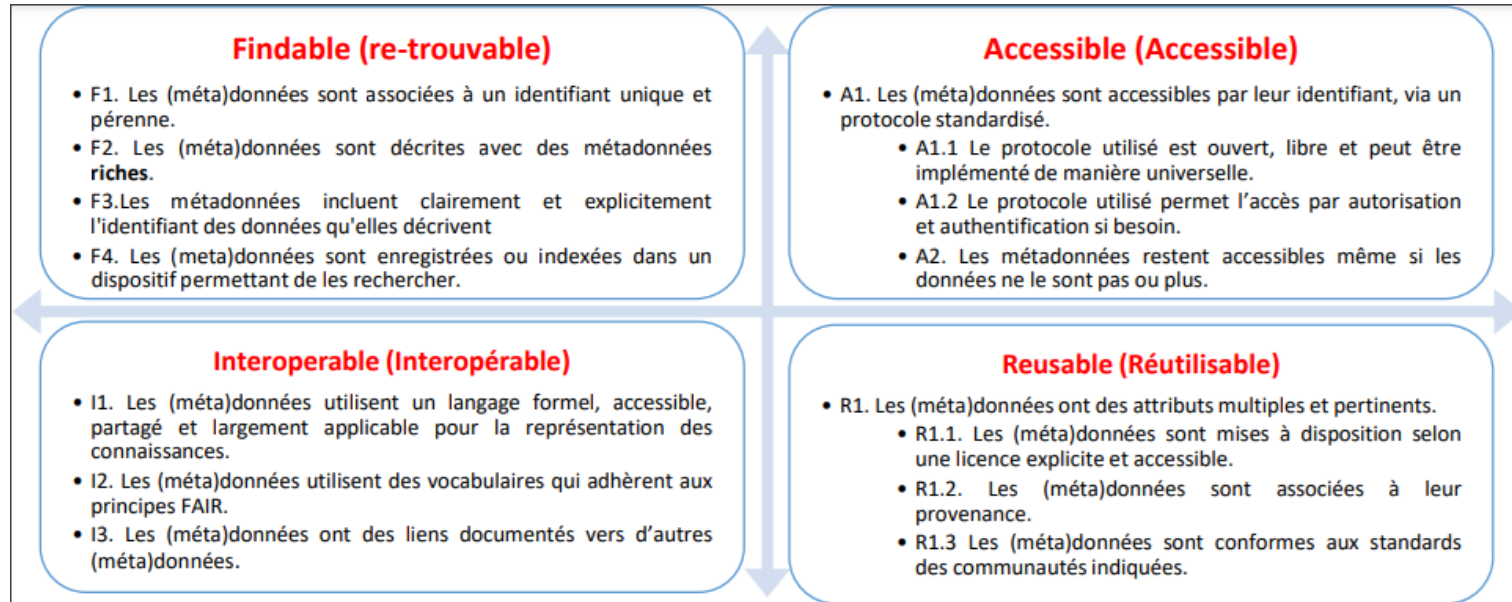
*Jeu de données Liste des stations SYNOP*

*Métadonnées pour décrire les stations SYNOP*

# Introduction

## Principes FAIR

Les principes FAIR ont été proposés pour répondre de façon globale à la problématique de partage des données en vue de leur réutilisation [\[1\]](#)





# Contexte

## Semantics4FAIR



### Domaines de recherche :

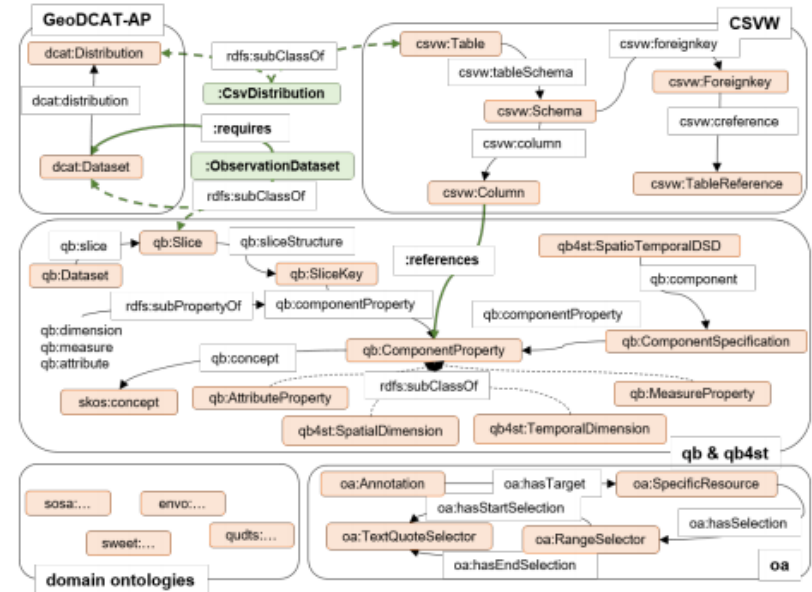
- Ingénierie des connaissances
- Ontologies
- Web Sémantique
- Traitement Automatique du Langage Naturel

### Approches de recherche :

- Etude théorique des fondements linguistiques de la sémantique et de sa représentation formelle
- Etude théorique des structures d'interaction et de discours
- Travaux expérimentaux basés sur des développements logiciels
- Construction de ressources et des expériences basées sur des corpus

# Contexte Semantics4FAIR

- Faciliter la tâche de recherche et d'accès aux données scientifiques résultant à la fois de la recherche et de la production par une communauté scientifique donnée
- Soutenir le développement de nouveaux usages par d'autres communautés



# Contexte Semantics4FAIR

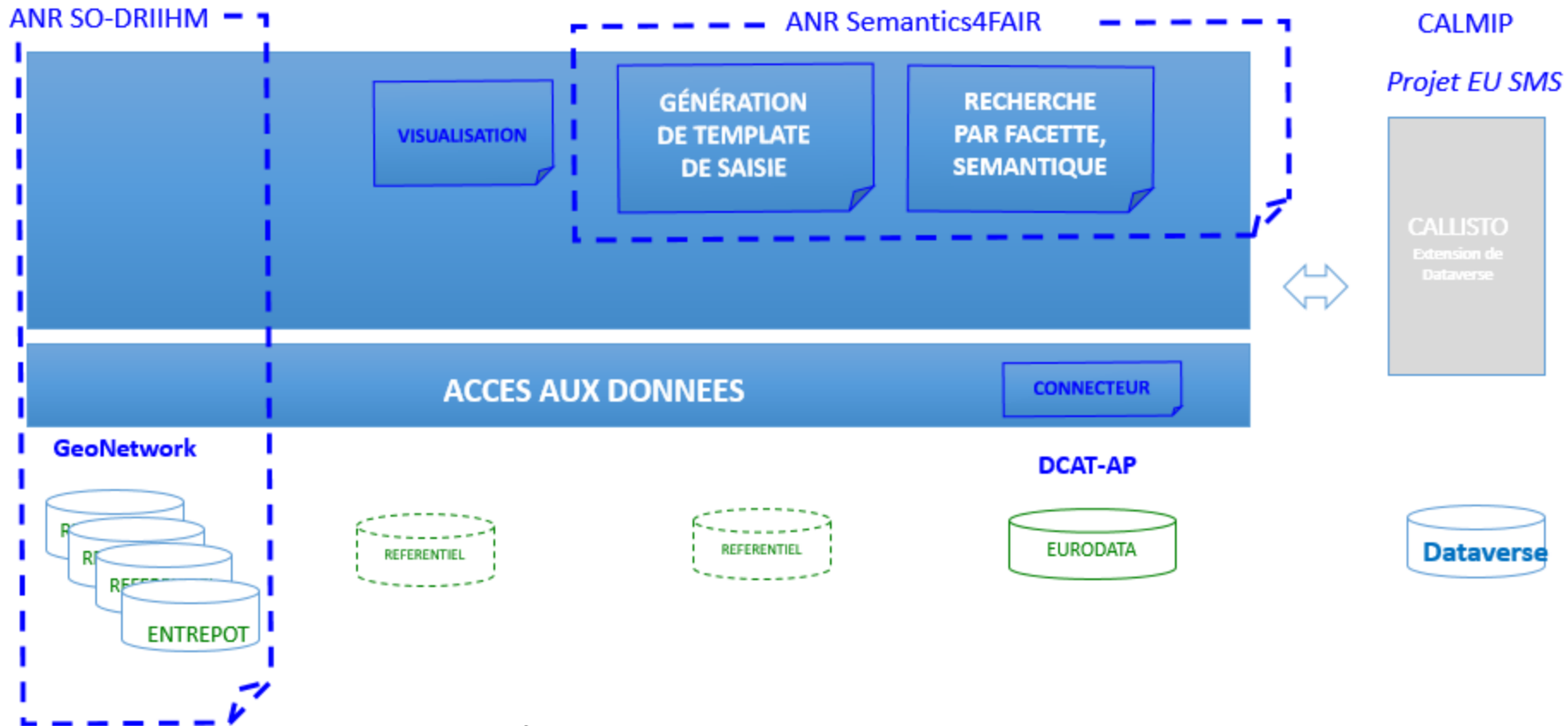






- Une alliance académique regroupant chercheur.e.s, enseignant.e.sc hercheur.e.s, ingénieur.e.s, postdoctorant.e.s, étudiant.e.s en Doctorat/Master/Licence autour du partage de ressources numériques et des pratiques de la connaissance
- Contribue à définir les conditions nécessaires pour que des jeux de données soient véritablement conformes aux principes FAIR, y compris pour des personnes d'autres disciplines que celle des données.

⇒ Plateforme DataNoos



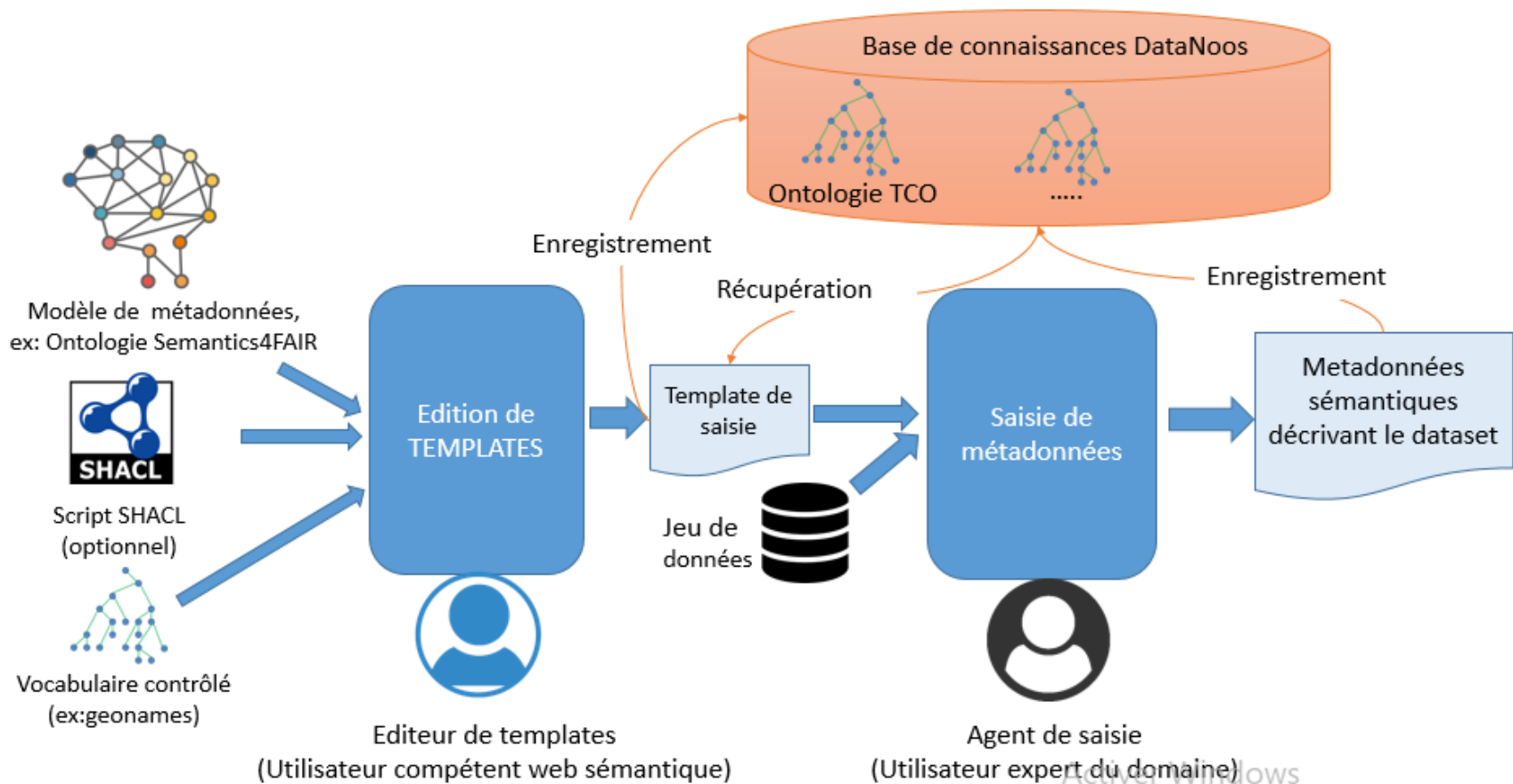
*Plateforme d'expérimentation DataNoos*

# Problématique

- Semantics4FAIR : FAIRisation de données == Ajout de métadonnées au jeux de données
  - Instancier un modèle ontologique
- Experts métier n'ayant pas de connaissance en web sémantique ⇒ Problème d'instanciation du modèle ontologique
- Solution : Développer une application avec des interfaces ergonomiques appelés **templates** pour faciliter la génération de métadonnées à base de modèle ontologique
- **Modèle pas unique ⇒ outil de génération semi-automatique de templates à base de modèle ontologique (module DN-Template)**

# PLAN

- Architecture globale du module DN-Template
- Etude de l'existant
- Documentation de la plateforme DataNooS
- Développement du module DN-Template
- Conclusion
- Perspectives



Architecture globale du module DN-Template

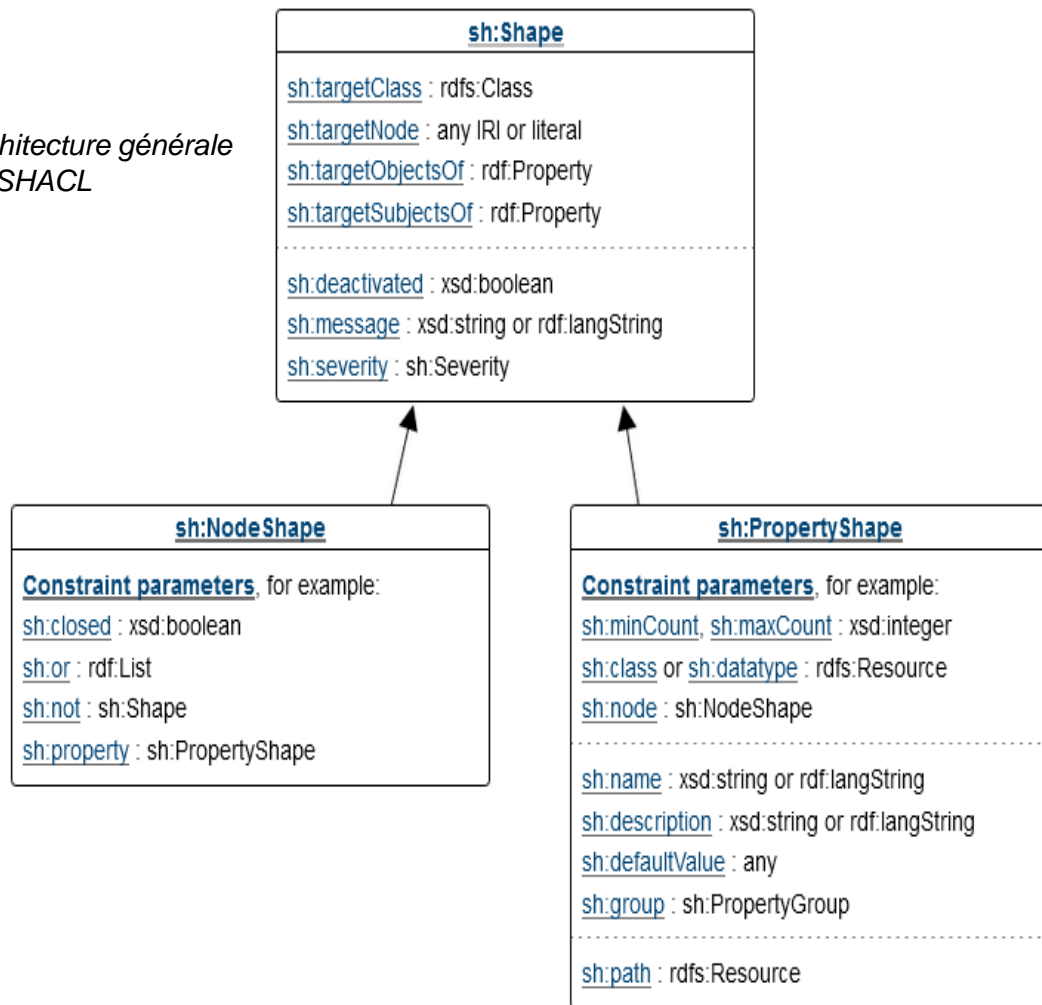
# Etude de l'existant

- Les principes FAIR
- Vocabulaires pour représentées les métadonnées dans Semantics4FAIR : DCAT, DCAT-AP, OWL, RDF DATA CUBE...
- Etude du modèle de représentation des métadonnées dans Semantics4FAIR
- Shacl Shapes Constraint (SHACL)

# Shapes Constraint Language (SHACL)

- SHACL est un langage et une recommandation du W3C, qui permet la représentation des contraintes et la validation des données saisies par rapport à ces contraintes.
- Utiliser pour extraire les propriétés associées aux concepts à décrire dans le template de saisie pour définir automatiquement des contraintes sur les champs de saisie

Architecture générale  
de SHACL



```
:Dataset_Shape
  a sh:NodeShape ;
  sh:name "Dataset"@en ;
  sh:property [
    sh:minCount 1 ;
    sh:nodeKind sh:Literal ;
    sh:path dct:description ;
    sh:severity sh:Violation
  ], [
    sh:minCount 1 ;
    sh:nodeKind sh:Literal ;
    sh:path dct:title ;
    sh:severity sh:Violation
  ], [
    sh:nodeKind sh:Literal ;
    sh:path dct:identifier ;
    sh:severity sh:Violation
  ], [
    sh:class vcard:Kind ;
    sh:path dcat:contactPoint ;
    sh:severity sh:Violation
  ]
```

Extrait d'un script shacl définissant les  
contraintes de la classe dcat:Dataset



# Documentation de la plateforme DataNooS

- Etude fonctionnelle
  - Diagramme de cas d'utilisation
  - Diagramme de séquence
- Etude technique
  - Etude du code existant pour identifier les différentes fonctions intervenant lors de l'exécution des use case

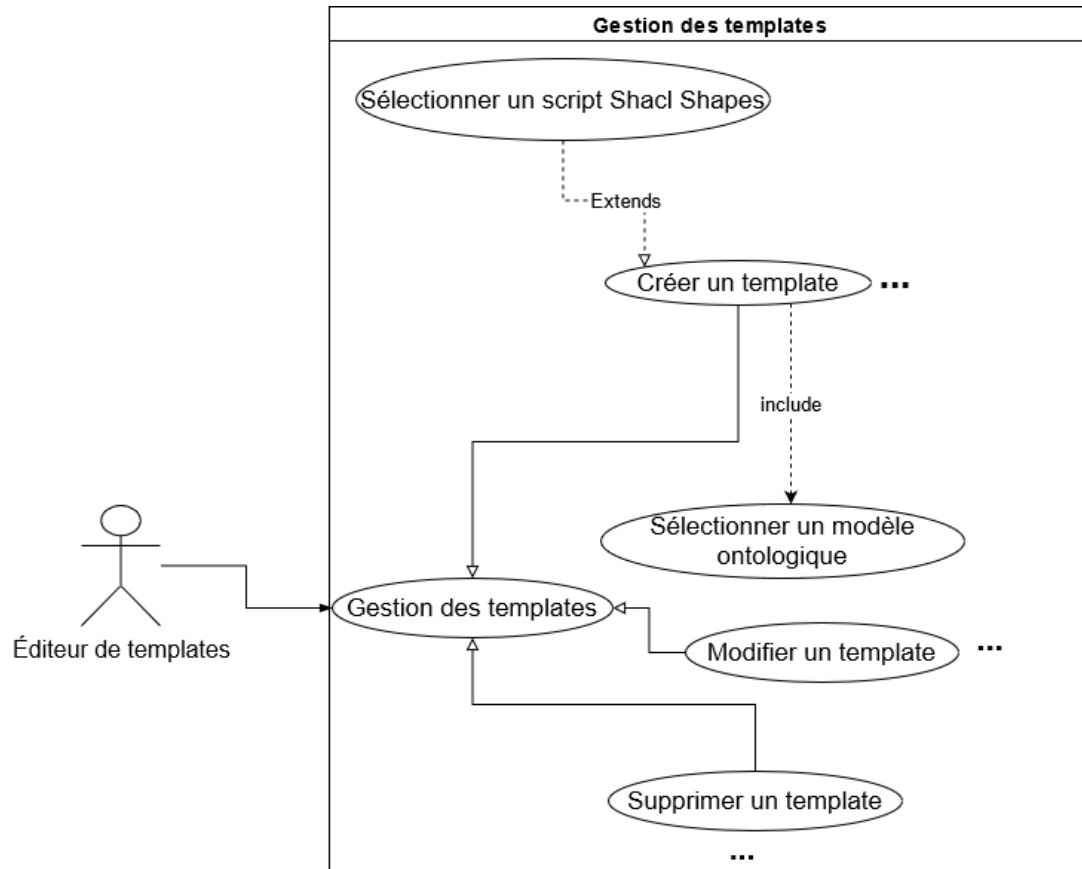
# Développement du module DN-Template

# Module DN-Template

Objectifs :

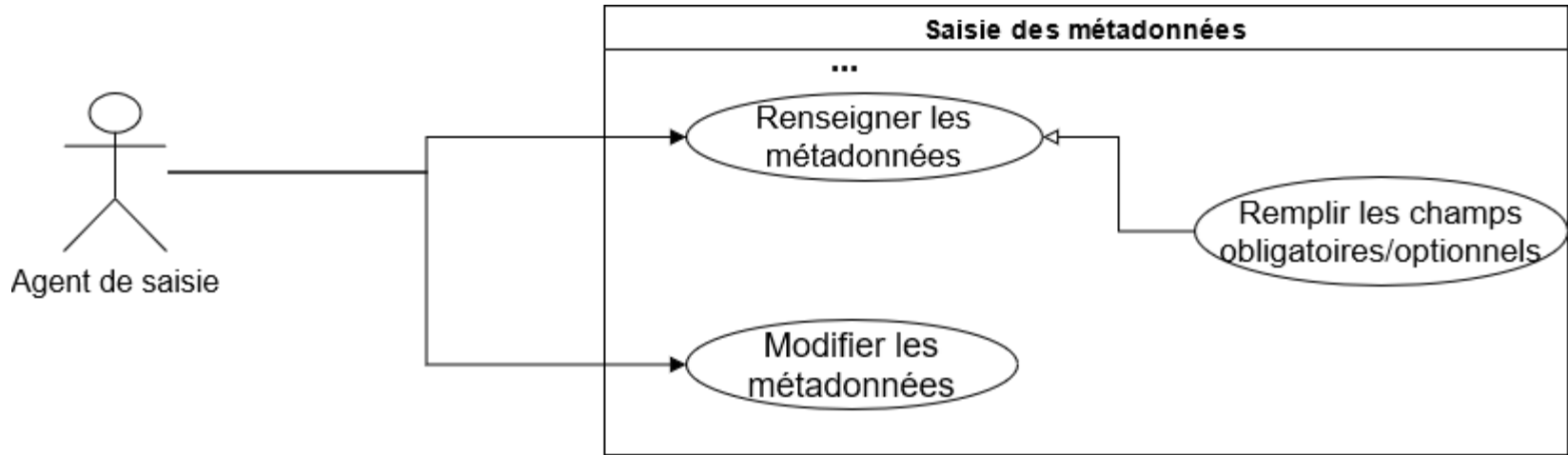
- Génération semi-automatique de template de saisie à base d'ontologie de représentation des métadonnées
- Saisie des métadonnées en utilisant les templates générés

# Identification des acteurs et des cas d'utilisation du module DN-Template



*Extrait du diagramme de cas d'utilisation pour la génération semi-automatique de template de saisie*

# Identification des acteurs et des cas d'utilisation du module DN-Template



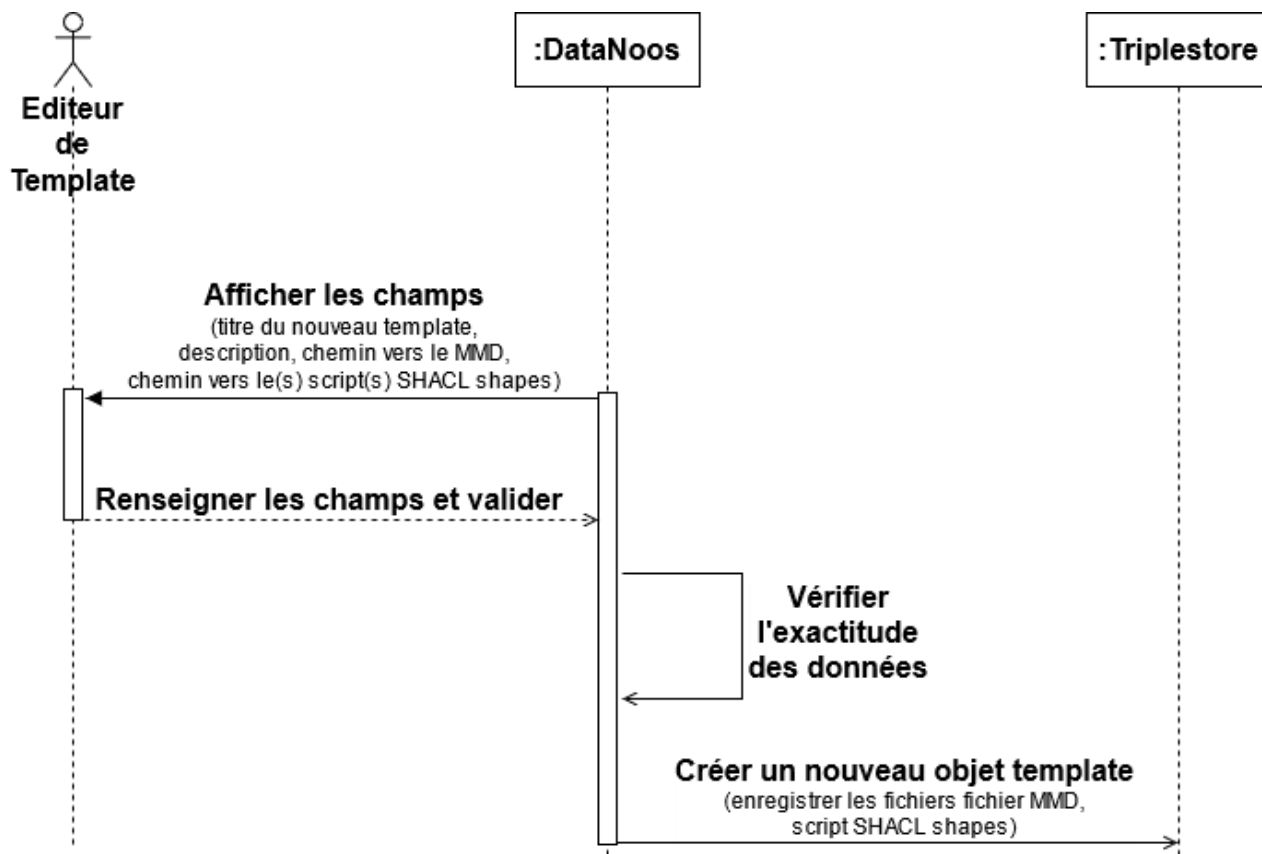
*Extrait du diagramme de cas d'utilisation pour la saisie des métadonnées*

# Spécifications du module DN-Template

Deux étapes pour la réalisation du module :

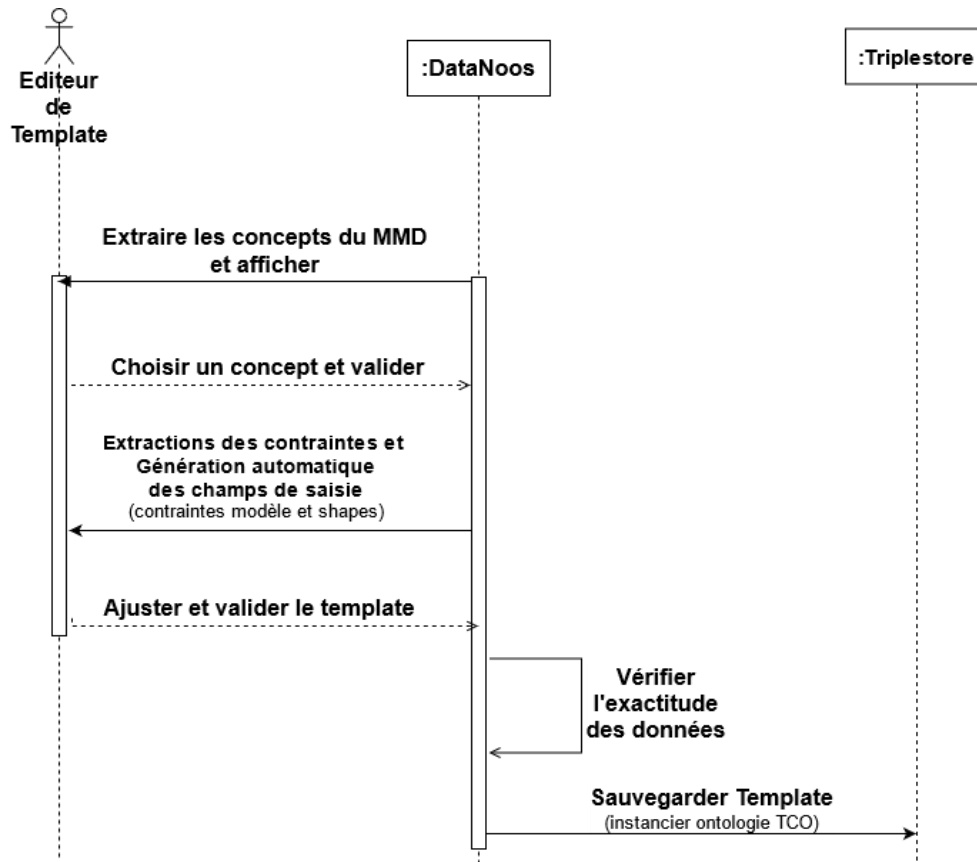
- Définition de template de saisie de métadonnées
- Saisie de métadonnées

→ Illustration avec des diagrammes de séquence



**MMD : Modèle de Métadonnées**

*Configuration d'un nouveau  
template*

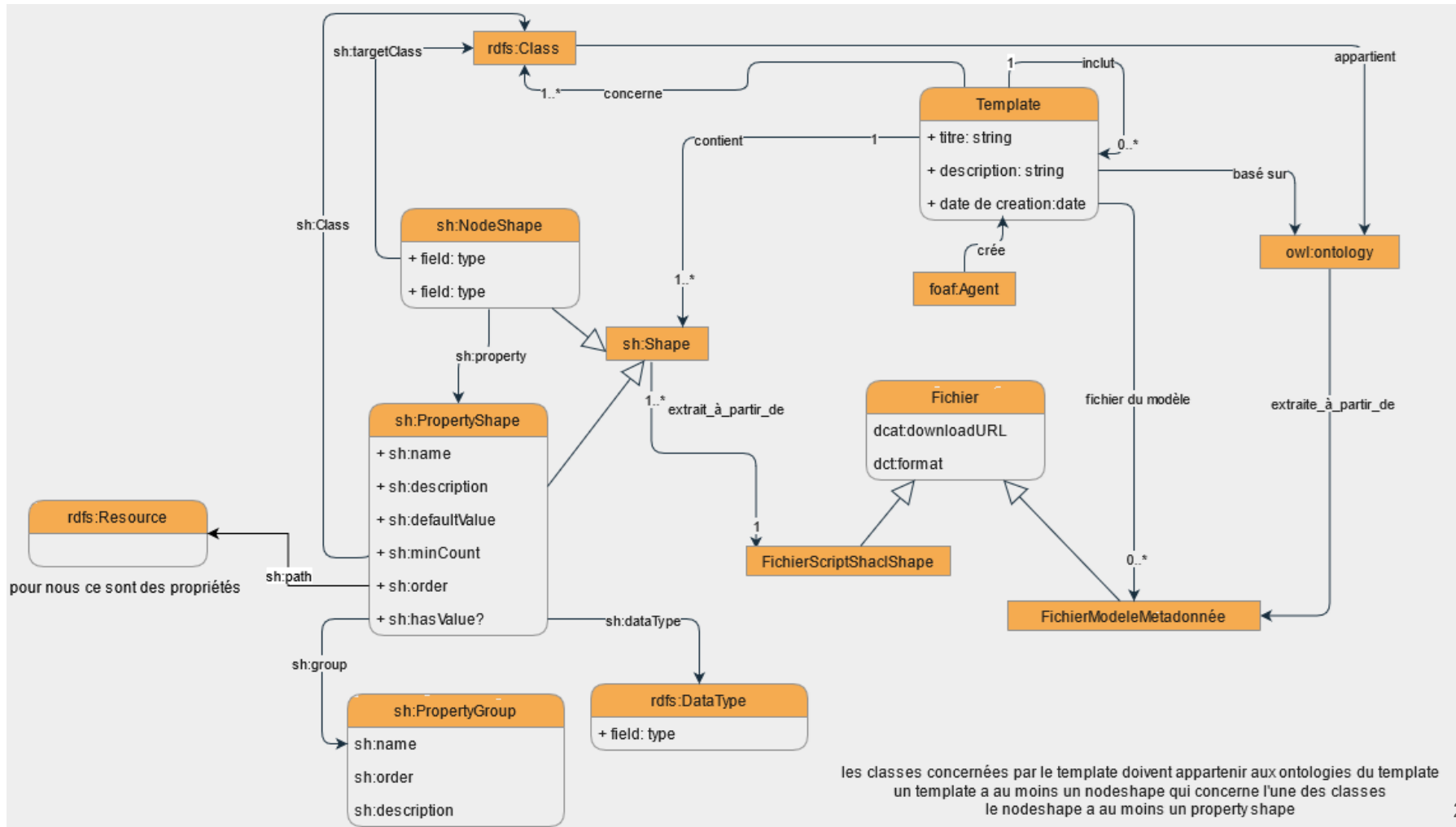


*Génération semi-automatique d'un nouveau template*



# Conception du module DN-Template

- Proposer une ontologie : TCO (Template and Constraint Ontologie)
  - Enregistrer les templates définis via l'éditeur de template
  - Garder les informations sur le template : date de création, chemin des fichiers (modèle de métadonnées et script shacl), IRI de l'ontologie du fichier modèle...
  
- Définir un modèle ontologique sous forme de diagramme de classe UML



# Intégration du module DN-Template à la plateforme DataNooS : Démo

Home Agent Claim Dataset Publication Service Template Workflow daaNooS Administration Tools

---

## Create a new template

URI  
  
Generated automatically

Title

Description

Metadata Model

Script SHACL Shapes

Activer Windows  
Accédez aux paramètres pour activer Windows.

Pour illustrer l'intégration on utilise :

- Fichier DCAT au format turtle comme modèle de métadonnées
- Script SHACL shapes de DCAT-AP

PREFIX TCO: <<https://www.irit.fr/recherches/MELODI/ontologies/TCO#>>

**TCO:Template30082021** a TCO:Template

dct:title " Template basé sur le modèle DCAT ";

dct:description " Ce template est créé pour décrire les jeux de données en utilisant le modèle DCAT";

dct:created "3008202100959"; #en datetime

TCO:fichierDuModèle <[https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument\\_dcat](https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument_dcat)>;

TCO:basedOn <<http://www.w3.org/ns/dcat>>; #iri de l'ontologie

TCO:contains <[https://www.irit.fr/recherches/MELODI/ontologies/TCO#shaclShapeScriptDocument\\_dcat-ap](https://www.irit.fr/recherches/MELODI/ontologies/TCO#shaclShapeScriptDocument_dcat-ap)>.

<<http://www.w3.org/ns/dcat>> TCO:isExtractedFrom <[https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument\\_dcat](https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument_dcat)>.

<[https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument\\_dcat](https://www.irit.fr/recherches/MELODI/ontologies/TCO#MetadataModelDocument_dcat)> a metadataFileModel;

dcat:downloadURL "D:\Semantics4FAIR\Files\_onto\dcat.ttl".

<[https://www.irit.fr/recherches/MELODI/ontologies/TCO#shaclShapeScriptDocument\\_dcat-ap](https://www.irit.fr/recherches/MELODI/ontologies/TCO#shaclShapeScriptDocument_dcat-ap)> a ScriptShaclShapeFile;

dcat:downloadURL "D:\Semantics4FAIR\Files\_onto\dcat-ap.ttl".

## Create a new template

Choose a concept

- http://www.w3.org/ns/dcat#Role
- http://www.w3.org/ns/dcat#Catalog
- http://www.w3.org/ns/dcat#Relationship
- http://www.w3.org/ns/dcat#Resource
- http://www.w3.org/ns/dcat#Distribution
- n50ef6ca2d95f430cb499ba95de3f5b04b32
- http://www.w3.org/ns/dcat#DataService
- http://www.w3.org/ns/dcat#CatalogRecord
- http://www.w3.org/ns/dcat#Dataset

## Create a new template

Choose a concept

les propriétés extraites automatiquement (Script Shacl Shapes + Modèle de métadonnées)



# Conclusion

## → Contributions principales de ce stage :

- la documentation de la plateforme DataNoos (aucune documentation détaillée n'existait auparavant)
- une ontologie qui permet de décrire les formulaires de saisie de métadonnées (ontologie TCO)
- le module DN-Template pour la gestion de la saisie de métadonnées sur la plateforme DataNoos
- la validation de la proposition en utilisant le modèle DCAT et son script SHACL existant

## → A titre personnel :

- ◆ Expérience enrichissante pleine d'apprentissage
- ◆ Mise en pratique des savoirs acquis lors de ma formation en DC
- ◆ Opportunité de travailler sur un projet innovant tel que DataNooS

# Perspectives

- Achever la phase d'implémentation
  - ◆ Générer le template final avec des champs plus compréhensibles pour les experts du domaine (les météorologues)
  - ◆ Terminer la phase de saisie de métadonnées
  
- Exporter les templates définis par l'utilisateur sous un format de script SHACL



# Références

- ★ Mark D Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: Scientific data 3.1 (2016), pp. 1–9.
- ★ Annika Jacobsen et al. “A Generic Workflow for the Data FAIRification Process”. In: Data Intelligence 2.1-2 (2020), pp. 56
- ★ Dan Brickley, Matthew Burgess, and Natasha F. Noy. “Google Dataset Search: Building a search engine for datasets in an open Web ecosystem”. In: The World Wide Web Conference, WWW, San Francisco, CA, USA. ACM, 2019, pp. 1365–1375. url: <https://doi.org/10.1145/3308558.3313685>
- ★ Amina Annane et al. “Un modèle sémantique en vue d’améliorer la FAIRisation des données météorologiques”. In: Journées Francophones d’Ingénierie des Connaissances (IC) Plate-Forme Intelligence Artificielle (PFIA 2021). Ed. by Maxime Lefrançois. Collège SIC (Science de l’Ingénierie des Connaissances) de l’AFIA. Bordeaux, France: AFIA : Association Française pour l’Intelligence Artificielle, June 2021, pp. 20–29. url: <https://hal-emse.ccsd.cnrs.fr/emse-03260061>.

**MERCI !!!**