RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

IGN
INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE

VALIDATION DE DONNÉES PRÉSENTATION QUADOGÉO

30/06/2022

Contexte de la réunion

Lors du temps d'échange du 24 mai sur la cartographie fonctionnelle de la Geoplateforme a été présenté le projet de généralisation du validateur du Géoportail de l'Urbanisme pour la Geoplateforme.

Lors de la dernière réunion de la commission règles et qualité a été évoqué le sujet du rapprochement de la méthodo CNIG avec celle de schema.data.gouv.fr pour formaliser les standards CNIG et notamment faciliter l'agrégation des jeux de données.

La présentation d'aujourd'hui vise à éclairer les travaux réalisés par l'IGN sur le validateur.

Sommaire

1. Le Validateur du Géoportail de l'Urbanisme

2. Le Validateur TRI

3. Le Validateur PCRS

4. L'API

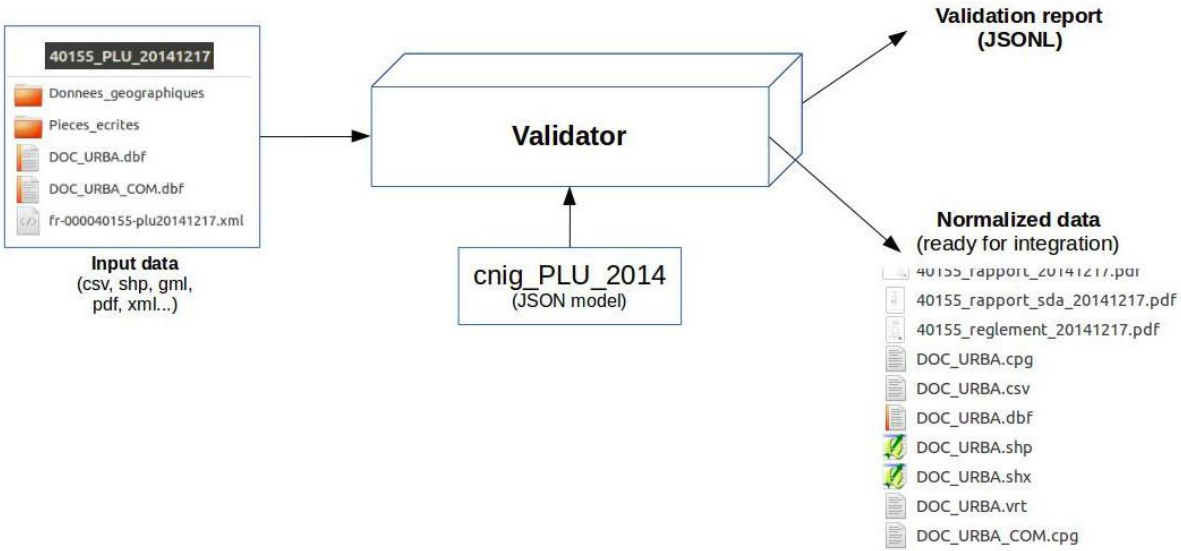
5. Le démonstrateur de l'API

1. Le Validateur du Géoportail de l'Urbanisme

Naissance du validateur

- Développement initié en 2014 pour valider les archives des PLU, POS, CC, PSMV et SUP en fonction des standards CNIG
- Validation en amont de l'**agrégation des données** dans les bases de données **France Entière** du **Géoportail de l'Urbanisme** (GpU)
- Travaux concentrés sur la **modélisation des standards CNIG au format PDF** dans un **méta-modèle** permettant la validation des données (http://cnig.gouv.fr/wp-content/uploads/2019/11/191115_Standard_CNIG_PLU.pdf) :
  - Description des tables « prisonnières » des PDF
  - Arborescence de fichiers à contrôler
  - Contraintes textuelles pas forcément formalisables dans un méta-modèle (ex : règles de nommage fonction du territoire concerné, de la version du standard,…)
- Code sous licence open-source CECILL-B depuis 2016 : https://github.com/IGNF/validator

Principe de validation des documents



=> En sortie du validateur, les données sont prêtes pour être livrées à un entrepôt Géoportail

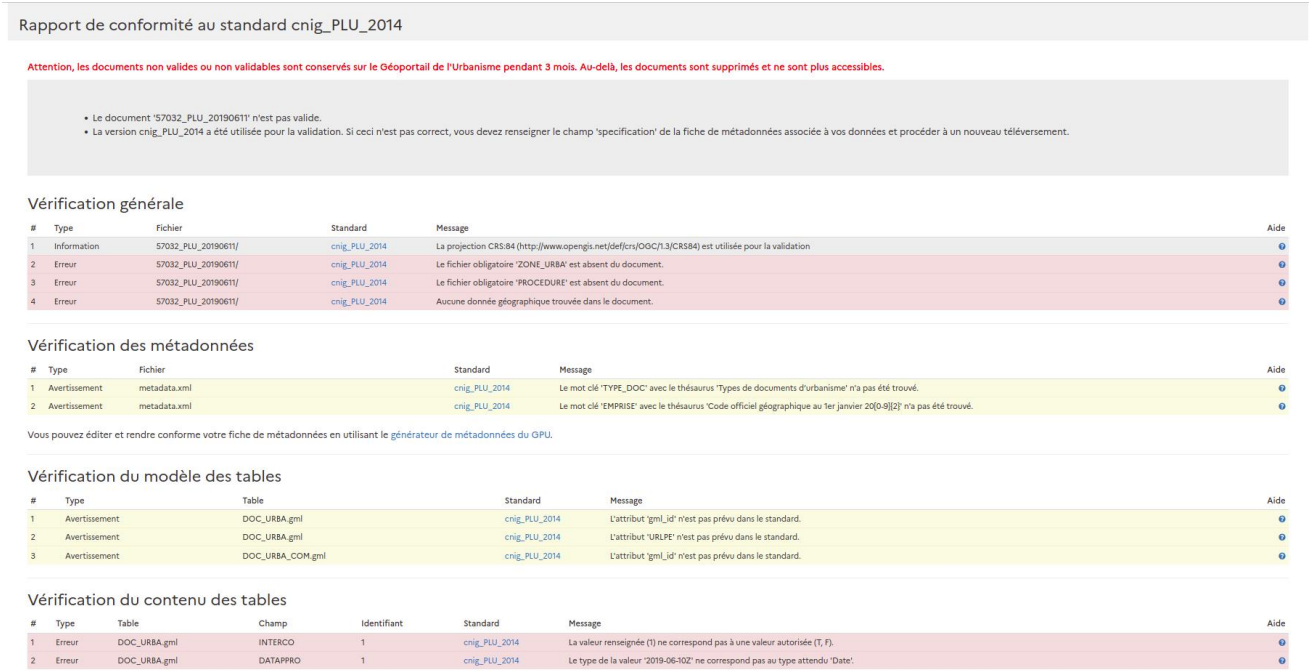
Intégration dans le portail Géoportail de l'Urbanisme

L'appel au validateur est intégré au niveau du backend de l'application GpU comme suit :

1. Un utilisateur téléverse une archive
2. L'archive est validée via un appel au validateur Les données normalisées sont livrées sur un entrepôt de données
3. L'entrepôt de données se charge d'agréger les données et les diffuser sous forme de flux WMS et WFS
4. L'application exploite ces flux pour mettre en œuvre les interfaces de consultation

Présentation des rapports de validation

Des efforts pour présenter les erreurs pour un public hétéroclite :



Rapport de conformité au standard cnig_PLU_2014

Attention, les documents non valides ou non validables sont conservés sur le Géoportail de l'Urbanisme pendant 3 mois. Au-delà, les documents sont supprimés et ne sont plus accessibles.

- Le document '57032_PLU_20190611' n'est pas valide.
- La version cnig_PLU_2014 a été utilisée pour la validation. Si ceci n'est pas correct, vous devez renseigner le champ 'specification' de la fiche de métadonnées associée à vos données et procéder à un nouveau téléversement.

Vérification générale

#	Type	Fichier	Standard	Message	Aide
1	Information	57032_PLU_20190611/	cnig_PLU_2014	La projection CRS:84 (http://www.opengis.net/def/crs/OGC/1.3/CRS84) est utilisée pour la validation	
2	Erreur	57032_PLU_20190611/	cnig_PLU_2014	Le fichier obligatoire 'ZONE_URBA' est absent du document.	
3	Erreur	57032_PLU_20190611/	cnig_PLU_2014	Le fichier obligatoire 'PROCEDURE' est absent du document.	
4	Erreur	57032_PLU_20190611/	cnig_PLU_2014	Aucune donnée géographique trouvée dans le document.	

Vérification des métadonnées

#	Type	Fichier	Standard	Message	Aide
1	Avertissement	metadata.xml	cnig_PLU_2014	Le mot clé 'TYPE_DOC' avec le thésaurus 'Types de documents d'urbanisme' n'a pas été trouvé.	
2	Avertissement	metadata.xml	cnig_PLU_2014	Le mot clé 'EMPRISE' avec le thésaurus 'Code officiel géographique au 1er janvier 20[0-9][2]' n'a pas été trouvé.	

Vous pouvez éditer et rendre conforme votre fiche de métadonnées en utilisant le générateur de métadonnées du GPU.

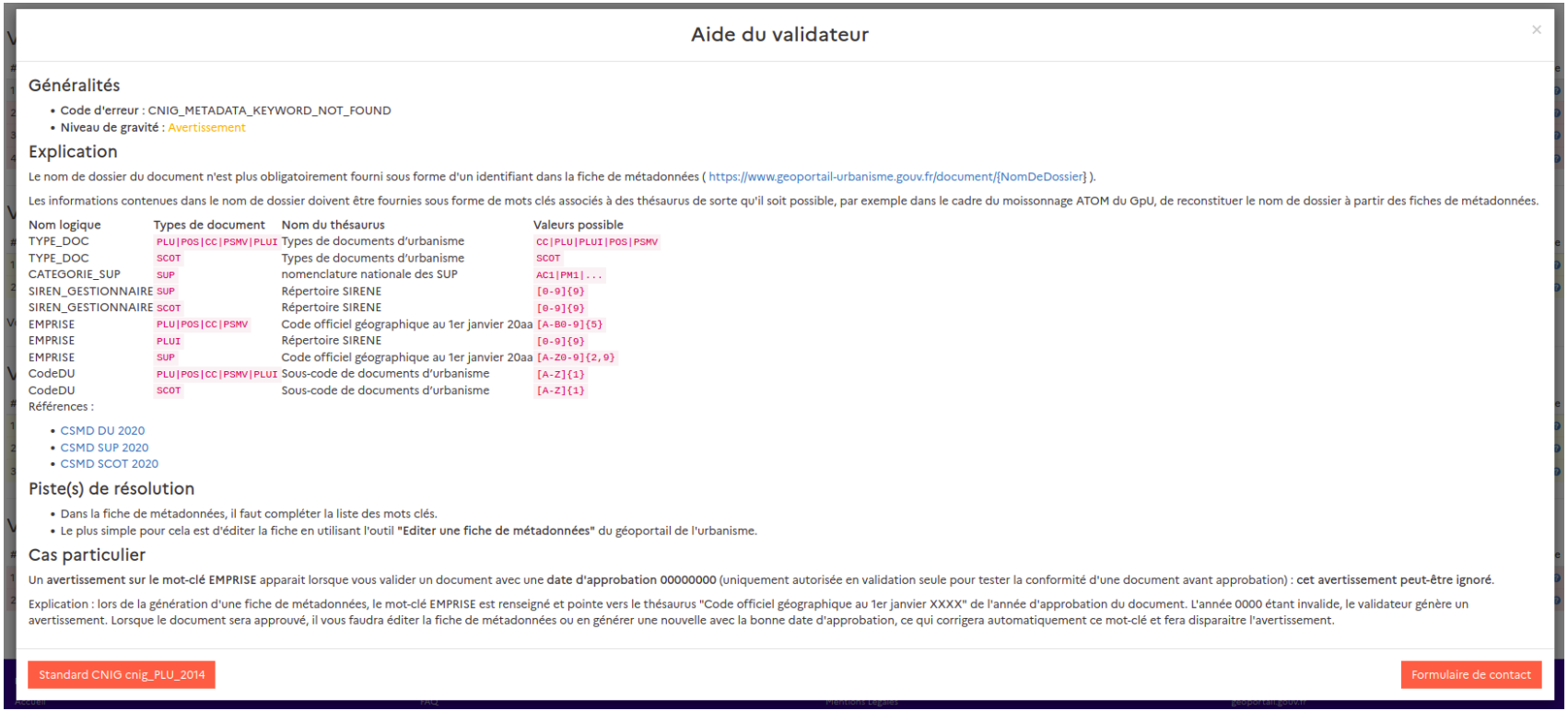
Vérification du modèle des tables

#	Type	Table	Standard	Message	Aide
1	Avertissement	DOC_URBA.gml	cnig_PLU_2014	L'attribut 'gml_id' n'est pas prévu dans le standard.	
2	Avertissement	DOC_URBA.gml	cnig_PLU_2014	L'attribut 'URLPE' n'est pas prévu dans le standard.	
3	Avertissement	DOC_URBA_COM.gml	cnig_PLU_2014	L'attribut 'gml_id' n'est pas prévu dans le standard.	

Vérification du contenu des tables

#	Type	Table	Champ	Identifiant	Standard	Message	Aide
1	Erreur	DOC_URBA.gml	INTERCO	1	cnig_PLU_2014	La valeur renseignée (1) ne correspond pas à une valeur autorisée (T, F).	
2	Erreur	DOC_URBA.gml	DATAPPRO	1	cnig_PLU_2014	Le type de la valeur '2019-06-10Z' ne correspond pas au type attendu 'Date'.	

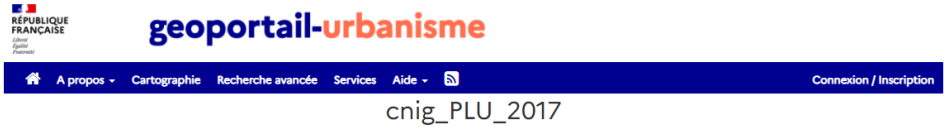
Une aide associée aux codes d'erreurs



Modélisation des documents

Modélisation de ce qui pouvait être formalisé :

- **Modélisation de l'arborescence des fichiers** : https://www.geoportail-urbanisme.gouv.fr/standard/cnig_PLU_2017#files / https://www.geoportail-urbanisme.gouv.fr/standard/cnig_PLU_2017.json
- **Modélisation de la structure des fichiers type "table"** : https://www.geoportail-urbanisme.gouv.fr/standard/cnig_PLU_2017#table-ZONE_URBA / https://www.geoportail-urbanisme.gouv.fr/standard/cnig_PLU_2017.json



cnig_PLU_2017

Organisation des fichiers

Nom	Type	Obligatoire?	Chemin
Donnees_geographiques	directory	ERROR	Donnees_geographiques
HABILLAGE_LIN	table	OPTIONAL	Donnees_geographiques/<INSEE_COMMUNE>_HABILLAGE_LIN_<YYYYMMDD>
HABILLAGE_PCT	table	OPTIONAL	Donnees_geographiques/<INSEE_COMMUNE>_HABILLAGE_PCT_<YYYYMMDD>
HABILLAGE_SURF	table	OPTIONAL	Donnees_geographiques/<INSEE_COMMUNE>_HABILLAGE_SURF_<YYYYMMDD>
HABILLAGE_TXT	table	OPTIONAL	Donnees_geographiques/<INSEE_COMMUNE>_HABILLAGE_TXT_<YYYYMMDD>

ZONE_URBA

Nom	Description	Type	Taille	Valeur requise?	Valeurs possibles	Expression régulière
LIB_IDZONE	identifiant d'objets	String	40	NON		
WKT	Geométrie associée	MultiPolygon		OUI		
URLFIC	URL ou URI qui pointe sur le fichier de règlement de la zone ou à défaut du règlement intégral indexé	Url	254	NON		
LIBELLE	Nom court de la zone tel qu'il apparaît sur le plan de zonage	String	12	OUI		
IDURBA	Identifiant du document d'urbanisme	String	30	OUI		
NOMFIC	Nom du fichier du règlement de la zone	Filename	80	OUI		
DATVALID	Date de la dernière validation de la zone ou de son règlement	Date		NON		
TYPEZONE	Type de la zone dans une nomenclature simplifiée	String	3	OUI	U AUc AUs A N	
LIBELONG	Nom complet littéral de la zone tel qu'il apparaît dans le chapitre du règlement écrit	String	254	NON		

Plugin CNIG du validateur IGN

Ce qui ne pouvait être formalisé sans faire un modèle « usine à gaz » a été implémenté dans un **plugin CNIG pouvant être activé à la demande** (--plugins cnig) :

- Contrôles **d'unicité** et de **références** entre **plusieurs tables** GENERATEUR_SUP_**P/L/S** et ASSIETTE_SUP_**P/L/S** où :
  - Les valeurs "IDGEN" doivent être uniques dans les tables `GENERATEUR_SUP_P/L/S`
  - Les valeurs "IDASS" doivent être uniques dans les tables `ASSIETTE_SUP_P/L/S`
  - Les valeurs référencées par "IDGEN" dans `ASSIETTE_SUP_P/L/S` doivent exister dans l'une des tables `GENERATEUR_SUP_P/L/S`
- Contrôles avancés sur les **formats du champ "IDURBA"** dans le cas des PLU
- Contrôles spécifiques au type de document sur les **fiches de métadonnées** (c.f. Consigne de saisie des métadonnées des DU)
- …

2. Le Validateur TRI (Territoires à Risque important d'Inondation)

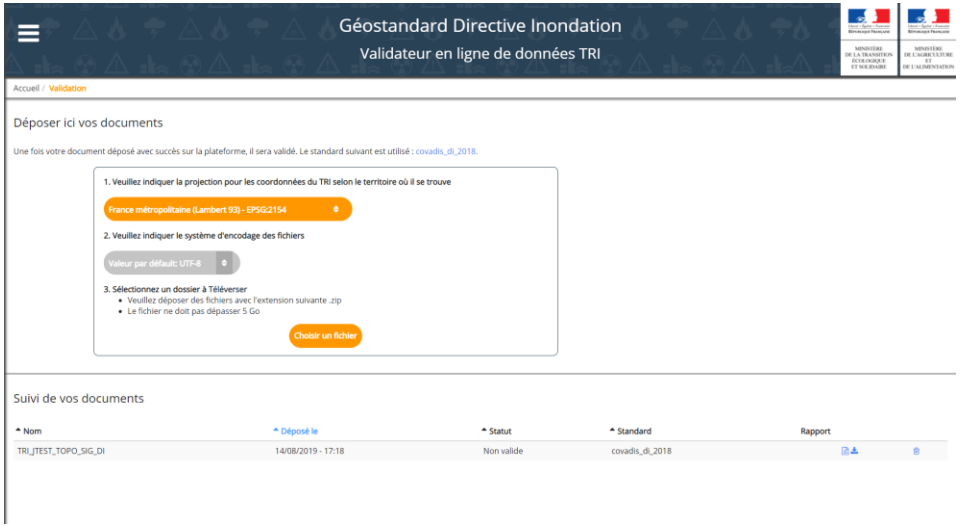
Première réutilisation du validateur GpU

- **Réutilisation du cœur de validation**
- Création d'un **plugin dédié pour les contrôles spécifiques** à ce nouveau modèle :
  - **Contrôles topologiques** (table devant former une partition de l'espace)
  - Contrôle de relation entre table

A l'origine d'améliorations qui ont profité au GpU (contrôle d'unicité, contrôle de référence, concept de base de validation)

Nouveau portail de validation…

- Création d'une application dédiée https://validateur-tri.ign.fr
- Même mode d'intégration du validateur que GpU :
  - avec **export GeoJSON des rapports de validation**
  - sans **fiches d'aide associées aux codes d'erreurs** (manuel PDF pour l'aide)



Les erreurs relatives au contenu des tables sont les suivantes :

Gravité	Message	Explication
ERROR	Une valeur vide n'est pas autorisée pour ce champ.	Cette erreur se produit lorsqu'un champ n'autorisant pas les valeurs nulles contient une valeur vide
ERROR	Le type de la valeur '{VALUE}' ne correspond pas au type attendu '{EXPECTED_TYPE}'.	Cette erreur se produit lorsqu'un champ contient une valeur de type non autorisé par le modèle
ERROR	La valeur ({VALUE}) ne correspond pas à l'expression régulière ({REGEXP}).	Cette erreur se produit lorsqu'un champ contient une chaine de caractère qui ne correspond pas à la règle d'écriture du modèle.

3. Le Validateur PCRS (Plan Corps de Rue Simplifié)

Validation d'un modèle GML complexe

Contexte : Volonté des producteurs PCRS d'avoir un validateur de référence

- Les **données vectorielles du PCRS** sont livrées en **GML** (contrairement au cas du GpU où elles sont livrées au format Shapefile, MapInfo ou GML)
- Les données GML du PCRS sont déjà modélisées à l'aide d'un fichier XSD au niveau du standard CNIG
- Un fichier GML PCRS contient plusieurs collections d'objets de types différents
- Les fournisseurs de données PCRS sont à priori à même d'interpréter des erreurs de **validation XSD**

Évolutions du validateur pour PCRS

Nouvelles améliorations réutilisables :

- **Validation des fichiers XML/GML en fonction d'un schéma XSD**
- **Concept de fichier « multi-table »** pour support des GML contenant plusieurs collections (exploitable au besoin pour GeoPackage)

(Voir modèle document.json résultant)

Un plugin PCRS à prévoir…

Les premières discussions montrent que là encore, **il ne sera pas possible de tout formaliser** et donc de tout intégrer dans le cœur du validateur :

- Contrôle d'unicité de l'emprise PCRS
- Contrôle d'unicité des identifiants sur des tables point, ligne, polygone
- …

La volonté de ne pas créer un n-ième portail...

=> Participation aux travaux de développement d'une API WEB de validation pour l'espace collaboratif.

4. L’API

Pourquoi l'APIsation du validateur

- Pour simplifier l'utilisation du validateur sur les différents sites qui en auront besoin (notamment sur l'espace collaboratif)
- Pour avoir une API WEB complémentaire à l'API en ligne de commande.
- Parce que dans le cadre de la Géoplateforme, les fonctionnalités sont prévues API-First et que le validateur est une des briques essentielles

  => Réorientation générale des développements à l'IGN vers des architectures API-First et microservices.
- Pour déployer un démonstrateur générique

Une volonté de s'aligner au maximum sur l'existant

Une API simple inspirée de validata.

Des modèles de table alignés au maximum sur TableSchema

- Focalisation sur les **modèles JSON** / arrêt des modèles XML
- Ajout d'une propriété "*constraints*", renommage de "*limit*" en "*maxLength*", …
- Un alignement incomplet :
  - Des difficultés pour s'aligner sur les **types de base très orientés JSON de TableSchema**
  - Concept de Point, LineString,… VS geopoint/geojson
  - Concept manquant pour les "noms de fichiers"
  - …

Type	`Point`	`geopoint`	Equivalence approximative
Type	`LineString`	?	Restriction de type possible sur `geojson` ?
Type	`Polygon`	?	Restriction de type possible sur `geojson` ?
Type	`MultiPoint`	?	Restriction de type possible sur `geojson` ?
Type	`MultiLineString`	?	Restriction de type possible sur `geojson` ?
Type	`MultiPolygon`	?	Restriction de type possible sur `geojson` ?
Type	`GeometryCollection`	?	Restriction de type possible sur `geojson` ?
Type	`Geometry`	`geojson`	Equivalence approximative

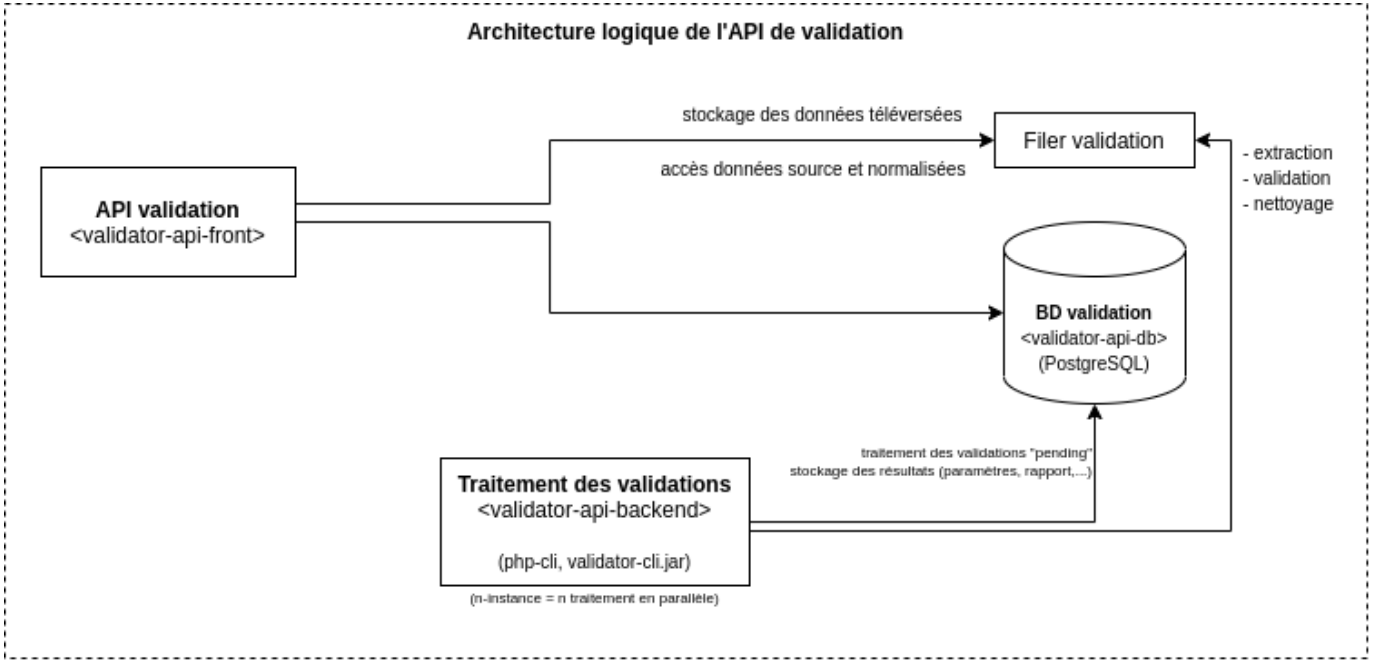
- Une **volonté de conserver des syntaxes simples (sql-like)** pour la définition des contraintes du type **référence**, **clé étrangère**,… (la gestion des modèles est plus chronophage que les développements du validateur)

Des modèles accessibles par URL

Inspiré par validata/schema.gouv.fr :


```
{
    "uid": "007544rj19blz74f9053dvv1",
    "dataset_name": "CARENE-20210510-3947",
    "arguments": {
        "srs": "EPSG:3947",
        "model": "https:\/\/ignf.github.io\/validator\/validator-core\/src\/test\/resources\/config-json\/CNIG_PCRS_v2.0\/document.json",
        "max-errors": 30,
        "normalize": true,
        "encoding": "UTF-8"
    },
    "date_creation": "2021-11-15T15:52:05+01:00",
    "status": "finished",
    "message": null,
    "date_start": "2021-11-15T15:52:14+01:00",
    "date_finish": "2021-11-15T15:54:41+01:00",
    "results": [
        {
            "code": "VALIDATOR_INFO",
            "scope": "DIRECTORY",
            "level": "INFO",
```

Une architecture simple



5. Le démonstrateur de l'API

Intégration d'un client par défaut dans l'API

Tout le monde n'étant pas à l'aise avec curl ou swagger, un client a été développé pour les besoins de démonstration PCRS :

- 1) Dépôt d'un jeu de données à valider :

Intégration d'un client par défaut dans l'API

- 2) Attendre quelques minutes

Validation de GeoVendee-Montaigu_20220203-3947

Statut : En attente

- **Modèle :** https://ignf.github.io/validator/validator-core/src/test/resources/config-json/CNIG_PCRS_v2.0/document.json
- **Projection :** EPSG:3948

Intégration d'un client par défaut dans l'API

- 3) Visualiser les erreurs

Statut : Terminée

- **Modèle :** https://ignf.github.io/validator/validator-core/src/test/resources/config-json/CNIG_PCRS_v2.0/document.j
- **Projection :** EPSG:3948

Rapport de validation

Fichier	Code	Message
	VALIDATOR_INFO	Validation avec le modèle : CNIG_PCRS_v2.0
	VALIDATOR_INFO	Version GDAL utilisée : GDAL 3.2.2, released 2021/03/05
GeoVendee-Montaigu_202202...	FILE_UNEXPECTED	Le fichier 'GeoVendee-Montaigu_20220203-3947' est mal nommé ou n'est pas prévu dans le standard.
GeoVendee-Montaigu_202202...	VALIDATOR_PROJECTION_INFO	La projection EPSG:3948 (http://www.opengis.net/def/crs/EPSG/0/3948) est utilisée pour la validation
GeoVendee-Montaigu_202202...	XSD_SCHEMA_ERROR	cvc-enumeration-valid : La valeur '' n'est pas un facet valide par rapport à l'énumération '[002, 005, 010, 040, 150, 999]'. Il doit s'agir d'une valeur provenant de l'énumération.
GeoVendee-Montaigu_202202...	XSD_SCHEMA_ERROR	cvc-type.3.1.3 : La valeur '' de l'élément 'precisionAltimetrique' n'est pas valide.

- **code :** XSD_SCHEMA_ERROR
- **scope :** DIRECTORY
- **level :** ERROR
- **message :** cvc-enumeration-valid : La valeur '' n'est pas un facet valide par rapport à l'énumération '[002, 005, 010, 040, 150, 999]'. Il doit s'agir d'une valeur provenant de l'énumération.
- **documentModel :** CNIG_PCRS_v2.0
- **fileModel :** DONNEES
- **file :** GeoVendee-Montaigu_20220203-3947/GeoVendee_PCRS_Montaigu_export_20220203.gml
- **id :** 24810

Limites de l'API

- Pas de téléversement par morceaux
- Articulation avec entrepôt Geoplateforme à exploiter à terme (livraison & traitement appelant le validateur)
- Déploiement à industrialiser

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

IGN
INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE

MERCI DE VOTRE ATTENTION