

VALIDATION DE DONNÉES ÉCHANGE TECHNIQUE ETALAB-IGN

24/11/2021

Contexte de la réunion

Lors de la réunion du GT QUADOGEO du 7 octobre 2021, M. Geoffrey ALDEBERT a présenté la mission de l'équipe data.gouv.fr d'Etalab en termes :

- d'accompagnement et d'animation des communautés de producteurs de schéma et de données
- de développement et de maintien en condition opérationnelle d'outils permettant l'amélioration de la qualité des données publiées en open data.

Suite à cette présentation, l'IGN a proposé de partager avec l'équipe d'Etalab sur leurs expériences réciproques concernant les validateurs.

Sommaire

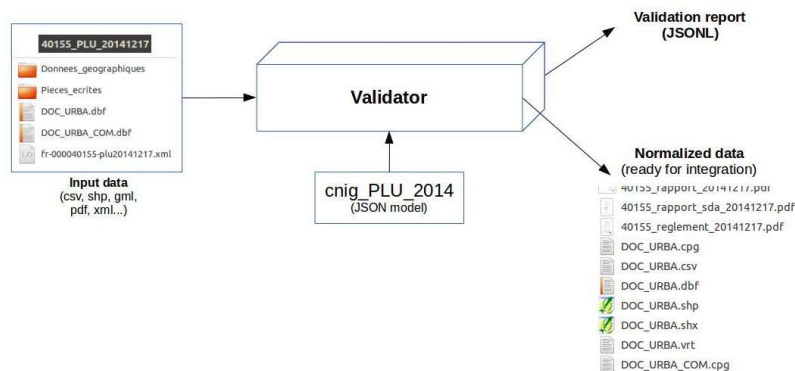
1. Le Validateur du Géoportail de l'Urbanisme
2. Le Validateur TRI
3. Le Validateur PCRS
4. L'API

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 formalisable 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

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	URI	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 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 :

- Contrôle du nombre de lignes présentes dans les tables en fonction des types de document (ex : plusieurs lignes dans la table DOC_URBA_COM dans le cas d'un PLUi / une seule dans le cas d'un PLU)
- Contrôles d'unicité et de référence 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** (ex : CSMD DU)

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
3. Les données normalisées sont livrées sur un entrepôt de données
4. L'entrepôt de données se charge d'agréger les données et les diffuser sous forme de flux WMS et WFS
5. L'application exploite ces flux pour mettre en œuvre les interface de consultation

(Note : Pas d'API-first / focalisation des efforts sur les téléversements par morceaux vu la présence d'archives de ~5Go)

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 S7032_PLU_20190617 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	S7032_PLU_20190617	cnig_PLU_2014	La projection CRS84 (http://www.opengis.net/def/ur/OGC/1.3/CRS84) est utilisée pour la validation.	Aide
2	Error	S7032_PLU_20190617	cnig_PLU_2014	Le fichier obligatoire S7032_URBA est absent du document.	Aide
3	Error	S7032_PLU_20190617	cnig_PLU_2014	Le fichier obligatoire S7032_URBA est absent du document.	Aide
4	Error	S7032_PLU_20190617	cnig_PLU_2014	Aucune donnée géographique trouvée dans le document.	Aide

Vérification des métadonnées

#	Type	Fichier	Standard	Message	Aide
1	Avertissement	metadatas.xml	cnig_PLU_2014	Le mot clé 'TYPE_DOC' avec le thésaurus 'Types de documents d'urbanisme' n'a pas été trouvé.	Aide
2	Avertissement	metadatas.xml	cnig_PLU_2014	Le mot clé 'SMPRIS' avec le thésaurus 'Code officiel géographique' au 1er janvier 2019-01-01 n'a pas été trouvé.	Aide

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.	Aide
2	Avertissement	DOC_URBA_gml	cnig_PLU_2014	L'attribut 'SMPRIS' n'est pas prévu dans le standard.	Aide
3	Avertissement	DOC_URBA_COP_gml	cnig_PLU_2014	L'attribut 'gml_id' n'est pas prévu dans le standard.	Aide

Vérification du contenu des tables

#	Type	Table	Champ	Identifiant	Standard	Message	Aide
1	Error	DOC_URBA_gml	INTENCO	1	cnig_PLU_2014	La valeur renseignée (1) ne correspond pas à une valeur autorisée (1, 2).	Aide
2	Error	DOC_URBA_gml	DATAPRO	1	cnig_PLU_2014	La type de la valeur '2019-06-10' ne correspond pas au type attendu 'Date'.	Aide

Une aide associée aux codes d'erreurs

Aide du validateur

Généralités

- Code d'erreur : CNIO_METADATA_KEYWORD_NOT_FOUND
- Niveau de gravité : **Avertissement**

Explication

Le nom de dossier du document n'est plus obligatoirement fourni sous forme d'un identifiant dans la fiche de métadonnées (<https://www.geoportail-urbanisme.gouv.fr/document/{NomDeDossier}>).

Les informations contenues dans le nom de dossier doivent être fournies sous forme de mots clés associés à des thésaurus de sorte qu'il soit possible, par exemple dans le cadre du moissonnage ATOM du Opu, de reconstituer le nom de dossier à partir des fiches de métadonnées.

Nom logique	Types de document	Nom du thésaurus	Valeurs possible
TYPE_DOC	PLU (POS) (CC) (PSRV) (PLUS)	Types de documents d'urbanisme	CC (PLUS) (PLUS) (PSRV)
TYPE_DOC	SCOT	Types de documents d'urbanisme	SCOT
CATEGORIE_SUP	SUP	Nomenclature nationale des SUP	AC1 (PM1) ...
SIREN_QUESTIONNAIRE_SUP	SUP	Répertoire SIRENE	(0-9) (0)
SIREN_QUESTIONNAIRE_SCOT	SCOT	Répertoire SIRENE	(0-9) (0)
EMPRISE	PLU (POS) (CC) (PSRV)	Code officiel géographique au 1er janvier 20aa	(A-00-0) (S)
EMPRISE	PLUS	Répertoire SIRENE	(0-9) (0)
EMPRISE	SUP	Code officiel géographique au 1er janvier 20aa	(A-20-0) (2, 0)
CodeDU	PLU (POS) (CC) (PSRV) (PLUS)	Sous-code de documents d'urbanisme	(A-2) (S)
CodeDU	SCOT	Sous-code de documents d'urbanisme	(A-2) (S)

Références :

- CSMD DU 2020
- CSMD SUP 2020
- CSMD SCOT 2020

Piste(s) de résolution

- Dans la fiche de métadonnées, il faut compléter la liste des mots clés.
- Le plus simple pour cela est d'éditer la fiche en utilisant l'outil "Editer une fiche de métadonnées" du géoportail de l'urbanisme.

Cas particulier

Un avertissement sur le mot-clé **EMPRISE** apparaît lorsque vous validez un document avec une date d'approbation 00000000 (uniquement autorisée en validation seule pour tester la conformité d'un document avant approbation) : cet avertissement peut-être ignoré.

Explication : lors de la génération d'une fiche de métadonnées, le mot-clé **EMPRISE** est renseigné et pointe vers le thésaurus "Code officiel géographique au 1er janvier XXXX" de l'année d'approbation du document. L'année 0000 étant invalide, le validateur génère un avertissement. Lorsque le document sera approuvé, il vous faudra éditer la fiche de métadonnées ou en générer une nouvelle avec la bonne date d'approbation, ce qui corrigera automatiquement ce mot-clé et fera disparaître l'avertissement.

Standard CNIS orig. PLU_2014

Formulaire de contact

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

Nouveau cas d'usage pour le validateur IGN

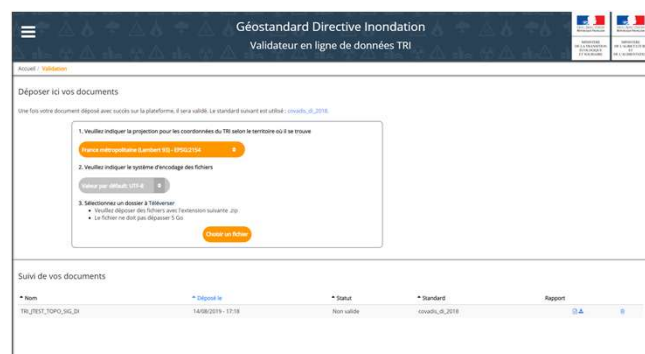
- Réutilisation du cœur de validation développé pour le GpU
- Création d'un plugin dédié pour les contrôles spécifiques au modèle :
 - Contrôle topologique (table devant former une partition de l'espace)
 - Contrôle de relation entre table

Génération du concept de base de données de validation avec :

- Chargement des données à valider dans une base SQLITE ou PostGIS
- Formalisation des **contrôles d'unicité**
- Formalisation des **contrôles de référence** (clé étrangère simple)
- Mise en œuvre au passage d'un cadre simple pour des contrôles métiers SQL (comptages, relations attributaires,...)

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 **fiche d'aide associées aux codes d'erreurs** (manuel PDF pour l'aide)



Géostandard Directive Inondation
Validateur en ligne de données TRI

Accueil - Validation

Déposer ici vos documents

Une fois votre document déposé avec succès sur la plateforme, il sera validé. Le standard concerné est celui-ci : [covid19_0_2019](#)

1. Veuillez indiquer la projection pour les coordonnées du TRI selon le territoire où il se trouve.
2. Veuillez indiquer le système d'encodage des fichiers.
3. Sélectionnez un dossier à télécharger.
 • Veuillez déposer des fichiers avec l'extension suivante : zip
 • Le fichier ne doit pas dépasser 5 Go

Suivi de vos documents

Nom	Déposé le	Statut	Standard	Rapport
TRI_0101_TOPO_SGL_ZIP	14/08/2019 - 17:18	Non validé	covid19_0_2019	D R

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 chaîne 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

Nous avons donc procédé aux évolutions suivantes au niveau du validateur IGN pour être en mesure de valider les données PCRS :

- Ajout de la **validation des fichiers XML/GML en fonction d'un schéma XSD** avec un code d'erreur dédié :
 - Intégration du code de validation fourni par un producteur PCRS
 - Utilisation du driver GMLAS de GDAL pour la lecture des données
- **Ajout du concept de fichier « multi-table »** pour support des GML contenant plusieurs collections

Le modèle résultant pour le validateur IGN est le suivant :

https://ignf.github.io/validator/validator-core/src/test/resources/config-json/CNIG_PCRS_v2.0/document.json

Des évolutions qui ne concernent pas que PCRS...

- La validation XSD peut être utilisée sur les fiches de métadonnées
- Le concept de fichier multi-table s'appliquerait pour valider des fichiers au format GeoPackage

Volonté de ne pas créer un n-ième portail

Pour l'heure, les tests sont réalisés via la fourniture par l'IGN de rapport au format Excel produit par l'IGN :

A	B	C	
code	scope	level	message
VALIDATOR_INFO	DIRECTORY	INFO	Validation avec le modèle : CNIG_PCRS_v2.0
VALIDATOR_INFO	DIRECTORY	INFO	Version GDAL utilisée : GDAL 2.4.0, released 2018/12/14
FILE_UNEXPECTED	DIRECTORY	WARNING	Le fichier 'CARENE-20210510-3947' est mal nommé ou n'est pas prévu dans le standard.
VALIDATOR_PROJECTION_INFO	DIRECTORY	INFO	La projection EPSG:3947 (http://www.opengis.net/def/crs/EPSG/0/3947) est utilisée pour la validation
XSD_SCHEMA_ERROR	DIRECTORY	ERROR	cvc-datatype-valid.1.2.3 : '00' n'est pas une valeur valide du type d'union 'NatureAffleurantPCRSTypeType'.
XSD_SCHEMA_ERROR	DIRECTORY	ERROR	cvc-type.3.1.3 : La valeur '00' de l'élément 'pcrs:nature' n'est pas valide.
MULTITABLE_UNEXPECTED	DIRECTORY	WARNING	La table 'AffleurantGeometriquePCRS_enveloppe' n'est pas prévue dans le modèle de validation.
MULTITABLE_UNEXPECTED	DIRECTORY	WARNING	La table 'AffleurantGeometriquePCRS_ligne' n'est pas prévue dans le modèle de validation.
MULTITABLE_UNEXPECTED	DIRECTORY	WARNING	La table 'AffleurantGeometriquePCRS_point' n'est pas prévue dans le modèle de validation.
ATTRIBUTE_GEOMETRY_INVALID	FEATURE	ERROR	La géométrie de l'objet n'est pas topologiquement correcte. Un contour intérieur (trou) ou extérieur s'auto-intersecte

L'équipe en charge des travaux sur la validation PCRS participe aux développements de l'API de validation afin d'éviter d'avoir à maintenir un nouveau portail.

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
- ...

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)
- 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

Spécifications

Une API simple inspirée de validata

API Validator 1.0.0 GA55

/api/validator-api.yml

API permettant d'appeler IGNF/validator pour valider et normaliser des dossiers contenant une arborescence de fichiers.

Contact IGNF

CECILL-B

validation Une demande de validation de documents

POST	/api/validations/	Téléverser un jeu de données à faire valider	⌵
GET	/api/validations/{uid}	Voir une validation	⌵
PATCH	/api/validations/{uid}	Préciser les arguments de validator	⌵
DELETE	/api/validations/{uid}	Supprimer une validation	⌵
GET	/api/validations/{uid}/files/source	Télécharger les données de source	⌵
GET	/api/validations/{uid}/files/normalized	Télécharger les données normalisées	⌵

Spécifications

Des modèles décrits par des URLs (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",
    }
  ]
}
```

Alignement des modèles sur TableSchema

- Focalisation sur les modèles JSON / dépréciation des modèles XML
- Alignement au maximum de la modélisation des tables avec TableSchema :
 - Ajout propriété "constraints", renommage "limit" => "maxLength", ...
 - Voir document DRAFT-interne-ign_diff-table-schema.pdf de juillet 2019
- Un alignement incomplet :
 - Des difficultés pour s'aligner sur les types de base très orienté JSON de TableSchema (concept de "Path", Point, LineString,... VS Geometry)
 - Une volonté de conserver des syntaxes simples (sql-like) pour la définition des contraintes du type référence, clé étrangère,...

Intégration d'un client par défaut

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

IGNFidemo-validator

Validation d'une archive

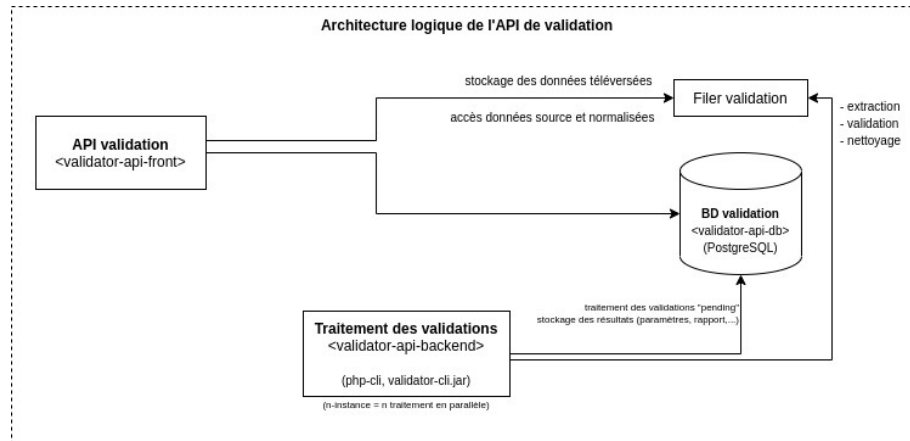
Modèle de données	CNG_PCRS_v2.0	
Projection des données	EPSG:3947 - CC47 - Conique Conforme Zone 6	
Archive		Browse

Valider

Propriété	Valeur
code	XSD_SCHEMA_ERROR
scope	DIRECTORY
level	ERROR
message	cvc-complex-type.2.4.a : Contenu non valide trouvé à partir de l'élément "[http://www.safes.com/gn:time]NumeroVairePCRS": L'une des valeurs "[http://www.opengis.net/gn:3.2]AbstractFeature" est attendue.
documentModel	CMG_PCRS_v2.0
fileModel	DONNEES
file	ENEDIS-GeoVendee_Fougeres-3947-20210608/GeoVendee_Fougeres.gml
id	46

[illegible]

Architecture simplifiée de l'API



Limites de l'approche API

- Pas de téléversement par morceaux
=> pas de traitement de fichiers aussi volumineux que sur le GpU

Reste à faire

- Déploiement en externe du démonstrateur (reste du travail sur l'éditorial et difficultés pour trouver un environnement de déploiement)
- Passage du dépôt en open-source public (reste du travail sur la documentation, la licence...)

MERCI DE VOTRE ATTENTION