

Qualification des jeux de données du catalogue PIGMA

Contrôle qualité, automatisation et labellisation



Contexte du stage

Besoins et objectifs

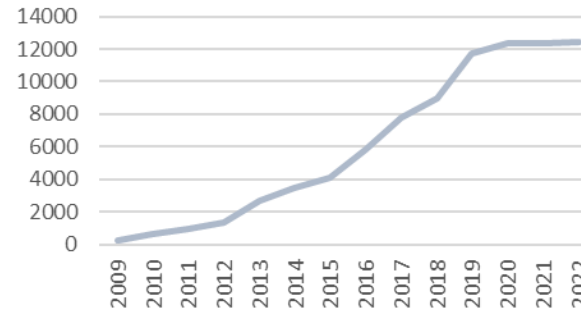
Réalisations

Bilan et perspectives



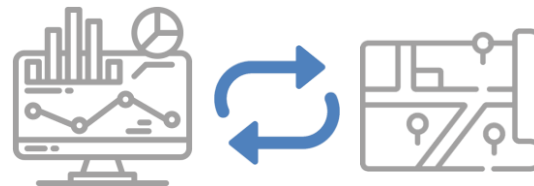
La Qualité des données : Quels enjeux ?

- Des données toujours plus nombreuses
 - + 600 contributeurs
 - + 11 500 jeux de données



- Alerter l'utilisateur sur la qualité des données

- Favoriser leur réutilisation



- Sensibiliser les producteurs à la qualité de leurs données





Problématique : Comment améliorer la réutilisation des jeux de données de la plateforme PIGMA ?



Datavisualisation

- Graphiques, chiffres clés
- Favoriser la « Découvrabilité »
- Permettre une meilleure connaissance des jeux de données



Qualité

- Déterminer la qualité
- Valoriser les JDD de qualité

De la qualité de la donnée ...

... à sa réutilisation et valorisation



Travaux antérieurs

2017 : Fiches techniques du Cerema > Synthèse de la norme ISO 19157

Déc. 2018 : Création du GT CNIG QuaDoGéo

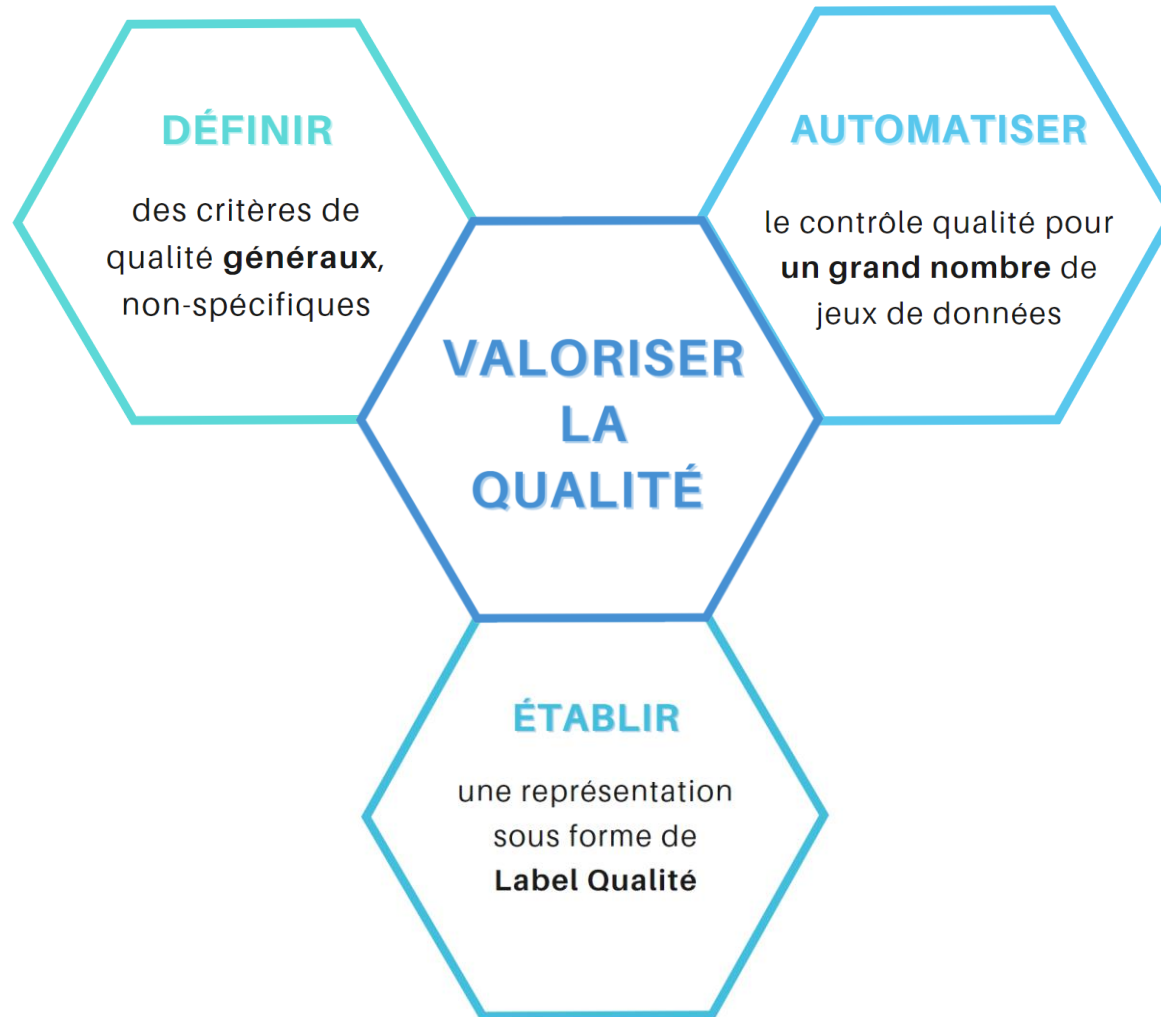
Mars 2022 : Stage CRIGE PACA
> Chaîne de contrôle qualité autonome sur une donnée

2022 : PIGMA > Travaux de qualification
> Critères de qualité, méthodes de travail, Label Qualité

2023 : Poursuite du travail de qualification avec automatisation



Objectifs de la qualification des données





Références

Qualité des métadonnées

SCORE DE MÉTADONNÉES



data.gouv.fr

Critère	Description
Description des données	La description des données est de qualité (la description du jeu de données suffisamment longue).
Mise à jour	- La fréquence de mise à jour est renseignée. - La fréquence de mise à jour est respectée.
Licence	- La licence est renseignée. - La licence est ouverte. Voir la page licence pour plus d'information .
Métadonnées des ressources	Présence d'au moins une ressource avec un format ouvert déclaré.
Couverture spatiale	- La couverture spatiale est renseignée. - La granularité spatiale est renseignée.
Couverture temporelle	La couverture temporelle des données est renseignée.

Qualité des données géo

FICHES TECHNIQUES



Les fiches « critères »



Cohérence logique



Exhaustivité



Précision thématique



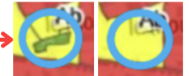
Précision de position



Qualité temporelle

Défauts concernés :

~~id~~



~~29 février 2007
25/13/2014~~



Travaux antérieurs du GIP ATGeRi

Vérification manuelle

métadonnée

CRITERES/ COUCHES		PLAGES	ES	EMS	DU	OCS
Description des données		1	1	1	1	1
Mise à jour (ressource)	fréquence mise à jour renseignée	1	1	1	1	1
	fréquence mise à jour respectée	1	1	1	1	1
Licence	licence renseignée	1	1	1	1	1
	licence ouverte	1	1	1	1	1
Métadonnées des ressources		1	1	1	1	1
Couverture spatiale	granularité spatiale renseignée	1	1	1	1	1

métadonnée



cohérence logique



précision position



qualité temporelle

Vérification par un script FME, pour chaque donnée

donnée

CRITERES/ COUCHES			PLAGES	ES	EMS	DU	OCS
COHERENCE CONCEPTUELLE	identifiant unique	nb total d'objets	184	605	2182	149128	1199119
		présence champ ID	1	1	1	1	0
		nb O avec ID doublon	0	0	0	143492	0
		nb O avec ID unique	184	605	2182	5683	0
		% doublon	0	0	0	96	0
	cohérence code géo/ zone géo	présence champ code géo	1	1	0	1	0
		nb O avec code géo différent	0	5	0	5002	0
		nb O avec code géo correct	0	600	0	144126	0
	géométrie correcte	% nb O avec code géo différent	0	0,8	0	3,4	0
		type géométrie	ponctuelle	ponctuelle	ponctuelle	surfacique	surfacique
		nb O avec géométrie non valide (NULL)	0	0	3	0	0
		nb O avec géométrie valide	184	605	2179	149128	1199119
TOT % géométrie valide		100	100	99,9	100	100	
TOT % géométrie non valide	0	0	0,1	0	0		
COHERENCE DE FORMAT	format fichier respecté		1	1	1	1	1
	encodage respecté		1	1	1	1	1
	projection respectée		1	1	1	1	1
	nom champ bien écrit		1	0	0	1	1
	format attribut respecté		1	1	1	1	1
COHERENCE TOPOLOGIQUE	TOT % superposition (ponctuel)		0	36	24,3	0	0
	TOT % auto-chevauchement		0	0	0	0	0
	TOT % auto-intersection		0	0	0	0,3	0

CRITERES/ COUCHES	PLAGES	ES	EMS	DU	OCS
donnée géo	1	1	1	1	1
mention échelle d'utilisation définie	0	0	0	1	1

CRITERES/ COUCHES	PLAGES	ES	EMS	DU	OCS
nb total d'objets	184	605	2182	149128	1199119
infos temporelles	1	0	0	1	0
nb O respectant le format date	174	0	0	148139	0
TOT % de non respect du format date	5,4	0	0	0	0
TOT % de non conformité chronologique	0	0	0	1,1	0
TOT % de non validité temporelle (2022-13-01)	0	0	0	0	0



CALCUL NOTE FINALE



Travaux antérieurs du GIP ATGeRi

Premiers critères retenus

- Qualité des métadonnées → Score de Data.Gouv
- Qualité de la donnée géographique → Critères Norme ISO 19157

Méthode de qualification de la donnée

- Script FME à exécuter sur chaque donnée
- Priorisation des jeux de données étudiés durant le stage sur **530 jeux de données**

Représentation sous forme de Label

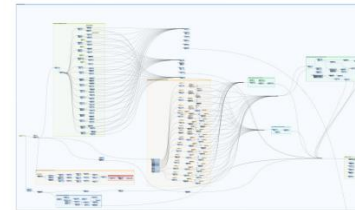
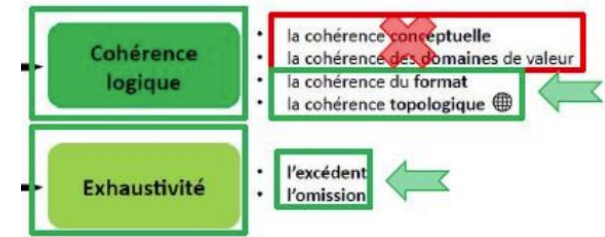
- **Label Qualité** qui valorise les jeux de données de qualité



Travaux du CRIGE PACA

Processus de qualification automatique

- Qualité de la donnée
- Analyse de 2 critères et 4 sous-critères
- Donnée au format Shapefile
- Donnée sur le territoire CRIGE PACA
- Production d'un rapport d'analyse au format HTML



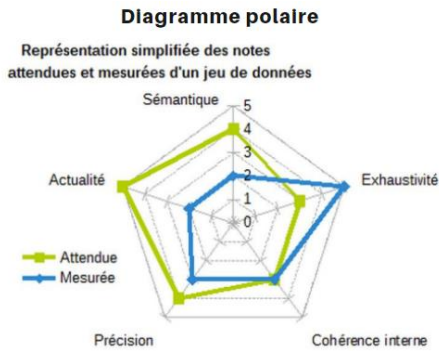
Crige Centre de Recherche en Information Géographique Provence Alpes Côte d'Azur		
RAPPORT D'ANALYSE DU FICHIER "LINE_TTE_ERREUR_ALL_OUTSIDE"		
INFORMATIONS GÉNÉRALES SUR LE FICHIER		
Projet :	LE PAYSAN DE CRISTAL - PACA	Mail : calpig@univmed.fr
Données :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
Path du fichier :	longueur - 1 à 6 (kilomètres)	Surface = 101258 m ²
Projet :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
Projet :	LES PAYSAN DE CRISTAL	
NOTATION GLOBALE = 2.5 sur 5		
CRITÈRE DE COHÉRENCE LOGIQUE = 4.5 sur 5		
TOUTS CRITÈRE DE COHÉRENCE DU FORMAT = 3 sur 5		
Projet :	Projet :	Projet :
Projet :	Projet :	Projet :
Projet :	Projet :	Projet :

Premiers travaux d'automatisation MAIS pas d'analyse des métadonnées, uniquement des données SHAPE, processus sur une seule donnée à la fois



Comparaison des modes de représentation (benchmark)

Différents modes de représentation



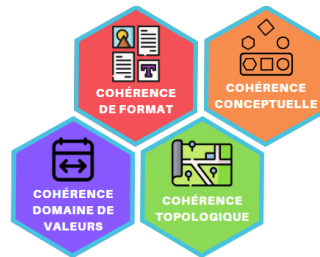
Notation par étoile



Notation par jauge moyenne des utilisateurs



Badges débloqués



Jauge de compétences / exigences respectées



Rapport de qualité



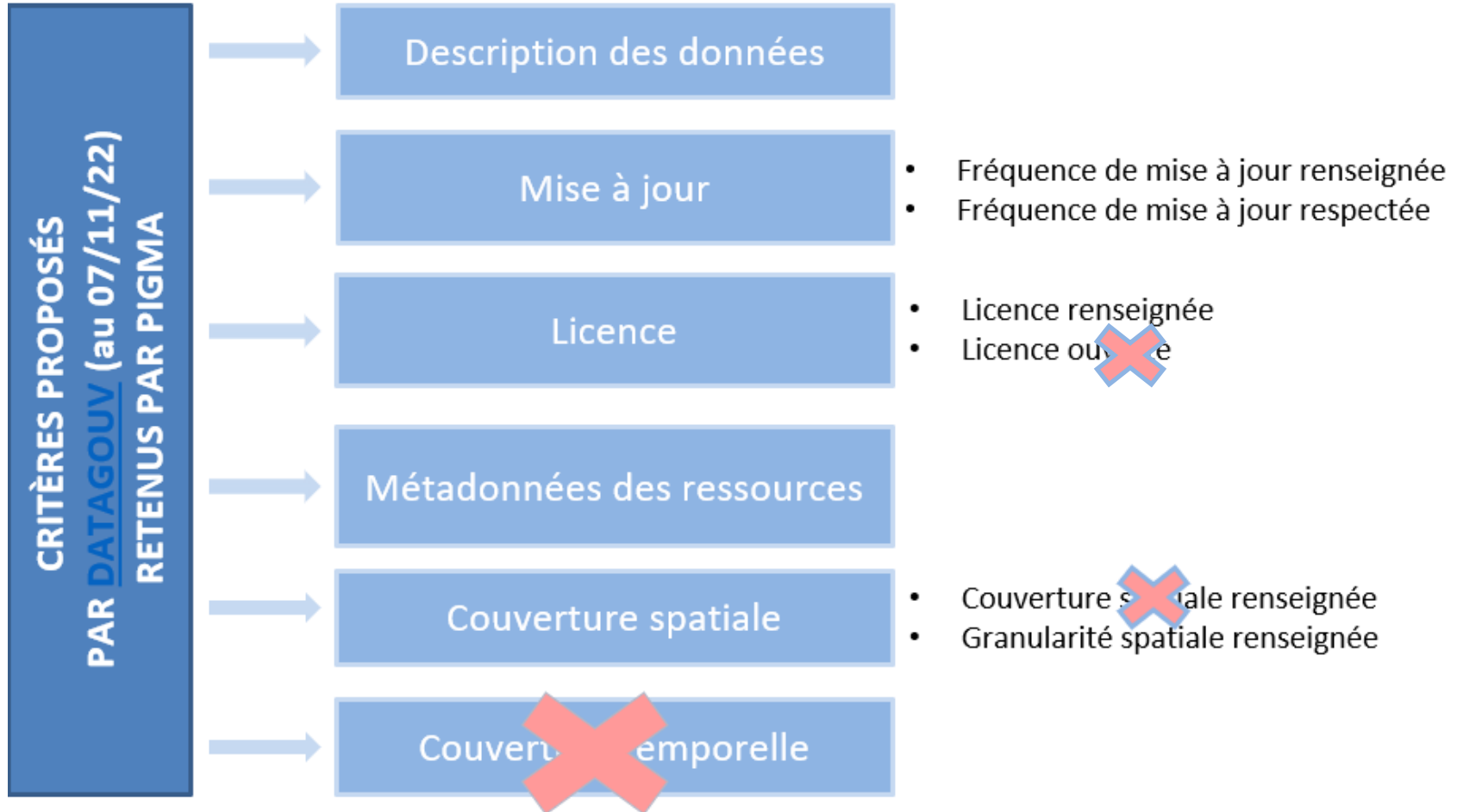
Label Qualité



Label Qualité : Seul mode qui identifie les éléments positifs sans dévaloriser les éléments négatifs (avec le système de badges)



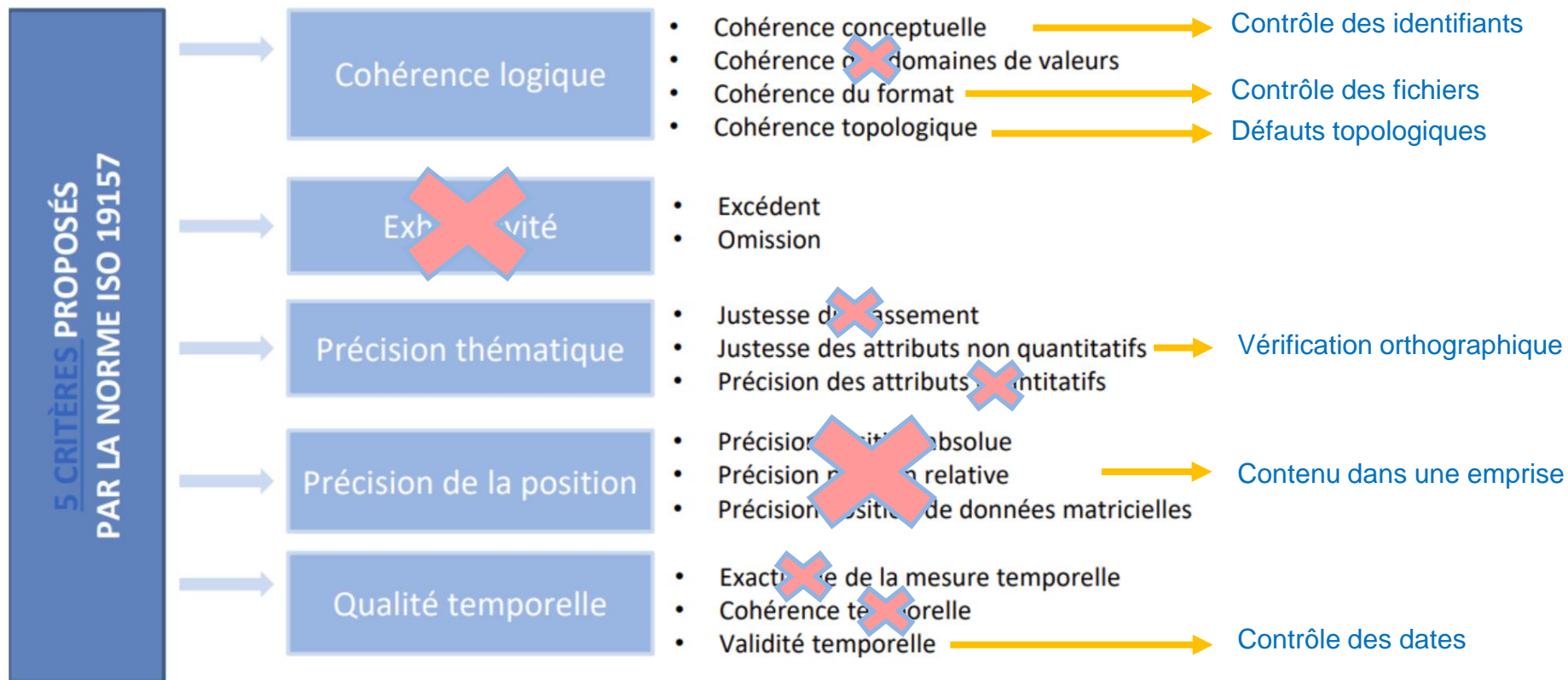
Qualité des Métadonnées





Qualité de la Donnée Géographique

Analyses possibles



Premiers critères retenus, analysables par PIGMA



On peut faire les observations suivantes :

L'utilisateur est d'abord confronté à la qualité des métadonnées

Une donnée ancienne ou inaccessible ne peut pas être aisément réutilisée

Les métadonnées des jeux de données sont décrites selon les mêmes champs



Un Label Qualité à 2 niveaux



Label Minimal

Champs indispensables remplis

Une ressource principale réutilisable

Une donnée fraîche ou millésimée

Qualité des Métadonnées



Label Optimal

Critères Label Minimal

Des informations détaillées et concordantes

Une donnée sans défaut topologique retenu

Qualité des Métadonnées avancée
+ Qualité Donnée



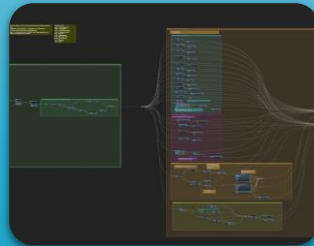
Notations et pondération

Label Minimal

- Aucune notation, ni pondération
- Tous les critères sont obligatoires
- L'absence d'un seul champ ou critère exclut le jeu de données du Label Minimal

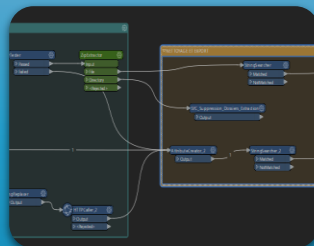
Label Optimal

- Les critères sont notés et sont pondérés
- 4 critères sont obligatoires :
 - Date de MAJ Ressource
 - Noms des champs corrects
 - Présence d'un champ identifiant
 - Absence de géométries nulles



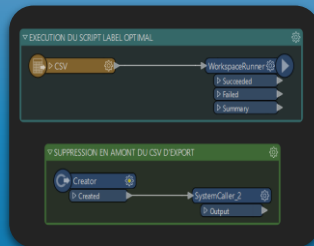
SCRIPT LABEL MINIMAL

- Pour l'analyse du Label Minimal



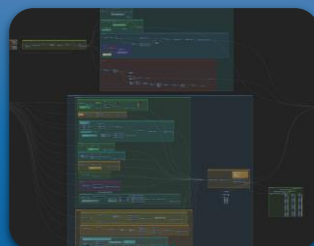
RECUPERATION DONNEE

- Pour le téléchargement des ressources principales des jeux de données obtenant le Label Minimal



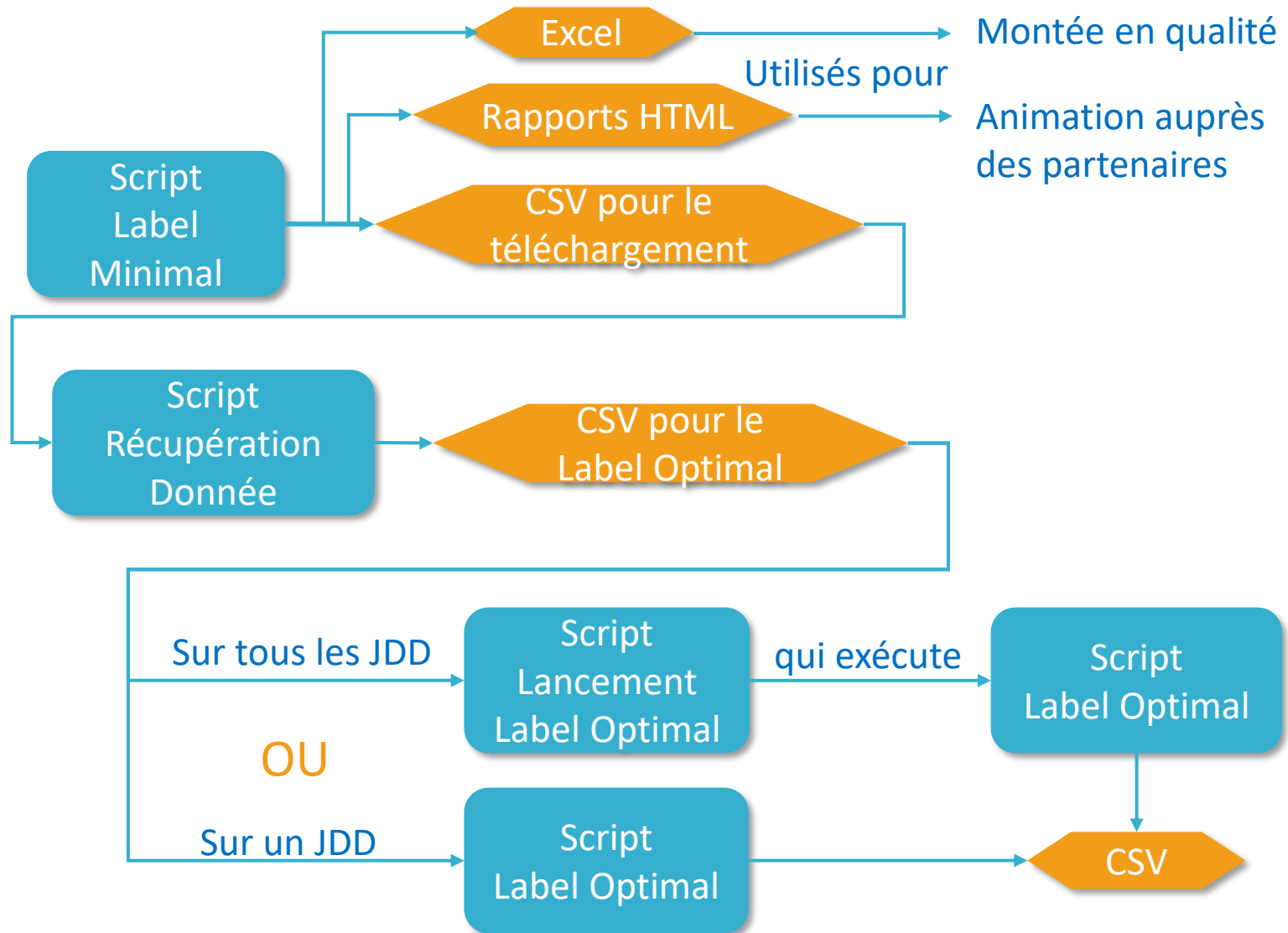
SCRIPT LABEL OPTIMAL

- Pour l'exécution du Label Optimal pour un jeu de données



LANCEMENT LABEL OPTIMAL

- Pour l'analyse du Label Optimal pour tous les jeux de données ou pour les JDD et organisations choisis





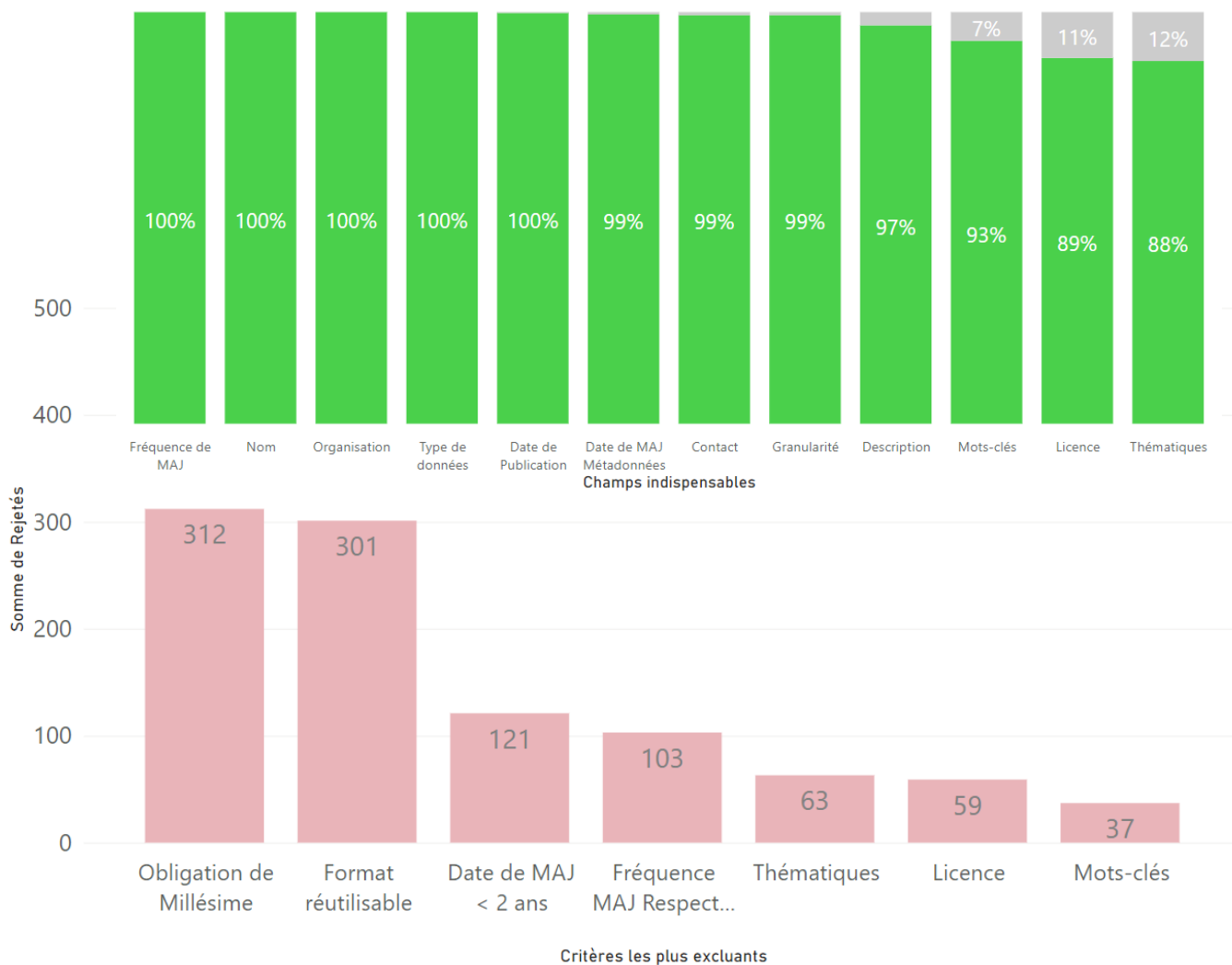
Jeux de données obtenant le Label Qualité Minimal



Soit **15%** des jeux de données étudiés

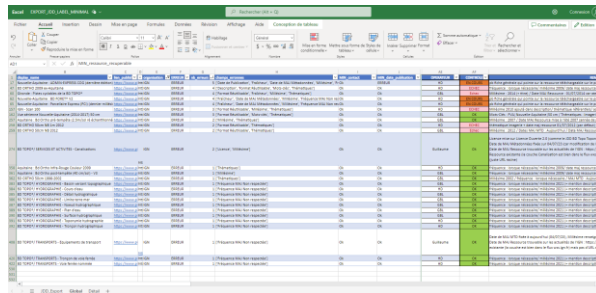


Des critères plus excluants que d'autres





Montée en qualité



Corrections manuelles à partir de l'export Excel



Modifications de portions du script Label Minimal

Fréquence d'erreurs

Résumé des corrections aux principaux types d'erreurs

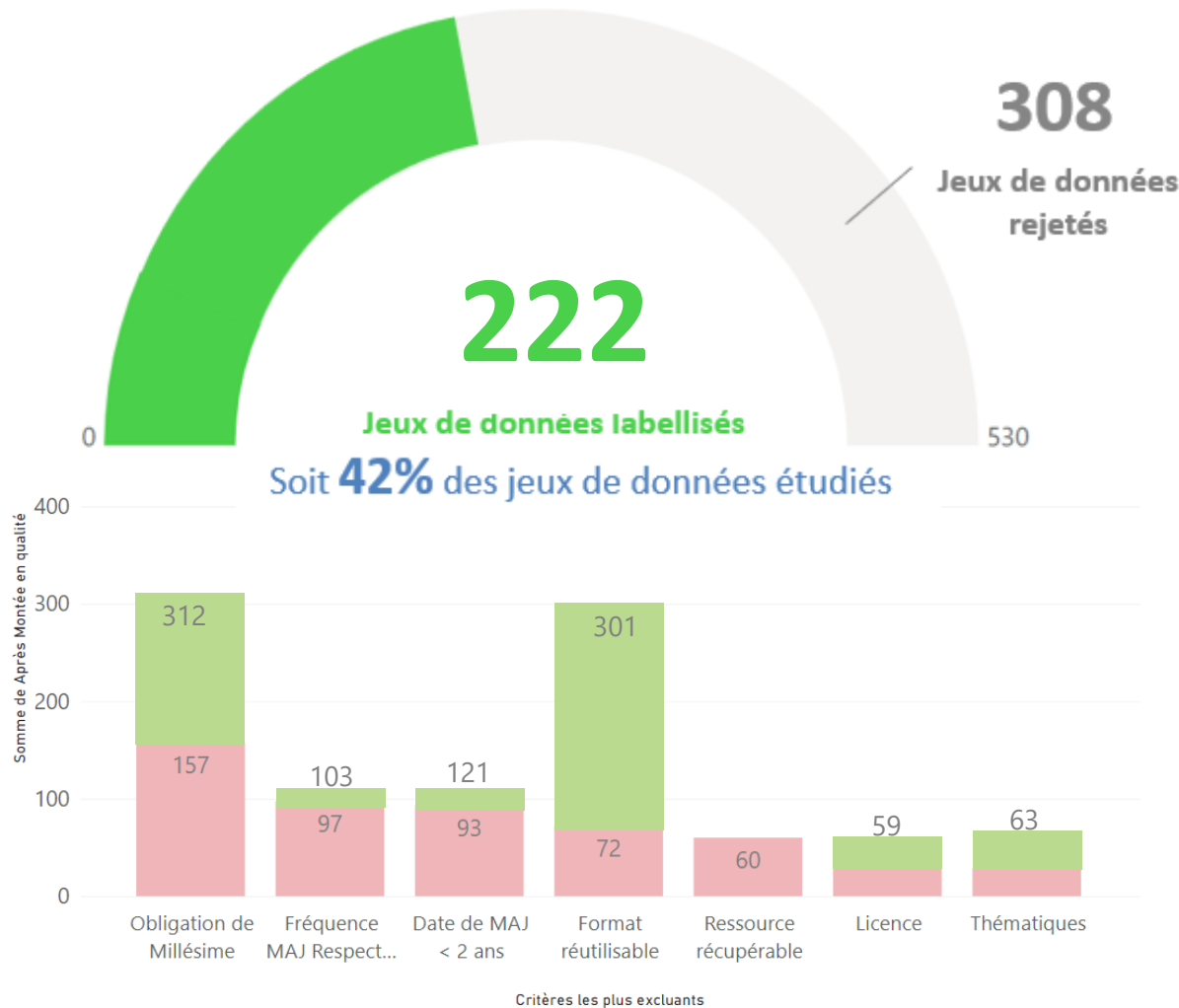
Type d'erreur	Cause	Corrections pour le Label Minimal	Corrections pour le Label Optimal
Format réutilisable	Le format de la ressource ne fait pas partie de la liste des formats facilement réutilisables ou ne génère pas de ressources associées réutilisables.	La ressource principale doit faire partie de la liste des formats réutilisables : CSV, JSON, XML, GeoJSON, SHP, WMS, WFS, TIF, GTFS et NetEs. Pour une ressource principale en WMS, les services de téléchargement aux formats CSV, GeoJSON, KML et GML doivent être activés. Dans le cas contraire, contacter le support PIGMA.	
Millésime	Le millésime du jeu de données n'est pas renseigné alors que : <ul style="list-style-type: none"> la date de mise à jour de la ressource est supérieure à 2 ans la fréquence de mise à jour n'est pas régulière (Irégulière, Inconnue, Non planifiée, Lorsque nécessaire) 	Renseigner le millésime de ressource principale (année de production de la ressource) dans le champ Millésime. Autres correctifs : Mettre à jour la ressource principale Fixer la fréquence de mise à jour sur une valeur régulière (Continue, Quotidienne, Hebdomadaire, Bi-mensuelle, Mensuelle, Trimestrielle, Semestrielle, Annuelle) Bonne pratique : Renseigner le millésime dans le champ Millésime ET dans le titre.	Le millésime doit être égal à l'année de la date de mise à jour de la ressource. Le millésime doit être antérieur ou égal à la date de mise à jour des métadonnées.
Fraisier	Les dates de mise à jour renseignées sont supérieures à 2 ans sans que la donnée ne soit millésimée.	A minima, mettre à jour la date de mise à jour des métadonnées Si possible, mettre à jour la ressource principale et sa date de mise à jour Si la ressource principale ne peut être mise à jour, renseigner la date de dernière mise à jour de la ressource et le millésime.	La date de mise à jour de la ressource est obligatoire pour le Label Optimal et doit être égale ou antérieure à la date de mise à jour des métadonnées.
Fréquence de mise à jour non respectée	La fréquence de mise à jour renseignée n'est pas respectée - les dates de mise à jour sont plus anciennes que le délai couvert par la fréquence	Mettre à jour la ressource principale et sa date de mise à jour Mettre à jour la date de mise à jour des métadonnées Bonne pratique : Mettre en place un processus de révision périodique de vos jeux de données Fixer la fréquence de mise à jour sur une valeur régulière	Fixer la fréquence de mise à jour sur une valeur régulière. Les fréquences régulières reçoivent une meilleure notation dans le Label Optimal que les fréquences irrégulières (Irégulière, Lorsque nécessaire, Non planifiée, Inconnue).
Ressource irrécupérable	La ressource principale n'est pas téléchargeable.	Déclarer une nouvelle ressource en tant que ressource principale Supprimer ou déclarer en tant qu'annexe l'ancienne ressource.	La ressource doit contenir tous les fichiers nécessaires à sa lecture (Ex : dans le cas d'un Shapefile, les fichiers SHP, SHX et DBF doivent être présents)



Correctifs de montée en qualité



Jeux de données obtenant le Label Qualité Minimal





Critères de qualité retenus

Des informations concordantes

Au sein du modèle PIGMA, une même information peut parfois se trouver répartie dans plusieurs champs. On dispose ainsi de 2 dates de mise à jour (Métadonnées et Ressource), la Licence est reliée aux préconisations d'usage et limites d'utilisations, etc. Un JDD de qualité doit posséder une certaine concordance entre ces différentes informations, au risque de désorienter l'utilisateur, qui ne saurait pas à quelle information se fier, quelle information croire plutôt que telle autre.

Nom du champ	Coef.	Description	Utilité et Qualité attendue	Notation
Dates de mise à jour	5 et 3	Concordance entre la date de MAJ de de Ressource et la date de MAJ des Métadonnées	La date de MAJ Ressource indique la fraîcheur de la donnée. Elle est indispensable /fortement recommandée pour le Label Optimal. Les trois dates (Publication, MAJ Métadonnées et MAJ Ressource) doivent avoir un format valide (YYYYMMDD) et être datée du XXIe siècle (ou du XXe pour la MAJ Ressource). La Date de MAJ Métadonnées doit être postérieure à la date de MAJ Ressource (car la MTD peut être mise à jour mais pas la ressource ; en revanche, si la Ressource est mise à jour, la MTD doit l'être aussi).	Note sur le format des dates : 0 : L'une des trois dates si n'a pas le bon format 10 : Les trois dates ont le bon format + Note sur la chronologie des Dates 0 : MAJ Métadonnées antérieure à MAJ Ressource 5 : MAJ Métadonnées postérieure à Date Publication 10 : MAJ Métadonnées postérieure ou égale à MAJ Ressource
Millésime	3	Concordance entre le millésime de la donnée et le millésime de la zone géographique mentionnée	Un millésime est exigé dans le niveau minimal si la fréquence de MAJ est inférieure à 10. Le millésime doit correspondre à l'année de la donnée, et donc sur la zone géographique et/ou l'échelle d'utilisation. Si une zone géographique est mentionnée dans le titre, la granularité doit correspondre avec cette zone (exemple : BD ORTHO Grande → Département)	0 : Millésime ne correspond pas à MAJ Ressource 5 : Millésime correspond à MAJ Métadonnées 10 : Millésime correspond à MAJ Ressource
Couverture géographique	1	Concordance entre la granularité et la zone géographique mentionnée	La granularité de la donnée doit correspondre à la zone géographique mentionnée, et donc sur la zone géographique et/ou l'échelle d'utilisation. Si une zone géographique est mentionnée dans le titre, la granularité doit correspondre avec cette zone (exemple : BD ORTHO Grande → Département)	0 : La donnée n'est pas à la Zone Géo 5 : Granularité indéfinie (donc aucune concordance analysable) OU sur Parcelles, POI, IRIS 10 : Granularité correspond à la Zone Géo
Licence	3 et 3	Concordance entre la Licence renseignée et les limites et préconisations	Assure que la donnée pourra être réutilisée et dans quel cadre (contraintes d'utilisation ou de partage, mention du producteur, etc.). Les informations contenues dans les champs Limites d'utilisation ou Préconisation d'usage doivent concorder avec la Licence renseignée. Si la donnée est de type Données ouvertes, le JDD doit être disponible en téléchargement.	Note sur la cohérence des informations 0 : Les informations ne concordent pas avec la Licence 10 : Les informations concordent avec la Licence + Note sur la cohérence de l'ouverture 0 : Le JDD est de type Données ouvertes sans être disponible en téléchargement OU le JDD est téléchargeable sans être de type Données ouvertes. 10 : Le JDD est ouvert et en téléchargement ou non-ouvert et indisponible en téléchargement

Il est possible de pondérer certaines de ces notations, notamment en ce qui concerne la Licence ou la Couverture géographique. Sans pondération : Score minimal = 0 ; score maximal = 60.

Qualité Métadonnées

2 documents, présentant :

- Les critères retenus
- Leur intérêt
- La notation de chaque critère
- Leur pondération

Notations et Pondération des critères retenus

Cohérence logique

Nom du critère	Coef.	Type de critère	Utilité et Qualité attendue	Notation
Identifiant unique		Cohérence conceptuelle	La donnée doit posséder au moins un champ identifiant. Les identifiants doivent être uniques et renseignés pour toutes les entités, sans exception ni doublon.	Critère exclusif : La présence d'un doublon ou l'absence d'identifiant pour une entité de la donnée exclut le jeu de données du Label Optimal
Noms des champs		Cohérence conceptuelle	Le nommage des champs exclut les caractères spéciaux, les espaces et les signes de ponctuation.	Critère exclusif : La présence d'un ou plusieurs champs mal nommés exclut le jeu de données du Label Optimal
Cohérence et Conformité du Format	5 et 5	Cohérence du format	Le format déclaré de la donnée correspond au format réel (Cohérence). La donnée est conforme aux spécifications du format (JSON valide, fichiers indispensables du Shapefile).	Note sur la cohérence : 0 : Les formats réel et déclaré ne concordent pas 5 : Les formats réel et déclaré concordent + Note sur la Conformité : ERREUR : La donnée n'est pas conforme 5 : La donnée est conforme mais ne contient que les fichiers indispensables 10 : La donnée est conforme et contient plus de fichiers que ceux indispensables
Codes valides	3	Cohérence du format	Toutes les valeurs des champs contenant des codes (type code culture, code usage, etc.) correspondent au format du code, avec un seul d'erreur inférieur à 5%.	Note sur la cohérence : 0 : Au moins un champ code comporte des erreurs, avec un taux d'erreurs supérieur à 5% 5 : Au moins un champ code comporte des erreurs, avec un taux d'erreurs inférieur à 5% OU La donnée ne comporte aucun champ code 10 : Tous les champs codes sont valides

Qualité Donnée



Correctifs de montée en qualité

Fréquence d'erreurs

Résumé des corrections aux principaux types d'erreurs

Type d'erreur	Cause	Corrections pour le Label Minimal	Corrections pour le Label Optimal
Format réutilisable	Le format de la ressource ne fait pas partie de la liste des formats facilement réutilisables ou ne génère pas de ressources associées réutilisables.	La ressource principale doit faire partie de la liste des formats réutilisables : CSV, JSON, XML, GeoJSON, SHP, WMS, WFS, TIF, GIFS et NetEx. Pour une ressource principale en WMS, les services de téléchargement aux formats CSV, GeoJSON, KML et GML doivent être activés. Dans le cas contraire, contacter le support PIGMA.	
Millésime	Le millésime du jeu de données n'est pas renseigné alors que : <ul style="list-style-type: none"> la date de mise à jour de la ressource est supérieure à 2 ans la fréquence de mise à jour n'est pas régulière (Irégulière, Inconnue, Non planifiée, Lorsque nécessaire) 	Renseigner le millésime de ressource principale (année de production de la ressource) dans le champ Millésime. Autres correctifs : Mettre à jour la ressource principale Fixer la fréquence de mise à jour sur une valeur régulière (Continue, Quotidienne, Hebdomadaire, Bi-mensuelle, Mensuelle, Trimestrielle, Semestrielle, Annuelle) Bonne pratique : Renseigner le millésime dans le champ Millésime ET dans le titre	Le millésime doit être égal à l'année de la date de mise à jour de la ressource. Le millésime doit être antérieur ou égal à la date de mise à jour des métadonnées.
Fraîcheur	Les dates de mise à jour renseignées sont supérieures à 2 ans sans que la donnée ne soit millésimée.	À minima, mettre à jour la date de mise à jour des métadonnées Si possible, mettre à jour la ressource principale et sa date de mise à jour Si la ressource principale ne peut être mise à jour, renseigner la date de dernière mise à jour de la ressource et le millésime.	La date de mise à jour de la ressource est obligatoire pour le Label Optimal et doit être égale ou antérieure à la date de mise à jour des métadonnées.
Fréquence de mise à jour non respectée	La fréquence de mise à jour renseignée n'est pas respectée : les dates de mise à jour sont plus anciennes que le décalé couvert par la fréquence	Mettre à jour la ressource principale et sa date de mise à jour Mettre à jour la date de mise à jour des métadonnées Bonne pratique : Mettre en place un processus de révision périodique de vos jeux de données Fixer la fréquence de mise à jour sur une valeur régulière	Fixer la fréquence de mise à jour sur une valeur régulière. Les fréquences régulières reçoivent une meilleure notation dans le Label Optimal que les fréquences irrégulières (Irégulière, Lorsque nécessaire, Non planifiée, Inconnue).
Ressource irrécupérable	La ressource principale n'est pas téléchargeable.	Déclarer une nouvelle ressource en tant que ressource principale Supprimer ou déclarer en tant qu'annexe l'ancienne ressource.	La ressource doit contenir tous les fichiers nécessaires à sa lecture (Ex : dans le cas d'un Shapefile, les fichiers SHP, SHX et DBF doivent être présents)

Résumé des corrections pour chaque Label, avec les erreurs triées par fréquence

Mots-Clés

Cause

Aucun mot-clé n'est renseigné alors qu'au moins un est requis.

Solution pour le Label Minimal

Renseigner au moins un mot-clé, par une saisie de texte libre ou par sélection parmi des valeurs existantes, en rapport avec le jeu de données.

Solution pour le Label Optimal

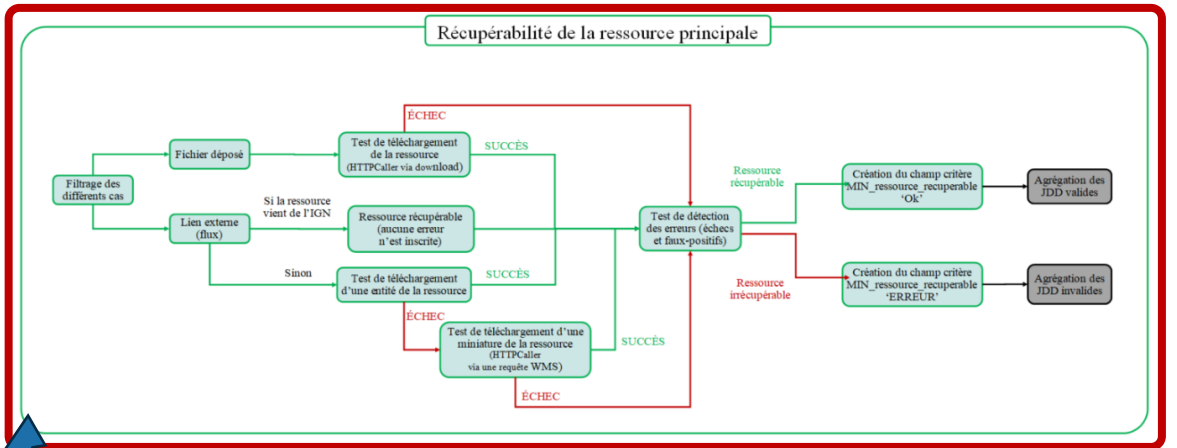
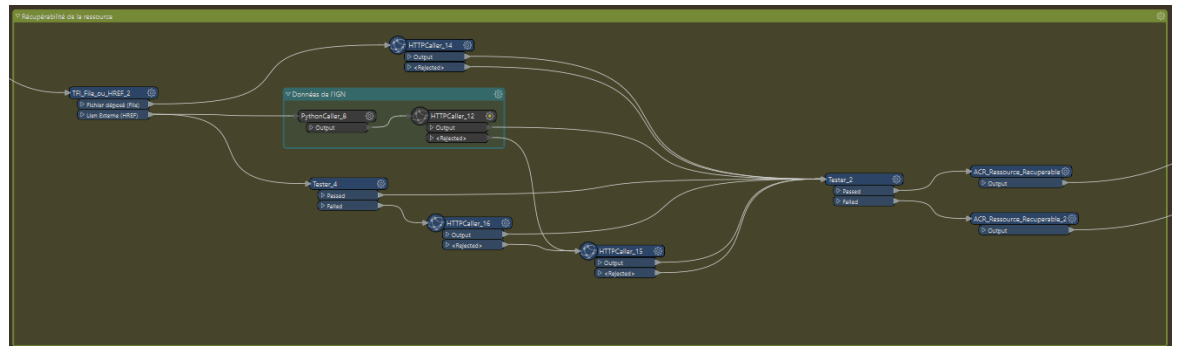
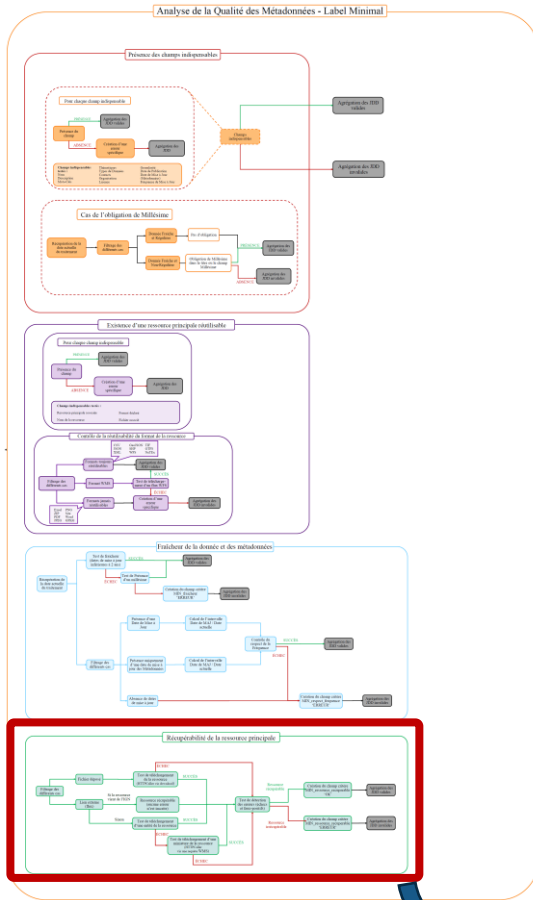
Le Label Optimal accorde de meilleures notations aux jeux de données avec des informations fournies et pertinentes.

- Renseigner de nombreux mots-clés.
- Renseigner au moins 1 mot-clé déjà présent dans le titre ou la description (ex : **Vitidata**)
- Renseigner au moins 1 mot-clé qu'on ne retrouve ni dans le titre, ni dans la description (ex : **pesticides**)

Guide de rédaction d'une fiche Métadonnées, avec les corrections pour chaque Label



Synthèse des scripts de qualification





Notice de lancement des scripts

PRÉSENTATION DES SCRIPTS

Le processus de qualification PIGMA se compose de 4 scripts :

SCRIPT_LABEL_MINIMAL

- Ce script analyse le niveau minimal de qualité des jeux de données, centré sur la qualité des métadonnées. Il produit notamment en sortie un fichier CSV EXPORT_POUR_TELECHARGEMENT_DONNEE.

RECUPERATION_DONNEE

- Ce script utilise le CSV précédent et télécharge la ressource principale de chaque jeu de données en fonction de la liste d'organisations.

SCRIPT_LABEL_OPTIMAL

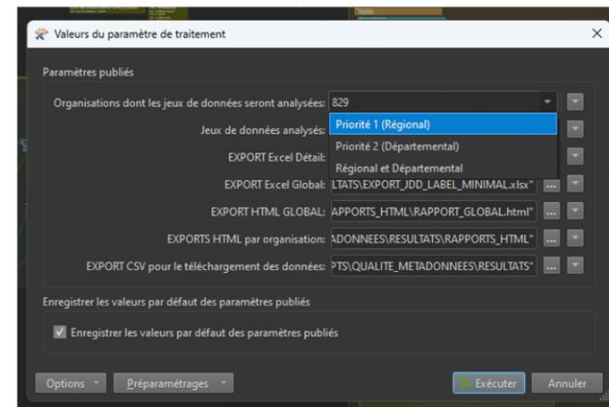
- Ce script utilise le CSV EXPORT_POUR_TELECHARGEMENT_DONNEE pour effectuer une analyse de qualité optimale des jeux de données sur les métadonnées et les données (identifiant, géométrie, topologie).

LANCEMENT_LABEL_OPTIMAL

- Ce script utilise le CSV EXPORT_POUR_TELECHARGEMENT_DONNEE pour lancer le script du Label Optimal par organisation, faisant transiter un à un dans le processus de qualification.

GUIDE PAS À PAS

Ouvrir le script FME nommé **SCRIPT_LABEL_MINIMAL.fmw**.
En cliquant sur  , la fenêtre suivante apparaît.

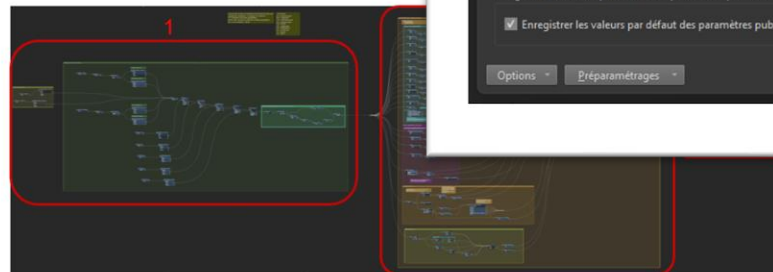


Paramètres publiés

Organisations prises en compte : Choisir un des éléments de la liste OU Renseigner l'identifiant du/des organisations

GUIDE PAS À PAS

Ouvrir le script FME nommé **SCRIPT_LABEL_MINIMAL.fmw**.
Il se compose de 3 parties.



- 1. Récupération des informations en base**
- 2. Analyses du Label Minimal**
 - Présence des champs indispensables
 - Ressource principale réutilisable
 - Contrôle de la fraîcheur
 - Récupérabilité de la ressource
- 3. Exports**
 - 1 CSV pour le téléchargement des données
 - 1 Excel en 2 feuilles pour la montée en qualité
 - 1 rapport HTML Global et 1 rapport HTML par organisation



Des critères de qualité définis

Définition des critères qualité

- Champs indispensables
- Format réutilisable
- Fraîcheur de la donnée
- Informations détaillées et concordantes
- Donnée sans défaut retenu

Label Qualité à 2 niveaux

- Label Minimal (Qualité Métadonnées)
- Label Optimal (Qualité Métadonnées avancée + Qualité Donnée)



Un processus fonctionnel avec des livrables précis

Un processus fonctionnel

- 4 scripts de qualification :
- Script Label Minimal
- Récupération Donnée
- Script Label Optimal
- Lancement Label Optimal
- 222 JDD obtenant le Label Minimal (3 min)
- 110 JDD obtenant le Label Optimal (1h 5min)

Des livrables précis

- Des critères documentés
- Une synthèse du script Label Minimal
- Un document de correctifs de montée en qualité
- Une notice de lancement des scripts



Limites du processus

Dépendance à l'architecture de données PIGMA

- Connexion à la base de données
- Noms des champs analysés
- Valeurs contrôlées par les analyses

Une qualité de la donnée en chantier

- Peu de critères analysés
- Temps de traitement long, même avec échantillonnage des données
- 5/10 : seuil d'obtention du Label Optimal
- Fonctionnement générique / spécificité de la donnée (nom du champ id)
- Détermination de la qualité limitée → rôle du producteur



Perspectives

Animation régionale

- Montée en qualité des jeux de données administrés
- Organisation d'animation grâce aux exports
- Validation des critères lors d'un GT régional

Extension de la qualification à l'ensemble de la plateforme

- Tous les jeux de données intégrés (modification des requêtes SQL)
- Tous les jeux de données indexés (passage par l'indexeur)

Participation à l'animation nationale

- Présentation des travaux lors du GT CNIG QuaDoGéo
- Se mettre d'accord sur les critères qualité retenus
- Une base pour des travaux futurs



MERCI

www.pigma.org
pigma@gipatgeri.fr
[@PlateformePIGMA](#)