



CRITÈRES DE QUALITÉ MÉTADONNÉES

Critères retenus pour les deux niveaux de la qualité des Métadonnées des jeux de données du Catalogue PIGMA

Membres de l'équipe PIGMA

Guillaume Blanchard – Administrateur de la plateforme PIGMA

Héloïse Deschamps – Animatrice technique de la plateforme PIGMA

1. Principes directeurs du 1^{er} niveau de la qualité des Métadonnées

Les jeux de données (ou JDD) du catalogue PIGMA se caractérisent par une fiche métadonnées, avec des champs obligatoires et d'autres optionnels et une ou plusieurs ressources associées. Certains jeux de données ont une ressource principale, des annexes et des informations très détaillées, d'autres n'ont que les informations obligatoires, présentées de façon succincte, sans ressource associée. Afin de déterminer la qualité des données du catalogue PIGMA, on définit 3 grands principes de la qualité des Métadonnées.

1.1 Des champs de base remplis

La Métadonnée associée à un jeu de données est de qualité si un certain nombre de champs minimaux sont remplis. Ces champs présentent des informations obligatoires et/ou importantes à afficher lors de la consultation de la fiche Métadonnées. Le **1^{er} niveau de qualité** s'attache à considérer si ces champs sont bien remplis (c'est-à-dire si une valeur valide est **présente** dans ce champ). Voici la liste des **champs minimaux** retenus, présentés dans l'ordre dans lequel ils apparaissent dans Publish, l'interface de publication des jeux de données.

Nom du champ	Description	Utilité du champ
Nom	Intitulé du jeu de données	Permet de nommer et d'identifier le jeu de données par rapport aux autres. Repris pour créer un identifiant interne unique.
Description	Description plus ou moins détaillée du jeu de données	Donne des informations sur le type de données, le contexte, la zone géographique concernée, etc.
Mots-clés	Saisie libre de mots-clés associés	Améliore le référencement interne du JDD. Permet de le retrouver via la barre de recherche
Thématiques	Thèmes Inspire associés à la donnée	Permet à l'utilisateur d'identifier les thèmes couverts par la donnée. Permet la recherche des JDD par thématiques.
Type de données	Grands types de données associés aux données (géographiques,	Indique si la donnée est géographique ou non, ouverte ou non et si elle correspond à

Y:



	ouvertes, tabulaires, satellites, etc.)	des données satellites. Oriente rapidement l'utilisateur.
Contact	Adresse(s) mail d'un contact de l'organisation productrice du JDD	Renvoie vers une personne à contacter si un souci émerge avec la donnée (dont PIGMA n'est pas responsable)
Organisation	Organisation ou groupe d'utilisateurs ayant publié le JDD	Rattache le JDD à une structure, qui a la charge de la donnée.
Licence	Licence de réutilisation et de partage de la donnée	Indique sous quelles conditions l'utilisateur peut ou non réutiliser et partager la donnée.
Granularité	Échelon territorial pertinent pour l'étude du JDD (commune, département, EPCI, POI, parcelle, etc.)	Donne à l'utilisateur une information immédiate de la zone couverte par les données ainsi qu'une indication de l'échelle
Date de publication	Date de publication du JDD	Indique quand la fiche a été publiée
Date de dernière mise à jour	Date de MAJ du JDD	Indique la date à laquelle la Métadonnée a été mise à jour (et non la Ressource). Indique la fraîcheur de la métadonnée et permet de contrôler le respect de la fréquence. Actuellement seule information de MAJ visible dans PIGMA
Fréquence de mise à jour	Fréquence à laquelle la donnée Ressource est mise à jour	Indique si une mise à jour est programmée ou non.
Millésime	Année de production de la donnée	Obligatoire dans le cas où la fréquence de MAJ est non-régulière (Inconnu, Irrégulière, Non-planifiée et Lorsque nécessaire)

Si un seul de ces champs n'est **pas rempli**, on considère que la métadonnée **ne passe pas le 1^{er} niveau de qualité**. À l'inverse, si l'ensemble de ces champs est présent et complété pour un jeu de données, cela peut ouvrir droit au Label Qualité Métadonnées de base, si les autres conditions sont remplies.

1.2 Une ressource principale réutilisable et récupérable

Pour être considéré comme de qualité, un JDD doit posséder une ressource principale qui lui est associée. Cela se définit comme l'existence d'une valeur rattachée au JDD dans la table reliant les ressources et le type de ressources. Cette ressource principale doit posséder les champs suivants remplis.

Y:





Nom du champ	Description	Utilité du champ
Nom	Intitulé de la ressource	Permet de nommer la ressource et de la distinguer des autres.
Date de dernière mise à jour	Date de MAJ de la ressource	Indique la date à laquelle la Ressource a été mise à jour (et non la Métadonnée). Indique la fraîcheur de la donnée et permet de contrôler le respect de la fréquence.
Fichier	Fichier associé à la ressource déclarée	Fournit le lien permettant d'accéder à la donnée (soit sous forme d'URL Racine, soit sous forme de lien vers la donnée en zip)
Format	Format du fichier associé à la ressource déclarée	Format de données réutilisable par un grand nombre d'utilisateurs et de postes. La liste des formats ouverts est fournie par DataGouv. À cette liste s'ajoute également les ressources déclarées en tant que flux OGC (WMS et WFS)
Type de données	Type de ressources associé au JDD (Ressource principale ou Annexe)	Qualifie le lien entre une ressource et un JDD. Sert à indiquer si la ressource est principale ou non. Seule valeur admise : 1 = Ressource principale

Cette ressource déclarée doit également être **recupérable** :

- Pour les ressources issues d'un fichier déposé sur la plateforme : l'URL de téléchargement doit permettre le téléchargement de la donnée
- Pour les ressources issues d'un flux OGC : le téléchargement d'un fichier via une requête de type *GetMap* ou *GetFeatures* doit être fonctionnel

Si le jeu de données possède une ressource principale où tous les champs précédent sont correctement renseignés, cela peut ouvrir droit au Label Qualité Métadonnées de base, si la dernière condition est remplie. À l'inverse, l'absence d'un seul champ constitue un défaut majeur de qualité qui empêche l'obtention du Label.

Y:



1.3 Une donnée fraîche ou millésimée

Pour être considéré comme de qualité, un JDD doit avoir une certaine fraîcheur, afin de conserver sa qualité et sa pertinence. Cela concerne les champs suivants.

Nom du champ	Critères à respecter
Dates de dernière mise à jour (Ressource et Métadonnées)	Inférieure à 2 ans (à la date du contrôle qualité)
Fréquence de mise à jour	<p>Fréquence de MAJ régulière (Continue, Quotidienne, Hebdomadaire, Bimensuelle, Mensuelle, Trimestrielle, Semestrielle et Annuelle) (Continue : ≤ 1 jour)</p> <p>Autres fréquences valides : Non planifiée, Lorsque nécessaire, Irrégulière, Inconnu si le JDD possède un millésime</p> <p>Fréquence respectée : La différence entre les dates de MAJ et la date du contrôle qualité ne doit pas excéder le seuil défini par la fréquence (ex : hebdomadaire = 7 jours maximum)</p>
Millésime	Millésime obligatoirement renseigné si la fréquence est définie sur les valeurs Non planifiée, Lorsque nécessaire, Irrégulière ou Inconnu

Si le jeu de données ne remplit pas un de ces trois critères, on considère que la qualité n'est pas suffisante. Inversement, si un jeu de données remplit tous ces critères et valide les deux autres conditions (Champs minimaux et Ressource principale réutilisable et récupérable), alors ce jeu de données obtient le **Label Minimal Qualité Métadonnées de Base de PIGMA**.

2. Principes directeurs du 2nd niveau de la qualité des Métadonnées

Le Label Minimal Qualité Métadonnées s'attache à caractériser la présence de champs indispensables, ainsi que le respect et la cohérence de certaines informations (Fréquence et Dates de mise à jour) afin de valider l'obtention de ce Label pour un JDD. Pour valoriser davantage les données de la plateforme, on définit un **2nd niveau de Qualité** qu'on nomme **Label Optimal**.

Pour obtenir le Label Optimal, les jeux de données doivent notamment posséder des informations de Métadonnées **détaillées** et **correspondantes** entre elles. Contrairement au 1^{er} niveau de qualité, le Label Minimal, l'obtention du Label ne dépend pas de la présence de chacun des éléments présents ci-dessous. Le critère de détail et de concordance des informations du Label Optimal est obtenu par un JDD lorsque celui-ci remplit dépasse une valeur seuil de notation, qu'on définira et pondérera.

Y:





Un Label Qualité à 2 niveaux

QUALITÉ
PIGMA

Label Minimal

Champs indispensables remplis

Une ressource principale réutilisable

Une donnée fraîche ou millésimée

Qualité des Métadonnées

QUALITÉ
PIGMA

Label Optimal

Critères Label Minimal

Des informations détaillées et concordantes

Une donnée sans défaut topologique retenu

Qualité des Métadonnées avancée
+ Qualité Donnée

Un jeu de données obtient le Label Optimal s'il remplit tous les critères du Label Minimal ainsi que les deux critères spécifiques au Label Optimal :

- Des informations de métadonnées détaillées et concordantes
- Une donnée sans défaut topologique ni logique retenu

Les critères de Qualité de la Donnée définis pour le Label Optimal sont décrits dans un autre document : *Critères Qualité Données retenus*.

2.1 Notations et Pondération

1. Notation

On définit de façon basique 3 notations possibles pour chaque élément contrôlé :

- 0 = Absence d'informations pour le JDD ou informations incohérentes
- 5 = Présence minimale d'informations (mais peu fournies ou peu précises)
- 10 = Présence d'informations fournies et/ou précises

Obtenir la note de 0 à l'un des critères contrôlés n'est pas rédhibitoire. De plus, cette notation fait l'objet d'une pondération

2. Pondération

La pondération intervient comme un coefficient à affecter à la notation d'un critère, afin d'attribuer des importances différentes aux critères du Label Optimal.

Pour simplifier la pondération des différents critères, on définit 3 valeurs de coefficients :

Y:





- **Priorité faible** : coefficient = 1
- **Priorité moyenne** : coefficient = 3
- **Priorité forte** : coefficient = 5

Ainsi, si un jeu de données obtient pour un critère A, de priorité moyenne, une note de 5/10 et pour un critère B, de priorité faible, une note de 10, la note finale pondérée sera :

$$\bar{x} = \frac{3 \times 5 + 1 \times 10}{3 + 1} = 6,25$$

2.2 Des informations détaillées

La fiche Métadonnées d'un jeu de données doit informer et orienter au mieux l'utilisateur sur la donnée souhaitée. Aussi, un JDD dont les métadonnées présentent des informations détaillées dans certains champs aura une qualité supérieure à ceux dont les informations sont manquantes ou laconiques.

Pour chaque champ des métadonnées, on rappelle le sens de ce champ, son utilité et les attendus en termes de qualité, ainsi que les coefficients et notations à appliquer aux différentes analyses concernant ce champ.

Exemple :

Nom du champ	Coef.	Description	Utilité et Qualité attendue	Notation
Description	5 et 5	Description plus ou moins détaillée du jeu de données	Description d'une longueur suffisante. Notation en fonction de la longueur. Pondération possible en fonction de la pertinence du contenu.	Note sur la longueur de la description 0 : < 50 caractères 5 : 50-199 caractères 10 : ≥ 200 caractères + Note sur la pertinence de la description 0 : Ratio FuzzyString ≥ 0,6 5 : Ratio FuzzyString entre 0,4 et 0,6 10 : Ratio FuzzyString < 0,4

Y:



Nom du champ	Coeff.	Description	Utilité et Qualité attendue	Notation
Description	5 et 5	Description plus ou moins détaillée du jeu de données	Description d'une longueur suffisante. Notation en fonction de la longueur. Pondération possible en fonction de la pertinence du contenu.	Note sur la longueur de la description 0 : < 50 caractères 5 : 50-199 caractères 10 : ≥ 200 caractères + Note sur la pertinence de la description 0 : Ratio FuzzyString ≥ 0,6 5 : Ratio FuzzyString entre 0,4 et 0,6 10 : Ratio FuzzyString < 0,4
Mots-clés	1 et 1	Saisie libre de mots-clés associés, facilitant la recherche d'un JDD	Au moins 1 mot-clé existant et 1 mot-clé absent du titre et de la description. 1 mot-clé existant signifie que les informations sont cohérentes ; 1 mot-clé absent signifie que ce mot-clé apporte une information supplémentaire	Note sur le nombre de Mots-Clés 0 : 1 MC 5 : 2 - 4 MC 10 : ≥ 5 MC + Note sur la pertinence des mots-clés 0 : 0 mots-clés 5 : Que des MC Existants ou que des MC Absents 10 : 1+ MC Existant et 1+ MC Absent
Contacts	3	Contacts renseignés pour le JDD	Orienté l'utilisateur vers une personne à contacter pour plus d'informations sur le JDD. Le mail ne doit pas être du type ...@...fr ou a@a.fr et doit correspondre à une adresse valide. On peut valoriser également la présence de plus d'un contact.	0 : Contact non-valide 5 : 1 Contact valide 10 : Plusieurs contacts renseignés valides
Mode de constitution	5	Généalogie de la donnée (comment la donnée a été produite)	Information pertinente pour déterminer son utilisation ou son domaine de validité.	0 : < 50 caractères 5 : 50-200 caractères 10 : ≥ 200 caractères
Limites d'utilisation	1	Grandes restrictions qui s'appliquent au JDD	Informe sur les restrictions et contraintes à la réutilisation de la donnée (Valeurs définies). Les limites doivent être renseignées et connues.	0 : Champ vide 5 : Conditions inconnues 10 : Autres valeurs

Y:



Préconisation d'usage	1	Recommandations avant toute utilisation de la donnée	Indique les précautions et contraintes à respecter pour réutiliser la donnée (Saisie libre). Doit être différent de « <i>Préconisations d'usage</i> » et être suffisamment fourni.	0 : < 50 caractères 5 : 50-200 caractères 10 : ≥ 200 caractères
Mention d'échelle	5	Mention d'une échelle d'utilisation ou d'acquisition des données dans la description ou le mode de constitution	Indique à quelle échelle la donnée a été acquise et/ou peut être utilisée.	0 : Aucune mention d'échelle 10 : Mention d'une échelle
Mention d'une zone géographique	5	Mention d'une zone géographique concernée dans le titre	Donne à l'utilisateur une information immédiate de la zone couverte par les données. Un JDD de données géographiques doit contenir la mention d'une zone géographique	0 : Aucune mention d'une zone géographique 10 : Mention d'une zone géographique
Fréquence de Mise à Jour	3	Fréquence à laquelle le JDD est mise à jour	Information pertinente, en lien avec la fraîcheur de la donnée. Les fréquences régulières (autre que <i>Non planifiée, Lorsque nécessaire, Irrégulière, Inconnu</i>) sont valorisées	0 : Fréquence non-régulière 10 : Fréquence régulière

Y:



2.3 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. On liste ci-dessous les champs et analyses prises en compte pour déterminer cette concordance.

Nom du champ	Coeff.	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 pour le Label Optimal . Les trois dates (<i>Publication, MAJ Métadonnées et MAJ Ressource</i>) doivent avoir un format valide (YYYYMMDD) et être datée du XXI ^e siècle (ou du XX ^e pour la MAJ Ressource). La Date de MAJ Métadonnées doit être postérieure ou égale à la date de MAJ Ressource (car la métadonnée peut être mise à jour mais pas la ressource ; en revanche, si la Ressource est mise à jour, la métadonnée doit l'être aussi).	Note sur le format des dates : 0 : L'une des trois dates 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 renseigné et l'année de la MAJ	Un millésime est exigé dans le niveau minimal si la fréquence de MAJ n'est pas régulière. Ce millésime doit correspondre à l'année de la Date de MAJ de la Ressource (à minima la date de MAJ des Métadonnées), sinon il y a contradiction et confusion possible pour l'utilisateur.	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 donne une information sur l'étendue maximale 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 Gironde → Département)	0 : Granularité ne correspond pas à la Zone Géographique 5 : Granularité Indéfinie (donc aucune concordance analysable) OU sur Parcelles, POI, IRIS 10 : Granularité correspond à la Zone Géographique
Licence	3 et 3	Concordance entre la Licence renseignée et	Assure que la donnée pourra être réutilisée et dans quel cadre (contraintes d'utilisation ou de partage, mention du producteur,	Note sur la cohérence des informations

Y:



		<p>les limites et préconisations</p>	<p>etc). Les informations contenues dans les <i>champs Limites d'utilisation</i> ou <i>Préconisation d'usage</i> doivent concorder avec la <i>Licence</i> renseignée. Si la donnée est de type <i>Données ouvertes</i>, le JDD doit être disponible en téléchargement.</p>	<p>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</p>
--	--	--------------------------------------	---	--

Ces critères de qualité des Métadonnées s'accompagnent, pour le Label Optimal, de critères sur la qualité de la Donnée géographique, décrit dans un autre document.

Y:

