

FAQ CATALOGUE DE MÉTADONNÉES DE LA PDS

Kit Documentaire de l'intégration au catalogue de métadonnées de la PDS à destination des responsables de données

Section 1 : Généralités sur le catalogue de métadonnées

Qu'est ce que les métadonnées ?	Les métadonnées sont des informations structurées qui décrivent un jeu de données (contenu, périmètre, producteur, conditions d'accès et d'usage), afin d'en permettre la découverte, la compréhension et la réutilisation, sans accéder aux données elles-mêmes. Elles peuvent être considérées comme des éléments clés de la <i>fiche d'identité</i> d'une base de données.
Est-ce que les métadonnées contiennent des données personnelles ou sensibles?	Les métadonnées peuvent décrire des jeux de données contenant des données personnelles ou sensibles, sans inclure elles-mêmes d'informations identifiantes. Leur objectif est d'expliquer la nature, le périmètre et les conditions d'accès aux données, conformément aux exigences de l'EHDS et du standard HealthDCAT-AP.
A quoi sert un catalogue de métadonnées?	Le catalogue de métadonnées est un outil permettant de recenser et de publier des descriptions standardisées de bases de données. Il facilite leur découverte, leur compréhension et leur comparabilité. Il s'appuie sur des standards européens permettant de rendre les données FAIR (faciles à trouver, accessibles, interopérables et réutilisables), et clarifie les conditions d'accès et de réutilisation, sans donner accès aux données elles-mêmes.
Quelles sont les grandes catégories de métadonnées ?	Dans le cadre des standards européens (DCAT-AP et HealthDCAT-AP), les métadonnées couvrent notamment : <ul style="list-style-type: none"> • la description du jeu de données (titre, description, thématique, période, population) • les conditions d'accès et d'utilisation (niveau d'accès, restrictions, cadre juridique) • les informations sur le producteur et le détenteur des données • des éléments techniques de description (formats, structures générales)
Quelles informations peuvent être retrouvées dans les métadonnées ?	Les métadonnées renseignent le contexte originel de la collecte et de la production des données, le périmètre et le contenu global, les responsabilités et la gouvernance des données, la traçabilité, les modalités d'accès et d'utilisation, et le type de tables / variables disponibles.
Est-ce que je peux utiliser le catalogue sans être expert technique ?	Oui, l'interface est conçue pour être simple, intuitive et est destinée aux utilisateurs. Vous pouvez rechercher des jeux de données et consulter leurs métadonnées sans compétences techniques.

Comment accéder au catalogue de métadonnées de la PDS ?	Le catalogue est disponible sur le site internet de la PDS.
Comment accéder à l'éditeur du formulaire de métadonnées de la PDS ?	L'éditeur du formulaire de métadonnées, spécifiquement conçu pour générer le formulaire de métadonnées selon les exigences européennes, est disponible en ligne .
Comment accéder au catalogue de métadonnées en marque blanche et à ses codes source ?	Toutes les informations nécessaires au déploiement du catalogue de métadonnées en marque blanche sont disponibles sur le Gitlab de la PDS.

Section 2 : La standard de métadonnées de santé utilisé au sein du catalogue de métadonnées : HealthDCAT-AP

Qu'est ce que le HealthDCAT-AP ?	<p>Le HealthDCAT-AP (Health Data CAtalog VocabulArY Application Profile) un profil d'application du standard DCAT-AP, destinée à la description standardisée de jeux de données de santé dans des catalogues de métadonnées.</p> <p>Il s'inscrit dans le cadre de l'Espace Européen des Données de Santé (EHDS) et vise à améliorer la découvrabilité, la compréhension et l'interopérabilité des jeux de données en vue de leur réutilisation secondaire (recherche, innovation, politiques publiques).</p>
Quelle est la version actuelle du HealthDCAT-AP utilisée dans le catalogue de la PDS ?	Le modèle utilisé par la PDS est aligné avec les informations et les exigences précisées dans la documentation HDCAT-AP officielle version 6 .
Quelles sont les exigences de vocabulaire utilisé dans le catalogue de métadonnées ?	<p>Le catalogue s'appuie sur les vocabulaires contrôlés définis par le standard HealthDCAT-AP (version 6) afin d'assurer l'interopérabilité et la comparabilité des descriptions de jeux de données au niveau européen.</p> <p>Des recommandations complémentaires issues des travaux européens (notamment TEHDAS) peuvent guider leur mise en œuvre, sans se substituer au standard.</p>
Disposez-vous d'une correspondance entre les dénominations des métadonnées dans HDCAT-AP et celles figurant dans l'outil/fichier du HDH ?	Le mapping est présenté dans le guide utilisateur , disponible dans le Gitlab contenant les codes source de la PDS pour le développement en marque blanche du catalogue de métadonnées. A noter que toutes les métadonnées présentes dans la version implémentée par le HDH proviennent du standard HDCAT-AP (pas de métadonnées "custom").
Comment être informé des changements de versions du modèle et	HealthDCAT-AP est un standard européen évolutif, dont les mises à jour sont publiées par la Commission européenne. À ce stade, la version actuellement utilisée constitue une base stable pour la description des

comment gérer ces évolutions ?	jeux de données. Les évolutions futures du modèle seront prises en compte dans le cadre des mises à jour du catalogue et communiquées via les liens officiels
Le modèle de la PDS intègre-t-il des métadonnées personnalisées ou spécifiques à la PDS ?	Nous nous conformons rigoureusement au format HDCAT-AP. Des adaptations ou enrichissements limités pourront être proposés (par exemple, pour la lisibilité ou les traductions), sans remettre en cause la conformité ni l'interopérabilité européenne du catalogue..
Les utilisateurs ont-ils été consultés afin de définir les besoins et les usages du modèle HDCAT-AP ?	Le modèle HealthDCAT-AP a été élaboré dans le cadre de travaux européens associant les autorités compétentes, certains experts nationaux et parties prenantes institutionnelles des États membres, notamment via les projets TEHDAS. Des versions intermédiaires ont fait l'objet de consultations publiques au niveau européen.

Section 3 : L'espace européen des données de santé EEDS (EHDS)

Qu'est ce que l'EEDS?	<p>L'Espace européen des données de santé EEDS est le premier espace de données constitué d'un ensemble de règles, d'infrastructures et d'un cadre de gouvernance harmonisés, sécurisés et transparents afin de faciliter l'utilisation primaire et secondaire des données de santé à l'échelle européenne.</p> <p>L'usage primaire des données de santé permettra d'améliorer les diagnostics et les traitements grâce aux données de santé européennes, la sécurité patient, de permettre une meilleure continuité et efficacité des soins.</p> <p>L'usage secondaire des données de santé permettra d'améliorer les politiques de santé publique, d'accélérer la recherche médicale et de développer l'innovation en santé.</p>
Quel est le lien entre l'EEDS et le catalogue de métadonnées ?	<p>L'Espace européen des données de santé (EEDS) vise à mettre en place un cadre commun et harmonisé pour l'utilisation et l'échange des données de santé au niveau européen. Le règlement EEDS prévoit des exigences de transparence sur les données disponibles et leur qualité via les catalogues de métadonnées.</p> <p>L'impact principal sur les détenteurs de données est que le partage de la documentation des données dans un catalogue devient une étape obligatoire pour garantir une réutilisation sécurisée des données d'ici 2029 (date d'entrée en application).</p> <p>L'obligation de mettre ces données à disposition à des fins de réutilisation pour certains détenteurs de données, avec les exigences de transparence via les catalogues de métadonnées, est prévue dans le règlement EEDS (Chapitre IV, Section 6, article 77).</p>

Section 4 : L'interface du catalogue de métadonnées (Front-End)

Comment obtenir un modèle de formulaire de métadonnées ?	Un formulaire de saisie des métadonnées est mis à disposition via l'éditeur en ligne . Il peut être complété hors ligne et téléversé ultérieurement dans l'éditeur pour validation une fois complété.
Est-il envisageable d'utiliser le formulaire XLSX comme référence ou représentation du modèle de métadonnées de la PDS ?	Le formulaire XLSX et l'éditeur en ligne constituent la référence pour comprendre le modèle de métadonnées. Ils intègrent l'ensemble des métadonnées obligatoires du modèle HDCAT-AP et une sélection des métadonnées recommandées et optionnelles. Le standard de référence demeure le modèle HealthDCAT-AP tel que défini au niveau européen.
Est ce que toutes les métadonnées du formulaire sont apparentes dans le catalogue de métadonnées de la PDS?	Certaines métadonnées saisies dans le formulaire sont utilisées à des fins techniques ou de gestion interne du catalogue et ne sont pas destinées à être exposées dans l'interface publique. C'est le cas des métadonnées relatives à la section "catalogue" dans le formulaire XLSX, gérées par la PDS, qui ne sont pas apparentes mais sont énumérées dans le guide partagé en marque blanche .
A quel point est-il possible de s'éloigner du modèle de métadonnées de la PDS dans le choix de ses métadonnées?	Il est recommandé de rester conforme le plus possible au modèle de métadonnées HDCAT-AP, nécessaire pour garantir l'interopérabilité des métadonnées au niveau européen. Afin de passer les tests d'interconnexion avec le catalogue européen (soumis à une validation stricte), il est nécessaire que l'ensemble du fichier respecte les champs obligatoires. Ce mécanisme permet de garantir la qualité et la conformité des métadonnées, tout en évitant les erreurs de saisie ou de structure. Des métadonnées complémentaires peuvent être ajoutées au niveau local pour répondre à des besoins spécifiques, mais elles ne sont pas prises en compte dans les échanges européens ni remontées au niveau EHDS si elles ne sont pas conformes au standard HealthDCAT-AP.

Section 5 : Le fonctionnement du catalogue de métadonnées (Back-End)

Quels sont les prérequis pour l'installation en local ?	Pour une installation en local ou en production, il est nécessaire de disposer d'un environnement permettant l'installation de Python, Docker, Docker Compose et Make & d'un stockage de fichiers (objet), qui peut être local pour les tests. L'utilisation d'un service de stockage objet de type S3 (cloud) n'est pas requise pour une installation en local.
Quels sont les prérequis pour l'installation en production ?	Il faudra un cluster kubernetes, un nom de domaine sécurisé (HTTPS), un service de stockage objet (type S3) et des blocs de stockage (pour la résilience de la base de données, keycloak ainsi que les fichiers de customisation).
La PDS fournit-elle une configuration S3 permettant simplement de lancer l'application ?	Non, chaque organisation devra configurer son propre service S3 (Amazon, MinIO, etc.). La documentation disponible décrit les variables d'environnement requises par l'application pour qu'elle puisse se connecter à un service de stockage de fichiers.

Quelles sont les étapes à suivre pour lancer l'application ?	<p>La procédure est la suivante :</p> <ul style="list-style-type: none"> • Créer votre propre bucket chez un fournisseur proposant un stockage compatible S3. • Obtenir les informations de connexion de ce bucket : la région, l'URL du service, le nom du bucket, ainsi qu'une clé d'accès et une clé secrète. • Utiliser ces informations pour définir les variables d'environnement (S3_REGION, S3_URL, S3_ACCESS_KEY, etc.) avant de lancer la commande make docker-compose-prod-run.
Comment les métadonnées sont-elles stockées ?	Elles sont enregistrées dans un format RDF (standard du web sémantique) et gérées par Fuseki, ce qui garantit l'interopérabilité et la conformité avec les standards DCAT.
Quelles sont les composantes techniques du catalogue ?	<p>Le catalogue repose sur plusieurs services :</p> <ul style="list-style-type: none"> • API (FastAPI) pour la logique métier • Front-office (React) pour la recherche et la visualisation • Back-office pour la saisie des métadonnées • Keycloak pour la gestion des utilisateurs et rôles • Fuseki pour le stockage RDF des métadonnées • SHACL Validator pour la validation des métadonnées • S3 (ou équivalent) pour le stockage des fichiers en production
Pourquoi Docker a été choisi ?	Docker a été choisi pour simplifier le déploiement et garantir un comportement homogène de l'application quel que soit l'environnement (local, test ou production).. Sans Docker, il faudrait installer et configurer chaque composant manuellement (API, interface, Keycloak, Fuseki), ce qui est complexe.
Quels sont les besoins en ressources pour la production ?	<p>Cela est dépendant du volume de métadonnées. Toutefois, il convient de prévoir :</p> <ul style="list-style-type: none"> • CPU : 2-4 cœurs • RAM : 8 Go minimum • Stockage : dépend des fichiers, mais 50 Go est une bonne base
Comment mettre à jour l'application ?	Avec Docker, il suffit de récupérer les nouvelles images et relancer les conteneurs. Pas besoin de réinstaller manuellement.
Est-ce que je peux personnaliser l'application ?	Oui, un dossier "personnalisation" (customization) permet de changer le logo, les couleurs, les mentions légales et la politique de confidentialité sans toucher au code, vous permettant une bonne flexibilité d'adaptation.