poster_utilisation_dcat_insee.qxp_Mise en page 1 15/11/2024 17:46 Page 1

UTILISATION DU STANDARD DCAT À L'INSEE

Guillaume Duffes

Institut national de la statistique et des études économiques (Insee)

RÉSUMÉ

L'Institut national de la statistique et des études économiques (Insee) joue un rôle crucial en collectant, produisant, analysant et diffusant de nombreuses informations sur l'économie et la société françaises et doit assurer la visibilité et la découvrabilité de ses données. L'Insee a recours au vocabulaire DCAT pour ses jeux de données qui constituent le point d'entrée principal des résultats statistiques, Ainsi l'Insee s'est doté de son propre outil de gestion de ces objets DCAT, l'application Bauhaus, publiée en open source.

L'Insee, diffuseur de données

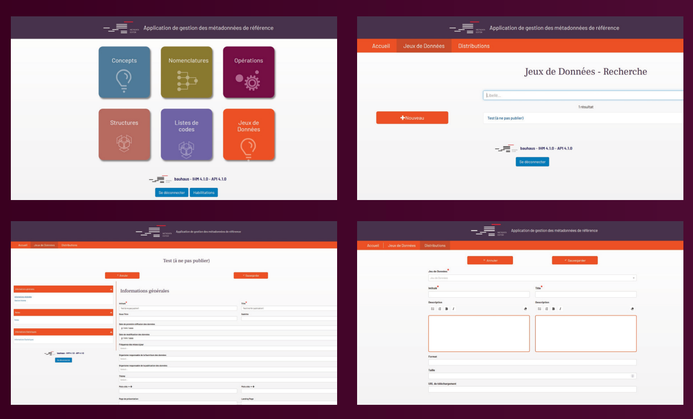
L'Insee collecte, produit, analyse et diffuse des statistiques et des études sur l'économie et la société françaises. Découvrez la représentation illustrée de la variété des enquêtes menées par l'Insee en 2023.

Bauhaus, outil de gestion des objets DCAT

Pour faciliter la transition vers l'interopérabilité avec les portails nationaux et européens, l'Insee s'est doté de son propre outil de gestion de métadonnées DCAT, l'application Bauhaus, publiée en open source.

Cet outil a été conçu pour que les étapes de conception et de construction de ces métadonnées ne soient plus réservées aux experts du standard. Grâce à Bauhaus, l'Insee peut gérer efficacement ses métadonnées, rendant la documentation et l'accès aux jeux de données plus simples et plus accessibles pour une plus large catégorie d'agents. En intégrant ces technologies avancées, l'Insee renforce son engagement à améliorer l'accessibilité et la visibilité de ses données statistiques, répondant ainsi aux besoins croissants des utilisateurs en quête d'informations structurées, fiables et interopérables.

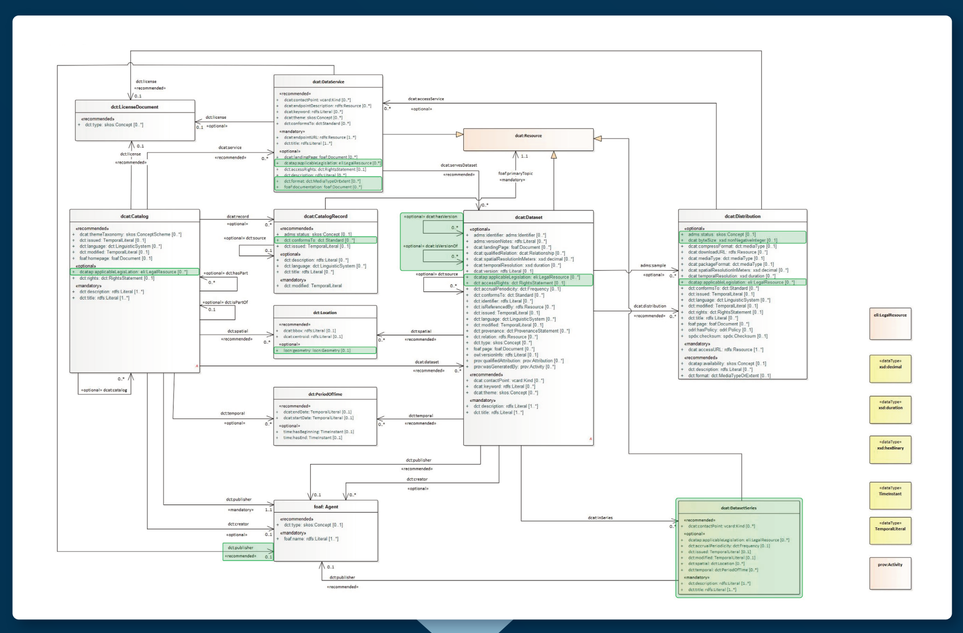
Démonstration :
https://gestion-metadonnees-front.demo.insee.io/

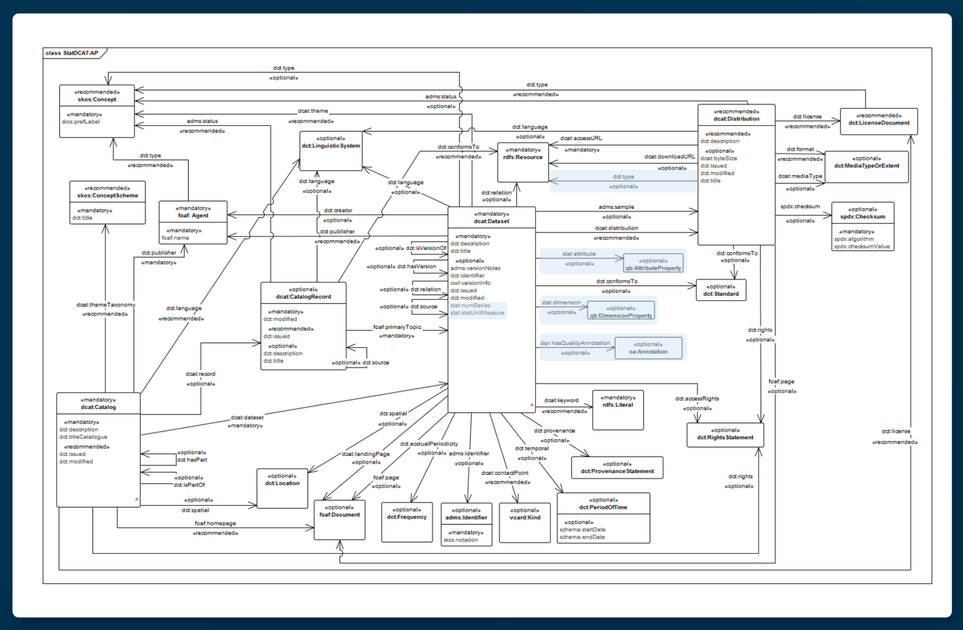


DCAT et les profils liés

L'Insee a recours aux vocabulaires RDF pour gérer une grande partie de ses métadonnées relatives aux processus métiers. Le jeu de données constituant le point d'entrée principal des résultats statistiques, il est donc essentiel que sa description soit facilement accessible et interopérable avec d'autres catalogues de données.

Afin d'atteindre cet objectif, l'Insee a adopté le Data Catalog Vocabulary (DCAT), un standard du web sémantique publié par le W3C. En s'appuyant sur des classes permettant de définir un catalogue (Catalog), un jeu de données (Dataset) ou encore un fichier (Distribution, représentant un moyen d'accéder à un Dataset, par exemple via un fichier CSV téléchargeable) et sur des extensions du standard DCAT, dont le profil d'application DCAT-AP, et plus particulièrement son extension pour les données statistiques StatDCAT-AP, publiées par la Commission Européenne, l'Insee vise des premières publications compatibles avec ces standards d'ici fin 2024. Cela lui permettra de rendre ses jeux de données interopérables avec des portails comme la plateforme ouverte des données publiques françaises (data.gouv.fr) ou le portail officiel des données européennes (data.europa.eu), augmentant ainsi la visibilité de ses travaux statistiques.





Triplets RDF publiés

Exemple de triplets DCAT

Prochaines étapes

- Le vocabulaire *DCAT* et ses profils d'application sont riches, seule une partie très minoritaire a été utilisée dans cette première mise en oeuvre.
- Utilisation des prédicats *stat:dimension* et *stat:attribute* de *statDCAT-AP*.
- Lier les rapports qualité au *dcat:Dataset* grâce à *dqv:hasQualityAnnotation* de *Data Quality Vocabulary (DQV)*.
- Ajouter les métadonnées géographiques essentielles dans de nombreux catalogues de jeux de données grâce à *geoDCAT-AP*.