

DataMed est un prototype de **moteur de recherche de données scientifiques dans le domaine biomédical**. Financé par le NIH dans le cadre du projet BioCADDIE, ce moteur de recherche open source facilite la découverte de données de recherche, en permettant, **en une seule requête**, de **chercher dans un large spectre d'entrepôts de données** en sciences de la vie.

DataMed indexe à ce jour **49 entrepôts ou agrégateurs de données**, l'objectif étant d'en indexer **environ 150**. La liste des entrepôts qui seront couverts par DataMed est déjà définie mais l'équipe est à l'écoute de toute suggestion. Un large éventail de types de données est déjà référencé par DataMed : essais cliniques, images, expression des gènes, spectrométrie de masse, séquences, structures, signaux physiologiques, protéomique...

**DataMed a pour ambition d'être, pour les données, ce que PubMed est pour la littérature scientifique**. La difficulté avec les données de recherche réside dans la multiplicité et la diversité des entrepôts de données et des standards de métadonnées. Le premier challenge du projet a donc consisté à définir un modèle unifié de métadonnées, appelé DatA Tag Suite (DATS).

Une fois la requête lancée dans le moteur de recherche simple ou avancé, DataMed fournit une page de résultats listant les **jeux de données pertinents** par rapport aux mots-clés saisis. Ces résultats peuvent être filtrés par année, type de données, nom de l'entrepôt source et condition d'accès. Cliquer sur un résultat ne donne pas accès directement à l'entrepôt source mais à une page intermédiaire affichant quelques métadonnées décrivant le jeu de données. Cette page fournit bien évidemment un lien pour **accéder à l'entrepôt source** et aux fichiers de données.

Pour aller plus loin :

- Ohno-Machado, L. *et al.* **Finding useful data across multiple biomedical data repositories using DataMed**. *Nat Genet* **49**, 816–819 (2017). <https://doi.org/10.1038/ng.3864>
- Chen, X. *et al.* **DataMed – an open source discovery index for finding biomedical datasets**. *Journal of the American Medical Informatics Association* **25**, 300–308 (2018). <https://doi.org/10.1093/jamia/ocx121>
- Lire aussi notre article : [Comment trouver des données et composants logiciels à réutiliser ?](#)