

POINTS FORTS

Pour tous types de données

- Données privées.
- Données publiques (*Open Data, Linked Open Data*).
- Tous formats : RDF, JSON, Excel, CSV, PDF, etc.
- Tous modes de distribution : SPARQL, API, Torrent, site Web, etc.

Pour tous types d'utilisateurs

- Les producteurs de données peuvent décrire leurs données au travers de requêtes.
- Les administrateurs de bases de données peuvent optimiser leurs requêtes et leurs bases de données.
- Les développeurs peuvent réutiliser leurs requêtes dans leurs applications en 1 clic.
- Les Data scientists peuvent obtenir sans effort un environnement de travail avec SPARK, Flink, Kafka, etc.

Pour toutes les phases de la vie des données

- Les utilisateurs des données et des requêtes sont alertés en cas de modifications.

Respect des standards du Web des données (*Linked Data, W3C*)

GAINS

Amélioration de la collaboration
entre les producteurs et les consommateurs de données.

Amélioration du partage des connaissances.

Augmentation de la productivité des *data scientists*.



BorderCloud

Centre de formation - Éditeur de logiciels - Conseils
215, rue Jean-Jacques Rousseau, 92130 Issy-les-Moulineaux - FRANCE
Tel. +33 (0)6 89 15 10 64 - contact@bordercloud.com

www.bordercloud.com

Plateforme LinkedWiki

Un seul logiciel pour la réutilisation des données



Depuis janvier 2016, le Center for Data Science de l'université Paris Saclay utilise la plateforme LinkedWiki afin d'aider les scientifiques à découvrir et à réutiliser les données.

<https://io.datascience-paris-saclay.fr>
Version ouverte : <http://linkedwiki.com>

Le logiciel de la plateforme **LinkedWiki** permet de découvrir, réutiliser, analyser et partager plus simplement les données au sein de votre organisation. Si les données respectent les standards du Web des données, le logiciel facilite le développement et le partage des requêtes d'accès aux données. Connecté à un cloud, il permet de gérer également les environnements de développement des *data scientists* afin de simplifier leur travail. Ce logiciel est l'outil indispensable au sein d'un système d'information, afin d'aider les utilisateurs à réutiliser leurs données.

LA PROBLÉMATIQUE

Des *Smart Cities* au *Big Data* en passant par l'*Open Data*, les données sont partout et elles sont de plus en plus accessibles. Ce raz-de-marée ouvre la voie à une nouvelle génération d'applications et de services, que les organisations doivent commencer à construire.

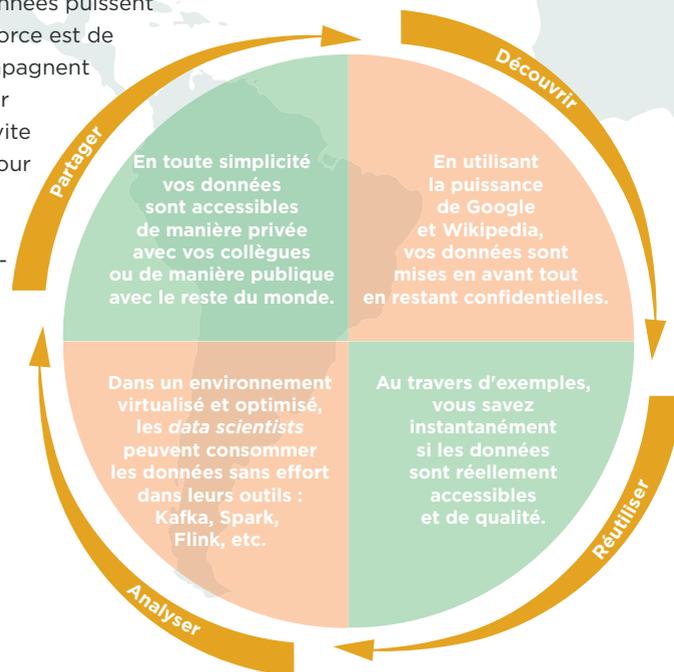
Si séduisantes que ces données puissent paraître de prime abord, force est de constater qu'elles s'accompagnent de contraintes quant à leur réutilisation, qui peuvent vite s'avérer insurmontables pour bon nombre d'organisations. Est-ce à dire qu'il faut renoncer à la réutilisation des données ? Non, mais une approche plus centrée sur l'utilisation des données est indispensable.

Quelles sont les difficultés auxquelles sont confrontés les utilisateurs qui doivent réutiliser les données disponibles au sein de leurs organisations ?

- 1 la capacité à découvrir les données disponibles.
- 2 l'utilisation de données avec peu de documentation.
- 3 transformer ces données en permettant la reproductibilité des traitements.
- 4 mettre à disposition ces données de manière à les réutiliser simplement.

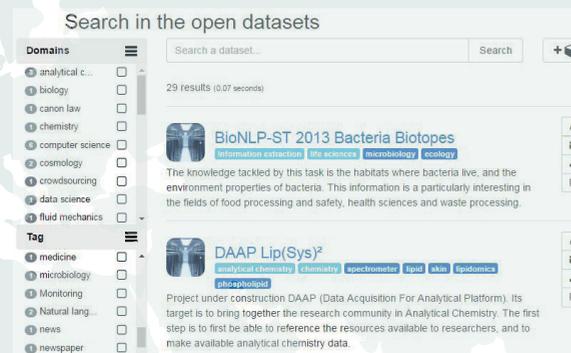
LA SOLUTION

Contrairement à tous les outils actuels du marché intégralement centrés sur l'analyse des données ou la mise à disposition de fichiers, la plateforme **LinkedWiki** est à l'inverse centrée sur la réutilisation des données. Ceci lui confère des caractéristiques uniques lui permettant de créer un cercle vertueux au sein d'une organisation.



PARTAGER

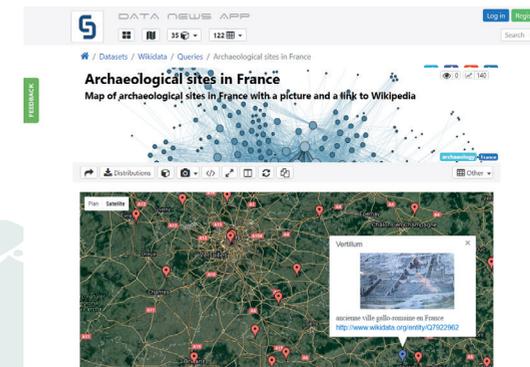
Le logiciel peut fonctionner au sein d'un intranet ou d'un extranet. Les utilisateurs du logiciel peuvent déclarer les sources des données qu'ils utilisent ainsi que leurs requêtes d'interrogation. L'utilisateur peut rendre visibles ou non aux autres utilisateurs du service les sources de données et les requêtes qu'il utilise.



RÉUTILISER

L'utilisateur peut rechercher et télécharger intégralement une source de données à travers sa description. Chaque source peut être utilisée à l'aide d'exemples qui permettent d'identifier l'accessibilité et la qualité de cette source. Après avoir sélectionné une source de données, l'utilisateur peut réutiliser simplement les données à l'aide d'exemples en Python, Java, R, Ruby, HTML, Javascript, PHP, etc.

Data Scientist software environment (V1.0)



DÉCOUVRIR

À travers un simple formulaire, les utilisateurs peuvent lier leurs sources de données entre elles. Le référentiel utilisé est celui de Wikidata, qui est la base de données généraliste la plus dynamique au monde. L'utilisateur peut ensuite retrouver ses sources de données et ses exemples de requêtes à travers une interface de recherche ou directement au travers de Google et Wikipedia.

Reuse these data in your code

Query, endpoint and code for reusing the same data



ANALYSER

Le logiciel propose aux *data scientists* un environnement virtualisé et optimisé afin de les aider à consommer les données sans effort au travers des standards du marché : Kafka, Spark, Flink, etc. De plus, la gestion virtualisée de leurs environnements facilite la sauvegarde et la reproductibilité des traitements au sein de leur organisation. Les *data scientists* peuvent ainsi se concentrer sur leur cœur de métier.