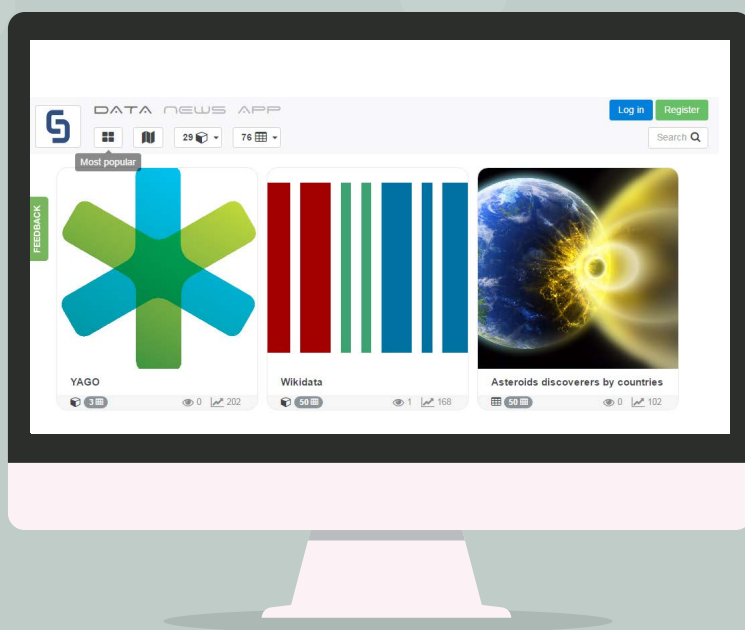
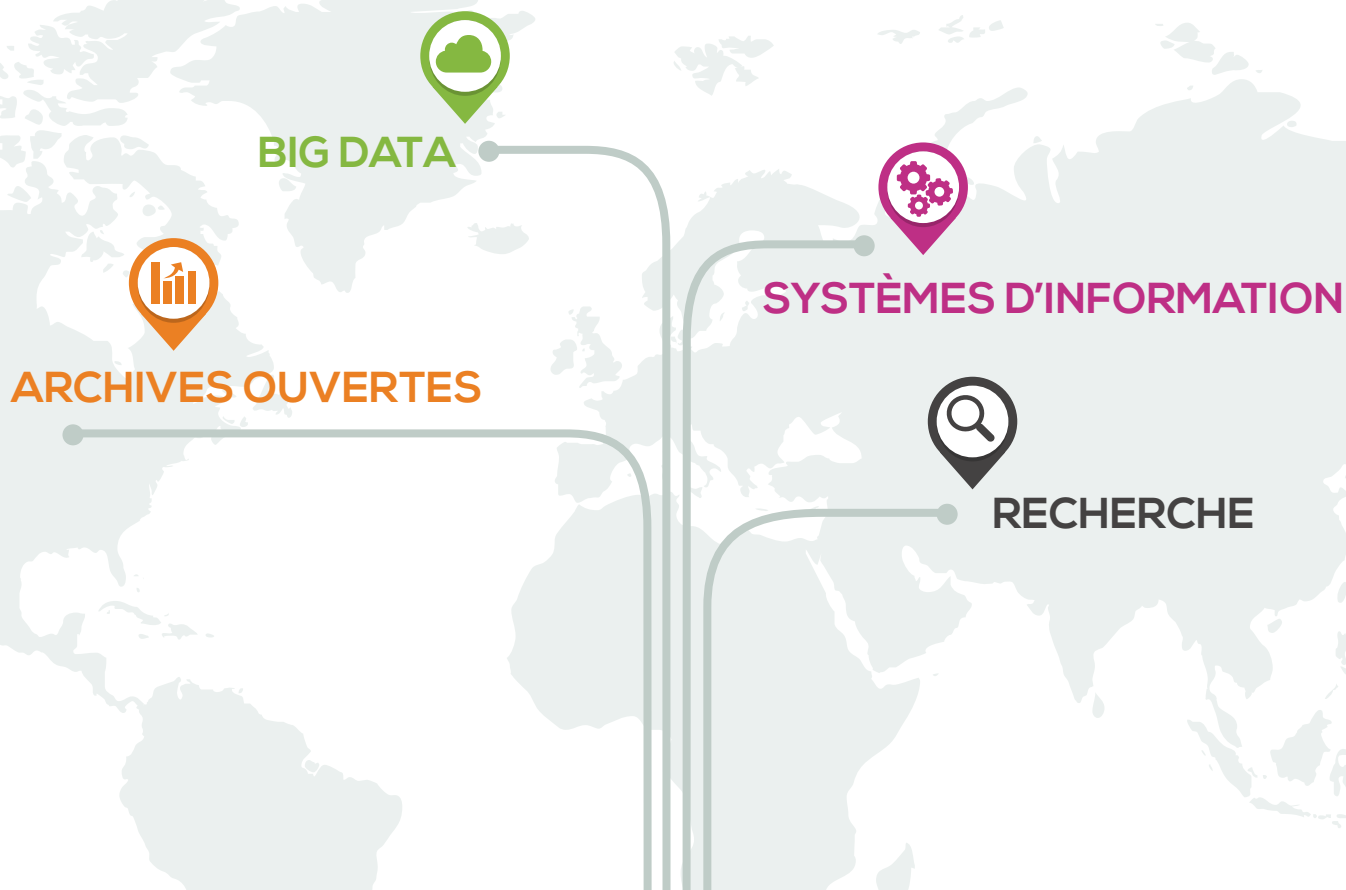


PLATEFORME LINKEDWIKI

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 de réutiliser les données entre les chercheurs :

<https://io.datascience-paris-saclay.fr>

Le logiciel de la plateforme **LinkedWiki** permet aux acteurs d'une organisation de réutiliser plus simplement les données qui y sont disponibles et présentes sur le Web. Ce logiciel permet de déclarer les sources de données accessibles, ainsi que la manière d'y accéder. Si les sources de données respectent les standards du Web des données, le logiciel facilite aussi le développement et le partage des requêtes d'accès aux données. 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 ?

- 01 la capacité à découvrir les données disponibles.
- 02 l'utilisation de données avec peu de documentation.
- 03 transformer ces données afin de les utiliser dans leurs applications.

LA SOLUTION

La révolution de la plateforme LinkedWiki se situe dans le fait que, 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 :

- Elle facilite la déclaration des données.
- Elle permet d'illustrer par l'exemple la manière de réutiliser les données.
- Elle facilite la recherche des données au travers d'un référentiel unique.
- Elle permet de tester concrètement si les données sont de qualité et si elles sont réellement accessibles.

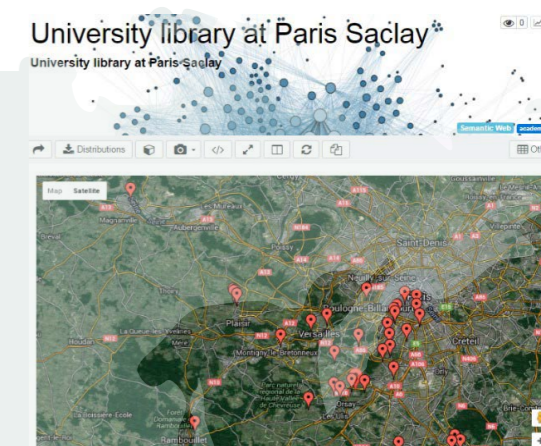
FONCTIONNEMENT

DÉCLARATION

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 afin de déterminer plus facilement les requêtes d'interrogation de ces sources de données. L'utilisateur peut rendre visibles ou non aux autres utilisateurs du service les sources de données et les requêtes qu'il utilise.

ILLUSTRATION PAR L'EXEMPLE

L'utilisateur peut facilement illustrer les sources de données à l'aide de requêtes d'interrogation qui peuvent générer des graphiques.



COMPARAISON

L'utilisateur peut facilement écrire des requêtes SPARQL ou rejouer les requêtes écrites par ses collègues. Le résultat d'une requête permet d'identifier immédiatement les problèmes de qualité de service (QoS) et la qualité réelle des données disponibles.

```

6 select ?countryLabel (COUNT(?asteroid) as ?nb)
7 where {
8   ?asteroid wdt:P31 wd:Q3863 ;
9   wdt:P61 ?discoverer .
10  ?discoverer wdt:P27 ?country .
11  ?discoverer wdt:P27(?pays de nationalité) ?country .

```

RÉUTILISATION

L'utilisateur peut rechercher et télécharger intégralement une source de données à travers sa description.

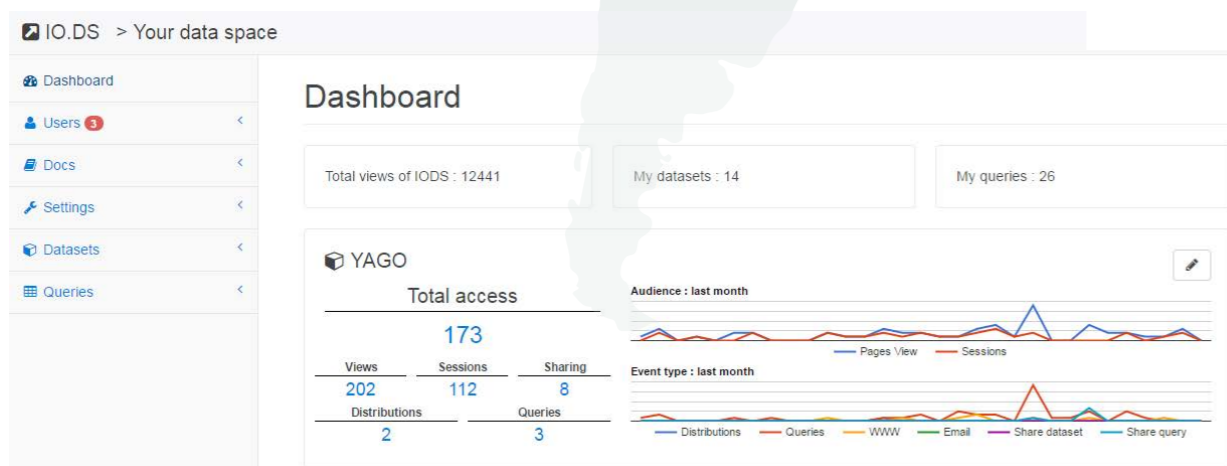
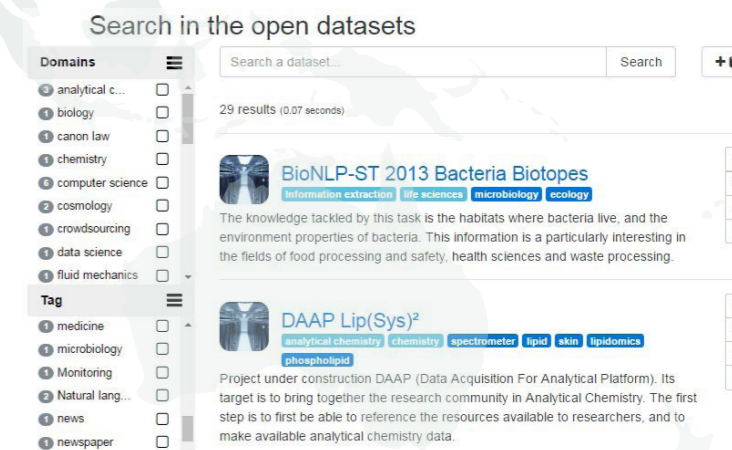
Si une source de données respecte les recommandations du W3C avec le protocole SPARQL, un utilisateur peut facilement rechercher des données ou des requêtes sur le service et réutiliser rapidement ces données dans son application, en recopiant les codes d'exemples en Python, R, Ruby, SPARQL, HTML, Javascript, PHP, etc.

À travers des notifications, l'utilisateur qui emploie une requête sera alerté si cette requête change.

DÉCOUVERTE

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.





POINTS FORTS

Pour tous types de données

- Données privées.
- Données publiques (Open Data, Linked Open Data).
- N'importe quel format : RDF, JSON, Excel, CSV, PDF, etc.
- N'importe quel mode 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.

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 dans une organisation.
- Partage plus efficace des connaissances sur les données (internes ou externes) et la manière de les utiliser.
- Augmentation de la productivité des analystes, qui n'ont plus à modifier le format des données avant de pouvoir les utiliser.



BorderCloud

Centre de formation – Editeur 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