DataSud Documentation

IDGO

avr. 25, 2019

Table des matières

1 Espace consultation **3**
- 1.1 Rechercher des données sur DataSud . . . 3

2 Espace utilisateurs **7**
- 2.1 S'inscrire sur DataSud . . . 7
- 2.2 Participer à DataSud . . . 9

3 Espace contributeurs **13**
- 3.1 Devenir Contributeur et Référent pour une organisation . . . 13
- 3.2 Créer une Organisation . . . 14
- 3.3 Editer la page d'une Organisation . . . 16
- 3.4 Publier un jeu de données . . . 17
- 3.5 Datastore et données intelligentes . . . 23
- 3.6 Géolocalisation des données tabulaires (XLS et CSV) . . . 24
- 3.7 Renseigner les métadonnées INSPIRE . . . 24

4 Espace développeurs **25**
- 4.1 Service API Ckan . . . 25
- 4.2 Requêter l'API CKAN Catalogue . . . 25
- 4.3 Requêter l'API CKAN DATA . . . 26
- 4.4 Construire une requête pour l'API . . . 27
- 4.5 Versions de l'API . . . 28
- 4.6 Authentification et clés . . . 28
- 4.7 Support JSONP . . . 29
- 4.8 Service WMS . . . 29
- 4.9 Service WFS . . . 29
- 4.10 Service CSW . . . 29
- 4.11 Marque blanche / Widget . . . 29

5 Cadre légal et réglementaire **33**
- 5.1 Le statut juridique des données publiées sur DataSud . . . 33
- 5.2 Connaître les Licences . . . 33
- 5.3 Faire une demande d'accès à un document administratif ou à des données . . . 34

6 Régles éditoriales **37**
- 6.1 Règles de nommage des ressources . . . 37
- 6.2 Amélioration des champs descriptifs avec le langage Markdown . . . 37

6.3 Les outils pour nettoyer vos données 39

Cette documentation est organisée en plusieurs guides, chacun correspondant à un usage particulier de la plateforme. Le catalogue des données et des ressources disponibles sont amenés à évoluer avec la version 2 de DataSud disponible depuis le 13 mars 2018.

CHAPITRE 1

Espace consultation

DataSud est un catalogue ouvert à tous : https://trouver.datasud.fr

Les consultations des données ouvertes, géographiques et intelligentes sont libres sur DataSud. Vous pouvez parcourir le catalogue, rechercher des jeux de données et télécharger des ressources dans différents format de fichier.

Ceci ne concerne pas les données diffusées sur accès retreints, pour lesquelles les producteurs ont volontairement limité leur téléchargement à certains utilisateurs. Toutefois ces jeux de données apparaissent au catalogue général pour porter à connaissance des publics l'existence de ces données. Pour les consulter il faut en faire la demande directement au producteur.

1.1 Rechercher des données sur DataSud

Pour de meilleurs résultats, DataSud permet de « filtrer » les données, d'effectuer des recherches par thématique, selon la fréquence de mise à jour, par format ou uniquement les jeux de données associés à une organisation.

L'ensemble de ces « filtres » peuvent être cumulés pour affiner les résultats avec un moteur de recherche "textuel"

- >> *Catalogue de données DataSud* https://trouver.datasud.fr/dataset
- >> *Liste des organisations DataSud* https://trouver.datasud.fr/organization
- >> *Liste des thématiques DataSud* https://trouverc.datasud.fr/group
- >> *Liste des réutilisations recensées à partir des données publiées DataSud* https://trouver.datasud.fr/showcase

Il n'y a pas d'inscription préalable pour accéder aux jeux de données et aux ressources diffusés en « Open Data ».

Dans le respect des conditions générales d'utilisation de DataSud, chaque jeux de données est publié avec une licence (licences ouvertes, licence odbl, etc…), choisie par le producteur de la donnée, dans le but de définir les conditions de leur réutilisation.

« Voir le passage sur le Cadre légal et réglementaire »

Si vous recherchez un jeu de donnée qui ne figure pas au catalogue de DataSud, vous pouvez utiliser le service de « demande de données » Toute demande sera étudiée et une réponse vous sera apportée. Nous relayerons le cas échéant votre demande à la collectivité ou à l'organisme concerné.

Vous pouvez déposer une demande d'accès à un document administratif ou à un jeu de données relatif au territoire régional à l'adresse suivante : https ://www.datasud.fr/demande-de-donnees

« Voir le passage sur la demande de documents administratif »

CHAPITRE 2

Espace utilisateurs

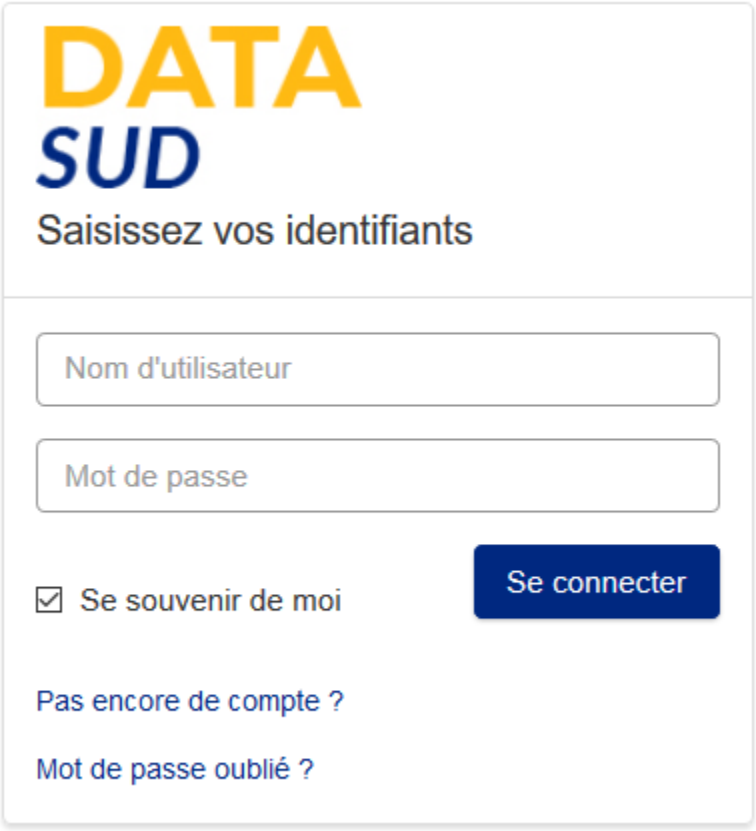
S'il n'est pas nécessaire de s'inscrire sur la plateforme pour consulter le catalogue et télécharger des données ouvertes, le fait de s'enregistrer sur DataSud https://publier.datasud.fr permet de disposer des fonctionnalités complémentaires par rapport à la consultation sans inscription.

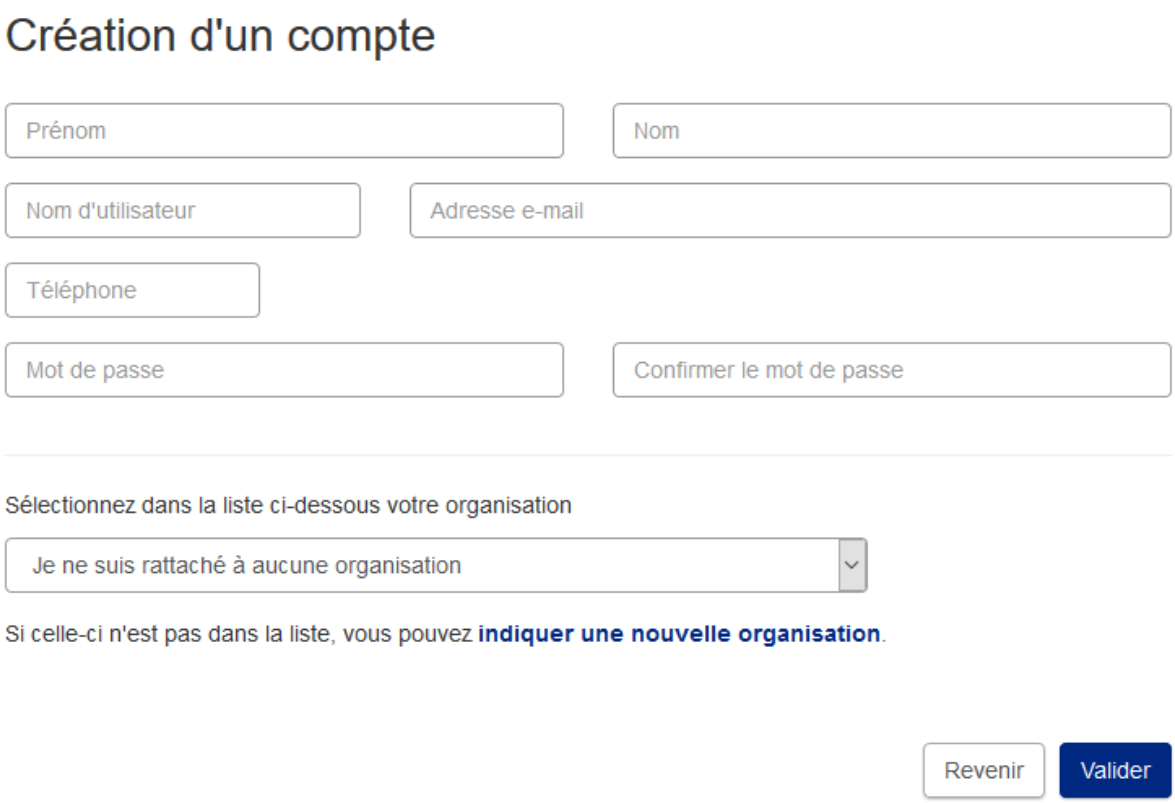
Note : Toute personne, morale ou physique, publique ou privée, peut s'inscrire sur DataSud et ainsi contribuer à l'ouverture et la mise en commun des données publiques ou privées, en publiant des jeux de données, des textes, des ressources et des commentaires.

2.1 S'inscrire sur DataSud

— L'utilisateur enregistre son identité qui est distincte de la personne morale qu'il représente.
— En s'inscrivant, l'Utilisateur crée un profil sur la Plateforme.
— En s'inscrivant, l'Utilisateur accepte les « conditions d'utilisation »

Note : *Le nom d'utilisateur doit contenir uniquement des caractères alphanumériques en minuscules (ascii) et ces symboles : -_*


DATA
SUD
Saisissez vos identifiants
Nom d'utilisateur
Mot de passe
Se souvenir de moi
Se connecter
Pas encore de compte ?
Mot de passe oublié ?

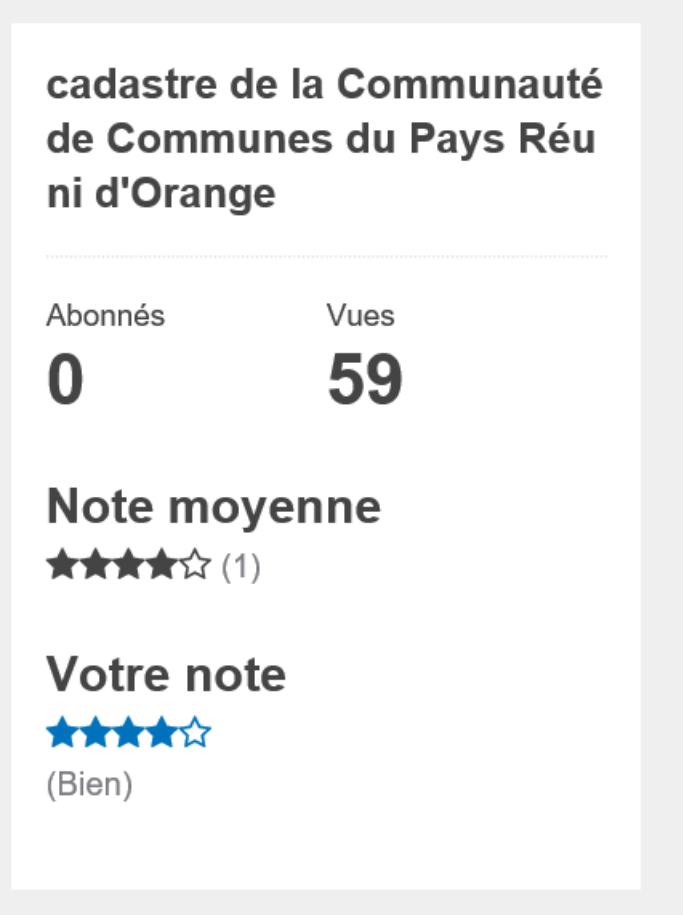


— l'Utilisateur doit ensuite valider son inscription en cliquant sur le lien reçu par courriel, sur sa boite aux lettres de courrier electronique.

2.2 Participer à DataSud

De nombreuses fonctionnalités participatives sont proposées :

— Faire une demande de donnée : https://www.datasud.fr/demande-de-donnees/
— Contacter les administrateurs de DataSud : https://www.datasud.fr/contactez-nous/
— Contacter directement le producteur ou le diffuseur d'un jeu de données
— Suivre/s'abonner à un jeu de données, une thématique ou une organisation,
— Noter la qualité d'un jeu de donnée,



— Demander l'accès aux administrateurs à des données réservées sur accès restreint,
— Accéder aux données et services autorisés pour une organisation.
— Commenter une organisation ou un jeu de données

— Partager un jeu de données ou une ressource sur un autre site ou via des réseaux sociaux
— Déclarer une réutilisation : https://trouver.datasud.fr/showcase/new
— Participer au contrôle de la qualité de DataSud en signalant les contenus n'ayant pas vocation à y figurer (illicites ou contraires aux CGU).
— Créer une nouvelle Organisation ou demander à être rattaché à une Organisation existante (voir rubrique dédiée),
— Demander à devenir contributeur d'une organisation et éventuellement référent d'une organisation pour mai-

triser l'ensemble des publications de ladite organisation (voir rubrique dédiée).
— Accéder à certaines fonctionnaltés de l'API nécessitant une clé d'authentification.
— Intégrer un catalogue de données en marque Blanche sur son propre site internet (Voir rubrique dédiée sur la marque blanche)

CHAPITRE 3

Espace contributeurs

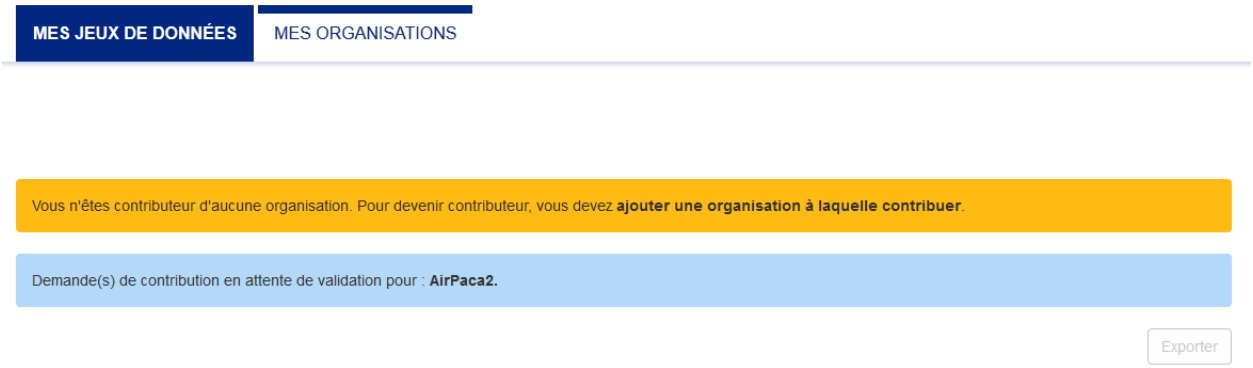
Ce guide est destiné aux producteurs de données, déjà inscrit en tant qu'Utilisateurs https://publier.datasud.fr/ et souhaitant contribuer à l'enrichissement des publications sur la plateforme. « voir la documentation sur les Utilisateurs »

Note : Toute personne, morale ou physique, publique ou privée, producteur de données publiques ou privées peut les publier sur la DataSud, sous reserve d'accepter les « conditions d'utilisation » et de respecter la réglementation sur les données à caractères personnelles.

3.1 Devenir Contributeur et Référent pour une organisation

Les organisations sont le plus souvent des personnes morales (autorités administratives, associations, entreprises) ou également des groupes informels.

Note : La création d'une nouvelle organisation peut-être effectuée soit au moment de l'inscription comme utilisateur de DataSud, soit après la validation du profil Utilisateur par les Administrateurs de DataSud.



Les demandes de statut de Contributeur ou de Référent sont soumises à la validation des Administrateurs de la plateforme. Il faut donc patienter un peu.

Note : Par défaut, un Utilisateur qui s'inscrit avec un email personnel (gmail, ymail, hotmail,...) et dont le nom de domaine ne peut correspondre à l'organisation pour laquelle il demande de contribuer, ne peut se rattacher, contribuer ou devenir référent d'une Organisation

---

Les Administrateurs de la Plateforme se réservent la possibilité de révoquer une inscription, une organisation, un statut de Contributeur ou de Référent, sans avis préalable.

Un Contributeur dispose des fonctionnalités suivantes :

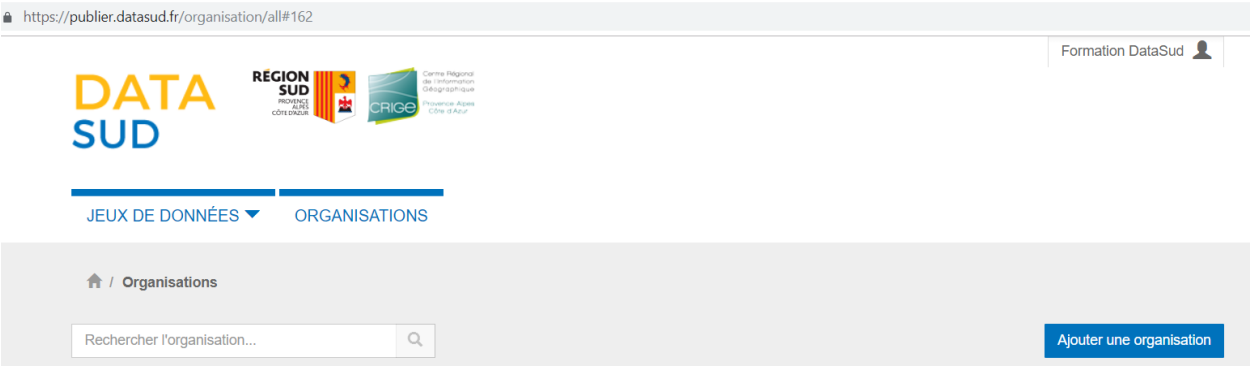
- Il peut publier un jeu de données et y ajouter des ressources, sous la forme d'un fichier téléchargeable, d'un lien URL ou d'une API,
- Il peut accorder le niveau d'accès aux ressources et jeux de données qu'il a crée pour son organisation : soit décider de les rendre accessible à tous, soit en restreindre l'accès uniquement à un ou plusieurs Utilisateurs inscrits ou bien à une Organisation choisie comme sa propre Organisation propriétaire du Jeu de données.

Un Référent des données de l'Organisation, à laquelle il appartient, dispose des fonctionnalités suivantes :

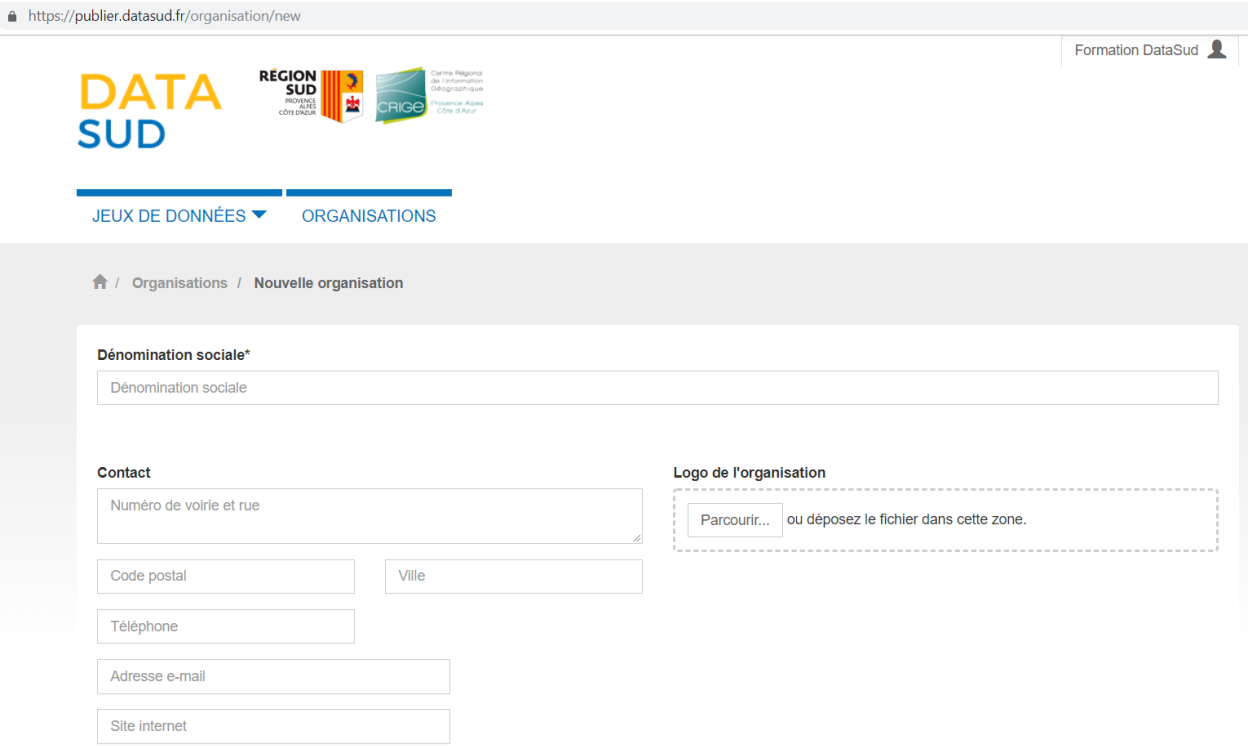
- Il peut éditer ou supprimer un jeu de données créé et publié par un autre Contributeur de l'Organisation,
- Il peut accorder le niveau d'accès aux ressources et jeux de données de toutes les publications de son Organisation,
- Il peut autoriser ou supprimer le statut de Contributeur aux Utilisateurs,
- Il recoit des notifications lorsque des modifications ont été apportées aux jeux de données et ressources de l'Organisation à laquelle il appartient.

3.2 Créer une Organisation

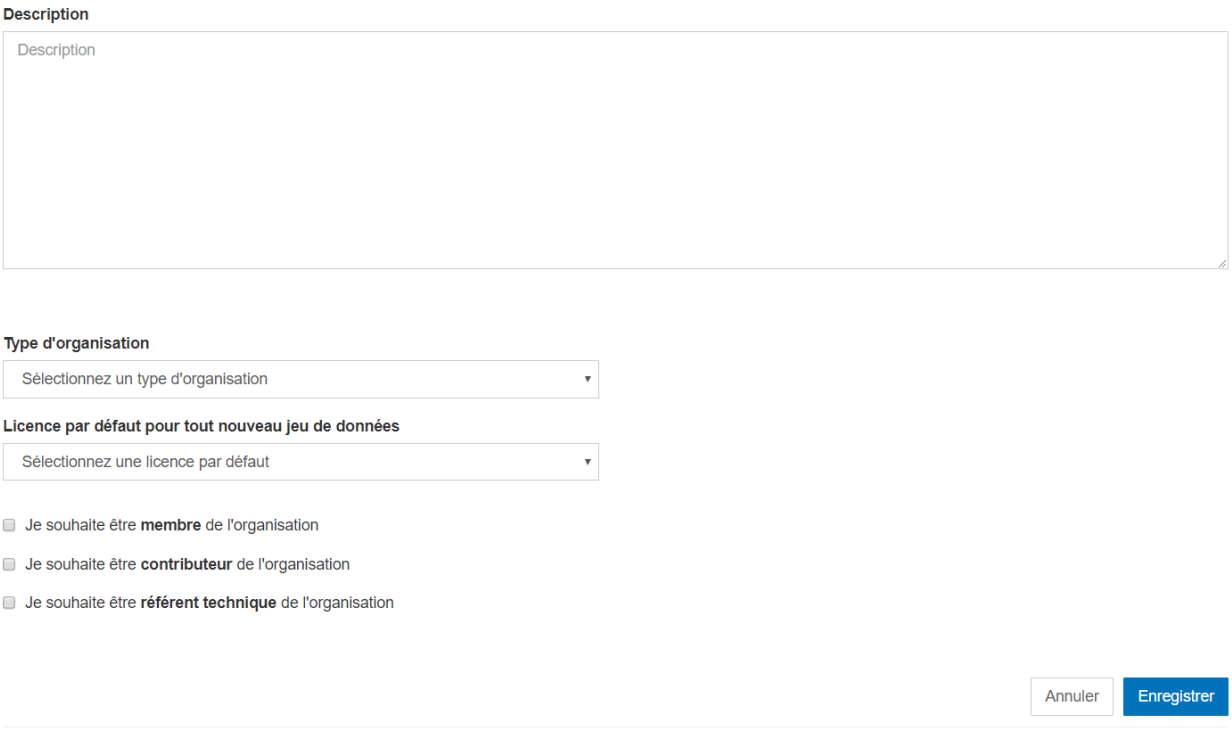
Toute demande de création d'une organisation est soumise à l'administrateur du site pour validation



La dénomination sociale est obligatoire

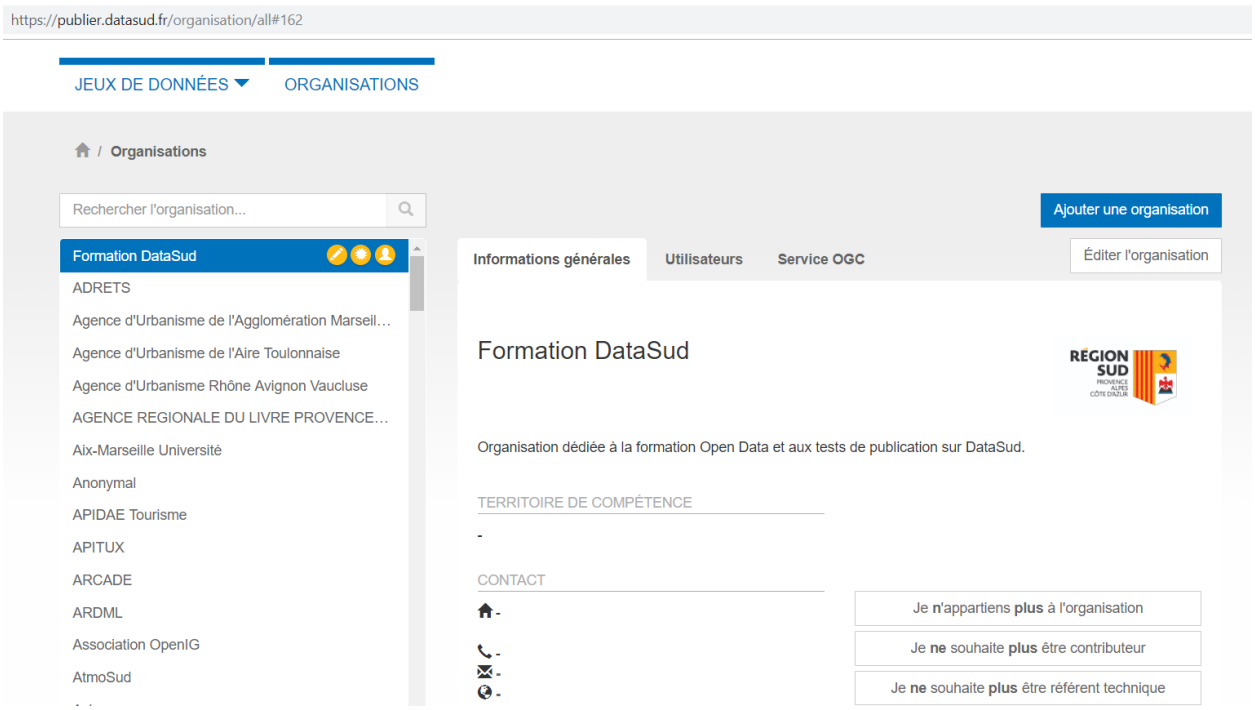


La description est facultative mais fortement conseillée, d'une part pour permettre de qualifier l'Organisation et sa démarche en matière d'ouverture des données publiques et géographiques et d'autre part pour permettre l'implementation automatique d'une page web spécifique à propos de l'organisation.



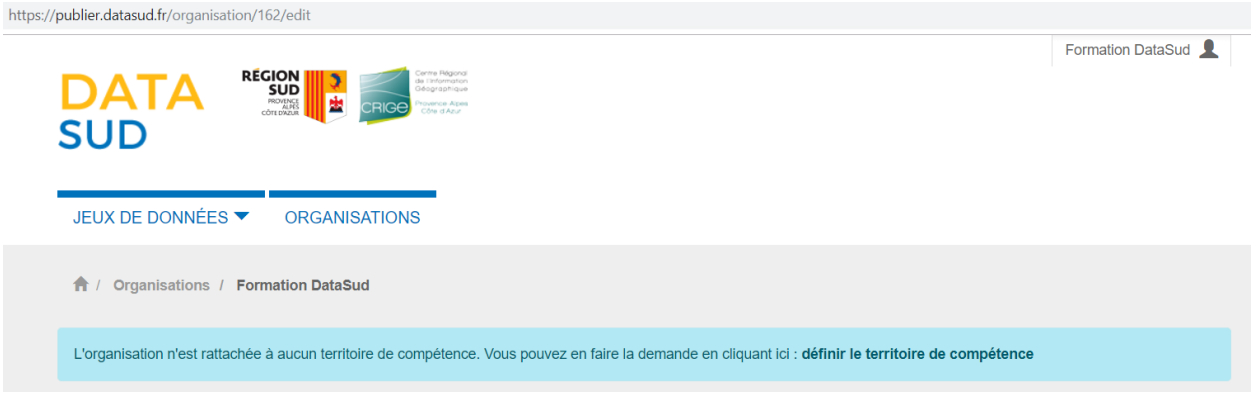
3.3 Editer la page d'une Organisation

Pour éditer la page de son organisation, le Contributeur clique sur l'onglet ORGANISATIONS dans son espace d'administration.



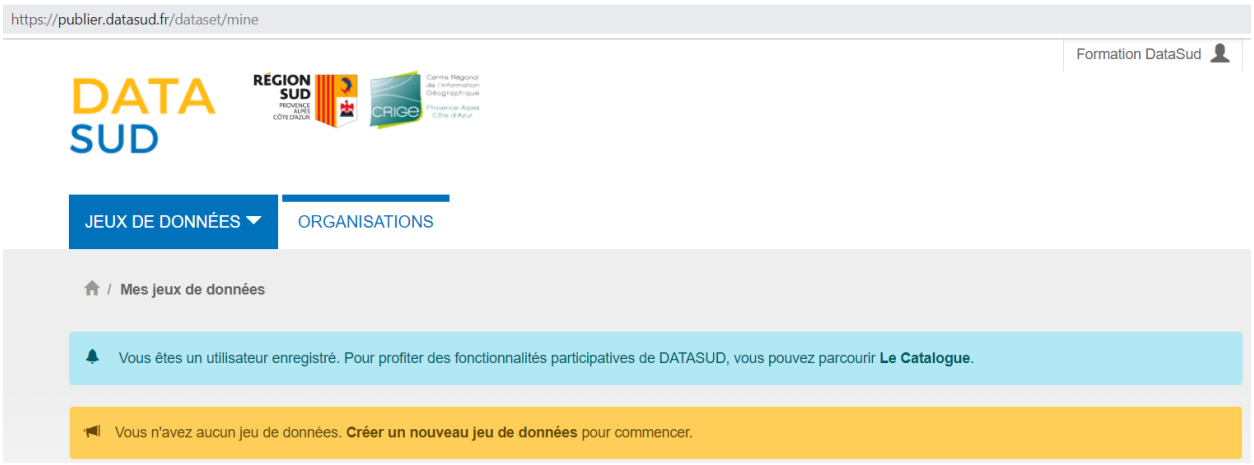
La première fois que le contributeur édite la page de son organisation, il lui sera demandé de définir le territoire de

compétence de l'organisation. La création de ce territoire de compétences permet de bénéficier de fonctionnalités spatiales supplémentaires dans DataSud. Cette demande est traitée par un administrateur du CRIGE.



3.4 Publier un jeu de données

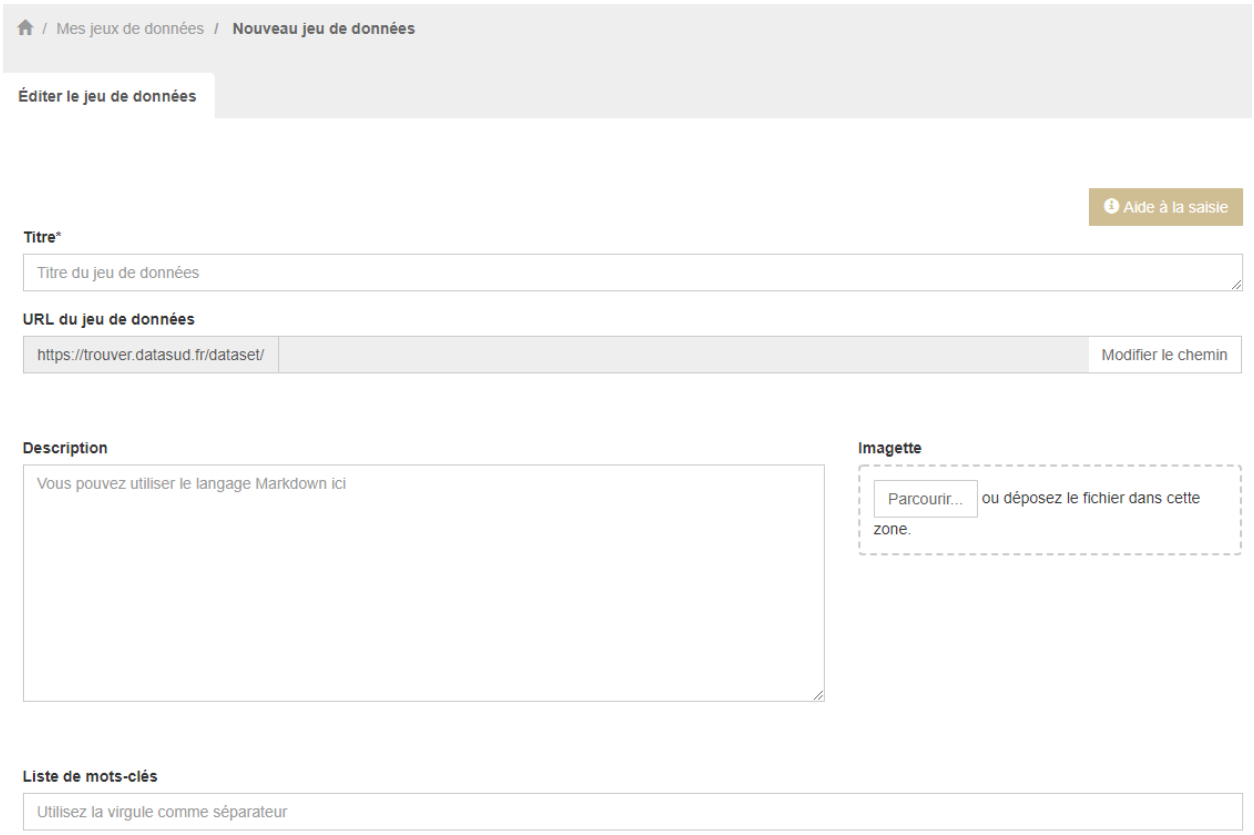
— Pour publier un jeu de donner le Contributeur se connecte avec son identifiant et mot de passe sur https://publier.datasud.fr



La publication se fait en deux étapes successives :

Tout d'abord on renseigne les métadonnées servant à définir ou décrire le jeu de données qui sera publié, puis on ajoute des jeux de données brutes ou des ressources complémentaires.

3.4.1 Etape n°1 : Renseigner les métadonnées



Note : De nombreux mots-clés sont déjà répertoriés dans la base. Ils apparaissent dans une liste déroulante lorsque vous saisissez les premières lettres du mot. Mieux vaut choisir un mot clés existant, plutot que d'en choisir un nouveau afin de permettre de relier votre jeu de donnée à d'autres jeux similaires inscrit au catalogue de DataSud.

Catégories (sélectionnez dans la liste ci-dessous une ou plusieurs catégories)

Administration et action publique | Agriculture, sylviculture et pêche | Aménagement et urbanisme | Culture, patrimoine et tourisme | Economie et entreprises | Energies et réseaux | Environnement et climat | Equipements, bâtiments et logement | Formation, éducation et emploi | Imagerie et occupation du sol | Limites administratives et référentiels | Mer, littoral et montagne | Mobilité et transports | Sciences, recherche et innovation | Citoyenneté et démocratie | Social, santé et sports

Date de création

29/01/2019 (valeur par défaut)

Date de dernière modification

29/01/2019 (valeur par défaut)

Date de publication

29/01/2019 (valeur par défaut)

Fréquence de mise à jour

Lorsque nécessaire

Couverture géographique

Indéfinie

Granularité de la couverture territoriale

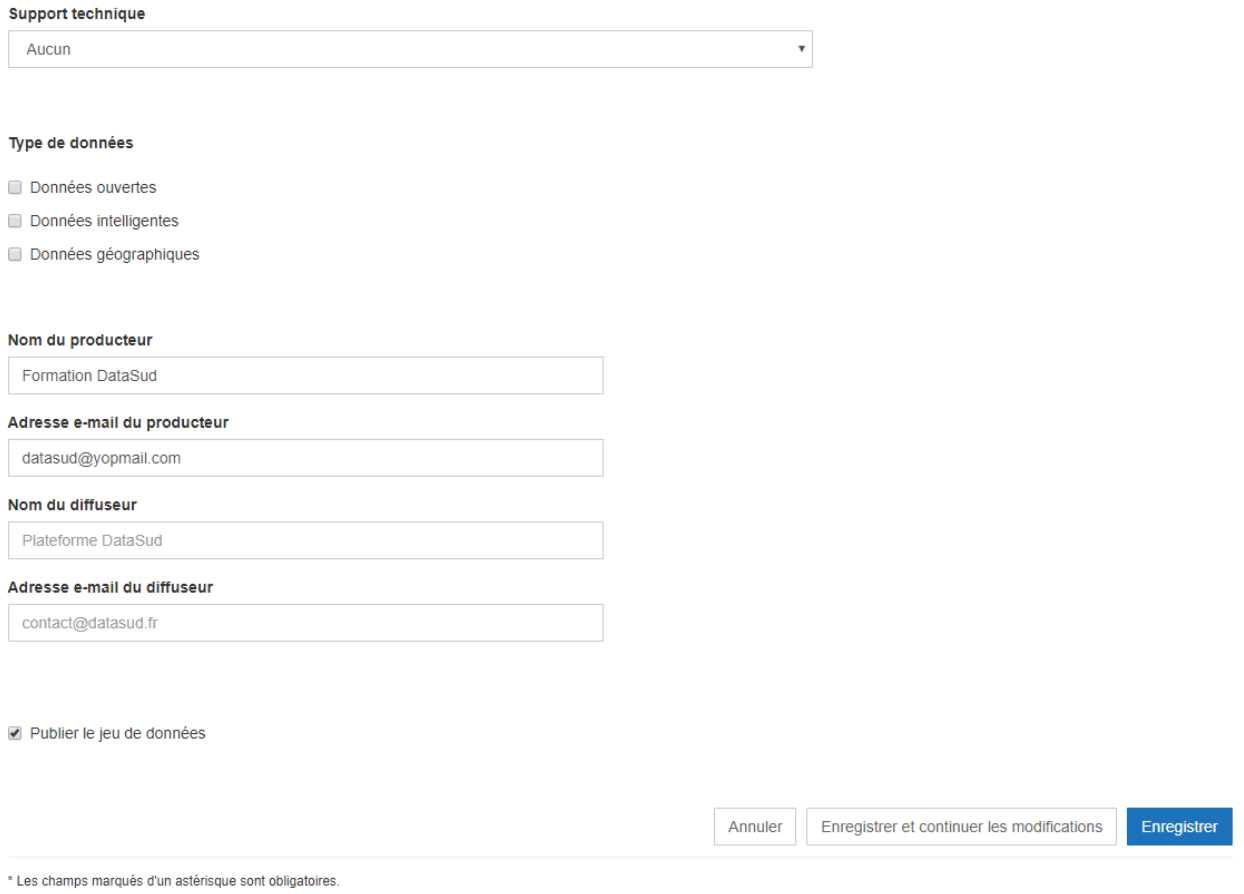
Indéfinie

Organisation à laquelle est rattaché ce jeu de données*

Sélectionnez une organisation

Licence*

Sélectionnez une licence



Les métadonnées obligatoires sont les suivantes :

- Titre
- Organisation à laquelle est rattaché ce jeu de données
- Descriptif : C'est un champ incontournable pour garantir une bonne réutilisation, car une donnée bien décrite est une donnée bien réutilisée !
- Dates de création, de dernière modification et de publication : la valeur par défaut indique la date du jour et la date de modification se met à jour automatiquement lorsque vous enregistrez des modifications sur les ressources.
- Licence : Selectionner une licence parmi celles qui sont proposées : Creative Commons attribution 4.0 ; Licence ouverte V2.0 ; Creative Commons cc-by-nc-nd 3.0, Open data base Licence V1.0 ou une Licence Spécifique.

Les métadonnées facultatives sont les suivantes :

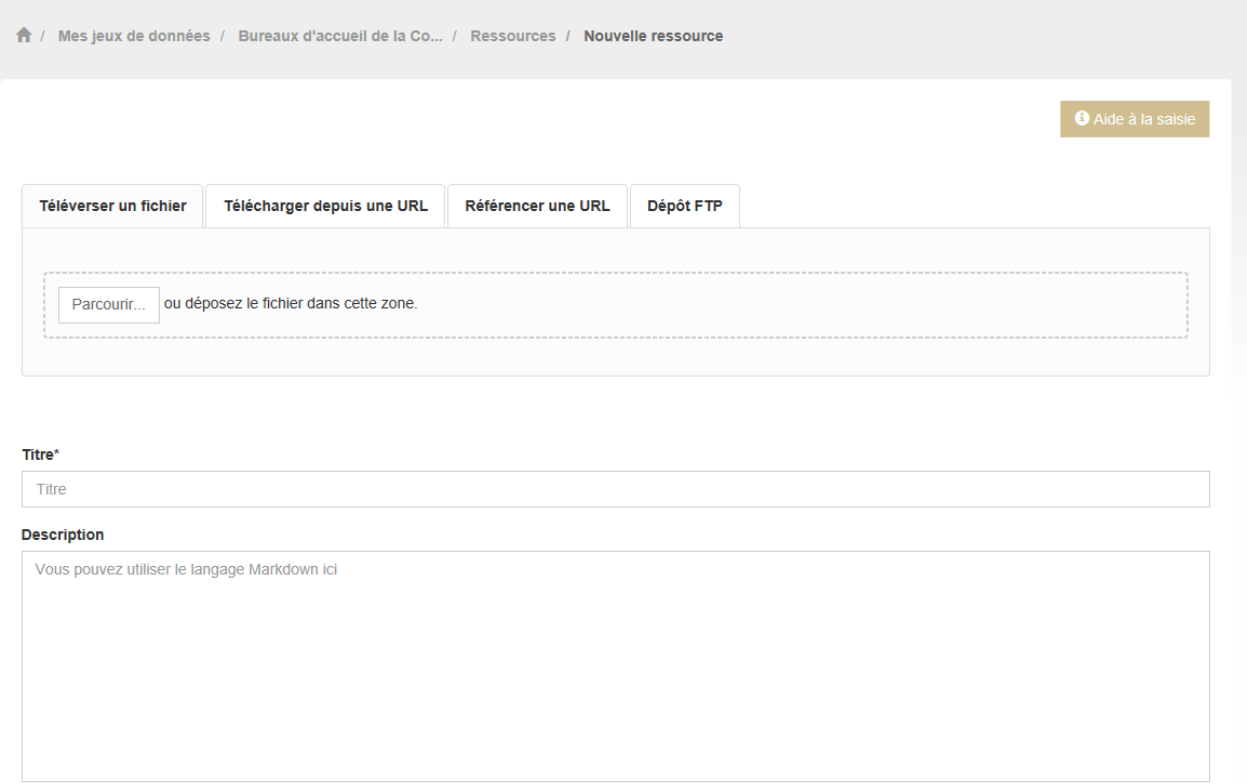
- Thématiques : un jeu de donnée peut-être associé à une ou plusieurs thématiques
- Mots-clés
- Fréquence de mise à jour à choisir dans le liste déroulante : Lorsque nécessaire ; Non planifiée ; Irrégulière ; Continue ; Temps réel ; Journalière ; Hebdomadaire ; Bi-mensuelle ; Mensuelle ; Trimestrielle ; Bi-annuelle ; Annuelle ; Inconnue.
- Type de données
- Meta-données INSPIRE
- Fréquence de mise à jour
- Couverture régionale

3.4.2 Etape n°2 : Publier une ressource

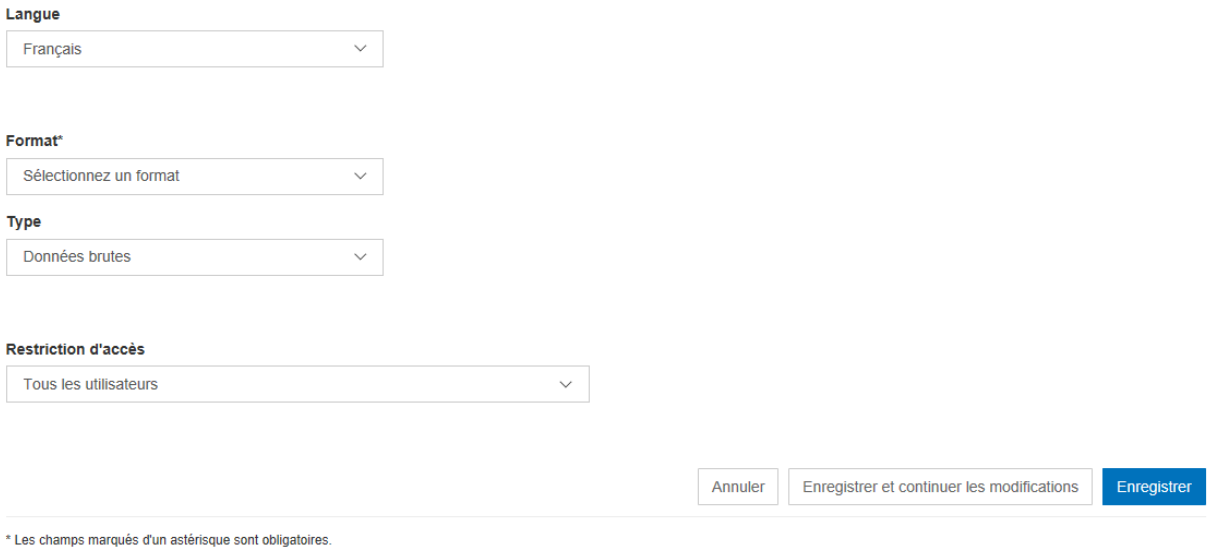
Il existe quatre manières différentes d'ajouter un jeu de données :

1. Téléverser manuellement un fichier depuis votre poste local :

A l'aide du bouton Parcourir, vous pouvez déposer le fichier qui s'ajoute dans l'entrepôt de données DataSud ;



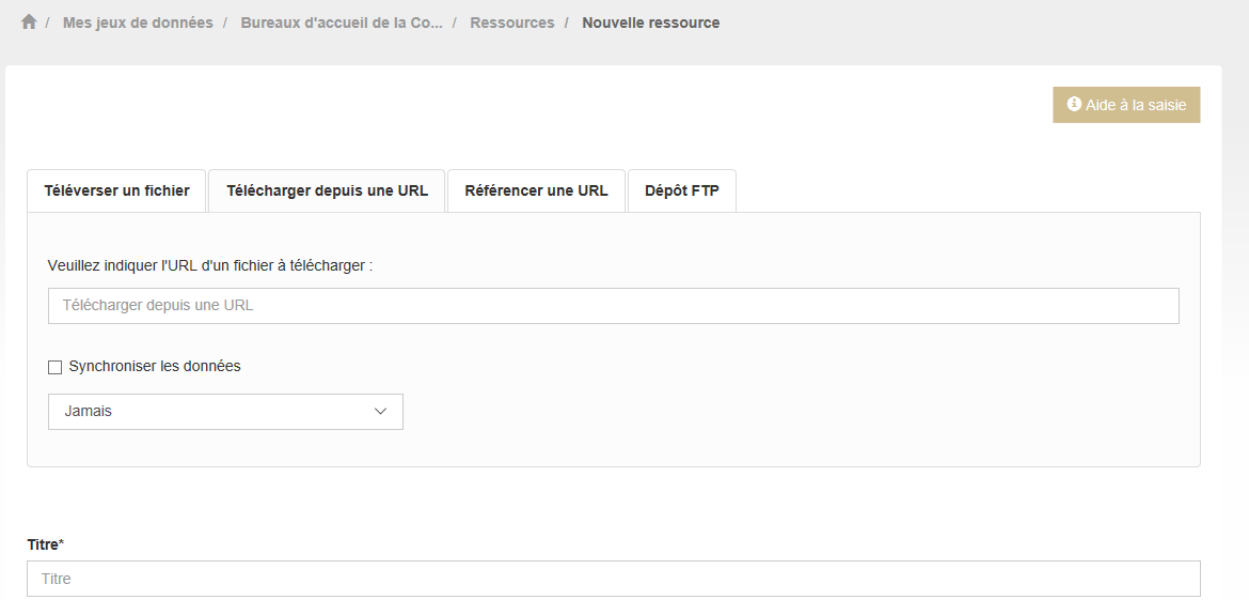
Le **Titre** de votre fichier est automatiquement recopié, mais il est possible de modifier manuellement le nommage de ce jeu de donnée.



Le format du fichier est automatiquement reconnu par DataSud. il faut préciser si le jeu de donnée est disponible en tant que Données brutes ou si c'est une documentation associée au jeu de donnée pour permettre aux visiteurs de DataSud d'avoir des informations complémentaires (plaquettes de communications, affiches, photographie, site internet. . . .)

2. Télécharger depuis une URL en indiquant une URL de téléchargement du jeu de donnée :

Dans ce cas, Datasud va télécharger la ressource pour l'ajouter dans l'entrepôt de données ;



Ce mode de publication permet de synchroniser la ressource distante, selon une périodicité régulière à indiquer :

- Jamais
- Quotidienne (tous les jours à minuit)
- Hebdomadaire (tous les lundis)
- Bimensuelle (1er et 15 de chaque mois)
- Trimestrielle (1er des mois de Janvier, Avril, Juillet et Octobre)
- Annuelle (1er Janvier)

Par exemple, un fichier transport.zip peut-être synchronisé sur DataSud directement grace à son URL de téléchargement.

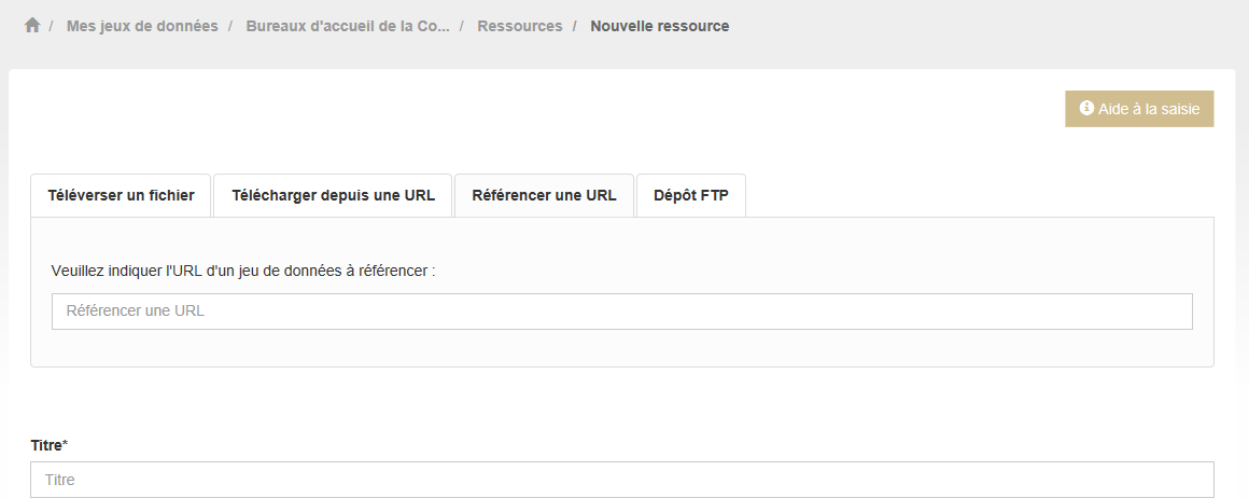
Note : Quelques précautions à prendre pour que la synchronisation s'active correctement :

- le nom de votre fichier doit avoir **exactement** le même nommage de fichier pour toute la synchronisation : si un script modifie le nom du fichier (pour rajouter une date ou autre par exemple), la synchronisation ne fonctionnera pas.
- votre fichier doit être accessible via une URL fixe : évitez les liens temporaires. :

En cas d'erreur, les Administrateurs de DATASUD se chargeront de vous indiquer que la synchronisation ne fonctionne pas ou plus.

3. Référencer une URL :

Dans ce cas, la ressource n'est pas téléchargée dans DataSud et vous indiquez précisement l'adresse URL de téléchargement de la donnée. qui reste hebergée chez son producteur. Cette donnée apparait au catalogue de DataSud mais elle n'est pas hébergée dans son entrepot.



4. Dépot FTP :

il faut pour activer au préalable le compte FTP en cliquant sur le lien « cliquez ici »

3.5 Datastore et données intelligentes

Datasud propose un **datastore**, c'est à dire un entrepôt de données qui offre des **services dits « intelligents » sur les données tabulaires aux formats CSV, XLS, GeoJSON, SHP**.

La publication des données sur Datasud, dans un format ouvert et interprétable par une machine, permet leur indexation dans le datastore afin notamment de proposer des apercus, de les filtrer par champs et de les parcourir sans utiliser de tableur dédiés.

Le format CSV est le format pivot à privilégier pour transformer vos données tabulaires en données semi-structurées dites « intelligentes » afin que le datastore génère des datavisualisations simples sous forme de grille, de graphe ou de carte.

Des données intelligentes permettent également d'en automatiser l'accès par API (Application Programming Interface) : L'accessibilité des données par interface de programmation est une condition nécessaire pour massifier et industrialiser les usages qui peuvent être fait de ces dernières. Les données indexées dans le datastore sont ensuite « requetables » directement à travers l'API à travers une série de fonctionnalités puissantes. (voir la présentation de l'API CKan : http://datasud.readthedocs.io/fr/latest/developpeurs/index.html#service-api-ckan)

Vos jeux de données doivent être préparés pour être proprement indexés dans le datastore :

- Le format CSV doit être privilégié avec un ; comme séparateur / délimiteur.
- Idéalement, passez tous vos jeux de données en UTF-8. Pour cela le programme Notepad++ fait cela très bien.
- Idéalement, exportez vos tableurs favoris (Microsoft, Libre et Open Office) au format CSV.
- Restreindre vos titres de colonnes à moins de 62 caractères.
- Ne pas doublonner le titre d'une colonne.
- En théorie les caractères spéciaux (" :.,(-") sont acceptés, mais c'est beaucoup mieux de les éviter dans les titres.
- Harmoniser le type de vos données (et oui vos données sont typées !) : en effet si une colonne ne comporte que des chiffres, le datastore autodéterminera le type de cette colonne comme étant un nombre. Or il suffit qu'une cellule de la colonne contienne l'entrée N/A, pour que le datastore génére une erreur.

Pour éviter les erreurs de type, il est préférable de les corriger avant d'indexer le jeu de donnée dans DataSud ou bien de transformer la valeur des cellules en cellules au format TEXTE. Cela n'est pas satisfaisant, mais ca fonctionne.

- ERREUR : En cas d'erreur supprimez complètement la ressource associée au jeu de données et ajoutez en une nouvelle.

Note : Attention avec Excel

- lorque le fichier contient plusieurs feuillet (ou onglet), seule la dernière feuille de calcul est indexée dans le datastore. Il est donc nécessaire de déplacer la feuille de calcul contenant les données que vous souhaitez indexer dans le datastore en dernière place de votre tableur.
- si vous ne voulez pas indexer vos données dans le datastore (pour plein de bonnes et mauvaises raisons), il suffit d'ajouter une feuille de calcul vide en dernière place de votre tableur.

3.6 Géolocalisation des données tabulaires (XLS et CSV)

Une carte peut automatiquement être générée à partir de vos données tabulaires geolocalisées. Pour cela vous devez intituler deux colonnes du tableau « latitude » et « longitude ».

Projections : en cours de rédaction.

3.7 Renseigner les métadonnées INSPIRE

Cette partie de la documentation est en cours de rédaction par le CRIGE

CHAPITRE 4

Espace développeurs

4.1 Service API Ckan

Le site **https ://trouver.datasud.fr** est construit à partir du système d'information OpenSource dédié à la gestion de catalogues de données CKAN.

4.2 Requêter l'API CKAN Catalogue

CKAN propose une API permettant d'interroger et de consulter le catalogue des données et leurs ressources. L'API permet également de requêter directement le contenu des ressources tabulaires (CSV, XLS) lorsque celles-ci ont été correctement intégrées au Datastore (http://datasud.readthedocs.io/fr/latest/producteurs.html#datastore-et-donnees-intelligentes).

Ainsi, il est par exemple possible de réaliser ce qui suit.

- Obtenir au format JSON, la liste des jeux de données, des groupes thématiques, des mots-clés utilisés ou des organisations du catalogue :
  http://trouver.datasud.fr/api/3/action/package_list
  http://trouver.datasud.fr/api/3/action/group_list
  http://trouver.datasud.fr/api/3/action/tag_list
  http://trouver.datasud.fr/api/3/action/organization_list
- Obtenir un flux des jeux de données récemment mis à jour :
  http://trouver.datasud.fr/api/3/action/recently_changed_packages_activity_list
- Obtenir une réprésentation détaillée d'un des objets (jeu de données, organisation, ressource), toujours au format JSON :
- Obtenir une représentation détaillée d'un jeu de données : https://trouver.datasud.fr/api/3/action/package_show?id=arbres-proteges-a-digne-les-bains
- Obtenir une représentation détaillée d'une organisation : https://trouver.datasud.fr/api/3/action/organization_show?id=air-paca
- Obtenir la liste de tous les jeux de données d'une organisation : https://trouver.datasud.fr/api/3/action/package_search?fq=organization:(ville-de-digne-les-bains)&rows=150

— Obtenir une liste de jeux de données « géographiques » : https://trouver.datasud.fr/api/3/action/package_list?datatype=donnees-geographiques
— Obtenier des informations sur la thématique « Environnement et Climat ». https://trouver.datasud.fr/api/3/action/group_show?id=environnement-et-climat
— Rechercher de jeux de données à partir d'un mot clé : https://trouver.datasud.fr/api/3/action/package_search?q=energies
— Rechercher des jeux de données « géographiques », au format CSV, associés à la thématique Culture, patrimoine et tourisme : https://trouver.datasud.fr/api/3/action/package_search?fq=+res_format:CSV+datatype:donnees-geographiques+groups:culture-patrimoine-et-tourisme

4.3 Requêter l'API CKAN DATA

DataSud.fr permet également de requêter directement le contenu des jeux de données, ou plutôt de leurs ressources. Cette mécanique est rendue possible à travers l'interrogation de l'API de données de CKAN (API CKAN DATA).

Comme expliqué plus haut, le Datastore propose un service d'indexation des données tabulaires (CSV et XLS). L'API CKAN DATA permet d'exposer le contenu des ressources indexées dans le Datastore dont on peut ainsi interroger tout ou partie sans avoir à télécharger le jeu de données. Il est alors possible de faire des opérations de recherche sur les différents champs de données.

— Afficher les cinq enregistrements du jeu de données des hôtels en région Provence-Alpes-Côte d'Azur :
Cette requête utilise la méthode datastore_search de l'API de CKAN avec la notion de filtres.

```
https://trouver.datasud.fr/api/3/action/datastore_search?
resource_id=9723b8ba-8379-4b1f-a85c-1f0efe916ce8&limit=5
```

Requête et résultats : https://bit.ly/2Lc3T12

— Trouvez toutes les entreprises de la base INFOGREFFE 2017 dont le champ ville est égal à MARSEILLE :
Cette requête utilise la méthode datastore_search de l'API de CKAN avec la notion de filtres.

```
https://trouver.datasud.fr/api/3/action/datastore_search?
resource_id=9723b8ba-8379-4b1f-a85c-1f0efe916ce8&filters={"Ville":"MARSEILLE"}
```

Requête et résultats : http://bit.ly/2BKn6VW

— Trouver toutes les entreprises de la base INFOGREFFE 2017 de la ville de MARSEILLE avec le code APE 6831Z, et afficher les résultats à partir du centième (série de 100 à 199)
Cette requête utilise la méthode datastore_search de l'API de CKAN avec la notion de filtres.

```
resource_id=9723b8ba-8379-4b1f-a85c-1f0efe916ce8&filters={"Ville":"MARSEILLE",
"Code APE":"6831Z"}&offset=100
```

Requête et résultats : http://bit.ly/2oliZId

— Production électrique régionale : trouvez les horaires ou le solaire est supérieur à 20MW (requête SQL)
Cette requête utilise la méthode datastore_search_sql de l'API de CKAN avec la notion de requête SQL .

```
https://trouver.datasud.fr/api/3/action/datastore_search_sql?sql=SELECT from
"52a8f5dd-758d-4e54-a837-8fc7ad57d378" WHERE "Solaire (MW)" > '20' AND "Date"
> '2018-07-10'
```

Requête et résultats : https://bit.ly/2N8JCKn

Documentation de l'API (catalogue et ressources) et de l'API Datastore (requête sur les ressources) en anglais :

http://docs.ckan.org/en/latest/api/ http://docs.ckan.org/en/ckan-2.7.2/maintaining/datastore.html#the-datastore-api

Note : le mot « package » qu'on trouve dans certaines requête et dans la documentation CKAN correspond à un jeu de donnée.

4.4 Construire une requête pour l'API

Pour appeler l'API CKAN, postez un dictionnaire JSON dans une requête HTTP POST sur l'une des URL d'API de CKAN. Les paramètres de la fonction API doivent être indiqués dans le dictionnaire JSON. CKAN retournera également sa réponse dans un dictionnaire JSON.

Une façon de publier un dictionnaire JSON sur une URL est d'utiliser le client HTTP en ligne de commande HTTPie. Il existe également d'autres outils comme Postman. Par exemple, pour obtenir une liste des noms de tous les jeux de données du groupe `environnment` sur le site, installez HTTPie, puis appelez la fonction API `group_list` en exécutant cette commande dans un terminal :

```
http http://trouver.datasud.fr/api/3/action/group_list
```

La réponse de CKAN ressemblera à ceci :

```
{
    "help": "...",
    "result": [
        "data-explorer",
        "department-of-ricky",
        "geo-examples",
        "geothermal-data",
        "reykjavik",
        "skeenawild-conservation-trust"
    ],
    "success": true
}
```

La réponse est un dictionnaire JSON avec 3 clés :

1. `"success"` : `true` or `false`.

   L'API est conçue pour retourner à chaque fois un `200 OK` dans le code statut de sa réponse, qu'il y ait une erreur ou non dans la requête, il est donc important de toujours vérifier la valeur de la clé `success` dans le dictionnaire de réponse, et si elle est à false, de vérifier la valeur de la clé `error`.

Note : S'il y a vraiment un gros problème de syntaxe dans la requête à l'API, CKAN pourra retourner une réponse HTTP avec un status code `409`, `400` or `500` (dans l'ordre croissant de gravité). Dans les prochaines versions de CKAN, il est prévu d'essayer de supprimer ce type de réponse pour n'avoirà la place que des retours `200 OK` et utiliser les valeurs `"success"` et `"error"`.

---

2. `"result"` : le résultat retournée par la fonction appelée. Le type et la valeur du résultat dépendent de la fonction appelée. Dans le cas de la fonction `group_list`, il s'agit d'une liste de chaînes, les noms de tous les jeux de données qui appartiennent au groupe.

   Si c'est une erreur qui est retournée à la requête, le dictionnaire contiendra une clé `"error"` avec le détail de l'erreur au lieu de la clé `"result"`. Un dictionnaire de réponse contenant une erreur ressemblera à ceci :

   ```
   {
       "help": "Creates a package",
       "success": false,
       "error": {
           "message": "Access denied",
           "__type": "Authorization Error"
           }
    }
   ```

3. `"help"` : le texte de documentation de la fonction appelée.

La même requête HTTP peut être effectuée en utilisant le module Python standard `urllib2` avec ce code Python

```
#!/usr/bin/env python
import urllib2
import urllib
import json
import pprint

# Make the HTTP request.
response = urllib2.urlopen('http://demo.ckan.org/api/3/action/group_list',
        data_string)
assert response.code == 200

# Use the json module to load CKAN's response into a dictionary.
response_dict = json.loads(response.read())

# Check the contents of the response.
assert response_dict['success'] is True
result = response_dict['result']
pprint.pprint(result)
```

4.5 Versions de l'API

Les API CKAN sont versionnées. Si vous faites une demande à une URL d'API sans numéro de version, CKAN choisira la dernière version de l'API :

```
https://trouver.datasud.fr/api/action/package_list
```

Vous pouvez également spécifier le numéro de version de l'API souhaité dans l'URL que vous envoyez :

```
https://trouver.datasud.fr/api/3/action/package_list
```

La version 3 est actuellement la seule version de l'API Action.

Nous vous recommandons de spécifier le numéro d'API dans vos demandes, car cela garantit que votre client API continuera à fonctionner si un jour le site est mis à niveau vers de nouvelles versions de CKAN).

4.6 Authentification et clés

Certaines fonctions de l'API nécessitent une autorisation, par exemple pour ajouter ou modifier des jeux de données et des ressources). L'API utilise la même fonction d'autorisation et la configuration en tant qu'interface web, donc si un utilisateur est autorisé à faire quelque chose dans l'interface web, il sera autorisé à le faire via l'API.

Lorsque vous appelez une fonction de l'API nécessitant une autorisation, vous devez vous authentifier vous-même en fournissant votre clé API avec votre requête HTTP. Pour trouver votre clé API, connectez-vous au site CKAN en utilisant son interface web et visitez votre profil utilisateur.

Pour fournir votre clé API dans une requête HTTP, incluez-la dans un En-tête ‘‘ Authorization‘‘ ou ‘‘ X-CKAN-API-Key‘‘.

Par exemple, pour demander si vous suivez actuellement l'utilisateur ‘‘ markw‘‘ sur demo.ckan.org en utilisant HTTPie, exécutez cette commande :

```
https://trouver.datasud.fr/api/3/action/am_following_user id = markw Autorisation: XXX
```

(Remplacer ‘‘ XXX‘‘ avec votre clé API.)

Par exemple, pour obtenir la liste des activités de votre tableau de bord utilisateur, on lance ce code Python : :
request = urllib2.Request(“http://trouver.datasud.fr/api/3/action/dashboard_activity_list”) request.add_header(“Authorization”, “XXX”) response_dict = json.loads(urllib2.urlopen(request, “{}”).read())

4.7 Support JSONP

Pour répondre aux scripts d’autres sites qui souhaitent accéder à l’API, les données peuvent être renvoyé au format JSONP, où les données JSON sont “complétées” avec une fonction call. La fonction est nommée dans le paramètre “callback”. Par exemple :

https://trouver.datasud.fr/api/3/action/package_show?id=adur_district_spending&callback=myfunction

Note : Cela ne fonctionne qu’avec les requêtes GET

4.8 Service WMS

En cours.

4.9 Service WFS

En cours.

4.10 Service CSW

En cours.

4.11 Marque blanche / Widget

DataSud permet d’intégrer un **catalogue de données en marque blanche sur un site web externe**. Cette fonctionnalité est également intitulée “widget”. Elle offre une solution technique pour valoriser le catalogue de données d’une organisation et plus largement de tout sous ensemble du catalogue de données DataSud filtré par une ou plusieurs facettes (organisations, thématiques, formats, licences, recherche par mot clé. . .).

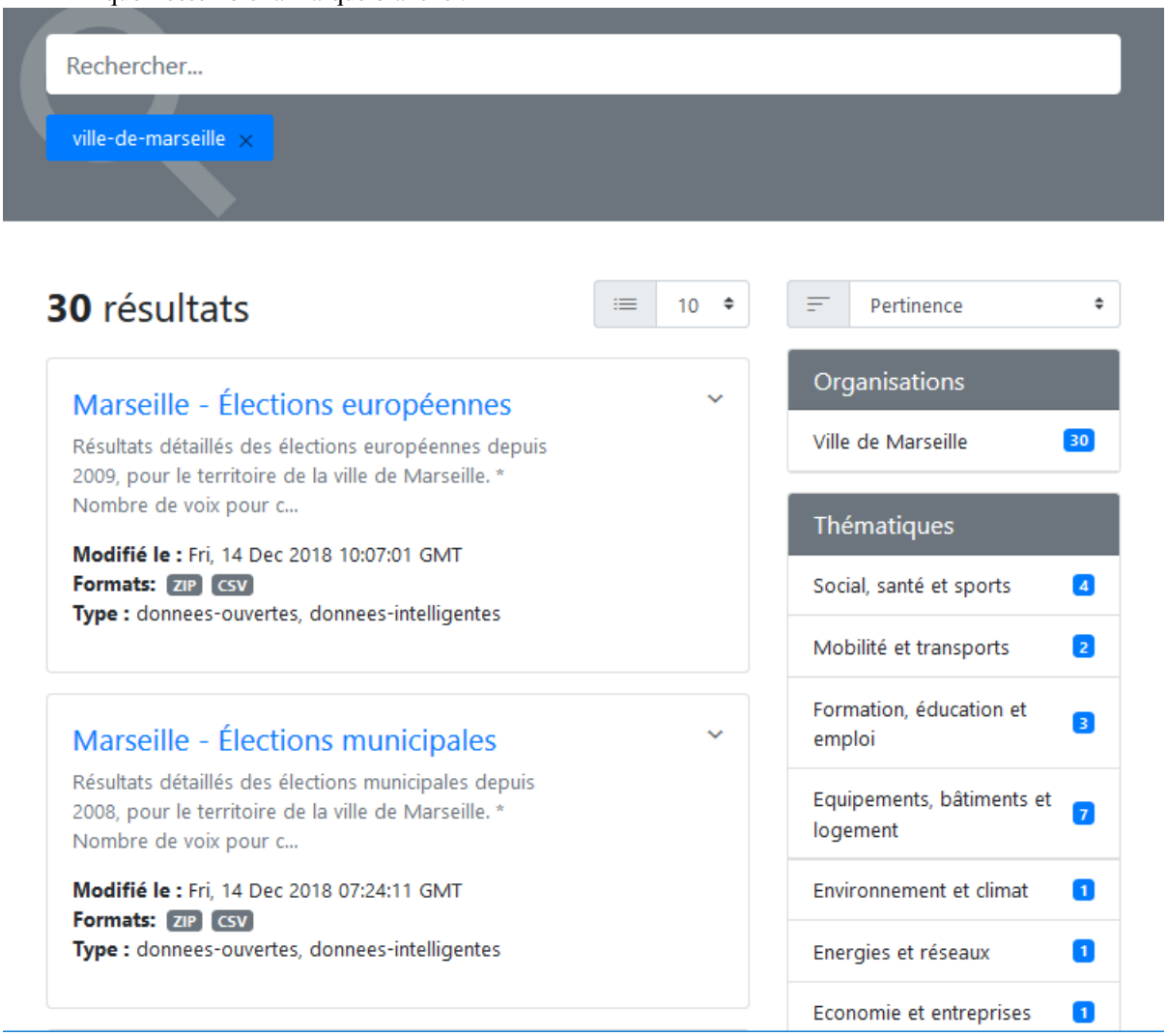
La marque blanche est accessible sans restriction et sans autorisation préalable à tout utilisateur, contributeur ou développeur de DataSud.

Techniquement, la marque blanche de DataSud passe par l’intégration de quelques lignes de code HTML à l’endroit souhaité sur une page web externe ainsi que deux appels à un fichier Javascrit (.JS) et une feuille de style CSS (.CSS).

La marque blanche DataSud a été développée par Neogeo Technologies. Elle est distribuée sur Gitub sous licence MIT. Le code source peut être utilisé pour afficher tout catalogue CKAN sur un site tiers.

— Code source :
https://github.com/neogeo-technologies/ckan-widget

— Licence :
https://github.com/neogeo-technologies/ckan-widget/blob/master/LICENSE
— Les fichiers à inclure et un exemple de code HTML sont disponibles ici :
https://github.com/neogeo-technologies/ckan-widget/tree/master/build
— A quoi ressemble la marque blanche ?



— Exemple de code d'implémentation :

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <link href="./app.css" rel="stylesheet">
    <link href="./static/css/main.css" rel="stylesheet">
    <title>Catalogue CKAN</title>
  </head>

  <body>
    <div id="ckan-widget"></div>
  </body>

  <script src="./static/js/main.js" type="text/javascript"></script>
```

(suite sur la page suivante)

(suite de la page précédente)

```
  <script type="text/javascript">
    var config = {
       // URL du catalogue CKAN cible
       ckan_api: 'https://trouver.datasud.fr',
      // Filtres complémentaires optionnels :

      //ckan_organizations: ['org1', 'org2'],
      //ckan_groups: ['group1'],
      //ckan_tags: ['tag1'],
      //ckan_facets: {
        //res_format: 'HTML',
    //    datatype: 'type'
    //  },

    // paramétrages de l'affichage :
      data_sort: 'title_string asc',
      result_page_size: 25,
      thumbnails_display: true
    }

    ckanWidget.init(config)
  </script>
</html>
```

Paramètres d'intégration de la marque blanche :

Le code d'inclusion html et son appel javascript permettent :

- 1/ De **charter l'interface graphique** à travers la modification de la feuilles de styles **app.css**.
- 2/ De **spécifier les facettes à filtrer** : les organisations (ckan_organizations), les thématiques (ckan_groups), les mots clés (ckan_tags) et plus généralement toute facette (ckan_facets) identifiable dans l'url des résultats d'une recherche effectuée sur DataSud.
- 3/ De **spécifier comment afficher les résultats** : tri (data_sort), nombre de résultats par page (result_page_size), et intégration d'un vignette (thumbnails_display : true).

Exemples d'intégration :

- Exemple pour le compte du Département des Hautes-Alpes :

http://beta.crige-paca.org/beta/widget/hautes_alpes/hautesalpes_catalogue.html

CHAPITRE 5

Cadre légal et réglementaire

5.1 Le statut juridique des données publiées sur DataSud

Dans le cadre de leurs démarches d'ouverture des données publiques, la Région Provence Alpes Côte-d'Azur et le CRIGE ont fait le choix de licences de réutilisation à titre gratuit. Ce choix est conforme au décret n°2017-638 prévu par l'article L323-2 du Code des Relations entre le Public et l'Administration (CRPA) dans lequel la licence Open Data Base License (ODbL) de l'Open Knowledge Foundation et la licence ouverte / open licence sont promues.

Conformément aux objectifs d'une politique d'accès libre aux données publiques, ces licences définissent les droits et les devoirs des réutilisateurs des données du portail DataSud. Le type de licence applicable (ODbL ou licence ouverte/open licence) est précisé sur chaque jeu de données. L'accès à un jeu de données vaut validation des conditions de la licence associé à ce jeu. A noter que seuls les textes complets des contrats de licence font foi.

5.2 Connaître les Licences

licence ODbL

Pour la licence ODbL, vous êtes libres de :

- De partager : copier, distribuer et utiliser la base de données.
- De créer : produire des créations à partir de cette base de données.
- D'adapter : modifier, transformer et construire à partir de cette base de données.

Aussi longtemps que :

Vous mentionnez la paternité : vous devez mentionner la source de la base de données pour toute utilisation publique de la base de données, ou pour toute création produite à partir de la base de données, de la manière indiquée dans l'ODbL. Pour toute utilisation ou redistribution de la base de données, ou création produite à partir de cette base de données, vous devez clairement mentionner aux tiers la licence de la base de données et garder intacte toute mention légale sur la base de données originaire. La source devra être mentionnée de la façon suivante : " [Nom de la base de donnée], [Producteur de la base de donnée], [date], sous licence ODbL." (par exemple, "Equipements publics - Nantes Métropole, 24/04/2018, sous licence ODbL").

Vous partagez aux conditions identiques : si vous utilisez publiquement une version adaptée de cette base de données, ou que vous produisez une création à partir d'une base de données adaptée, vous devez aussi offrir cette base de données adaptée selon les termes de la licence ODbL.

Gardez ouvert : si vous redistribuez la base de données, ou une version modifiée de celle-ci, alors vous ne pouvez utiliser de mesure technique restreignant la création que si vous distribuez aussi une version sans ces restrictions.

Licence ouverte / open licence

Pour la licence ouverte (v1 ou v2) https://www.etalab.gouv.fr/licence-ouverte-open-li… , vous êtes libres de :

- De reproduire, copier, publier et transmettre « l'information »
- De diffuser et redistribuer « l'information »
- D'adapter, modifier, extraire et transformer à partir de « l'information », notamment pour créer des « informations dérivées »
- D'exploiter « l'information » à titre commercial, par exemple en la combinant avec d'autres « informations « , ou en l'incluant dans votre propre produit ou application.

Aussi longtemps que :

Vous mentionnez la paternité : vous devez mentionner la paternité de « l'information » : sa source – a minima le nom du « Producteur » - et la date de sa dernière mise à jour, ou indiquer le ou les liens hypertextes (URL) renvoyant vers « l'information » et assurant une mention effective de sa paternité.

Avec cette licence, vous êtes détenteur d'un droit personnel, non exclusif et gratuit de réutilisation des données mises à disposition, pour une durée illimitée dans le monde entier, exempte de droits de propriété intellectuelle appartenant à des tiers. Le droit français est applicable.

Vous êtes le seul responsable de la réutilisation de « l'information » conformément aux libertés et conditions définies par la Licence Ouverte/Open Licence. La réutilisation ne doit pas induire en erreur des tiers quant au contenu de « l'information », sa source et sa date de mise à jour.

Si vous débutez en matière de réutilisation de données publiques, ouvertes, personnelles, sensibles, géographiques… Alors nous vous invitons à commencer par découvrir la documentation produite par OpenDataFrance :

[Ressources Opendatalocale](http://opendatalocale.net/ressources)

[module de eLearning proposé par le portail européen de l'OpenData](https://www.europeandataportal.eu/elearning/fr/module1/#/id/co-01 https://player.vimeo.com/video/153271129)

5.3 Faire une demande d'accès à un document administratif ou à des données

Présentation du cadre juridique de l'ouverture des données : un [guide pratique de la publication en ligne et de la réutilisation des données publiques](https://www.cnil.fr/sites/default/files/atoms/files/guide_open_data.pdf) a été élaboré par les services de la CADA et de la CNIL en association avec les services d'Etalab.

L'article 15 de la Déclaration Universelle des droits de l'homme et du citoyen précise que « la société a le droit de demander compte à tout agent public de son administration ».

Ce droit d'accès aux documents administratifs est régie par la loi CADA du 17 juillet 1978 qui considère que « les données produites ou détenues par les administrations, dans le cadre de leurs missions de service public, doivent être mises à disposition du public». Cela ne concerne pas les informations personnelles, ni celles touchant à la sécurité nationale, ou celles couvertes par les différents secrets légaux.

Le formalisme d'accès ainsi que vos droits à ce sujet sont précisés sur le site officiel de l'administration française : https://www.service-public.fr/particuliers/vosdroits/F2467

Conformément à la mise en oeuvre de la loi Numérique, la Commission d'accès aux documents administratifs (CADA) considère que tous les fichiers dont la communication a été sollicitée à partir du 8 avril 2017 doivent être publiés en Open Data « par défaut ».

De cette manière, il est possible de solliciter une administration, pour avoir accès à de nombreux documents administratifs, sans avoir à motiver votre demande, meme si ces données ne sont pas encore publiés sur les sites web des administrations ou sur les portails OpenData.

Afin de vous permettre de nous saisir à ce sujet, DataSud propose un service de « demande de données ».

> **Vous pouvez déposer une demande d'accès à un document administratif ou à un jeu de données relatif au territoire régional à l'adresse suivante : https ://www.datasud.fr/demande-de-donnees/**

Toute demande sera étudiée et une réponse vous sera apportée. Nous relayerons le cas échéant votre demande à la collectivité ou à l'organisme concerné.

- Si l'administration accepte votre demande de communication, elle doit normalement le faire dans un délai d'un mois.
- L'administration peut rejeter, par décision motivée, votre demande de communication.

Le refus de communication opposé par l'administration peut être :

- **Exprès** : il doit alors être motivé, en vertu de l'article 25 de la loi du 17 juillet 1978.

  Le refus de communiquer des informations relatives à l'environnement doit obligatoirement donner lieu à une décision expresse motivée (I de l'article L. 124-6 du code de l'environnement).
- **Tacite** : la décision de rejet naît du « silence gardé pendant plus d'un mois par l'autorité compétente, saisie d'une demande de communication de documents » (1er alinéa de l'article 17 du décret 2005-1755 du 30 décembre 2005). Ces décisions tacites sont dispensées de l'obligation de motivation (article 5 de la loi du 11 juillet 1979), sauf si un texte en dispose autrement.

La CADA ne peut être saisie qu'à la suite d'un refus de communication, qui peut d'ailleurs ne porter que sur un désaccord quant aux modalités de la communication. Une saisine formée avant l'expiration du délai d'un mois imparti à l'administration pour répondre est donc irrecevable.

En cas de communication insatisfaisante pour le demandeur (document tronqué, dossier ne contenant pas le document recherché…) sans refus exprès, la CADA exige que le demandeur attende l'expiration du délai d'un mois à compter de sa demande (voir site de la CADA : http://www.cada.fr/).

CHAPITRE 6

Régles éditoriales

6.1 Règles de nommage des ressources

Selon les recommandations en matière de nommage des fichiers électroniques et de plan de classement, nous vous proposons de respecter les règles suivantes relatives aux intitulés des ressources (fichiers) associées à vos jeux de données :

- Le nom d'un fichier doit être succinct : éviter de dépasser 30 caractères (sans compter l'extension).
- Le nom d'un fichier doit être précis : il contiendra idéalement : le nom du producteur, le sujet, le type de document, la date de création, éventuellement la version.
- Date : pour le 20 décembre 2018 => 20181220 (norme ISO 8601).
- Ne pas utiliser des articles ou mots vides : le, la, les, de, etc…
- Préférer le caractère _(underscore, tiret du 8) à un espace
- Eviter les lettres accentuées
- Le nom d'un fichier ne doit pas contenir : espace, ponctuation (sauf le point avant l'extension), caractères accentués ou spéciaux (ùé+'@à°[] :</* »& !$, etc.).
- La gestion des versions permet de suivre l'évolution et les étapes de l'élaboration d'un fichier. Il faut les distinguer soigneusement en les numérotant pour obtenir une suite logique exemple V01, V02, etc.

6.2 Amélioration des champs descriptifs avec le langage Markdown

Pour les champs descriptifs des jeux de données, des ressources et des organisations vous pouvez utiliser le langage Markdown dans le but est d'offrir une syntaxe facile à lire et à écrire.

Voici quelques exemples de syntaxe Markdown.

Cette liste n'est pas exhaustive.

=== Formatage ===

Mettre du texte en italique

```
*quelques mots*
```

quelques mots

Mettre du texte en gras

```
**plus important**
```

plus important

Pour mettre du code dans le texte :

```
``Mon code``
```

`Mon code`

=== Listes ===

Sauter une ligne avant le début de la liste.

Pour créer une liste non ordonnée

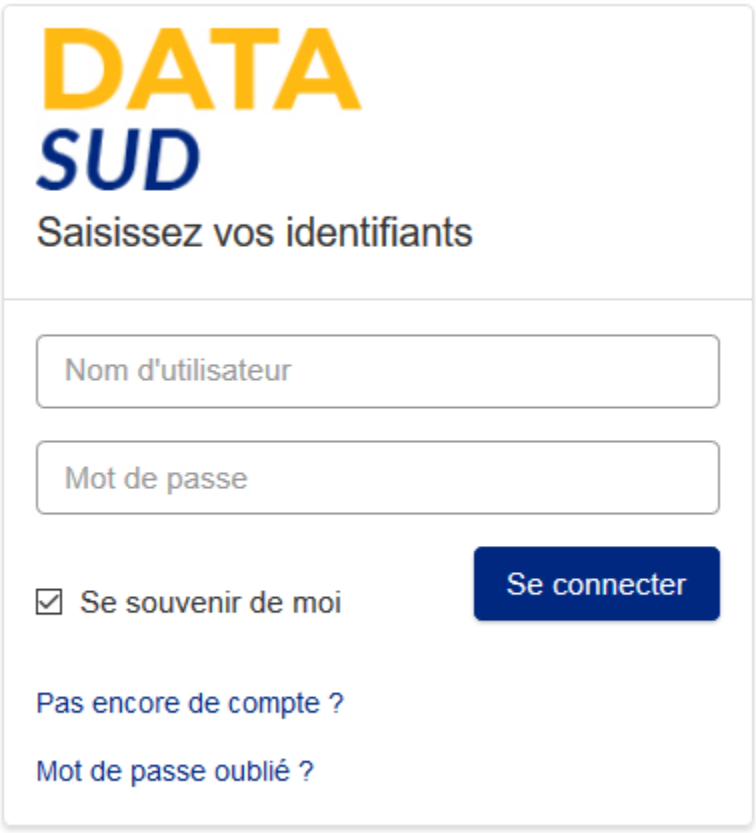
```
* Pommes
* Poires
```

- Pommes
- Poires

=== Image ====

Vous pouvez afficher une image dans vos descriptifs. Attention, la taille n'est pas paramétrable et l'image doit déjà être disponible en ligne quelque part

```
.. image:: CaptureDataSudConnect.PNG
```



=== Liens ===

Pour créer des liens

```
[texte du lien](url_du_lien "texte pour le titre, facultatif")
https://trouver.datasud.fr (automatique si mon url commence par http ou https).
```

[Trouver des données sur Datasud.fr](https://trouver.datasud.fr)

=== Aller plus loin ===

https://fr.wikipedia.org/wiki/Markdown

https://guides.github.com/features/mastering-markdown/

https://guides.github.com/pdfs/markdown-cheatsheet-online.pdf

6.3 Les outils pour nettoyer vos données

[La méthode infolabs pour produire un CSV de qualité](http://infolabs.io/prod-csv)

[l'Outil de validation des données ouvertes Validata](https://validata.fr/)

[Nettoyer les CSV avec CSVLint (en Anglais)](http://csvlint.io)

https://goodtables.io/

http://openrefine.org/

Voir aussi :

contents

Note : Ces guides sont maintenus par l'équipe Datasud.