



Data Fair
Version 4.4.1

Présentation fonctionnelle

22 avril 2024

Table des matières

1 - Introduction

2 - Portail de données (Front Office)

2.1 - Page d'accueil

2.2 - Catalogue de données

2.3 - Détails d'un jeu de données

2.4 - Catalogue des visualisations

2.5 - Consultation d'une visualisation

2.6 - Pages de contenus

2.7 - Compte utilisateur

2.8 - Notifications

2.9 - Réutilisations

2.10 - Accès par API

2.11 - Formulaire de contact

3 - Gestion des contenus (back-office)

3.1 - Administration des jeux de données

3.2 - Paramétrage des visualisations

3.3 - Portails de données et pages de contenus

3.4 - Gestion des utilisateurs

3.5 - Traitements périodiques

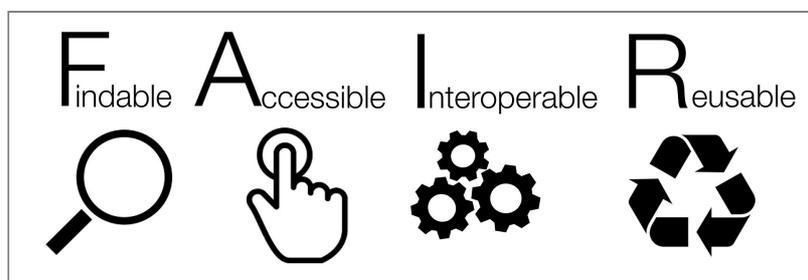
3.6 - Connecteurs de catalogues

3.7 - Métriques d'utilisation

1 - Introduction

Data Fair et son écosystème permettent de mettre en œuvre une plateforme de partage de données (interne ou open data) et de visualisations. Cette plateforme peut être mise à disposition du grand public, pour lui permettre de consulter les données au travers de **visualisations interactives** adaptées, mais également à un public plus expert qui accède aux données au travers des API.

Le mot **FAIR** fait référence à de la donnée « **Facile à trouver, Accessible, Interopérable et Réutilisable** ». Ceci est rendu possible grâce à **l'indexation** des données sur la plateforme. Elle permet de réaliser des recherches complexes dans des volumes de plusieurs millions d'enregistrements et d'accéder plus facilement et rapidement à ce qui nous intéresse. L'accès aux données au travers **d'API normalisées et documentées** permet d'interfacer la plateforme avec d'autres systèmes d'information et facilite la réutilisabilité des données.



Les utilisateurs accèdent à la plateforme au travers **d'un ou plusieurs portails de données**, qui permettent d'explorer le catalogue de jeux de données de différentes manières. Il est possible de consulter directement les jeux de données, que ce soit avec des vues génériques (tableaux, cartes simples...) ou des visualisations plus spécifiques préconfigurées. Les données sont diffusées au travers de pages qui les présentent sous la forme d'histoires (data storytelling), **permettant à n'importe qui de les comprendre plus facilement**. Les utilisateurs peuvent s'abonner à des notifications sur les mises à jour et les développeurs peuvent accéder à la documentation interactive des diverses API de la plateforme. Les portails peuvent être agrémentés de **pages de contenu** présentant, par exemple, les démarches, contributeurs ou réutilisations mises en avant.

Les administrateurs et contributeurs de données ont accès à un **back-office** qui permet de gérer les différents éléments de la plateforme : comptes utilisateurs, jeux de données, visualisations, traitements automatiques, portails de données et métriques. Les administrateurs peuvent paramétrer l'environnement et gérer les permissions d'accès aux données et visualisations. Selon leur profil, les utilisateurs du back-office peuvent créer, éditer, enrichir, supprimer les jeux de données, les cartes et les graphiques.

Fonctionnement général

Les **jeux de données** sont, en général, créés par les contributeurs en **chargeant des fichiers tabulaires ou géographiques** : le service stocke le fichier, l'analyse et déduit un schéma de données. Les données sont ensuite indexées suivant ce schéma et peuvent être requêtées au travers d'une API web propre. En complément des jeux de données basés sur des fichiers, Data Fair permet de créer des jeux de données **éditables par formulaires** et des jeux de données virtuels qui sont des **vues configurables d'un ou plusieurs jeux de données**.

L'utilisateur peut sémantiser les champs des jeux de données en leur **attribuant des concepts**, par exemple, en déterminant qu'une colonne contenant des données sur 5 chiffres est un champ de type code postal. Cette sémantisation permet 2 choses :

- les données peuvent être **enrichies** à partir de données de référence ou elles-mêmes devenir **données de référence** pour en enrichir d'autres ;
- les données peuvent être utilisées dans des **visualisations adaptées à leurs concepts**.

Les **visualisations** permettent d'exploiter au maximum le potentiel des données. Quelques exemples : un jeu de données contenant des codes de commune peut être projeté sur une carte du découpage administratif français, un jeu de données contenant des codes de parcelles peut être projeté sur le cadastre, etc.

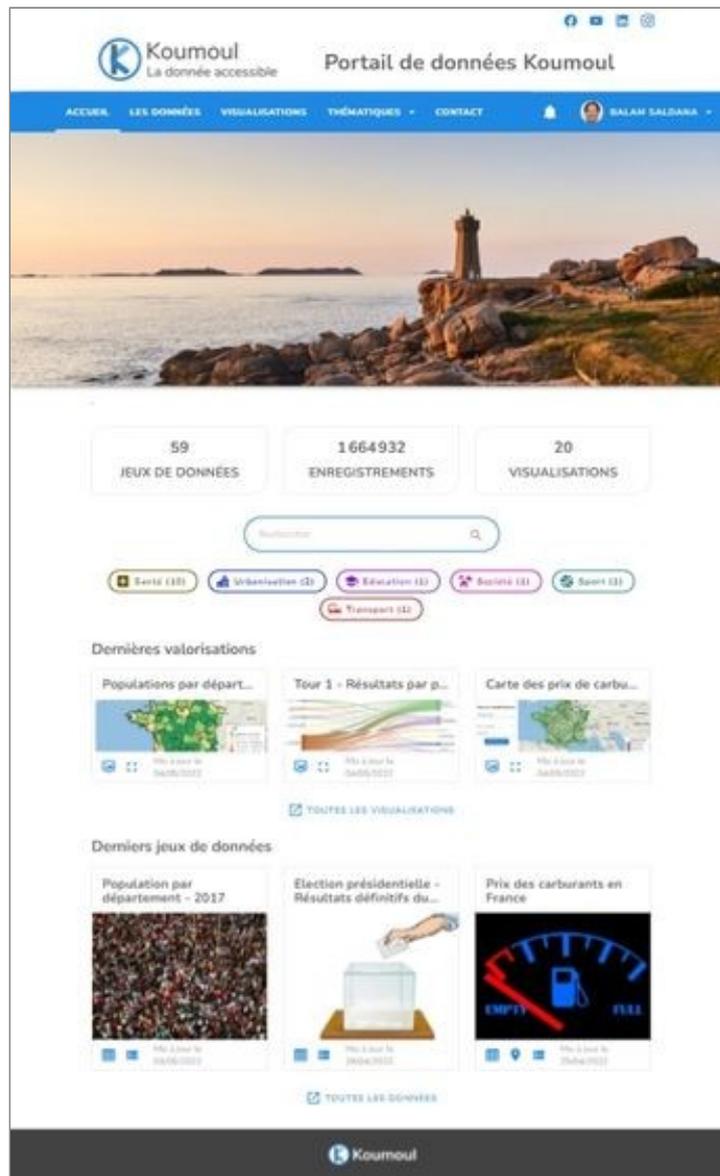
Principaux atouts de la plateforme

Data Fair permet de mettre en place une organisation centrée autour de la donnée :

- Possibilité de charger des données sous différents formats de fichiers ou par saisie via formulaire, permettant même de faire du *crowdsourcing* ;
- Consultation des données au travers d'un large choix de visualisations interactives (graphiques, cartes, moteurs de recherche...) ;
- Possibilité de créer plusieurs portails suivant les cas d'usage (open data, échanges internes...) ;
- Création facilitée d'API de données et enrichissement des données pour leur donner encore plus de valeur ;
- Mise en place de traitements périodiques permettant d'alimenter automatiquement la plateforme en données ;
- Cadre sécurisé, code source ouvert et utilisation de standards.

2 - Portail de données (Front Office)

Les utilisateurs des données accèdent à la plateforme au travers **d'un ou plusieurs portails de données** dédiés à différents cas d'usage. Il peut y avoir, par exemple, un portail de données publiques (open data), un portail pour les données internes et un portail de **préproduction** présentant les données en cours de consolidation ou qui ne sont pas encore prêtes à être publiées.



Un portail de données est composé de différentes pages (accueil, contact...), de moteurs de **recherche multicritère**, de jeux de données, de visualisations et de pages de contenu (actualités, thématiques, contributeurs, démarche...) Il est possible de présenter des chiffres clés et **d'éditorialiser des contenus** mis en avant.

Le portail de données a un **responsive design** et son affichage est adapté à une grande diversité de terminaux, permettant une utilisation sur **poste fixe ou mobile**. Il est possible de configurer différents

éléments pour personnaliser au maximum le portail (titre et description, choix des couleurs principales et secondaires, logo, favicon et image d'accueil, liens dans le footer...)

Une **mire d'authentification** est accessible sur toutes les pages du portail pour pouvoir se connecter ou créer un compte. Le compte permet d'accéder aux données privées selon ses droits dans le cas de portails internes, et de s'abonner à des **notifications** sur les jeux de données.

2.1 - Page d'accueil

La page d'accueil est le point d'entrée principal du portail de données. Elle permet **d'accéder rapidement aux données** à partir des thématiques et d'un champ de recherche textuelle. Des **chiffres clés** permettent de quantifier les données présentes sur le portail.

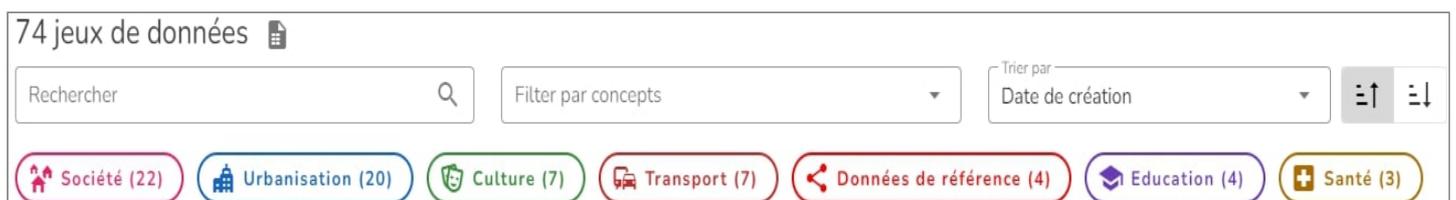
Il est possible de mettre en avant du **contenu éditorial** (texte expliquant la démarche ou présentant le portail), ainsi qu'une visualisation de données qui est directement utilisable sur cette page. On peut, par exemple, mettre en avant une carte présentant la localisation de centres de vaccination, ou les menus des cantines pour la semaine. Dans le cadre d'une démarche de communication, il est également possible d'afficher sur cette page un fil Twitter ou un fil d'actualités.

Pour les visiteurs réguliers, des sections présentent les **derniers jeux de données ajoutés**, ainsi que les dernières visualisations réalisées, avec une navigation rapide pour voir plus en détail ces nouveaux contenus.

Il est possible de personnaliser quels éléments sont affichés sur cette page : une bannière plutôt qu'un texte éditorialisé, par exemple, ou même d'utiliser une visualisation à la place de la bannière. Cette dernière option permet de présenter un **carrousel avec des liens de navigation**.

2.2 - Catalogue de données

Le catalogue de données est un moteur de recherche permettant d'**accéder rapidement** aux jeux de données susceptibles d'intéresser l'utilisateur. Outre le champ de recherche textuelle, il est possible d'accéder aux jeux de données par thématique ou par concept présent dans les données. Il est, par exemple, possible de lister tous les jeux de données géographiques en filtrant par concept « Latitude », ou toutes les données liées à des entreprises en filtrant par SIREN.



The screenshot shows the top section of the data catalog interface. At the top left, it displays '74 jeux de données' with a document icon. Below this is a search bar with the placeholder text 'Rechercher' and a magnifying glass icon. To the right of the search bar is a dropdown menu labeled 'Filter par concepts'. Further right is another dropdown menu labeled 'Trier par' with 'Date de création' selected. To the right of the sorting menu are two icons for ascending and descending sort order. Below these elements is a horizontal row of six category buttons, each with an icon and a count in parentheses: 'Société (22)' (house icon), 'Urbanisation (20)' (city icon), 'Culture (7)' (globe icon), 'Transport (7)' (train icon), 'Données de référence (4)' (document icon), and 'Santé (3)' (plus sign icon).

La liste des jeux de données est parcourue au moyen d'un mécanisme de scroll infini, aussi bien adapté à une utilisation **bureautique ou mobile**. Il est également possible de trier les résultats suivant

différents critères (alphabétique, date de création...) La liste des résultats obtenus peut être exportée au format CSV en un clic.

Les résultats sur cette page sont présentés sous forme de vignettes qui affichent des informations, comme le titre du jeu de données, sa date de mise à jour ou les thématiques qui lui sont associées. Une partie de la description est également affichée, mais elle peut être remplacée par une image pour faire un catalogue plus « visuel ».

En plus de permettre de naviguer vers la page de détails d'un jeu de données, les vignettes présentent des boutons d'action permettant de :

- visualiser les données avec une **vue tabulaire** dans laquelle on peut trier les colonnes, paginer, effectuer des recherches fulltext et télécharger les données filtrées ;
- éventuellement visualiser les données avec une **vue cartographique** quand les données le permettent ;
- accéder à la **documentation interactive de l'API** ;
- consulter le **schéma des données**.



Le catalogue présente les jeux de données que l'utilisateur a le droit de voir. S'il n'est pas connecté, il ne verra que les jeux en open data ; s'il est connecté et membre de l'organisation propriétaire du portail, il pourra, en plus, voir des jeux de données privés.

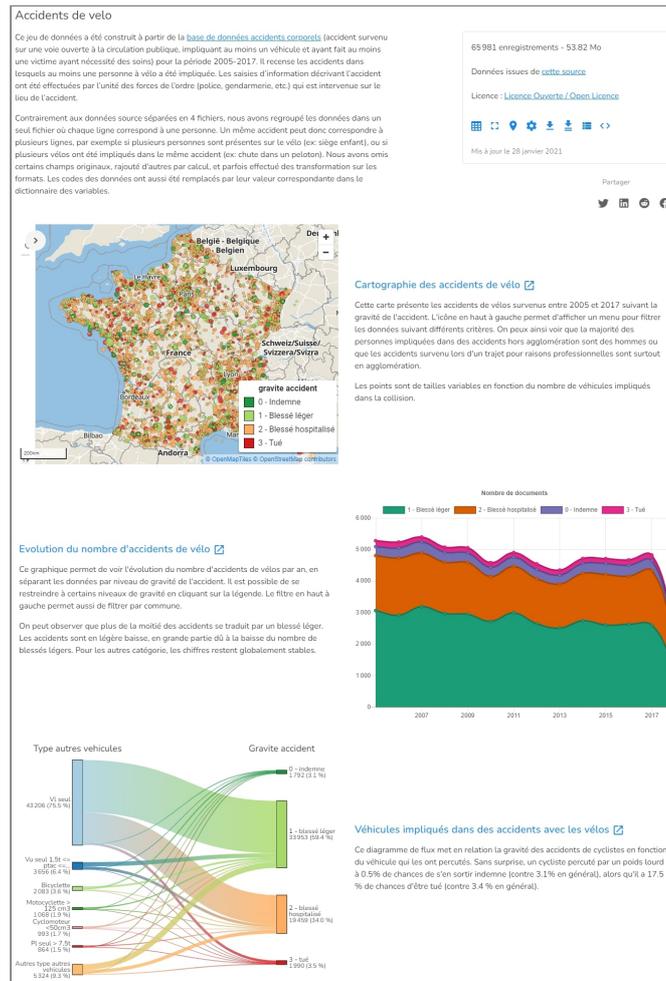
2.3 - Détails d'un jeu de données

Pour pouvoir être accessibles au plus grand nombre, les jeux de données sont présentés avec des **visualisations adaptées aux données**. Cela permet de raconter une histoire et de mettre en avant différents aspects du jeu de données ou des cas d'utilisations possibles. Les visiteurs du portail peuvent rapidement s'imprégner des données et **interagir** avec elles.

Différentes **métadonnées** (producteur, licence, date de dernière mise à jour...) sont présentées à l'utilisateur, et on retrouve les boutons d'action des vignettes du catalogue (vue tableau, accès au schéma, documentation d'API). D'autres actions supplémentaires sont offertes à l'utilisateur. Il peut **télécharger les données sources** qui ont été uploadées par le producteur. Si l'utilisateur est authentifié, il peut s'abonner aux notifications liées à ce jeu de données, ce qui lui permet d'être informé quand celles-ci sont mises à jour.

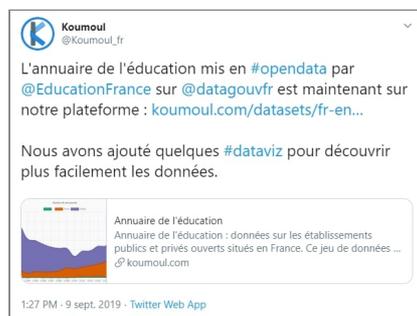
Il a accès à une vue tableau en plein écran qui lui permet de voir plus de lignes et de colonnes, et surtout d'obtenir une URL qui correspond aux filtres qu'il a appliqués et qu'il peut **partager avec d'autres utilisateurs**. Il peut ainsi partager des données qui seraient, par exemple, filtrées sur une commune particulière. Il est possible de n'afficher que certaines colonnes et de télécharger le jeu de données filtré dans plusieurs formats (CSV, ODS, XLSX, GeoJSON...)

L'utilisateur peut également accéder aux **pièces jointes** du jeu de données. Ce sont en général des fichiers chargés en complément des fichiers de données, comme un manuel d'utilisation des données au format PDF.



En plus des visualisations du portail, qui sont des réutilisations des données « internes », d'autres réutilisations de données peuvent être mentionnées. Celles-ci peuvent être présentées au travers d'une vignette cliquable, ou directement embarquées dans la page en iframe.

Différents boutons permettent de partager la page du jeu de données sur les réseaux sociaux (Facebook, Twitter, LinkedIn...) La page de données possède des **métadonnées enrichies** qui permettent une meilleure indexation par les moteurs de recherche et un affichage d'une miniature sur les réseaux sociaux. Cette miniature présente la première visualisation associée au jeu de données, et ce côté graphique permet d'**augmenter l'engagement** des utilisateurs des réseaux sociaux.



2.4 - Catalogue des visualisations

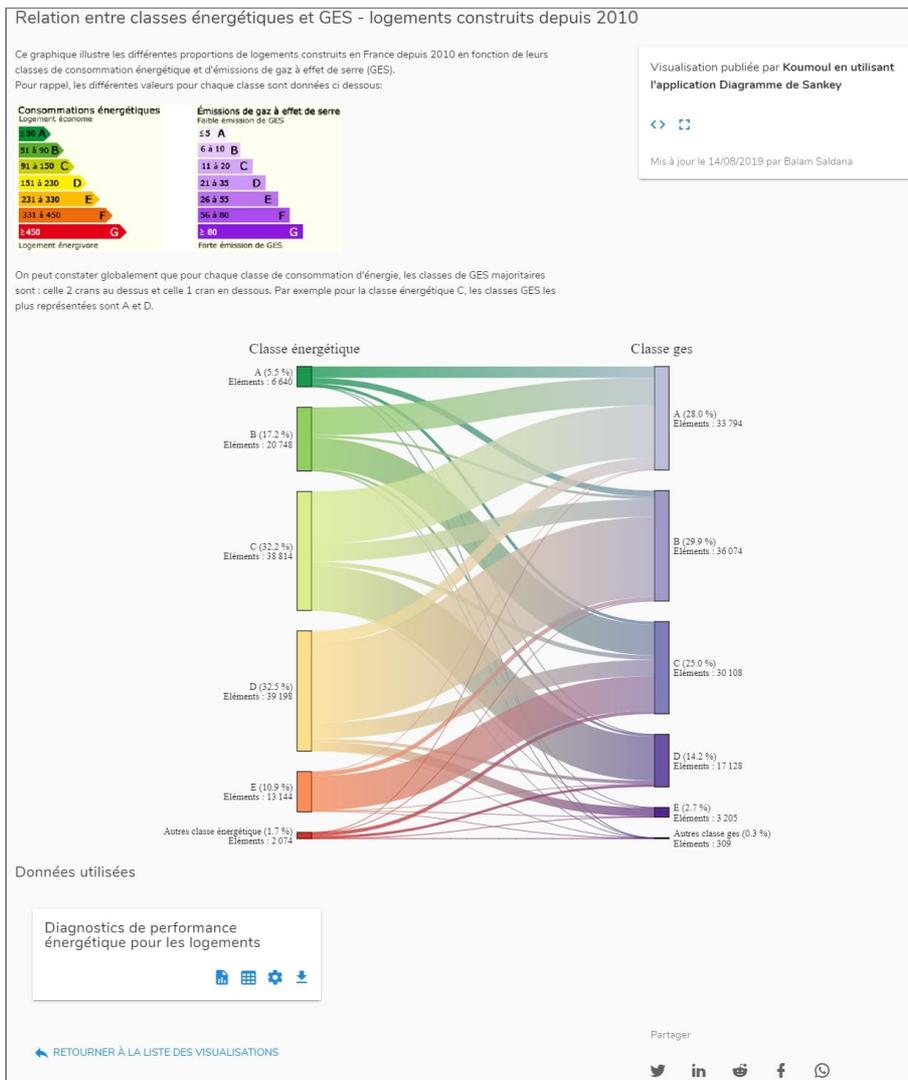
Tout comme pour les jeux de données, un catalogue des visualisations est disponible avec un **moteur de recherche** et des outils pour filtrer ou trier les visualisations. On peut ainsi accéder rapidement aux visualisations correspondant à certaines thématiques ou liées à des données ayant des concepts particuliers.

Les visualisations sont présentées au travers de **vignettes imagées**. Comme pour les jeux de données, ces vignettes possèdent différents boutons d'action pour pouvoir **interagir** rapidement avec la visualisation ou naviguer vers une vue plein écran de celle-ci. La liste des résultats est parcourue au moyen d'un mécanisme de scroll infini, aussi bien adapté à une utilisation **bureautique ou mobile**.

La plateforme offre une grande variété de visualisations, et de nouvelles sont régulièrement ajoutées. Il y a des cartes interactives pour présenter des données géographiques, des graphiques, visualisations animées, nuages de mots, etc. Il est même possible de réaliser des moteurs de recherche, plus adaptés que la vue tabulaire pour présenter des données avec des textes longs, ou des mini-jeux qui permettent de découvrir les données en jouant avec elles. L'ensemble des visualisations disponibles est présenté plus en détail dans la section 3.2.

2.5 - Consultation d'une visualisation

Chaque visualisation est présentée au travers d'une **page dédiée**, ce qui permet de lui associer contexte, titre et description, ainsi que d'offrir différentes actions à l'utilisateur. Il est possible d'accéder à la visualisation en plein écran (les visualisations sont consultables aussi bien sur **poste fixe que smartphone**) et de récupérer un code HTML qui permet d'embarquer la visualisation dans un autre site web.



Il est également possible de générer une image de la visualisation ou de partager la page à quelqu'un d'autre dans l'état de navigation : on peut, par exemple, partager une carte zoomée sur une certaine commune ou avec certains filtres déjà appliqués.

Tout comme pour les jeux de données, il y a, en bas de page, différents boutons pour **partager** la page de la visualisation sur différents réseaux sociaux (Facebook, Twitter, LinkedIn...)

La page de la visualisation possède des **métadonnées enrichies** qui permettent une meilleure indexation par les moteurs de recherche et l'affichage d'une miniature sur les réseaux sociaux.

2.6 - Pages de contenus

Les **pages de contenus** peuvent être de différents types : articles, pages thématiques autour de plusieurs jeux de données, pages d'actualités, data storytelling, licences, conditions d'utilisation, etc. Il est ainsi possible de **mettre en avant certaines données** et de leur donner encore plus de contexte, ou de faire des tableaux de bord intégrant différentes données.

En plus de la saisie de texte libre, l'utilisateur peut intégrer différents types d'éléments : tableau d'un jeu de données, visualisation, liste de jeux de données, intégration de contenus externe, etc.

Pour accéder aux pages de contenus créées, il est possible de renseigner des liens dans la barre de navigation. Les liens peuvent apparaître directement dans la barre, ou dans un menu ajouté à celle-ci. Les pages d'actualités sont accessibles soit depuis un fil d'actualités en page d'accueil, soit depuis la page qui liste toutes les actualités et qui est visible depuis la barre de navigation.

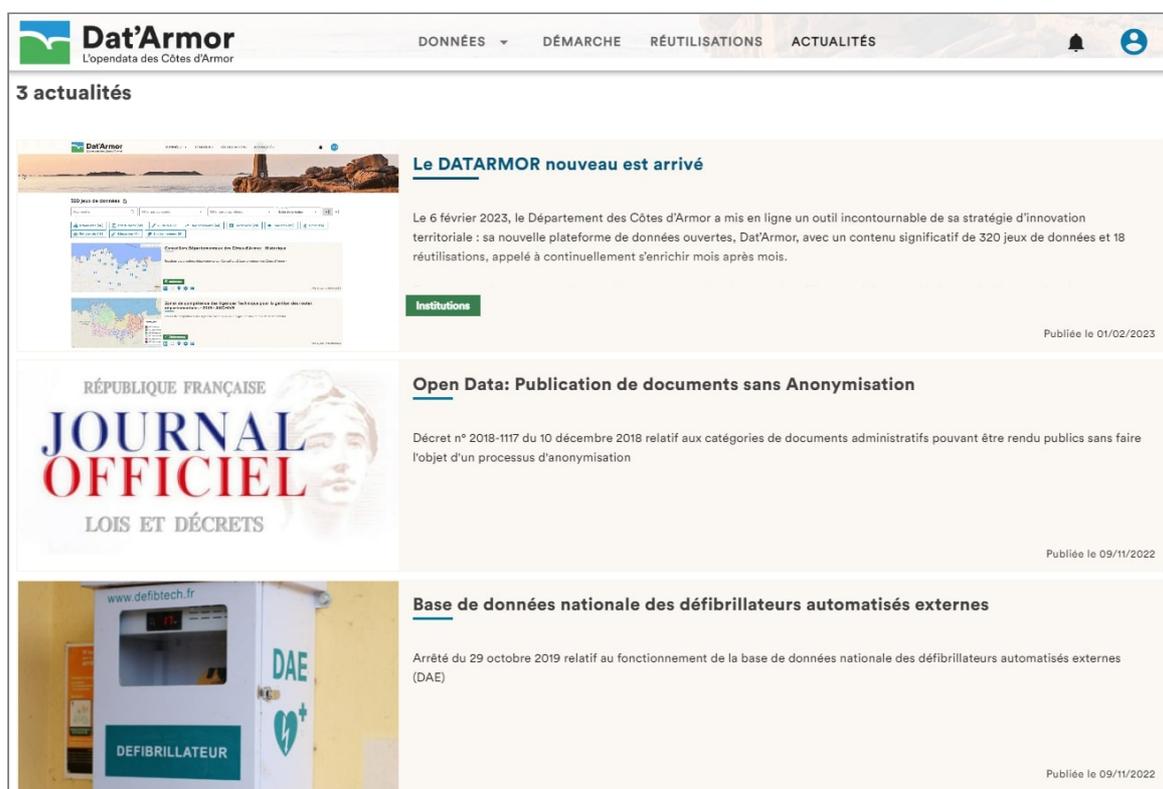
Il est possible de créer des pages **publiques** ou des pages **privées**.

Pages d'actualités

Les pages d'actualités permettent de communiquer les informations sur les événements concernant votre portail ou l'open data. Elles permettent d'établir une première phase de communication avec les visiteurs de votre portail.

La liste d'actualités est disponible sur la page des actualités sous forme de fiche contenant une image et le début de l'actualité (ou son résumé). En cliquant sur une fiche, les visiteurs accèdent aux informations concernant l'actualité.

Comme pour les pages de contenus, différents éléments peuvent être intégrés, tels que des images, vidéos, formulaires, tableaux de jeux de données, PDF, liens, etc.



The screenshot shows the Dat'Armor website interface. At the top, there is a navigation bar with the logo 'Dat'Armor L'opendata des Côtes d'Armor' and menu items: 'DONNÉES', 'DÉMARCHE', 'RÉUTILISATIONS', and 'ACTUALITÉS'. Below the navigation bar, the page displays '3 actualités'. The first news item is titled 'Le DATARMOR nouveau est arrivé', dated 01/02/2023, with a thumbnail image of a coastal landscape. The second item is 'Open Data: Publication de documents sans Anonymisation', dated 09/11/2022, with a thumbnail image of the 'JOURNAL OFFICIEL' cover. The third item is 'Base de données nationale des défibrillateurs automatisés externes', dated 09/11/2022, with a thumbnail image of a defibrillator machine.

Pour faciliter l'accès aux informations, un fil d'actualités est également disponible en page d'accueil.

Dernières actualités

Le DATARMOR nouveau est arrivé

Le 6 février 2023, le Département des Côtes d'Armor a mis en ligne un outil incontournable de sa stratégie d'innovation territoriale : sa nouvelle plateforme de

Publiée le 01/02/2023

Open Data: Publication de documents sans Anonymisation

Décret n° 2018-1117 du 10 décembre 2018 relatif aux catégories de documents administratifs pouvant être rendu publics sans faire l'objet d'un processus

Publiée le 09/11/2022

Base de données nationale des défibrillateurs automatisés externes

Arrêté du 29 octobre 2019 relatif au fonctionnement de la base de données nationale des défibrillateurs automatisés externes (DAE)

Publiée le 09/11/2022

TOUTES LES ACTUALITÉS

2.7 - Compte utilisateur

Si le portail de données est public, il n'y a pas d'obligation de créer de compte pour l'utiliser. Les utilisateurs peuvent, s'ils le souhaitent, créer un compte pour **s'abonner à des notifications** et créer des **clés d'API** pour utiliser les API avec moins de restrictions. Dans le cas où le portail est privé, les utilisateurs auront besoin d'un compte, mais il leur faudra aussi des autorisations données par un administrateur de l'organisation propriétaire du portail.

Pour limiter les problématiques liées au RGPD, un **minimum de données est collecté** et la seule donnée requise est l'e-mail de l'utilisateur. Il peut renseigner un prénom et un nom s'il le souhaite, ou mettre un pseudonyme à la place. Si l'utilisateur ne se connecte pas à son compte pendant trois ans, celui-ci est automatiquement supprimé. Les utilisateurs peuvent également **supprimer leur compte** à l'aide d'un bouton, sans à avoir à formuler une demande par e-mail ou autre.

La création de compte se fait en renseignant un e-mail et un mot de passe, il y a aussi la possibilité de passer par un compte Gmail, Facebook, LinkedIn ou Github via le **protocole oAuth2**. Un mécanisme de renouvellement de mot de passe est disponible pour les utilisateurs ayant perdu celui-ci ou désirant le changer.

Les utilisateurs qui créent des comptes par eux-mêmes ont leur données stockées dans des bases de données. Leur mot de passe est encrypté avec sels et multiples hachages pour garantir une sécurité maximale. Des règles de saisie empêchent la création de mots de passe faibles.

Il est également possible de configurer une connexion à un annuaire d'utilisateurs externes au travers du protocole LDAP.

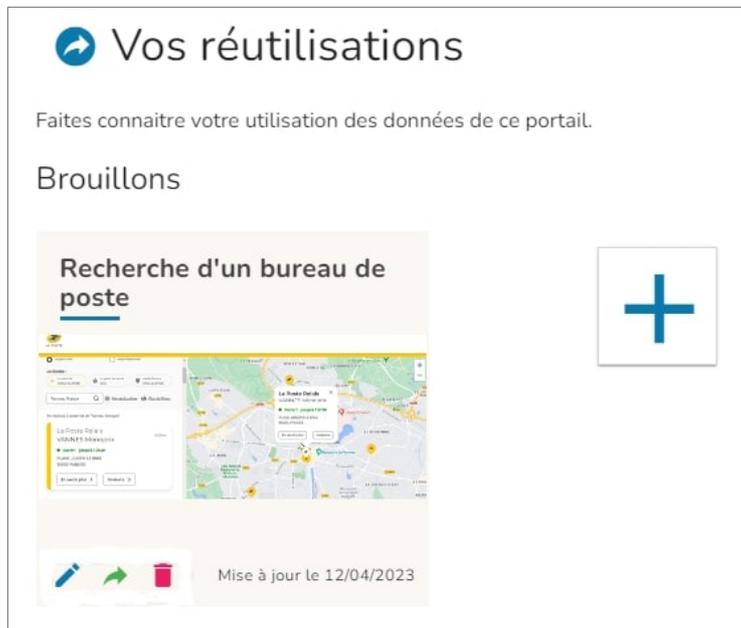
Les comptes créés sur le portail peuvent être utilisés pour créer des partenariats.

Un administrateur du portail peut donner les droits de contribution sur un ou plusieurs jeux de données à un compte utilisateur (ou un compte organisation) qui a été créé sur le portail.

Les partenaires pourront ainsi modifier les jeux de données en remplaçant le fichier au complet ou en éditant les lignes du jeu de données à partir de leur espace personnel.

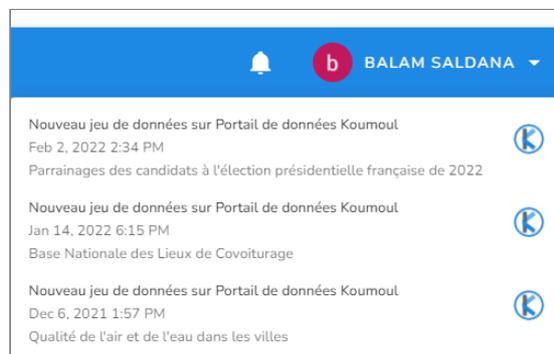
Les portails peuvent être configurés pour accepter les soumissions de réutilisations.

Celles-ci sont effectuées à partir du compte personnel de l'utilisateur et elles sont soumises à modération : c'est l'administrateur qui choisit de publier la réutilisation ou non.

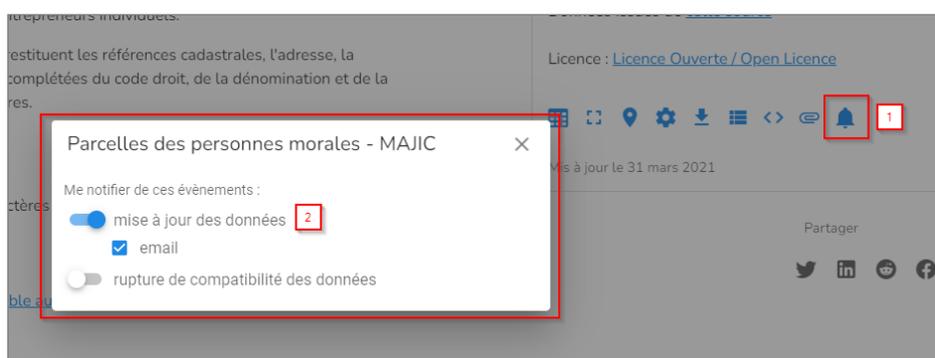


2.8 - Notifications

Les notifications permettent aux utilisateurs enregistrés de recevoir des alertes dans le portail ou par e-mail lorsque certains événements se produisent. Les alertes sont visibles sur la cloche de la barre de navigation et si l'utilisateur l'a choisi, il peut recevoir un **e-mail d'alerte**.



Les utilisateurs peuvent **s'abonner** à l'ajout d'un jeu de données sur le portail ou dans une thématique en particulier. Il est aussi possible de s'abonner à la mise à jour d'un jeu de données en particulier.



2.9 - Réutilisations

Un des principaux intérêts d'un portail de données est de permettre à différentes données de mieux circuler et donc d'être réutilisées au maximum. Pour mettre en avant cette démarche, il est possible de publier des pages listant les différentes utilisations des données publiées.

L'intérêt de ces pages est double. Pour le producteur de données, cela **valorise son action** de mise à disposition de ses données et permet de le conforter dans cette démarche de partage. Pour l'utilisateur des données, cela lui offre un **moyen d'être référencé** et de mettre en avant son projet qui utilise les données du portail.

The screenshot displays a web interface for data reuse. At the top left, there is a search bar labeled 'Rechercher' and a 'Trier par' dropdown menu set to 'Date de publication'. Below this, a grid of 18 reuse examples is shown, arranged in three rows and six columns. Each example consists of a title, a thumbnail image, and a publication date. The examples include:

- Carambole & Confinement**: A thumbnail showing people in historical attire with the text 'Commencez chaque jour par un sourire, après c'est selon l'humeur...'. Published le 10/11/2022.
- DATASPOT**: A thumbnail showing a website interface with various data points and charts. Published le 10/11/2022.
- HYBLAB - Datajournalisme - La culture en Bretagne**: A thumbnail showing a dashboard with multiple charts and maps. Published le 10/11/2022.
- Guingamp News22.fr**: A thumbnail showing a photograph of a stone building with a spire. Published le 24/10/2022.
- Qui dit miam !**: A colorful thumbnail with text 'Qui dit miam!' and 'C'est ce qu'on mange à la cantine?'. Published le 24/10/2022.
- Site de la commune de Goudelin**: A thumbnail showing a building facade with a sign that says 'MAIRIE'. Published le 24/10/2022.

Les pages de réutilisations de données peuvent être créées par les administrateurs du portail, mais on peut aussi permettre aux utilisateurs enregistrés de soumettre de nouvelles pages. Ces soumissions sont ensuite mises en attente de modération et c'est l'administrateur qui choisit de publier ou non la réutilisation.

2.10 - Accès par API

L'ensemble des fonctionnalités de la plateforme est disponible au travers d'**API Rest documentées**. Ces API peuvent être appelées en dehors du portail, mais pour les accès restreints, il faut passer par l'utilisation d'une **clé d'API**. Lors de l'ajout d'une clé d'API, il est possible de restreindre l'accès à une seule fonction. On peut ensuite limiter l'accès à une IP ou à un nom de domaine précis.

La documentation des API est réalisée en suivant la spécification **OpenAPI 3.0**. Cela permet une description claire et compréhensible au travers d'une interface interactive. La prise en main des APIs par les développeurs est ainsi plus rapide.

Un autre avantage d'utiliser cette spécification est l'**interopérabilité** accrue, certains systèmes informatiques (par exemple, les gateway API) étant capables de comprendre cette spécification. Les API réalisées avec Data Fair peuvent, par exemple, être directement intégrées par des sites comme <https://api.gouv.fr>.

SINOE® - Annuaire 2015 des déchèteries de déchets ménagers et assimilés (DMA)

API du jeu de données : SINOE® - Annuaire 2015 des déchèteries de déchets ménagers et assimilés (DMA)

Requêter les lignes du jeu de données.

GET <https://koumoul.com/s/data-fair/api/v1/datasets/sinoe-annuaire-decheteries-dma-2015-ok/lines>

DOCUMENTATION MAKE REQUEST

Security

Name	Type	Location
x-apiKey	apiKey	header
id_token	apiKey	cookie

Parameters

Name	Description	Type	Values
page	Le numéro de la page (indice de la pagination). Débute à 1.	integer	
format	Le format de la donnée. json par défaut, geojson et pbf pour tuiles vectorielles.		json, pbf, geojson

Il est possible d'utiliser l'API pour **moissonner les données d'un portail**. Par exemple, le site <https://opendatarchives.fr/> moissonne régulièrement les données du portail <https://data.ademe.fr/> propulsé par Data Fair.

2.11 - Formulaire de contact

Pour collecter les remarques, demandes ou questions des utilisateurs, un formulaire de contact peut être mis à disposition par simple paramétrage de l'adresse e-mail de contact. Cette adresse n'est pas visible sur la page du formulaire pour éviter d'être collectée par des sites malintentionnés.

Il y a également des mécanismes de protection qui sont mis en place pour éviter la soumission du formulaire par des robots.

3 - Gestion des contenus (back-office)

Le **back-office** permet de gérer différents éléments sur la plateforme : les données, les visualisations, les portails, les membres des organisations, les permissions, les connecteurs vers d'autres catalogues de données (en entrée ou en sortie) et les traitements périodiques. Il permet également d'accéder à des métriques d'exploitation.

L'accès au back-office est, bien sûr, restreint à des utilisateurs authentifiés, en particulier à ceux qui ont les rôles d'administrateur ou de contributeur.

3.1 - Administration des jeux de données

Les jeux de données permettent de mettre à disposition de l'utilisateur des données, ainsi que des informations sur celles-ci (métadonnées) comme la licence qui leur est associée, la date de mise à jour, leur propriétaire, etc.

Types de jeux de données

Il existe plusieurs types de jeux de données sur la plateforme : les fichiers, les jeux de données éditables, les jeux de données virtuels et des jeux de données externes.

- Les **jeux de données fichiers** correspondent à des données sous format tabulaire ou cartographique chargées sur la plateforme. Plusieurs formats de fichiers sont supportés, tels que CSV, TSV, OpenDocument, XLS, XLSX, GeoJSON, KML, KMZ, ESRI Shapefile, GPX et iCalendar. Suivant les besoins, de nouveaux formats de fichiers sont régulièrement ajoutés. Quand un fichier est chargé, il est converti en interne dans un format plus standard, puis analysé automatiquement pour déterminer le schéma du jeu de données. Le contributeur peut ensuite modifier ce schéma, par exemple, déterminer qu'une colonne aura sa donnée indexée en tant que chaîne de caractère plutôt que nombre entier.
- Les **jeux de données éditables** sont des données stockées en base et sont plutôt adaptés à des données qui évoluent régulièrement, ou mises à jour par des personnes métier qui veulent juste modifier quelques lignes. La création de ce type de jeu de données se fait en éditant son schéma de données : on définit chaque colonne et son type. Pour des données produites par des systèmes informatiques (données IOT, par exemple), ces jeux de données sont généralement mis à jour par API. Dans le cas de mises à jour manuelles par des agents, cela se fait via un formulaire de saisie.
- Les **jeux de données virtuels** correspondent à des vues d'un ou plusieurs jeux de données. Ils permettent d'avoir un contrôle d'accès plus poussé. Ils peuvent, par exemple, servir à créer une vue publique, restreinte à certaines lignes et certaines colonnes, d'un jeu de données plus complet qui reste privé. Cela permet aussi de faire de l'anonymisation de données. On peut, par exemple, restreindre un jeu de données national à un seul département. L'autre cas d'usage de ce type de

jeu de données est d'opérer avec des données millésimées ou territorialisées : on peut ainsi publier des données chaque année via un fichier qui a toujours le même format, puis faire une vue qui regroupe les différentes années.

- Les **jeux de données externes** n'ont pas de données indexées sur la plateforme. Ils permettent de renseigner des métadonnées (titre, description, licence...) et d'y associer des données dans des formats qui ne sont pas exploitables par la plateforme (PDF, archive ZIP...) ou de cataloguer des données présentes sur d'autres plateformes en renseignant un lien dans la description.

Schéma des données

La plateforme supporte l'indexation de données tabulaires. Chaque jeu de données (excepté les jeux de données externes) possède un schéma qui est la description des différentes colonnes (ou champs) qui le composent. *A minima*, chaque colonne a un identifiant et un type, mais il est possible de renseigner des informations complémentaires.

Un libellé et une description permettent d'avoir des entêtes de colonnes plus lisibles et compréhensibles. Le champ peut avoir un groupe qui permet de le retrouver plus rapidement quand il y a beaucoup de colonnes. Si le champ est de type texte, on peut opter pour du formatage riche : il sera alors possible de mettre du HTML ou Markdown dans ce champ. Le champ peut également être défini comme étant multivalué : dans ce cas, on spécifie le séparateur utilisé dans la colonne entre les différentes valeurs.

The screenshot shows the 'Structure' interface with a search bar for '20 colonnes'. A list of columns is displayed, with 'Code Commune' selected. The configuration panel for 'Code Commune' includes:

- Libellé: Code Commune
- Description: A rich text editor with a toolbar.
- Groupe: A dropdown menu.
- Clé dans la source: Code Commune
- Type: Texte
- Nombre de valeurs distinctes: 34967
- Concept: Code commune (highlighted with a red box)
- Séparateur: A dropdown menu.
- Ignorer la détection de type:
- Formaté:

Le dernier élément qui peut être renseigné et qui a une importance considérable est le **type métier** associé au champ. Cela se fait en sélectionnant un **concept issu d'un thésaurus**. Il y a une base de concepts communs à toute la plateforme et il est possible de **rajouter ses propres concepts**. Ceux-ci sont, en général, liés à du vocabulaire issu du Web sémantique, le concept de code postal a, par exemple, l'identifiant `http://schema.org/postalCode`.

Ce typage métier **augmente la réutilisabilité** des données et permet deux choses au sein de la plateforme : **l'enrichissement à partir d'autres données**, et la proposition de **visualisations adaptées** (en simplifiant le paramétrage de celles-ci) : les concepts « latitude » et « longitude » permettent, par exemple, de paramétrer des cartes avec des marqueurs.

Référentiel de schémas

Pour améliorer la saisie et la réutilisation des données, nous proposons un référentiel de schémas issus de <https://schema.data.gouv.fr/schemas.html> qui recense différents schémas de données publiques pour la France.

Lors de la création d'un jeu de données éditable, il est possible de choisir un des schémas proposés. Les données saisies seront standardisées, ce qui augmentera leur réutilisabilité, tant au niveau de la remontée de différentes données vers une destination commune pour fusionner les données, que dans l'utilisation de visualisations.

Par exemple, une application permettant de rechercher et visualiser des délibérations publiées au format SCDL pourra être utilisée par toute collectivité qui publie ses données de délibérations avec le schéma adapté.

Métadonnées et pièces jointes

Certaines métadonnées sont préremplies, comme les dates de mise à jour des métadonnées ou des données et l'utilisateur qui les a créées ou modifiées. La page d'édition d'un jeu de données permet de modifier les différentes métadonnées de ce jeu. Il est possible de modifier le titre et la description, de définir une **licence d'utilisation** et d'associer des **thématiques**. Les listes de licences et thématiques utilisables sont communes à toute l'organisation et peuvent être éditées par les administrateurs.

Il est possible d'associer des **pièces jointes à chaque ligne** d'un jeu de données. Cela se fait en associant une archive au format zip qui contient les fichiers à associer. Il faut aussi qu'il y ait dans le jeu de données une colonne contenant les noms des fichiers à associer à chaque ligne. Deux types de fichiers peuvent être liés aux lignes : des images (png, jpg, etc.) ou des documents (pdf, docx, xlsx, etc.) Dans le cas des documents, ils peuvent être indexés **fulltext** par la plateforme pour que les recherches tiennent compte du contenu de ces documents.

Les pièces jointes peuvent aussi être **directement attachées à un jeu de données**. On peut, par exemple, ajouter des fichiers de documentation ou des métadonnées riches. On peut également s'en servir pour publier des données qui ne peuvent pas être indexées par la plateforme dans le cas d'un jeu de données de type externe.

Données maîtres et enrichissement

Certaines données peuvent être utilisées à différents endroits et dans différents processus par les organisations. Il est possible de définir un jeu de données comme étant des **données maîtres** (*master data*, en anglais). Ce sont des données qui font référence à des concepts particuliers, et la plateforme met à disposition de toutes les organisations des données maîtres mutualisées.

Les données de la **base Sirene** sont rattachées aux concepts de « code SIREN », « code SIRET » et « code APE », par exemple. Il y a également des données mises à disposition à partir du **cadastre** (via des « codes parcelle ») ou des données de l'INSEE (via les « codes commune », « codes département »...) Il y a enfin des données d'adresses, issues de la **BAN**, qui permettent de faire du géocodage. De nouvelles données de référence sont régulièrement ajoutées à la plateforme, et chaque organisation peut elle-même créer ses propres jeux de données pivots.

Ces données maîtres sont d'une grande valeur, car elles permettent de **compléter facilement les autres données**. Dans le cadre des formulaires de saisie des jeux incrémentaux, on peut faire référence à des données maîtres en assignant un concept à un certain champ, et le **formulaire proposera une liste de valeurs** (avec un moteur de recherche si elle est grande) pour sélectionner ce qui sera mis dans le champ. Il est ainsi possible de contraindre la saisie dans un champ et de s'assurer que les valeurs dedans soient toutes valides suivant certaines règles métier.

La deuxième possibilité pour compléter les données est de mettre en place des **enrichissements** : des colonnes sont alors automatiquement ajoutées au jeu de données et les valeurs renseignées à partir d'une ou plusieurs autres colonnes. Par exemple, des colonnes qui ont les concepts « numéro de rue », « libellé de rue » et « code postal » peuvent être complétées par les données d'adresse et être géocodées, ce qui permet de projeter les données sur une carte. Quand les données sont actualisées, les enrichissements sont automatiquement mis à jour, et un jeu de données peut avoir plusieurs enrichissements provenant de données maîtres différentes.

Les données de référence peuvent également être utilisées via les jeux virtuels pour proposer les données de référence sur son portail. Il est alors possible de créer un jeu virtuel sur le portail de son organisation qui va être une vue du jeu de données de référence. Ainsi, lorsque les données de références seront mises à jour, le jeu virtuel sera également actualisé.

Les jeux virtuels peuvent posséder des filtres sur les valeurs de colonnes. Il est, par exemple, possible de filtrer les données de la **base Sirene** sur un territoire et sur certains **code APE** pour proposer un jeu de données sur un secteur d'activité de votre territoire. Il sera alors très simple de configurer une visualisation cartographique pour explorer ces données.

Permissions et publication des données

Un administrateur peut contrôler finement les permissions d'**accès aux données**. Les données ne sont, de base, visibles que par les administrateurs et les contributeurs de l'organisation. Il est ensuite possible de les partager en interne, c'est-à-dire que seuls les membres de l'organisation authentifiés peuvent les consulter, ou de les rendre **publiques** : dans ce cas, tout le monde pourra y accéder, y compris les utilisateurs non enregistrés.

On peut également définir des droits d'accès à certains utilisateurs ou à des organisations partenaires. Un mode avancé permet de définir les permissions pour chaque point d'accès de l'API d'un jeu de données : on peut, par exemple, rendre l'accès aux métadonnées public alors que l'accès aux données reste restreint.

Editer un ensemble de permissions

Portée
Organisation

Nom
Koumoul

Rôles (tous si aucun coché)
user

Actions

- Lister
- Lecture

ANNULER VALIDER

Quand un jeu de données vient d'être créé, il n'est pas encore disponible dans les différents portails de données de l'organisation. Il doit d'abord être **publié dans un ou plusieurs portails**. Dans le cadre de portails open data, le jeu de données doit aussi avoir une permission d'accès public en plus d'être publié.

Ce mécanisme de publication permet de travailler de manière agile : on peut, par exemple, avoir un **portail de recette** sur lequel on publie les jeux de données que l'on souhaite ouvrir prochainement, en les accompagnant de visualisations. Celles-ci peuvent mettre en avant un problème dans les données ou une mauvaise structuration de celles-ci, problème qui peut être vu par plusieurs personnes, car les données sont déjà publiées sur un portail. Une fois que l'on atteint la **qualité de publication souhaitée**, on dépublie le jeu de données du portail de recette et on le publie sur un ou plusieurs portails de production.

Partage

PERMISSIONS PORTAILS CATALOGUES

Configurez des portails pour mieux partager vos données au public ou en interne.

Publiez ce jeu de données sur un ou plusieurs de vos portails.

- Portail de données Koumoul
<https://opendata.koumoul.com/datasets/prenom-france-1900-2017>
- Portail de pré-production
- Portail interne

Il est aussi possible de publier un jeu de données sur des portails ou catalogues de données externes à la plateforme, ceci est décrit plus en détail à la section sur les **connecteurs de catalogues**.

Journal des événements

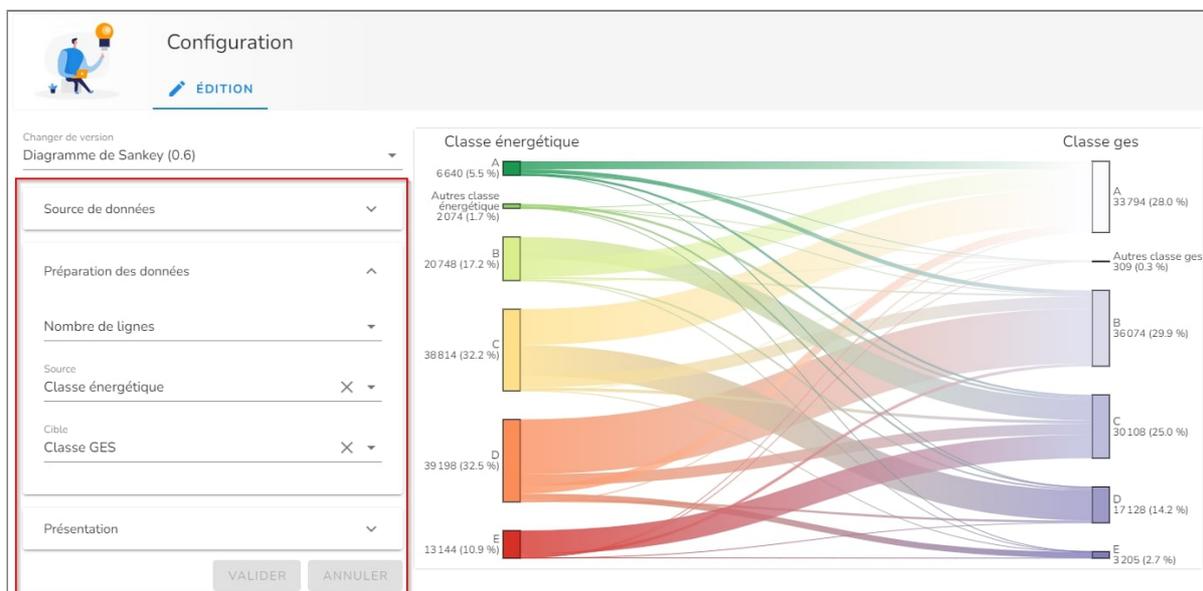
Chaque étape du traitement d'un jeu de données laisse des traces dans le journal associé. On peut ainsi retracer les actions, leurs dates, leurs durées et les éventuelles erreurs rencontrées.

3.2 - Paramétrage des visualisations

Les visualisations interactives permettent de présenter les données de manière synthétique et offrent à l'utilisateur la possibilité de les manipuler pour les comprendre plus rapidement. Elles sont servies par des applications web légères, ce qui permet une consultation directe sur tout type de terminal web, ou une intégration iframe dans des portails ou sites institutionnels.

Paramétrage intuitif

Les visualisations peuvent être configurées grâce à une interface graphique qui ne **requiert pas de compétences en programmation**. Le menu de configuration se compose de différentes sections qui diffèrent selon les applications. La plupart du temps, le menu est composé de trois sections : la source de données, les options de rendu et les éléments liés à la navigation. Il est souvent possible de filtrer les données si l'on ne souhaite pas représenter le jeu de données en entier.



Un aperçu donne un rendu de la visualisation. Lorsque l'on réalise des modifications sur le menu de configuration, elles sont directement représentées sur l'aperçu. Il est ainsi possible de modifier et tester rapidement différents rendus de la visualisation, en **mode brouillon**. Lorsque le rendu est satisfaisant, l'enregistrement permet de valider les modifications réalisées. Celles-ci sont alors visibles sur toutes les publications de la visualisation.

Types de visualisations

Data Fair offre une grande variété de visualisations en constante progression, il y a aujourd'hui près d'une **trentaine de visualisations différentes**. ([Cette section](#) décrit en détail la manière de les créer).

Visualisations cartographiques

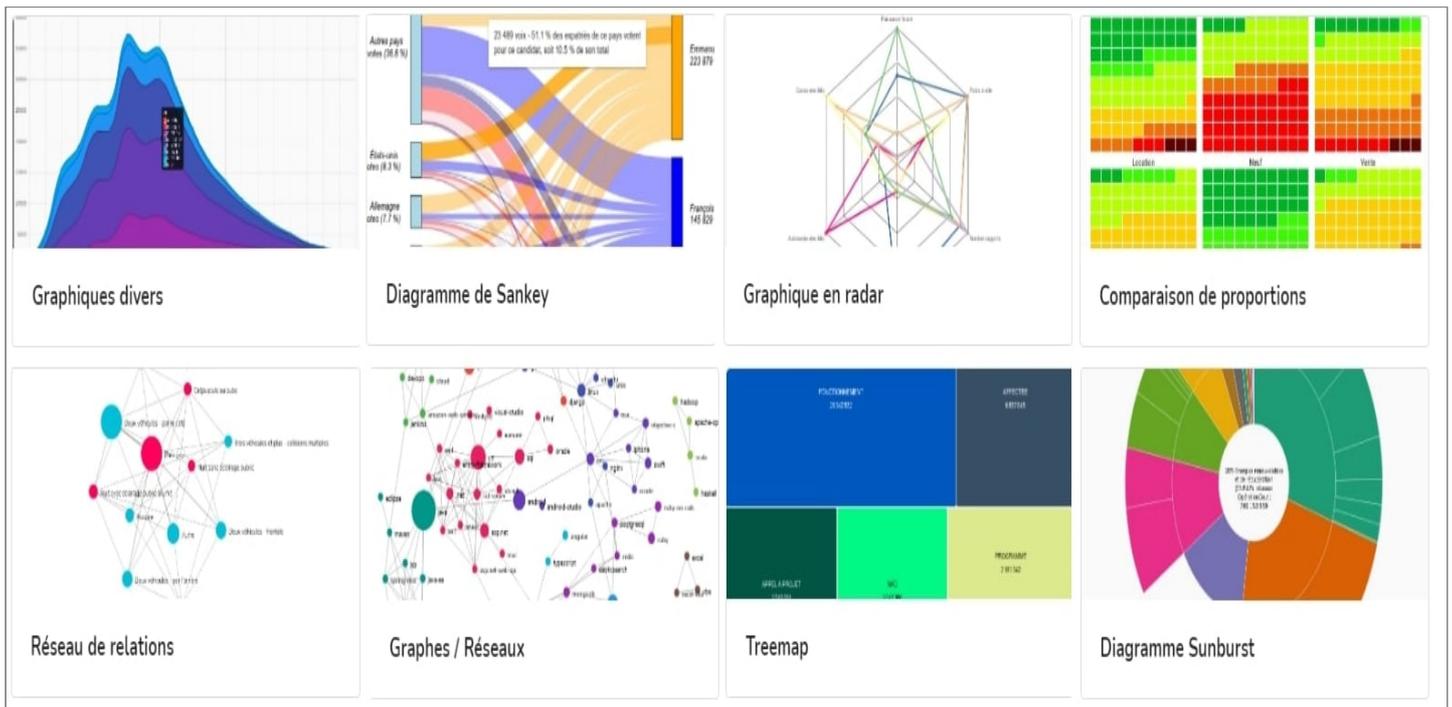
Plusieurs visualisations permettent d'afficher des données géolocalisées. L'une d'elle permet d'afficher des **marqueurs** qui présentent au clic une **fiche détaillée**, éventuellement accompagnée d'une image. Une autre affiche les données sous forme de **cercles colorés et proportionnels** à une donnée dans une colonne, elle se prête bien à la visualisation d'un **grand nombre de points**. Une troisième est, quant à elle, adaptée à l'affichage de **géométrie complexes**, par exemple, un réseau routier ou un plan local d'urbanisme.



D'autres visualisations sont adaptées aux données territorialisées ; ce sont les données possédant au moins un code territoire : code commune, département ou même code parcelle. On peut, par exemple, projeter les données sur le **plan cadastral** français. Il est également possible d'afficher une **carte choroplèthe** des territoires allant de l'IRIS à la région, adaptée à la présentation des données, comme les résultats d'élections ou d'autres indicateurs. Une application permet de **fusionner les géométries des territoires** pour avoir des zones de couverture, par exemple, le périmètre d'action des gendarmeries.

Visualisations en diagrammes

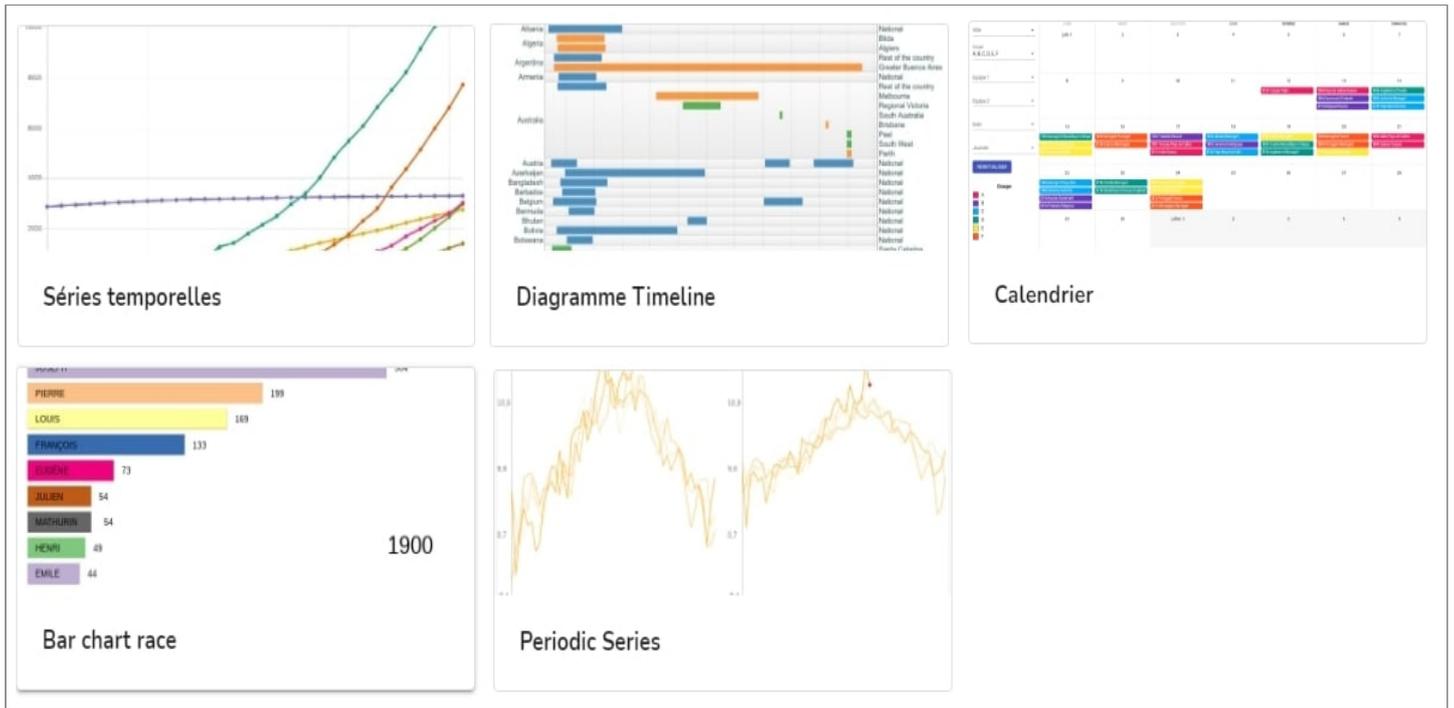
Dès que les données possèdent des colonnes avec des types numériques, il est possible de paramétrer pour elles des visualisations en diagrammes. Une visualisation permet de faire différents types de graphiques communs : **histogrammes**, **lignes**, **aires**, **camemberts**, etc. Une autre est adaptée à la **visualisation de flux**. Une troisième permet de comparer les lignes de données sur **différents critères**. Une autre encore produit des **diagrammes en gaufre**, par exemple, pour comparer la répartition des éléments suivant différentes catégories.



Des visualisations permettent de faire des **graphes**, en utilisant des nœuds et des liens, ou bien en se basant sur deux critères dans deux colonnes différentes d'un jeu de données. Pour les **données hiérarchiques**, deux visualisations sont disponibles et sont bien adaptées pour afficher des données de budget ou d'allocation d'aides.

Visualisations temporelles

Les données ayant des concepts liés aux dates peuvent être mises en valeur avec des visualisations temporelles. Certaines sont interactives, l'une d'elles permet de comparer des **courbes au cours du temps** (décès liés au covid, par exemple), d'autres de visualiser les données qui ont des dates de début et dates de fin.



D'autres visualisations temporelles sont animées : elles possèdent un bouton lecture et l'utilisateur peut les regarder de manière passive. Ces visualisations sont intéressantes pour les réseaux sociaux, car elles permettent d'attirer l'attention des utilisateurs qui n'ont pas à cliquer sur le lien pour voir le résultat. Une visualisation permet de voir des courses de barres horizontales et est adaptée quand il y a un **nombre important d'éléments qui ont la valeur maximale**, par exemple pour voir le prénom le plus donné au cours du temps. Une autre permet d'analyser les **données cycliques** pour voir comment elles se comportent d'une période à l'autre (par exemple, des températures ou du trafic de véhicules).

Visualisations textuelles

Ce type de visualisation présente les données sous forme de texte, y compris les données numériques. Il est possible de mettre en avant des **chiffres clés** sous la forme de tuiles. Une visualisation calcule des **agrégats sur un ou deux axes** et peut, par exemple, montrer des moyennes d'une valeur par départements. Une autre permet d'afficher quelques colonnes du jeu de données en proposant des **filtres ou une arborescence** pour accéder plus rapidement aux données (par exemple région > département > commune). Une troisième visualisation analyse la fréquence des différents termes d'un jeu de données ou de l'une de ses colonnes pour mettre en avant ceux qui sont les plus présents.

de collecter des retours directement stockés dans un jeu de données, de faire du *crowdsourcing* ou d'offrir une interface de mise à jour des données plus légère que l'intégralité du back-office pour une personne qui ne serait responsable de la mise à jour que d'un jeu de données.

Il y a aussi des visualisations qui sont spécifiquement **adaptées à des schémas de données**, comme certaines qui sont publiées suivant les schémas du **Socle Commun des Données Locales (SCDL)**.

Permissions et publication des visualisations

Les mécanismes de permission et de publication des visualisations sont les mêmes que pour les jeux de données. Par défaut, une visualisation est privée. Elle peut être rendue publique par la suite. Dans tous les cas, il est recommandé de renseigner une description. Celle-ci est visible sur la page de consultation de la visualisation, mais aussi sur la page du jeu de données associé.

Il est aussi possible de partager une visualisation privée avec des utilisateurs non authentifiés. Cela se fait en générant un lien de partage qui contient un code secret. Toute personne connaissant ce lien peut accéder à la visualisation, même si les données qu'elle utilise sont privées. Si le lien est compromis, il peut être supprimé puis régénéré avec un code secret différent. Cela permet d'embarquer des visualisations dans des sites privés, sans avoir à créer un ou plusieurs comptes utilisateur et transmettre des identifiants d'accès.

3.3 - Portails de données et pages de contenus

Paramétrage des portails

Data Fair permet de configurer plusieurs portails de données qui sont des lieux de publication des jeux de données, visualisations et pages de contenus que les utilisateurs vont pouvoir consulter. Il peut y avoir **des portails pour différents cas d'usage** : open data, partage de données en interne, partage avec des partenaires, données en cours de consolidation (portail de préproduction), etc.

Tout comme pour les visualisations de données, la configuration d'un portail de données se fait **de manière graphique** en deux temps : on travaille sur un brouillon que l'on publie ensuite en version courante, qui est la version présentée aux différents visiteurs de votre portail. Cela permet de mettre à jour le portail sans impacter les utilisateurs tant que l'on n'a pas validé le brouillon. **Aucune connaissance en HTML ou CSS** n'est requise et un portail s'administre à la manière d'un CMS comme Wordpress.

De **nombreux éléments sont configurables** : le logo, l'image d'accueil, la favicon, la couleur de la barre de navigation, la couleur du pied de page, des éléments de contenu (titre, description, visualisation à mettre en avant, visibilité publique ou privée) et divers éléments de communication (site web, e-mail de contact, et comptes sur les réseaux sociaux).

Il est possible de renseigner un compte [Google Analytics](#) ou [Matomo \(ex Piwik\)](#) pour le suivi d'activité et ainsi avoir des **statistiques** sur les pages les plus consultées et les données les plus téléchargées.

Édition des pages de contenus

Les pages de contenu sont de différents types : articles, pages thématiques autour de plusieurs jeux de données, pages d'actualités, data storytelling, licences, conditions d'utilisation, etc.

Il est ainsi possible de **mettre en avant certaines données** et de leur donner encore plus de contexte, par exemple, ou de faire des **tableaux de bord** intégrant différentes données.

The screenshot shows the Koumoul web editor interface. At the top, there's a navigation bar with 'portails / Portail de données Koumoul / pages / Foncier'. Below this, there are checkboxes for 'Publiée' and 'Publique'. The page model is set to 'Page thématique'. On the right, there's a 'Disposition dans le menu' dropdown set to 'Sous-menu' and a 'Nom du bloc' field set to 'Thématiques'. The main content area is divided into two columns. The left column contains a title 'Foncier', a subtitle 'Documents relatifs aux données foncières', a URL for the banner image, and a rich text editor with a description. The right column features a large image of a village, a title 'Documents relatifs aux données foncières', a description, and a map titled 'Base de permis de construire et autres autorisations d'urbanisme (SITADEL)'. The map shows France and surrounding countries with a legend for 'nouvelle construction' and 'travaux sur construction existante'.

La création d'une page se fait en trois étapes : on choisit d'abord le **modèle de page**, puis on renseigne les différents éléments au moyen d'un formulaire adapté au modèle de page choisi avec une **prévisualisation du résultat**. On peut enfin publier la page, ce qui permet de **préparer des pages en avance** et de les publier plus tard. En plus de la saisie de texte libre, il est possible d'intégrer différents type d'éléments : tableau d'un jeu de données, visualisation, liste de jeux de données, intégration de contenu externe, etc.

Pour accéder aux pages de contenu créées, il est possible de renseigner des liens dans la barre de navigation. Les liens peuvent apparaître directement dans la barre, ou dans un menu ajouté à celle-ci. Il est possible de créer des pages **publiques** ou des pages **privées**.

3.4 - Gestion des utilisateurs

Par défaut, il existe quatre rôles différents : utilisateur, contributeur, administrateur et super administrateur.

Les **super administrateurs** de la plateforme peuvent gérer l'ensemble des organisations, membres et tout le contenu de la plateforme. Ils ont la possibilité de configurer les visualisations disponibles, publier les portails sur des noms de domaine particuliers, paramétrer les traitements périodiques, ou définir les jeux de données maîtres. Il est prévu de transférer prochainement la gestion des deux derniers éléments aux administrateurs des organisations.

Les trois autres rôles sont définis par organisation : il est, par exemple, possible d'être administrateur dans une organisation et simple utilisateur dans une autre.

Rôles et permissions associées

Les administrateurs d'une organisation peuvent gérer les membres :

- **Inviter** de nouveaux membres par e-mail ;
- **Changer** les rôles des membres ;
- **Exclure** un membre.

Permissions par défaut des différents rôles d'une organisation :

Actions	Utilisateur	Contributeur	Administrateur
Ajout d'un jeu de données		x	x
Lecture d'un jeu de données	x	x	x
Modification d'un jeu de données		x	x
Administration d'un jeu de données			x
Ajout d'une visualisation		x	x
Lecture d'une visualisation	x	x	x
Édition d'une visualisation		x	x
Administration d'une visualisation			x
Accès et modification des paramètres			x
Création et modification du portail			x

Départements

En plus de leur rôle, les utilisateurs peuvent être assignés à un département de l'organisation. Cela permet une **forme de cloisonnement** et d'avoir des groupes d'utilisateurs qui gèrent chacun leurs données de leur côté. Les utilisateurs qui ne sont pas restreints à un département peuvent voir (ou modifier s'ils ont un rôle de contributeur ou administrateur) toutes les ressources de l'organisation.

Un contributeur d'un département ne peut mettre à jour que les jeux de données de ce département et, quand il crée un jeu de données, celui-ci est rattaché à son département. De la même manière, un administrateur rattaché à un département ne peut publier des jeux de données que sur un portail rattaché à son département. Par contre, un administrateur général de l'organisation peut publier ce même jeu de données sur un portail plus global à l'organisation.

3.5 - Traitements périodiques

Les traitements périodiques vont **chercher** des données à certains endroits, les **transformer** et les **importer** sur la plateforme. Ils sont réalisés sous la forme de plugins open source, et la liste des traitements disponibles est en évolution constante.

Par exemple, le plugin *download-file*, qui est assez générique, permet de couvrir plusieurs cas d'usage : ce traitement va chercher des fichiers à un endroit pour les publier sur la plateforme. Il est capable d'accéder à des fichiers via les protocoles **FTP, SFTP ou HTTP(S)**. Il fonctionne en général à la suite de traitements sur les données réalisés par des ETL qui déposent leurs résultats sous la forme de fichiers accessibles à distance.

Chaque plugin a son propre paramétrage (code d'accès, jeu de données à mettre à jour...), mais tous les traitements ont les mêmes options de planification. Les traitements peuvent être déclenchés toutes les heures, jours, semaines, mois ou être paramétrés pour être déclenchés manuellement.

Il est possible de générer une clé d'API spécifique au traitement pour créer un **webhook permettant de le déclencher** : un traitement dans un ETL peut, par exemple, créer un fichier sur un espace de partage puis appeler l'URL du webhook pour que le traitement d'import se déclenche.

The image shows three side-by-side configuration cards for data processing jobs. Each card has a title, associated data source, owner, type of treatment, frequency, last execution status, active status, and a webhook key. The first card is 'Données agrégées par pays' with a 'finalized' status. The second is 'Covid ECDC Our world in data...' with a 'finalized' status. The third is 'Covid ECDC' with a 'finalized' status. Each card also has a 'Clé webhook' field with a question mark icon. At the bottom of each card is a toolbar with icons for play, calendar, list, edit, and delete.

Un contributeur peut accéder à l'**état de réussite** des différentes exécutions d'un traitement, ainsi qu'aux logs détaillés de ses exécutions. Il peut s'abonner aux notifications d'un traitement pour être informé quand un traitement est en échec ou que des alertes sont présentes dans les logs.

3.6 - Connecteurs de catalogues

Les connecteurs permettent d'interagir avec d'autres plateformes ou services de données, en lecture comme en écriture.

En **écriture**, l'idée est de pouvoir pousser des métadonnées dans d'autres **catalogues**. Un exemple de catalogue est le catalogue de données ouvertes national data.gouv.fr : les jeux de données publiés à l'aide de Data Fair peuvent être synchronisés automatiquement et toute modification dans les métadonnées est propagée vers le catalogue distant.

Le fait de pousser les métadonnées vers un catalogue, plutôt que de se faire moissonner par lui, offre plusieurs avantages, dont le fait de propager immédiatement les modifications. De plus, s'il y a des modifications de l'API de Data Fair, le connecteur continuera de fonctionner, alors qu'un moissonneur pourrait devenir inopérant.

Les connecteurs peuvent éventuellement pousser les données vers ces catalogues, mais il est préférable d'éviter cela, à cause des problèmes de duplication et synchronisation de données. Comme mentionné précédemment, les données sont indexées de manière très performante avec Data Fair et il est préférable de consommer les données directement à partir des API qu'il offre.



En ce qui concerne la **lecture**, l'approche est, par contre, différente et les connecteurs se comportent plutôt comme des **moissonneurs de métadonnées et de données**. On peut ainsi, pour chaque connecteur, paramétrer les fréquences de collecte et les types de sources que l'on souhaite moissonner.

Le fait d'intégrer les données à la plateforme permet de les indexer et de pouvoir centraliser les contrôles d'accès, prérequis indispensable si l'on souhaite pouvoir **fusionner** les données ou consulter différentes sources sur une visualisation.

Il est possible de rajouter de nouveaux connecteurs de catalogues en suivant les instructions dans [cette section](#).

Par rapport aux traitements périodiques, les connecteurs de catalogue ont pour principale différence de pouvoir traiter **plusieurs sources de données référencées via une API** qui les liste. La fréquence de synchronisation est, en général, plus faible que pour les traitements périodiques.

3.7 - Métriques d'utilisation

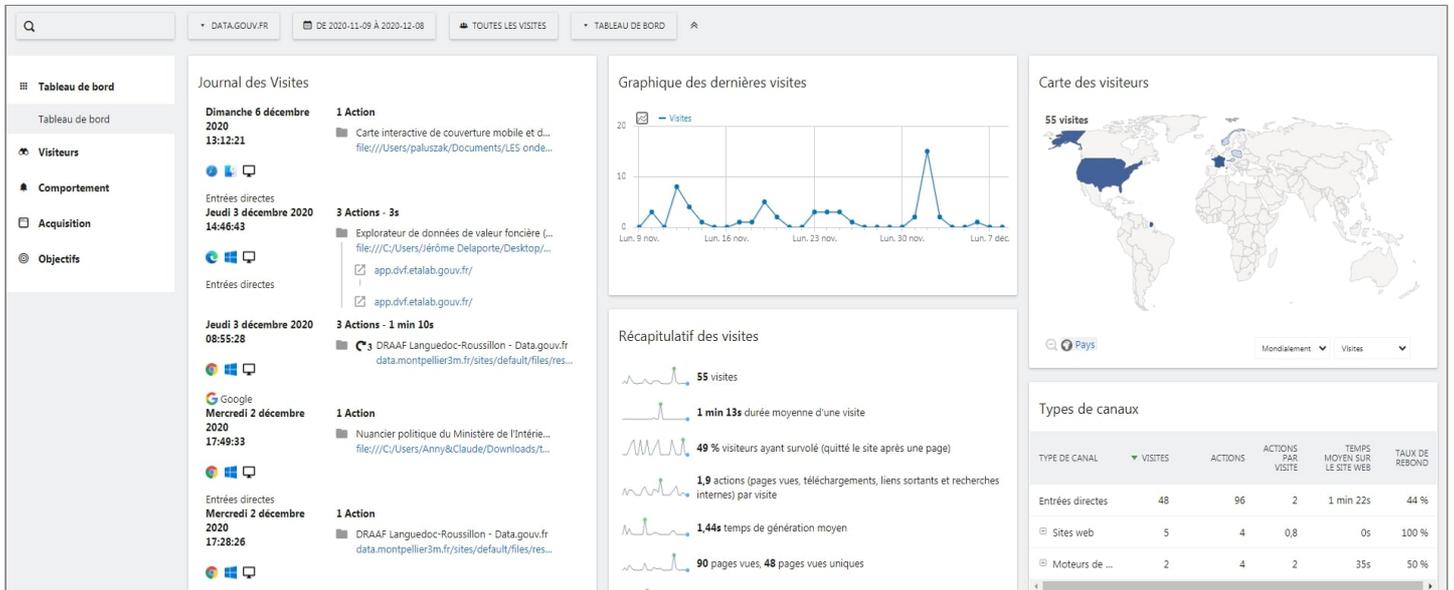
Il y a deux modules pour suivre l'utilisation de la plateforme. Le premier est les **analytics** et correspond au suivi des parcours des utilisateurs sur le portail de données. Cela permet de voir quelles pages sont consultées, d'où viennent les utilisateurs, le temps qu'il passent sur les pages, etc. Le second correspond à des **mesures de consommation des API** et permet de voir comment la plateforme est utilisée par les autres systèmes d'information ou les sites externes.

Analytics

Il est possible d'utiliser **Matomo Analytics** (anciennement Piwik) ou **Google Analytics** comme système de suivi. Cela se fait simplement par configuration du portail de données en renseignant quelques champs dans un formulaire.

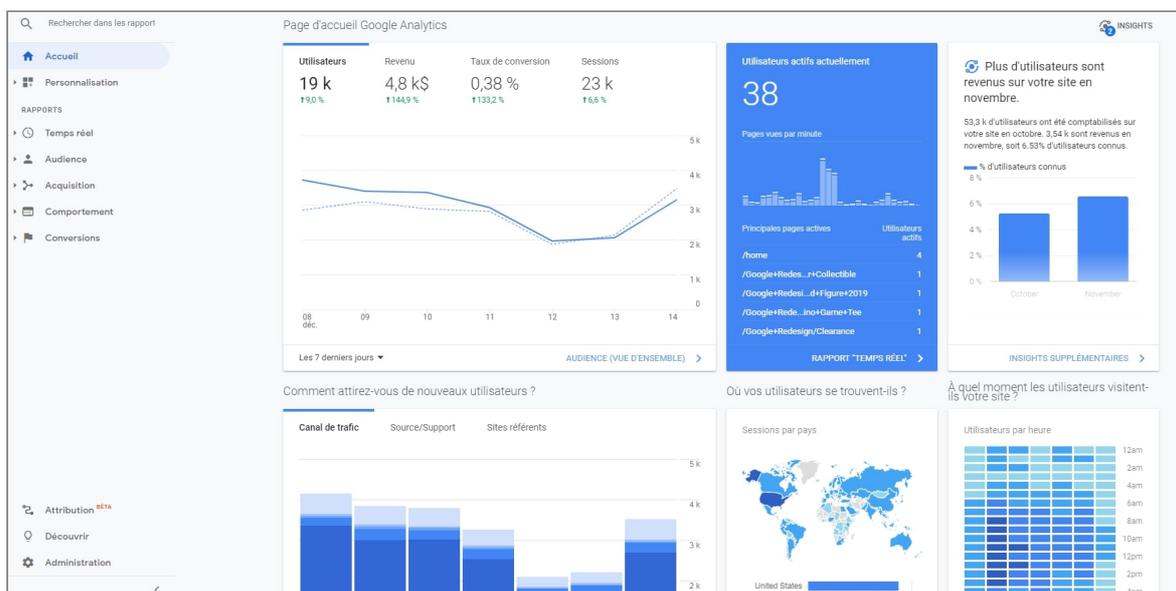
Matomo Analytics

La configuration se fait avec l'**URL du tracker** et l'**identifiant de votre site**. Les statistiques avec **Matomo Analytics** sont disponibles sous différentes formes : tableaux, graphiques et cartes. En sélectionnant les différentes représentations des statistiques, il est possible de personnaliser ses tableaux de bord. On peut également **anonymiser les données** et enregistrer les parcours utilisateurs tout en étant en conformité avec les recommandations de la **CNIL**.



Google Analytics

La configuration se fait grâce au **numéro d'identifiant**. Les statistiques sous **Google Analytics** sont aussi disponibles sous différentes formes : tableaux, graphiques et cartes. Il est aussi possible de personnaliser ses tableaux de bord.



Utilisation des API

Dans la mesure ou Data Fair et les différents services associés utilisent beaucoup les mécanismes de cache pour améliorer les temps d'accès aux ressources, les statistiques précises d'utilisation des différents points d'accès de la plateforme ne peuvent être collectées que par un service spécifique associé au reverse proxy de la plateforme.

Dans un souci de **respect du RGPD**, les données collectées sont anonymisées et agrégées à la journée. On peut accéder à des statistiques pour chaque jeu de données : **nombre d'appels d'API** et **nombre de téléchargements**. Les métriques sont agrégées par groupes d'utilisateurs (organisation propriétaire, utilisateurs authentifiés externes, anonymes...) ou par domaine d'appel. Des chiffres clés sont présentés sur la période demandée, avec une comparaison sur la période précédente, ce qui permet de voir si l'utilisation de certaines données augmente ou diminue.

