

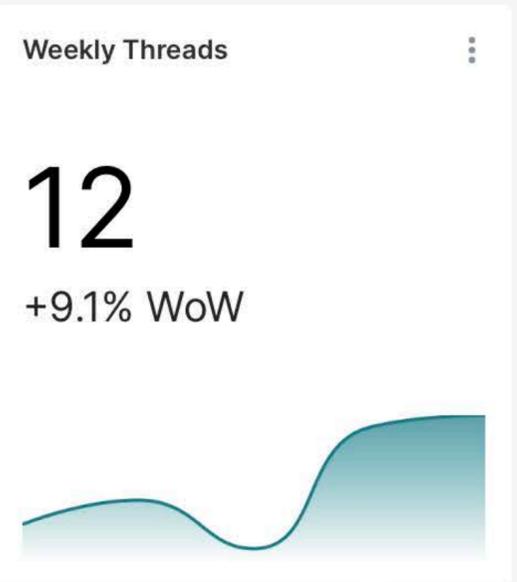
Exploration et visualisation de données d'observation

Avec Apache Superset



Plateforme d'exploration et de visualisation de données open- source.



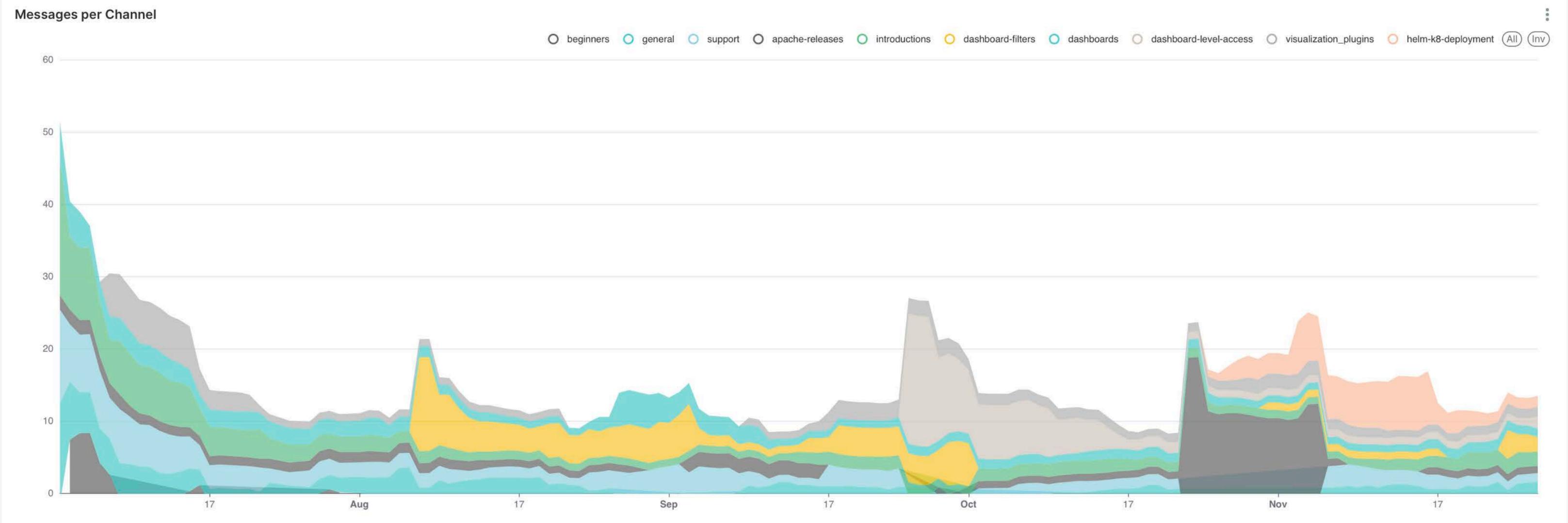


Top Timezones for Members

tz	COUNT(*)
America/Los_Angeles	416
Asia/Kolkata	354
America/New_York	315
Europe/Amsterdam	240
Asia/Chongqing	168
Europe/London	118
America/Chicago	102
Europe/Brussels	79
Europe/Moscow	49
America/Sao_Paulo	49

Number of Members

2.51k
Slack Members





Vue d'ensemble



Vue d'ensemble

Un peu d'histoire

- Initié en 2015 par Maxime Beauchemin, ingénieur data chez Airbnb
- Open-source en 2016
- Adoption croissante en 2017 et 2018
- Incubé par la Fondation Apache en 2019
- Sortie de la version 1.0 en 2021
- Dernière version 4.0.1 en mai 2024



Vue d'ensemble

Fonctionnalités clés

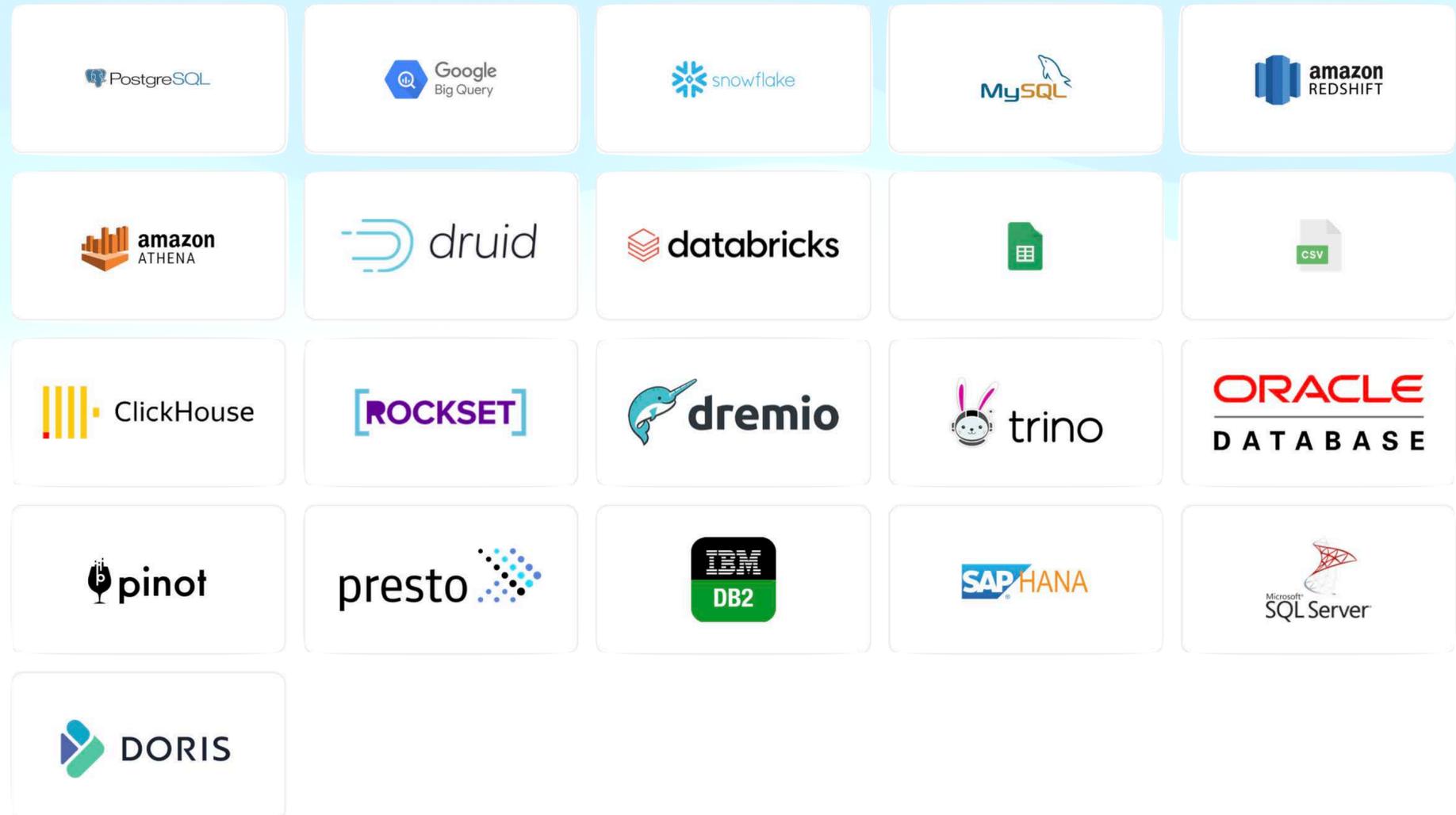
- Création de tableaux de bord personnalisés
- Couche sémantique simple
- Visualisation Apache ECharts
- Editeur SQL intégré
- Customisation CSS
- API intégrée



Vue d'ensemble

Quelques chiffres

- 59K 🌟 sur GitHub
- 10K+ forks
- 56 releases depuis 2019
- 1000+ contributeurs
- 40+ dataviz pré-installées
- 40+ base de données



Vue d'ensemble

Ecosystème

- Metabase (<https://www.metabase.com/>)
- Redash (<https://redash.io/>)
- Grafana (<https://grafana.com/>)
- Plotly Dash (<https://dash.plotly.com/>)
- Kibana (<https://www.elastic.co/fr/kibana>)



Grafana



kibana



Metabase



Vue d'ensemble

Une approche centrée Dataset



- Structure de données tabulaires que la majorité des utilisateurs connaissent déjà.
- Adoption progressive de la sémantique au fur et à mesure que les ensembles de données sont de plus en plus utilisés
- Gestion des données par du code possible avec l'API
- Intégration possible avec des couches sémantiques autonomes
- Combinaison des approches de visualisation centrées requêtes et centrées sémantique.

Vue d'ensemble

Pros & cons



👍 Visualisation riche (~40 graphiques)

👍 Large support bases de données

👍 On-premises

👍 Performant & scalable

👍 Communauté active

👎 UI un peu technique

👎 Déploiement parfois complexe

👎 Personnalisation par défaut limitée

Vue d'ensemble

Workflow



1. Connexion d'une ou plusieurs sources de données
2. Création des datasets
3. Création des graphiques (Map, Line Chart, Bar Chart...)
4. Création des tableaux de bords

+ Paramétrage général (gestion des utilisateurs, droits d'accès)



Démonstration



Démonstration

Sources de données d'observation



Observatoire du Larzac

Étude des processus qui contrôlent les variations spatio-temporelles du stock d'eau dans les milieux karstiques.

oreme.org/observation/gek/



ScaraB'Obs

Améliorer l'état des connaissances sur les *Scarabaeoidea Laparosticti* coprophages

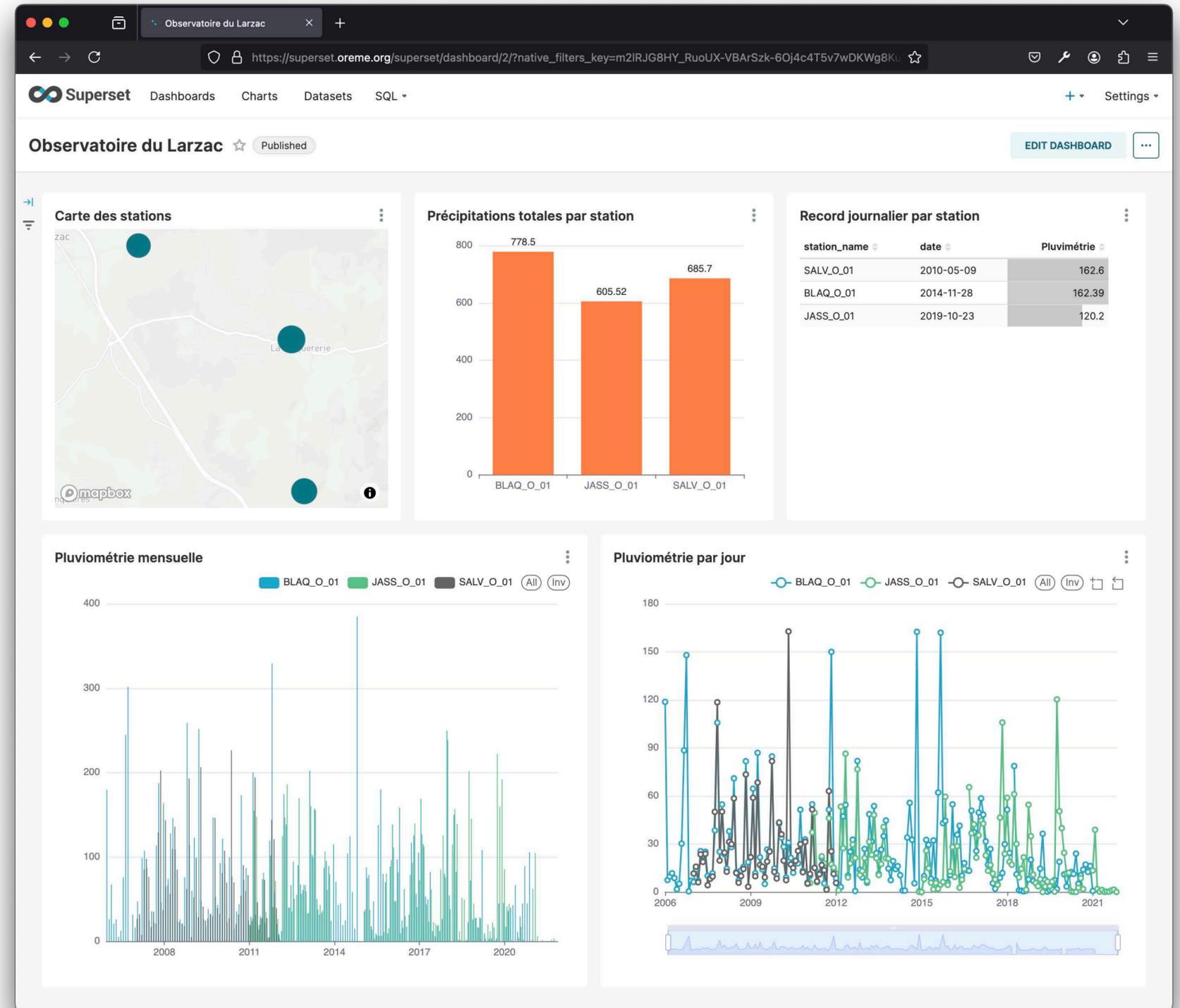
oreme.org/observation/ocoa/scarabees/

Démonstration

Observatoire du Larzac

Visualisations :

- Stations (cartographie)
- Précipitations totales (histogramme)
- Record journalier (table)
- Pluviométrie mensuelle (diagramme linéaire)
- Pluviométrie journalière (diagramme linéaire)

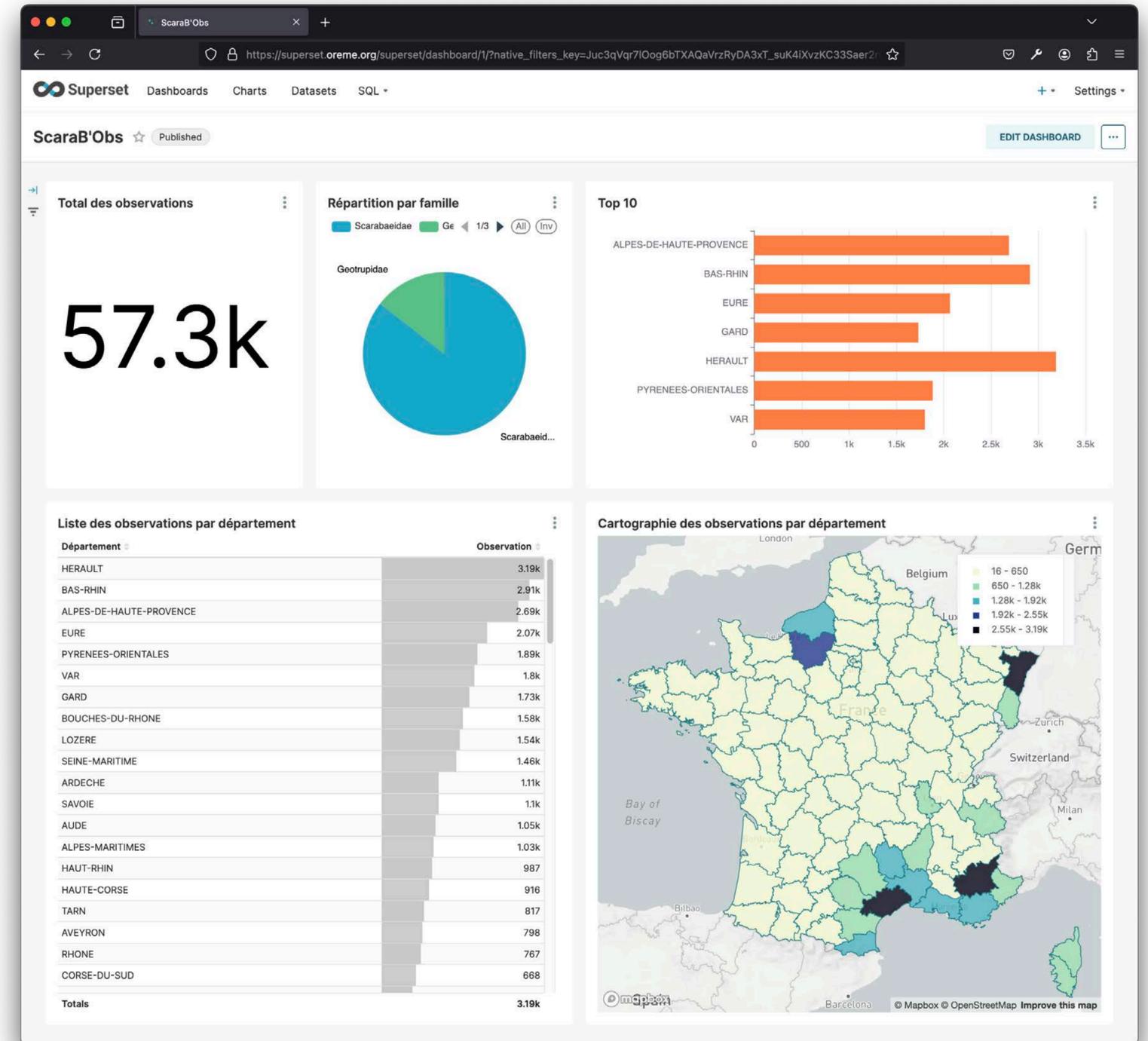


Démonstration

ScaraB'Obs

Visualisations :

- Total des observations (KPI)
- Répartition par famille (Pie chart)
- Top 10 (Histogramme)
- Liste des observations (table)
- Observations par département (cartographie)



Démonstration

Apache Superset OREME



<https://superset.oreme.org>

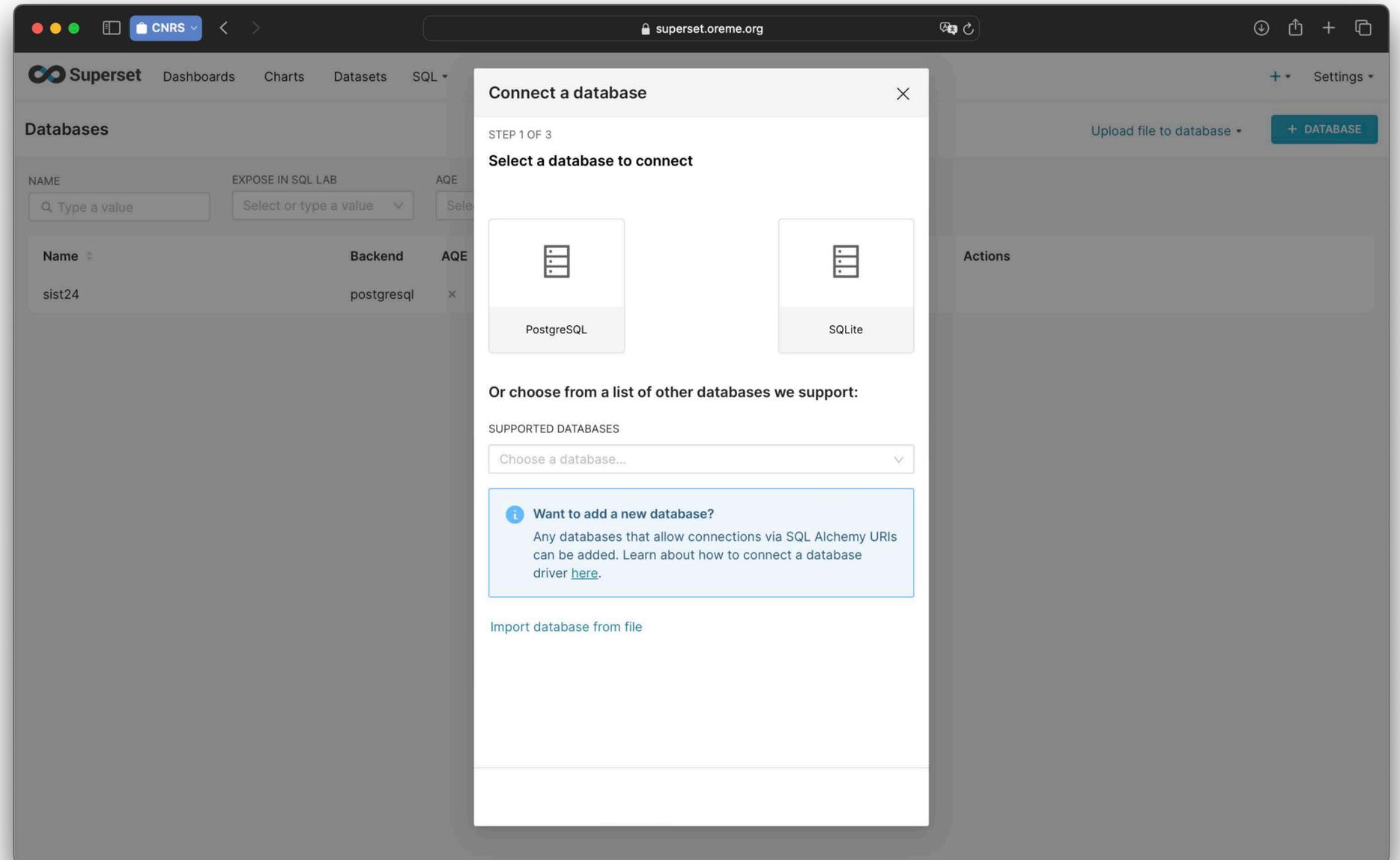
Nom d'utilisateur : superset1, superset2, superset3...

Mot de passe : Sist24#SupersetDemo

Démonstration

1/4 Base de données

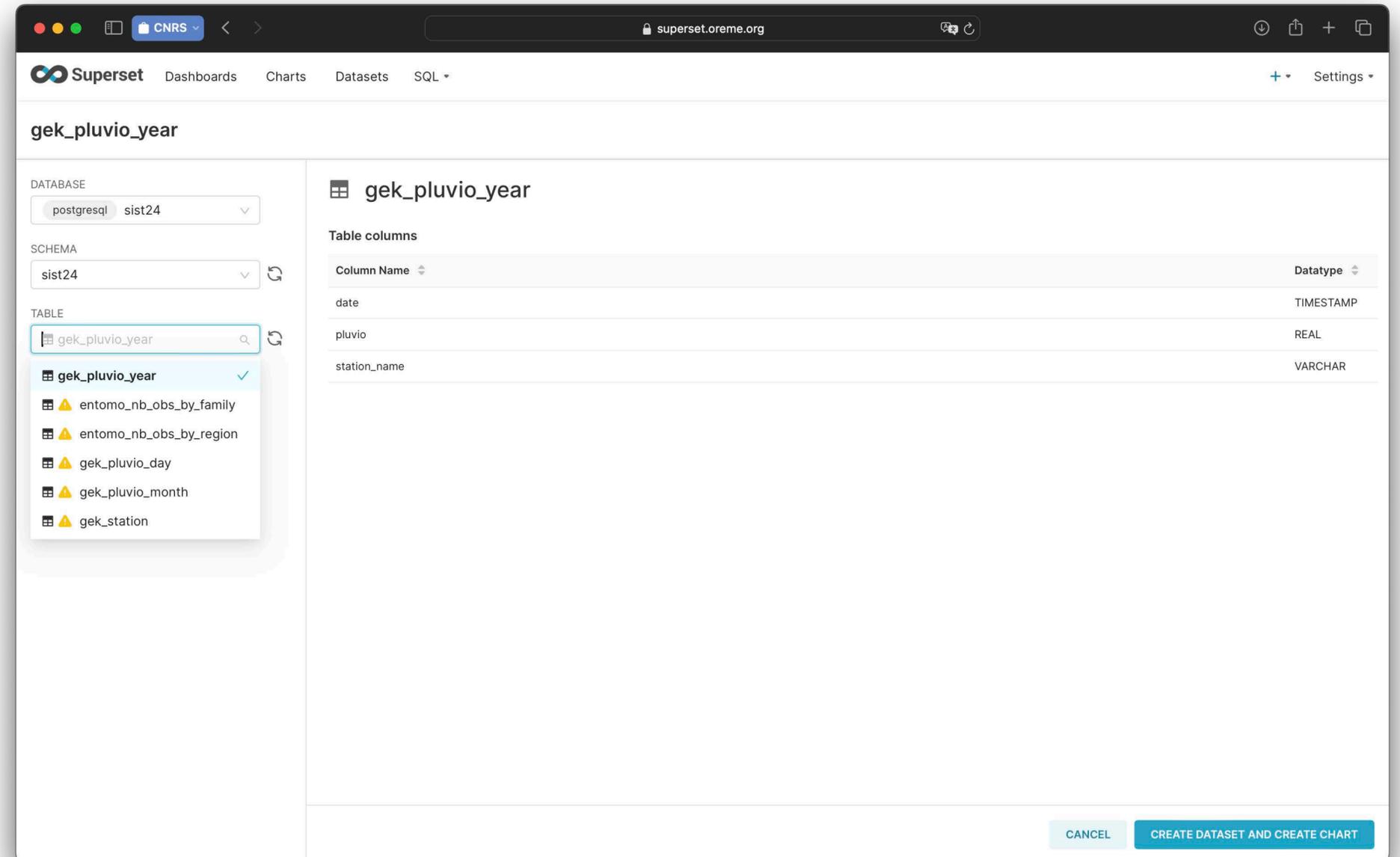
Connexion vers les différentes sources de données disponibles.



Démonstration

2/4 Dataset

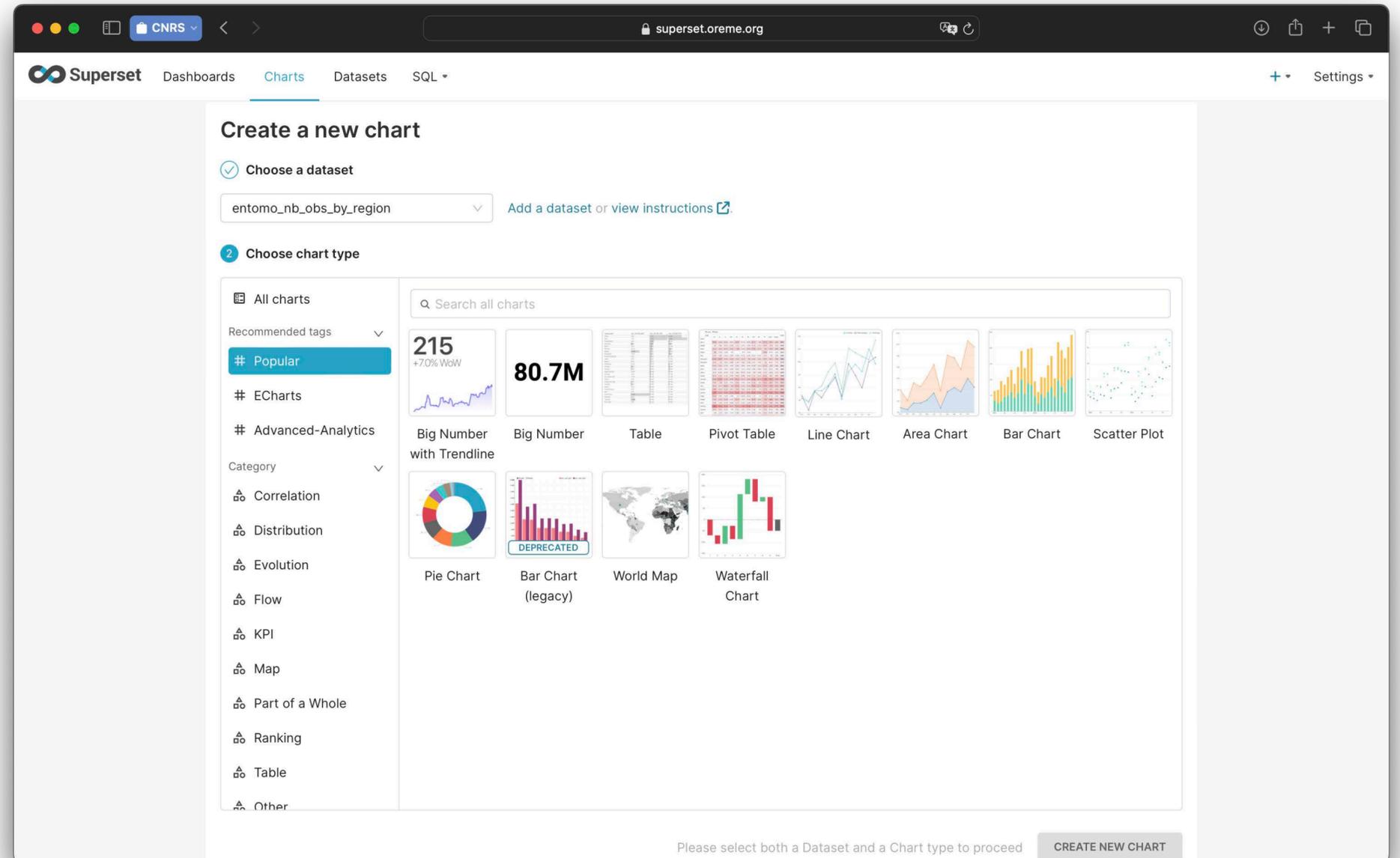
Connexion vers les sources de données



Démonstration

3/4 Charts

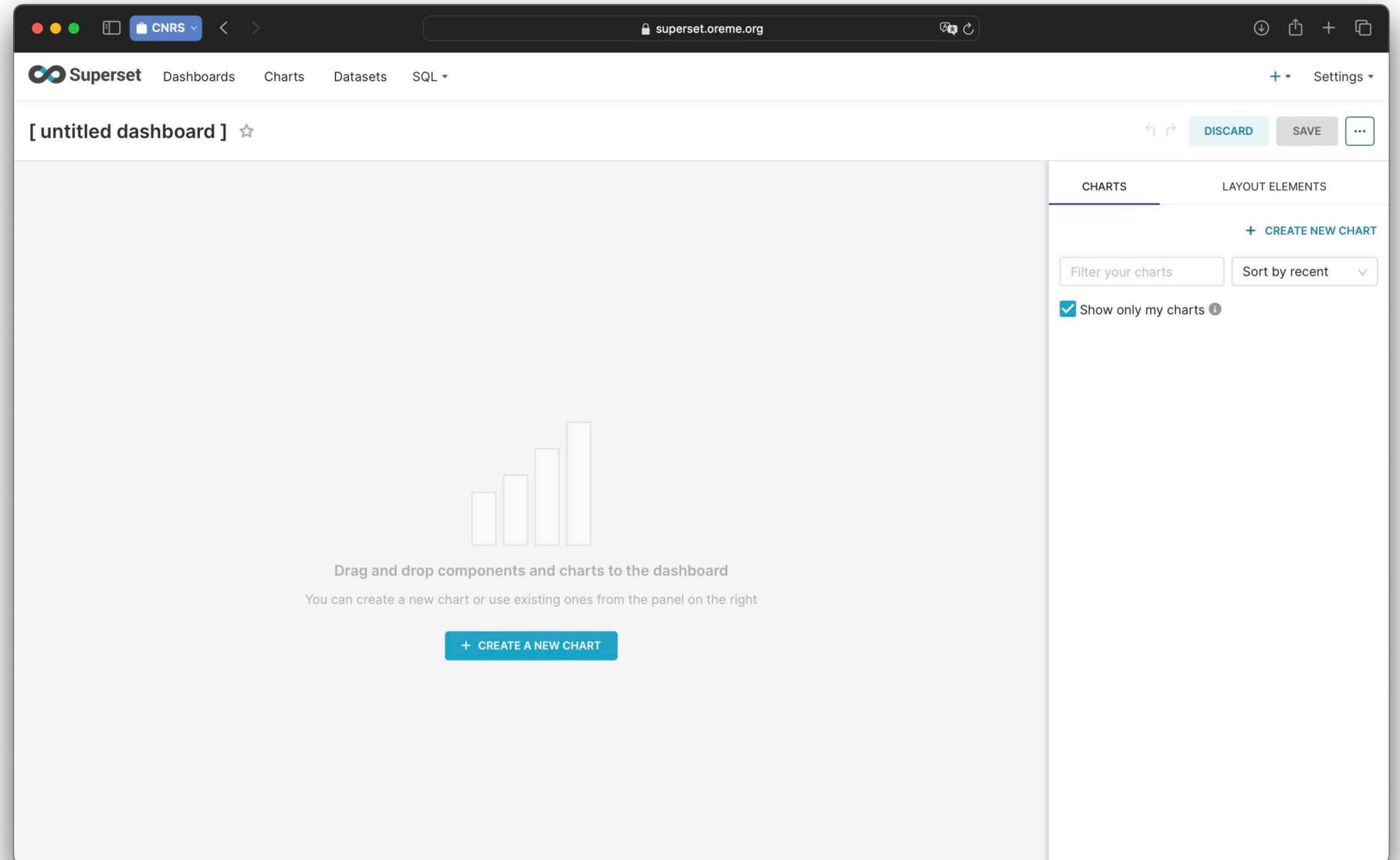
Sélection et configuration des modèles de visualisation.



Démonstration

4/4 - Dashboards

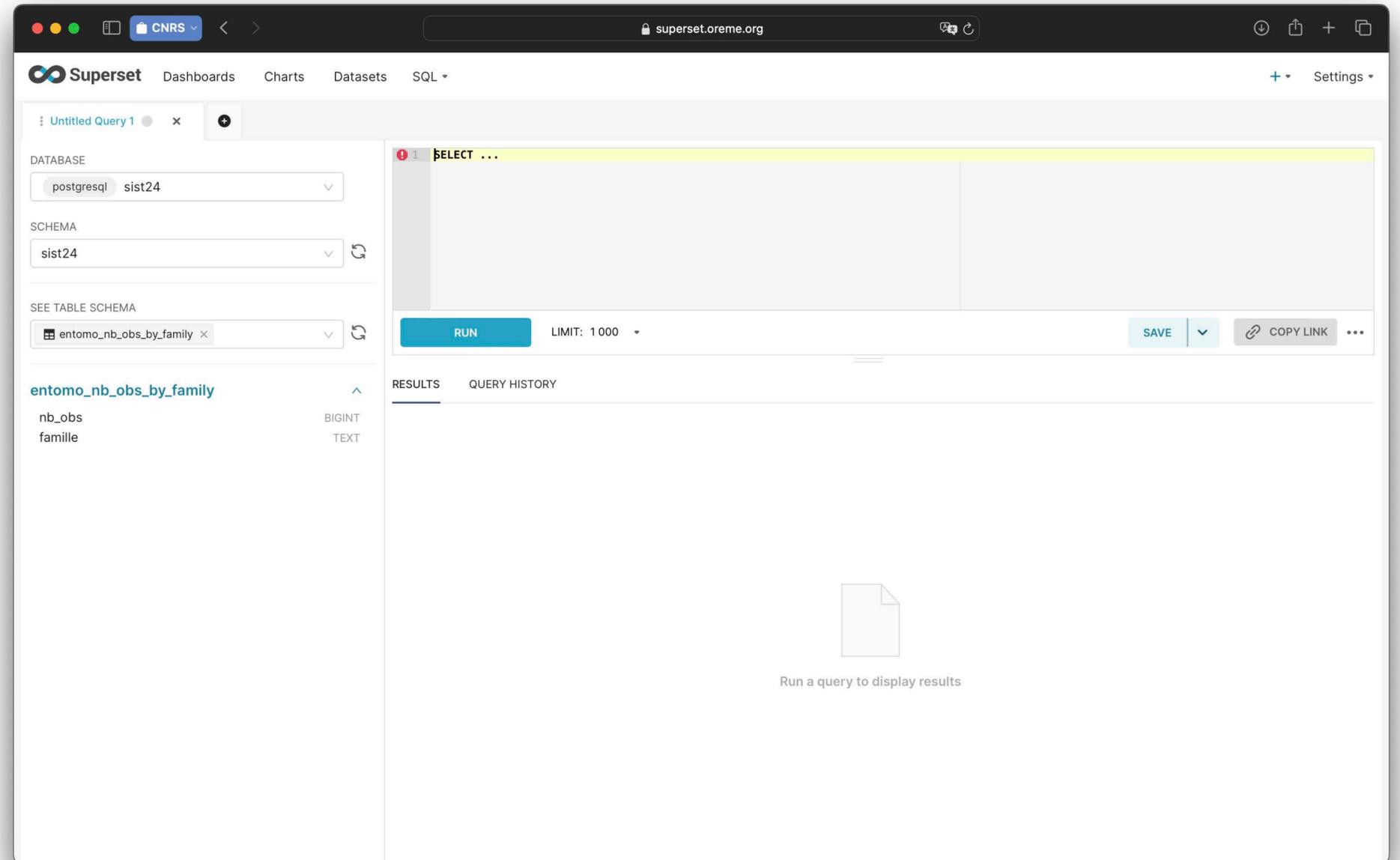
Création et édition des tableaux de bord.



Démonstration

Bonus - SQL Editor

Création de dataset spécifiques via des requêtes SQL personnalisées depuis l'éditeur intégré.



SIST 24 - mercredi 5 juin

Merci