

**neogeo**

## Rencontres régionales **PIGMA**

**Le chatbot PIGMA :  
première brique d'une  
architecture intelligente  
et interopérable des  
géodata**

Laurent MER, Neogeo  
Sébastien DA ROCHA, Neogeo

# neogeo



## neogeo

UNE **ENTREPRISE INDUSTRIELLE**  
AU SERVICE DE L'INNOVATION  
POUR FAIRE DES **DONNÉES**  
**GÉOGRAPHIQUES** DES LEVIERS  
DE DÉCISION PUBLIQUE ET DE  
**TRANSITION TERRITORIALE**

Nos piliers :

- › Interopérabilité et souveraineté
- › Valorisation territoriale des données
- › Innovation ouverte et durable

## › Qui sommes-nous ?

Spécialiste des plateformes open data et close data, avec plus de 50 plateformes déployées et 10 000 utilisateurs accompagnés.

Neogeo c'est :

- › Un éditeur de solutions open source avec notamment OneGeo Suite, solution modulaire et interopérable pour la collecte, l'agrégation, la publication et la valorisation de données.
- › Des experts en géomatique : gestion de données spatiales, formats standards et métiers, traitement cartographique avancé.
- › Des spécialistes en développements sur mesure (connecteurs, modules métiers, interface utilisateur...).
- › Une organisation structurée : équipe projet dédiée, coordination continue avec les MOA/MOE.
- › Une entreprise engagée dans l'innovation et l'open source, avec des contributions actives et un service de R&D intégré.
- › Un acteur spécialisé en data science : IA, machine learning, Chatbot, big data appliqués aux données spatiales.



# Introduction



1 Introduction

---

2 Chatbot PIGMA

---

3 Architecture data intelligente

---

4 Bonne pratiques et nouveaux services

---

5 Echanges

## Objectifs de l'atelier :

- Présenter un retour d'expérience concret sur le développement du chatbot PIGMA.
- Montrer comment ce service s'inscrit dans une vision d'architecture data moderne.
- Mettre en lumière les conditions techniques et sémantiques d'une exploitation efficace des géodonnées.



# Retour d'expérience : le chatbot PIGMA

02

1 Introduction

---

2 Chatbot PIGMA

---

3 Architecture data intelligente

---

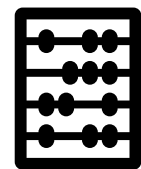
4 Bonne pratiques et nouveaux services

---

5 Echanges

# La Genèse

- Des **constats partagés** par PIGMA et la communauté OneGeo



Recherche **exacte** de termes  
insuffisante  
→ ex : Établissements  
scolaires ≠ écoles

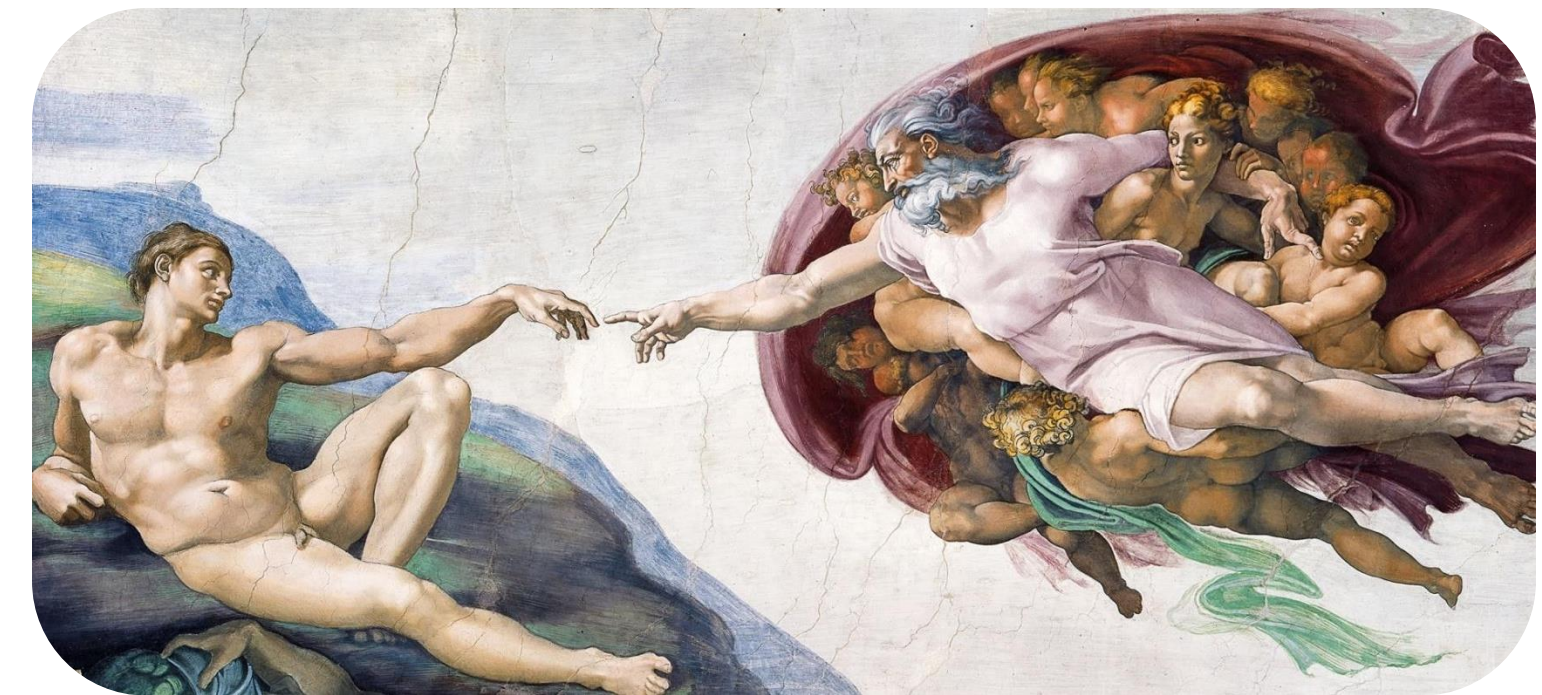


Trop de résultats ou...  
aucun



Pertinence fluctuante  
Pas de possibilité d'affiner  
sémantiquement sa recherche

- Et des **problématiques à venir**
  - De plus en plus de jeux de données
  - Métadonnées pas toujours bien renseignées
  - Plus d'interopérabilités entre les plateformes



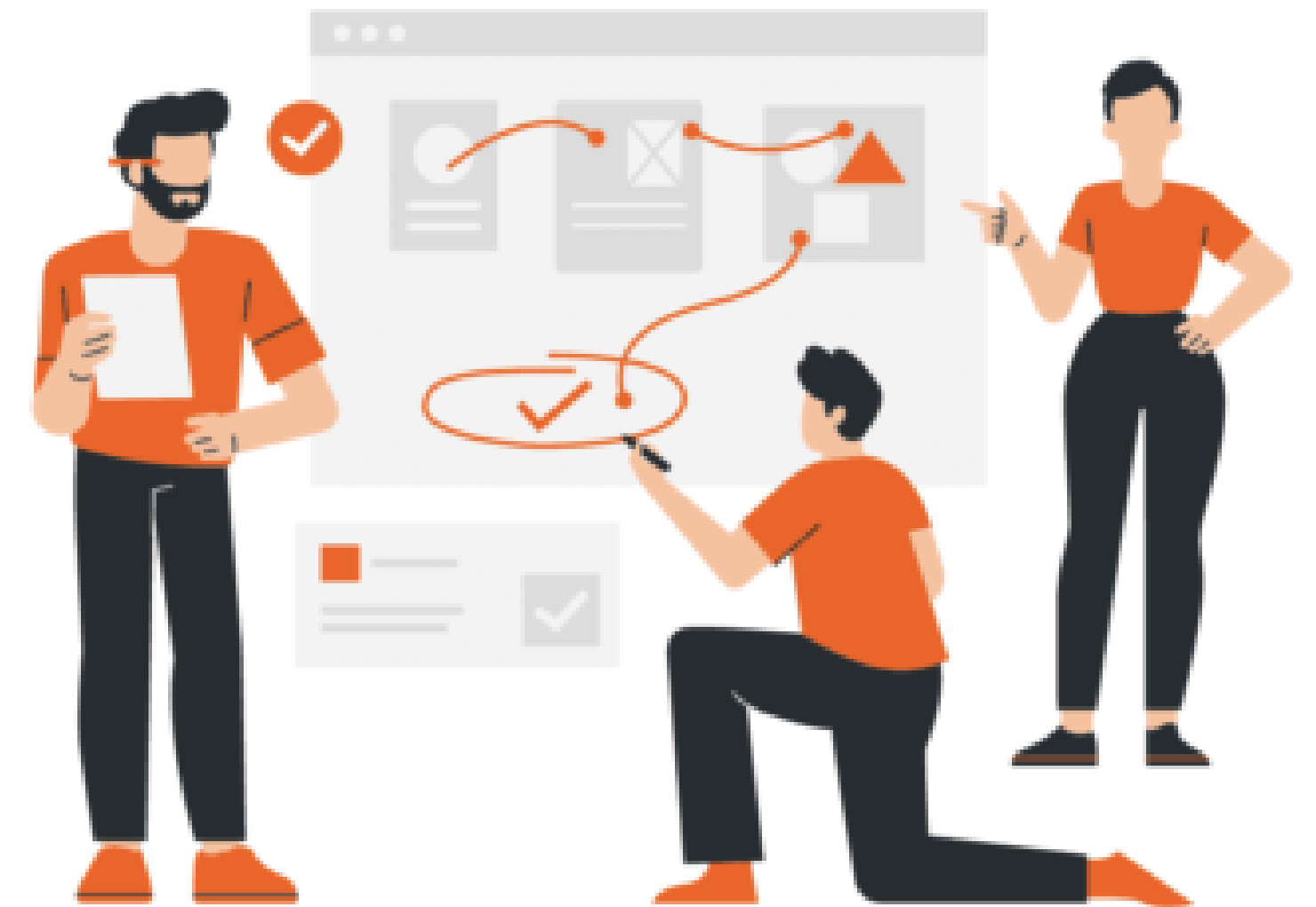
## Dans un **contexte existant**

- Besoin croissant de rendre la donnée plus accessible, exploitable, interopérable
- En lien avec le collectif CICCLO et l'ouverture des plateformes avec la connexion ProConnect et l'API feature
- Intégré dans la Road Co-développé avec PIGMA

## Les Objectifs

Disposer d'une recherche améliorée des jeux de données, au travers d'un assistant de recherche capable :

- d'interagir naturellement avec l'utilisateur
- d'aider à trouver des nouveaux termes de recherche et d'affiner sa connaissance du domaine
- de lancer une recherche sémantique (par sens plutôt que textuelle)
- de tenir compte des recherches précédentes



# Retour d'expérience : le chatbot PIGMA

- **Présentation :**

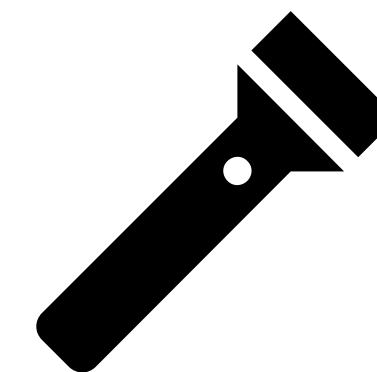
- Assistant conversationnel en langage naturel.
- Accès guidé à la donnée territoriale, via l'interface web PIGMA.
- Pensé pour accompagner **tous types d'utilisateurs** : élus, techniciens, agents publics.

- **Fonctionnalités clés :**

- Questions libres ou guidées.
- Moteur de recherche basé sur les métadonnées PIGMA.
- Orientation vers les fiches jeux de données ou services de consultation.

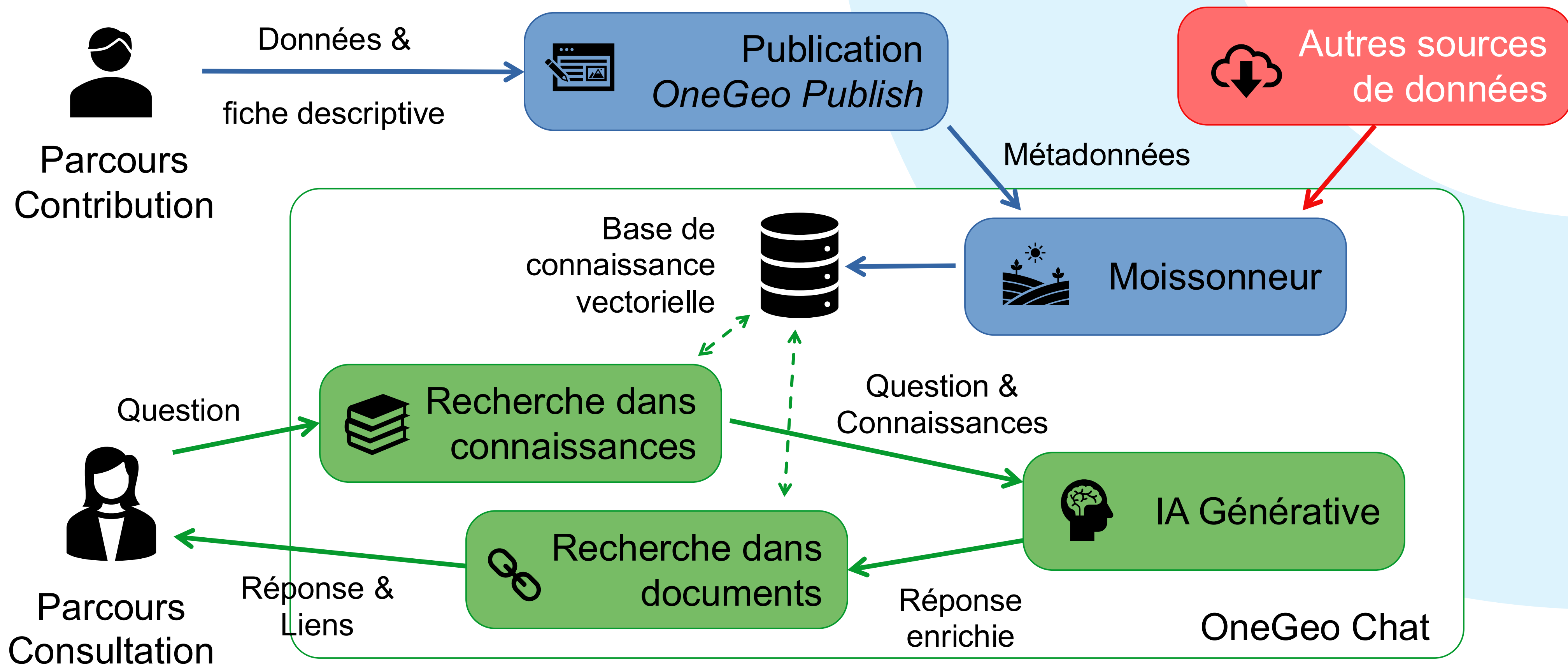
The screenshot shows the neogeo chatbot interface. The chatbot has responded to the user's question: "Bonjour, quelles données sont disponibles sur la fibre ?". The response includes a list of data types and their scores:

Donnée	Description	Score
Photographies aériennes du Bassin d'Arcachon en 1952	Photos aériennes noir et blanc (287 photos) prises à 10 000 pieds d'altitude	Score: 83%
Tronçon de fibre du réseau haut débit "CUB-DSI" sur le territoire de Bordeaux Métropole	Ce jeu de données représente sous forme de tronçons le réseau haut débit interne à Bordeaux Métropole qui concerne uniquement certains immeubles métropolitains.	Score: 91%
Nomad de fibre du réseau haut débit "CUB-DSI" sur le territoire de Bordeaux Métropole	Ce jeu de données ponctuelles renseigne sur les lieux de convergence (noeuds) des lignes qui composent le réseau haut débit interne à Bordeaux Métropole. Ce réseau concerne uniquement certains immeubles métropolitains.	Score: 91%
ADACIL: Plan	Plan cadastral informatisé à la norme du standard EIGAD0 sous du plan source IGN19 au même jour de	Score: 82%



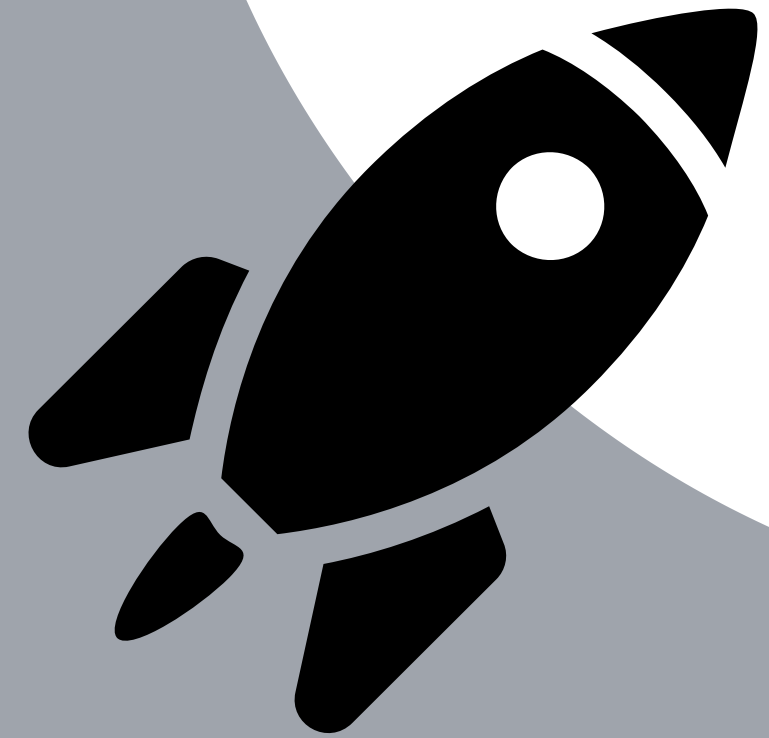
Démonstration

## Sous le capot...



## Dans la roadmap

- Etendre à la recherche sur les données et valoriser le contenu éditorial des plateformes (articles, études...),
- Utilisation du chatbot pour piloter des traitements sur la donnée et des analyses au travers la génération automatique de cartes ou de dataviz,
- Etude d'une solution mutualisée (SaaS) pour une vision multiplateforme et une mutualisation des serveurs de calcul



# Vision : une architecture data intelligente

03

1 Introduction

---

2 Chatbot PIGMA

---

3 Architecture data intelligente

---

4 Bonne pratiques et nouveaux  
services

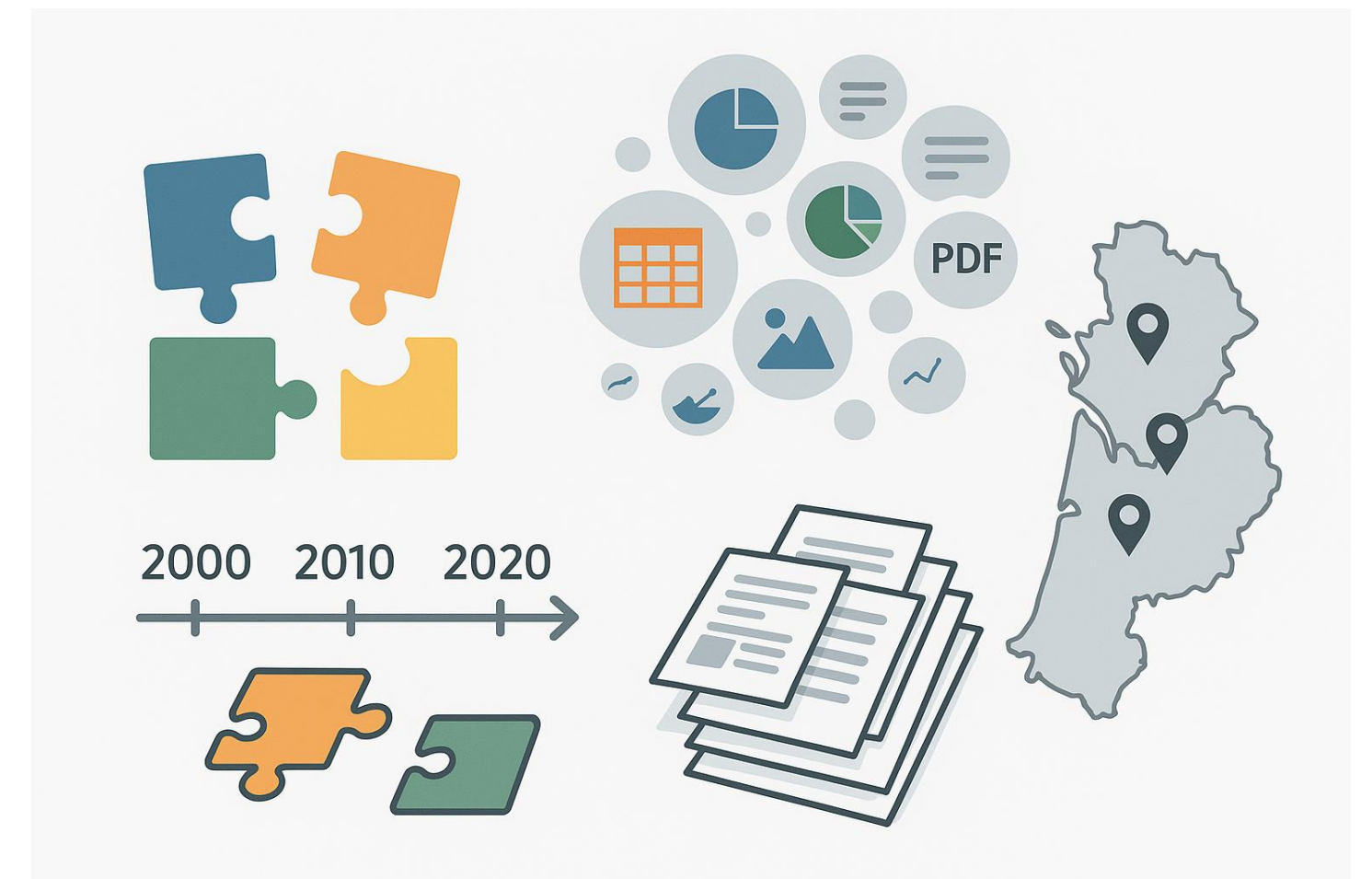
---

5 Echanges

## Problématique : Une richesse de données difficile à exploiter

- La **multiplication des jeux de données** offre un potentiel considérable, mais engendre une **complexité croissante** :
- Les données sont souvent **hétérogènes, dispersées** (formats, sources, usages), rendant leur croisement difficile.
- La **dimension temporelle** (séries chronologiques, données temps réel) ajoute un **niveau supplémentaire de complexité** dans la collecte, la structuration et l'analyse.

**Sans structuration adaptée**, cette richesse reste **sous-exploitée**, voire **inutilisable** pour des usages transverses ou automatisés.



## Vision : une architecture data intelligente et interopérable

Le chatbot peut être une **porte d'entrée intuitive**, mais son efficacité est démultipliée par une architecture structurée derrière **Un socle data robuste**

C'est là que le concept de datawarehouse intervient...

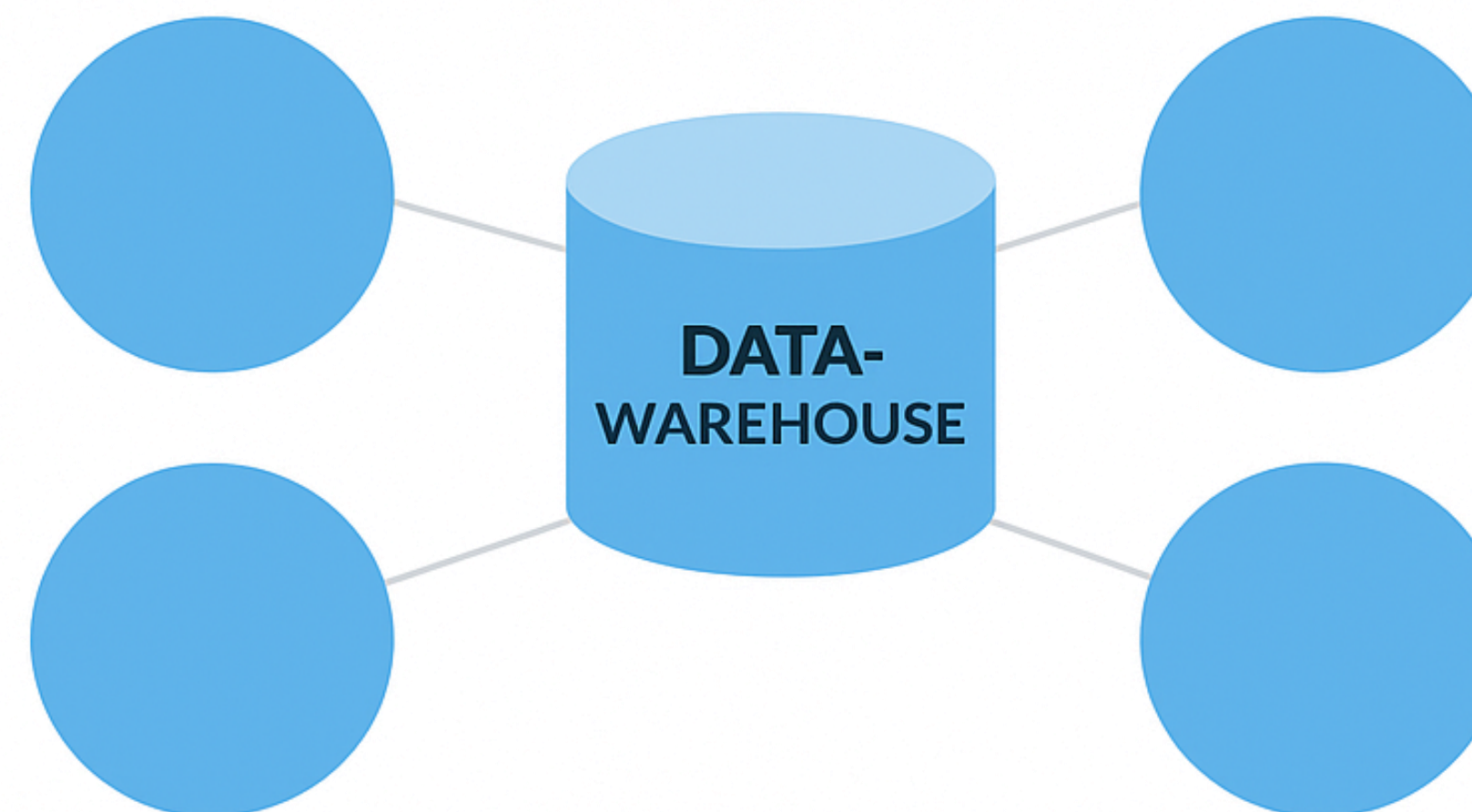
Organisé pour interconnecter les flux, les outils, les visualisations

### Interfaces d'accès

- Chatbot conversationnel
- API / Portails utilisateurs
- Dashboards / dataviz

### Services de traitement

- Moteurs ETL
- Services d'analyse spatiale
- Algorithmes IA (LLM, ML)



### Sources de données

- Jeux de données PIGMA
- Flux métiers / capteurs / PDF
- Bases externes (INSEE, OpenData...)

### Services de structuration

- Génération de métadonnées
- Validation par schémas (schema.data.gouv.fr)
- Sémantique et référentiels (OGC TJS, thésaurus)

# Qu'est qu'un datawarehouse

Un datawarehouse, ou entrepôt de données, est une [base de données](#) servant à stocker, combiner et archiver à long terme des données provenant de différentes sources. Les données y sont organisées de manière à préserver leur stabilité contextuelle et à faciliter les requêtes et prises de décision *data-driven*.

## Pourquoi créer un datawarehouse ?

- Pour consolider ses données au sein d'un référentiel unique
- Pour obtenir rapidement des informations stratégiques
- Pour assurer une cohérence et une traçabilité des données

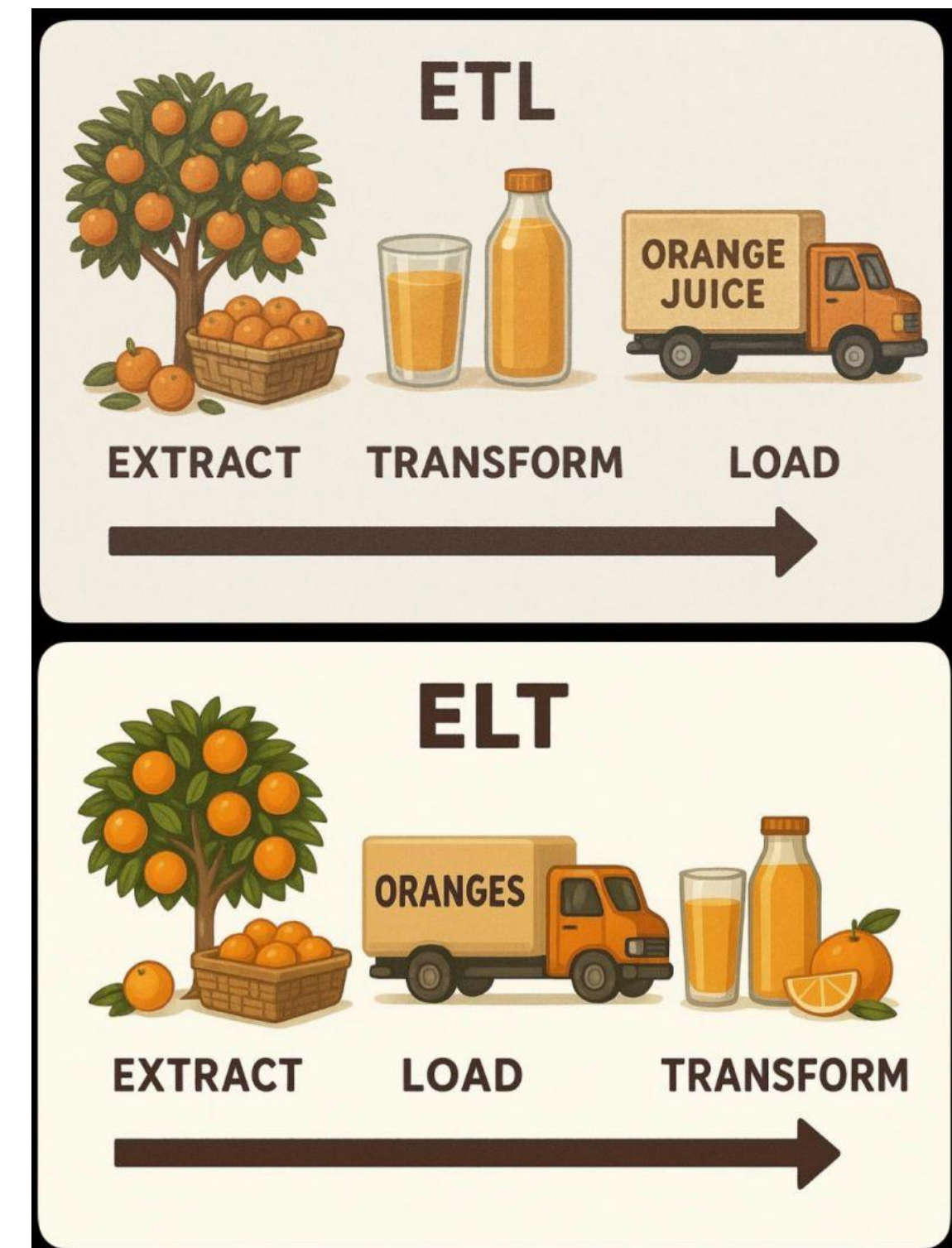


Mais attention, tout n'est pas parfait ni universel. Car parfois, on a affaire à des données non structurées, ou à des volumes tellement énormes que le Data Warehouse classique commence à montrer ses limites. Et c'est là qu'arrive un autre concept, un peu plus « sauvage », un peu plus « libre », le **Data Lake**.








## Permet une approche ELT plutôt que ETL

Intérêt de faire des traitements en fin de chaîne :

- **Performance & scalabilité** : les transformations s'exécutent dans le Datawarehouse, conçu pour des calculs massifs.
- **Flexibilité & traçabilité** : conservation des données brutes, facilitant réanalyse, audits et contrôles qualité.
- **Ingestion plus rapide** : les données sont chargées immédiatement, sans attendre les traitements en amont.
- **Sobriété numérique** : évite la duplication inutile des données et limite les traitements redondants en amont, en s'appuyant sur des moteurs mutualisés et optimisés.



# Intérêt de stocker les processus de traitement plutôt que la data

-  **Sobriété** : minimiser les coûts de transferts des données au sein d'infrastructures multi-fournisseurs
-  **Rejouabilité** : On peut rejouer le traitement à tout moment avec de nouvelles données ou paramètres.
-  **Moins de stockage** : Les résultats intermédiaires ou finaux peuvent être recalculés à la volée, évitant des téraoctets de duplication.
-  **Mise à jour dynamique** : Lorsque les données sources changent, les résultats sont régénérés automatiquement.
-  **Transparence et traçabilité** : Le traitement est documenté, versionné, et auditable.
-  **Paramétrabilité** : On peut adapter les traitements à différents cas d'usage ou jeux de données.
-  **Réutilisabilité** : Un processus générique peut être utilisé dans plusieurs projets ou cas métiers.

**Combiné à un datawarehouse spatial** (ex : PostgreSQL/PostGIS), on peut :

- **Déporter le traitement lourd** sur un cluster distant.
- **Automatiser** certains traitements (via Airflow ou cron).
- **Rendre les traitements accessibles** via une API publique (avec authentification).
- **Documenter les workflows** dans un catalogue de métadonnées.

**Pour aller plus loin...**

**ZOO-Project** : framework libre de traitements **géospatiaux** via des **services WPS**, standardisé par l'OGC

**PyWPS** : équivalent Python de ZOO, conforme au standard OGC WPS.

**GeoServer WPS** : extension de GeoServer pour exécuter des traitements simples via WPS.

**Actinia (GRASS GIS)** : API REST pour exécuter des traitements à distance sur des données spatiales.

**OpenEO** : standard API unifié pour lancer des traitements sur des plateformes cloud (données Copernicus/Sentinel).

# Bonnes pratiques et nouveaux services rendus possibles

04

1 Introduction

---

2 Chatbot PIGMA

---

3 Architecture data intelligente

---

4 Bonne pratiques et nouveaux  
services

---

5 Echanges

## Bonnes pratiques

Grâce à cette architecture, il devient plus facile de mettre en œuvre des bonnes pratiques :

- Structuration selon des **standards reconnus** (ex. [schema.data.gouv.fr](https://schema.data.gouv.fr))
- Utilisation de **référentiels interopérables** (ex. OGC TJS pour lier des données par des dimensions communes)
- Définition claire des **publics cibles et des usages associés**
  - Ex : visualisation publique / exploitation métier / extraction IA
- **Sémantique métier unifiée**
  - Pour garantir un langage commun entre producteurs et utilisateurs de données

## Bonnes pratiques rendues opérationnelles



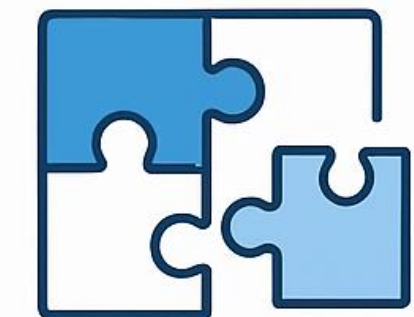
Respecter les standards



Respecter les référentiels



Déterminer les destinataires

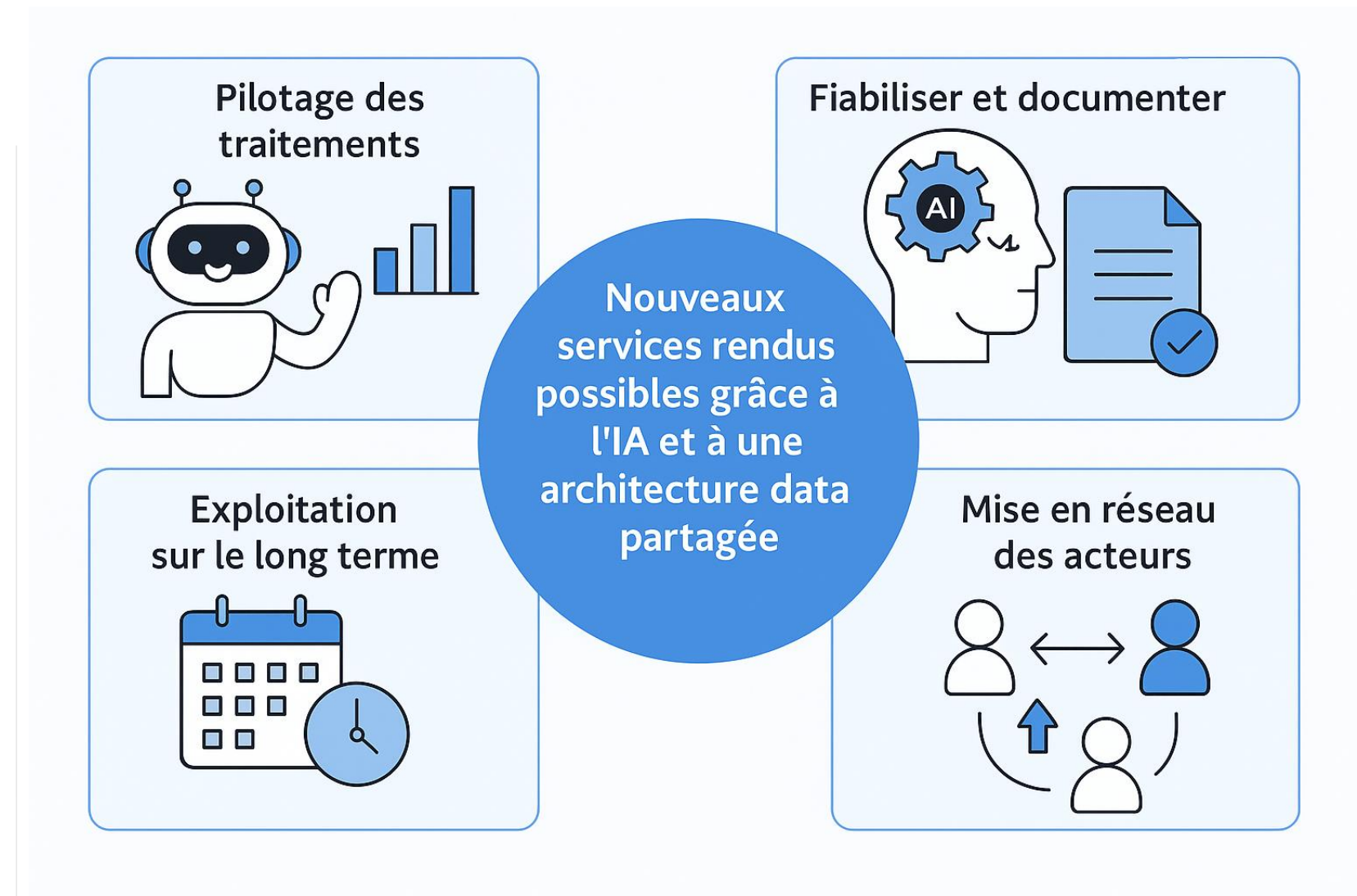


Utiliser une sémantique unifiée

## Nouveaux services

Nouveaux services autour de la mutualisation des traitements entre différents acteurs facilités par l'IA et cette architecture :

- Utilisation du chatbot pour **piloter des traitements** de données et la génération de cartes ou dataviz
- Utilisation de l'IA pour **fiabiliser et documenter** les données (détection d'erreurs, segmentation des données..., génération automatique ou assistée de métadonnées pour faciliter le catalogage et l'interopérabilité)
- **Meilleure exploitation** des données sur le long terme (archivage, données temporelles...)
- Mise en contact et **partages** des cas d'usages entre les consommateurs de données et **mutualisation** des traitements entre différents acteurs".

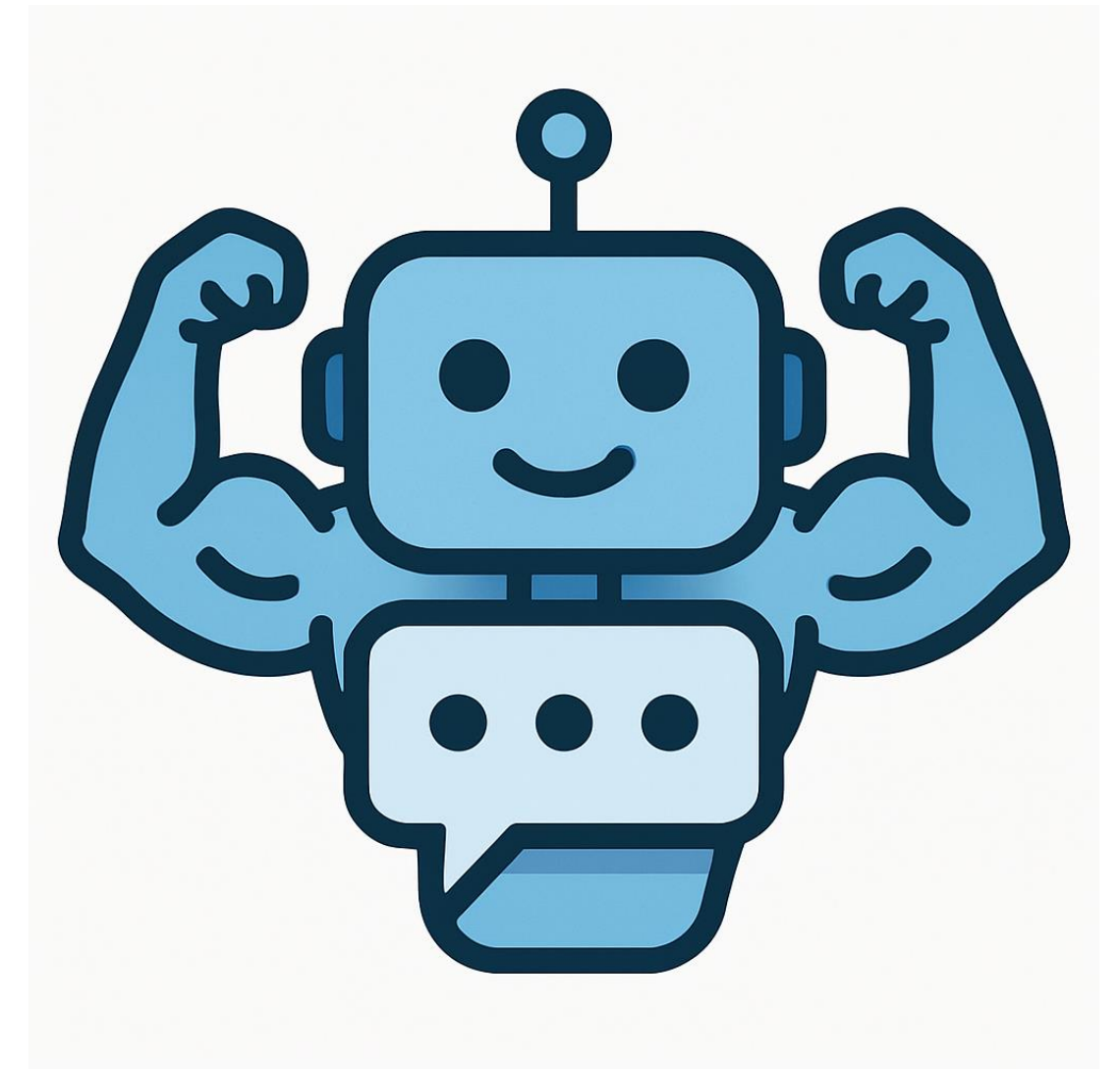


## Un chatbot augmentée

C'est aussi dans cette architecture que le chatbot peut apporter une vraie valeur ajoutée :

Par exemple rêvons un peu :

On pourra demander au chatbot de **créer un espace de travail** pour une problématique spécifique, il proposera des **jeux de données pertinents**, il discutera avec vous de la **structuration de la donnée**, de quelles informations sont **disponibles** en entrée, il vérifiera les jeux de données qui existent déjà sur la plateforme pour **éviter les doublons**, il pourra vérifier si des **schémas existent** déjà pour ce type de traitement, il pourra créer une **ébauche de traitement** de données, voir **mutualiser une partie** des traitements, et enfin, une fois le résultat validé, il pourra vous aider à **publier la donnée** en remplissant la **fiche de métadonnées** selon les règles de la plateforme ou selon un **standard...**



# Échange avec les participants / conclusion

05

1 Introduction

---

2 Chatbot PIGMA

---

3 Architecture data intelligente

---

4 Bonne pratiques et nouveaux services

---

5 Echanges

## Échange avec les participants / conclusion

- Questions-réponses
- Appel à collaboration :
  - Quelles attentes métiers autour d'un chatbot ?
  - Quels types de traitements / automatisations seraient utiles à votre collectivité / structure ?



neogeo

**Merci pour votre  
attention**

<https://neogeo.fr>

