

Smart Mountain for tomorrow



20<sup>th</sup> OCOVA FORUM

# GESTION DE L'EAU : QUALITÉ DES DONNÉES & VALORISATION PROJET SEDIMARK



Marianne MAROT

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them



# QUI SOMMES NOUS ?

Nous offrons des solutions **open source + standardisées**



**1.2M**  
CA de prestations intellectuelles



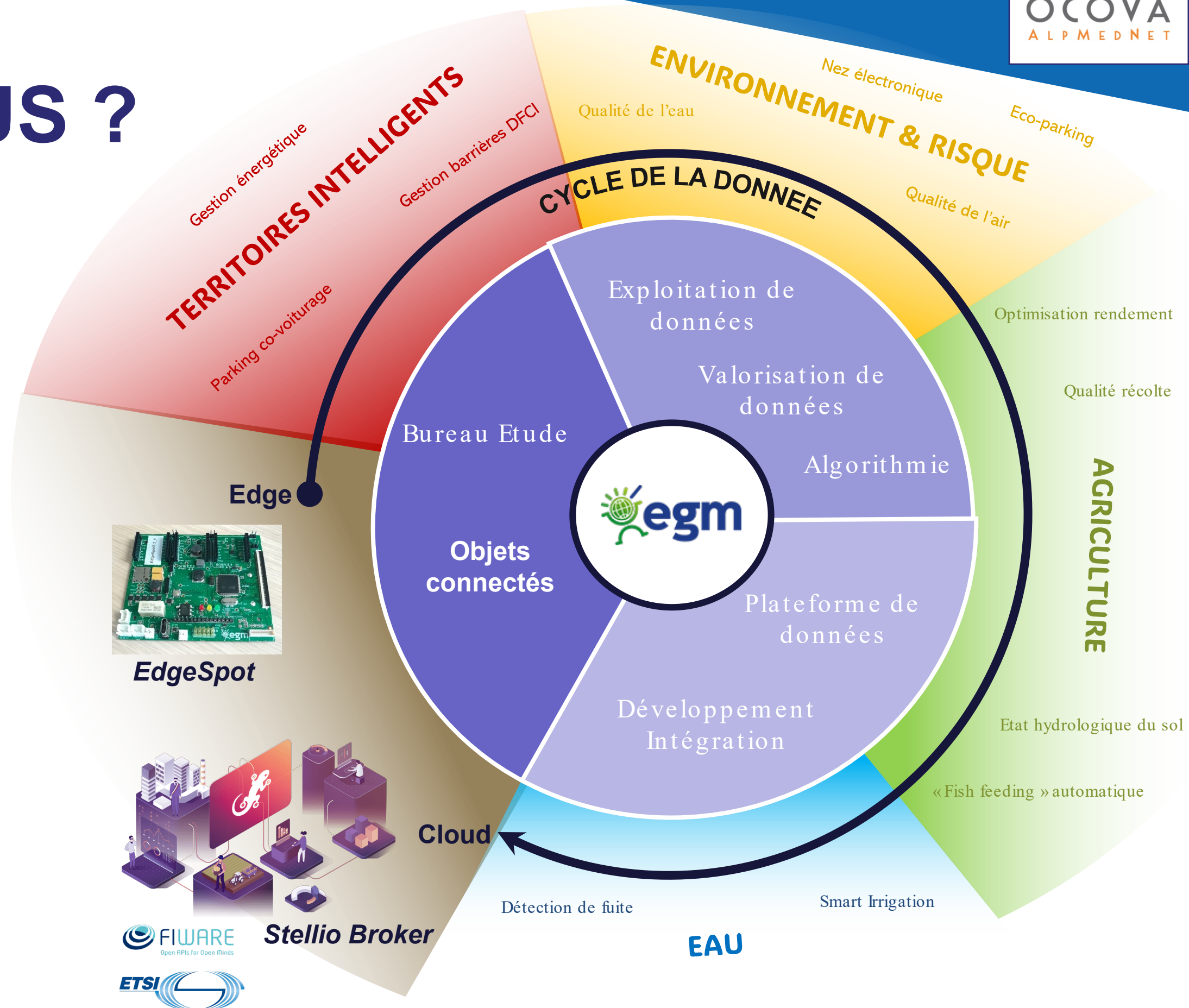
**14** années d'expérience

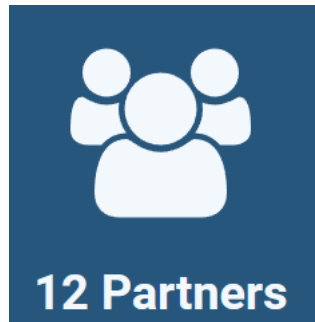


**15** collaborateurs

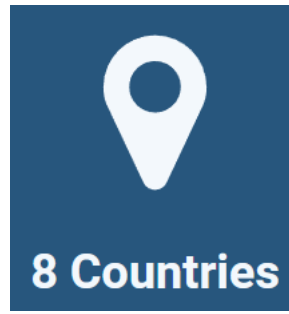


**+ 30** projets internationaux de recherche FIWARE, ETSI





12 Partners



8 Countries



4 Use Cases



Sept. 2025

# SEDIMARK

## SEcure Decentralised Intelligent Data MARketplace

### Enjeux

- EU: Devenir un centre de données massif et sûr
- Données : qualité élevée + partage sûr + FAIR
- Promouvoir la découverte de données multi-domaines, l'économie et la recherche transversale

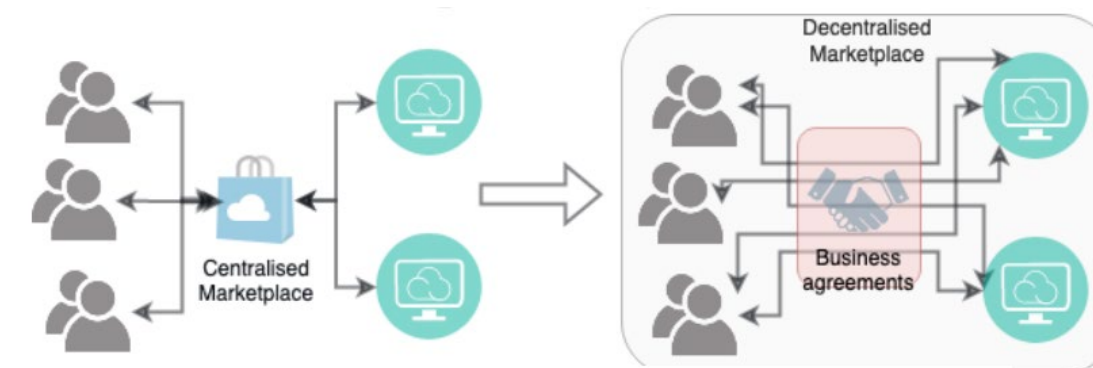
### Verrous

- Places de marché **centralisées**
- Données en **silos** + **non-FAIR**
- Absence de **contrôle qualité** + **modèles génériques** des données
- Manque de **sécurité** de partage + **protection** des données à caractère personnel
- **Intégration** inadaptée avec l'IA

### Objectifs

Place de marché **décentralisée** + **open source** :

- **Catalogue de données** : FAIR, certifiées propres + sûres + conformes RGPD, enrichies, complétées, annotées, ...
- **Outils** + **services distribués** intelligents (dont AI et analytiques)
- Techniques de gestion de **confiance**



**F**indable **A**ccessible **I**nteroperable **R**eusable  
Facilement trouvable, Accessible, Interopérable, Réutilisable



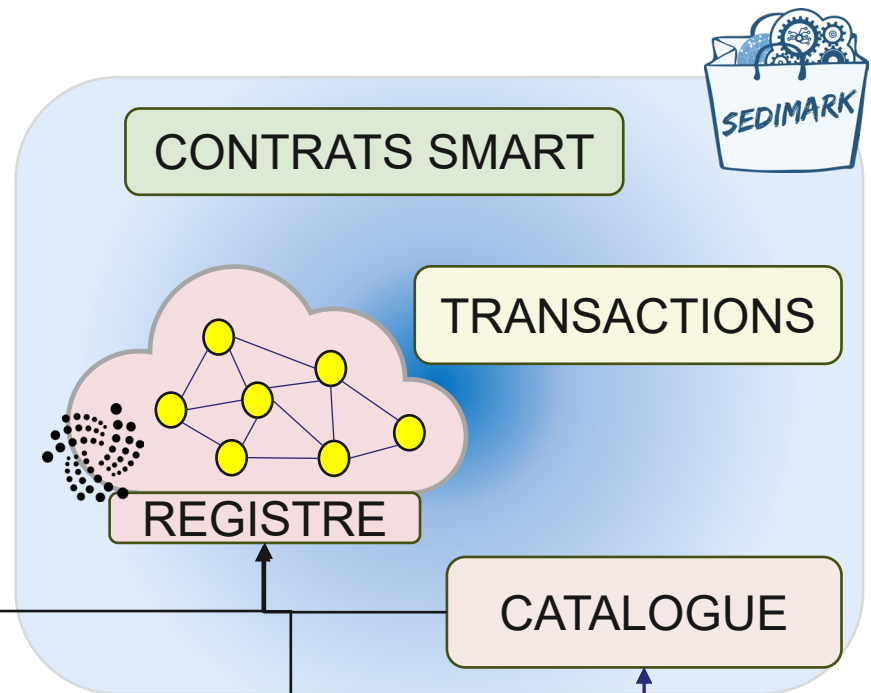
# SEDIMARK



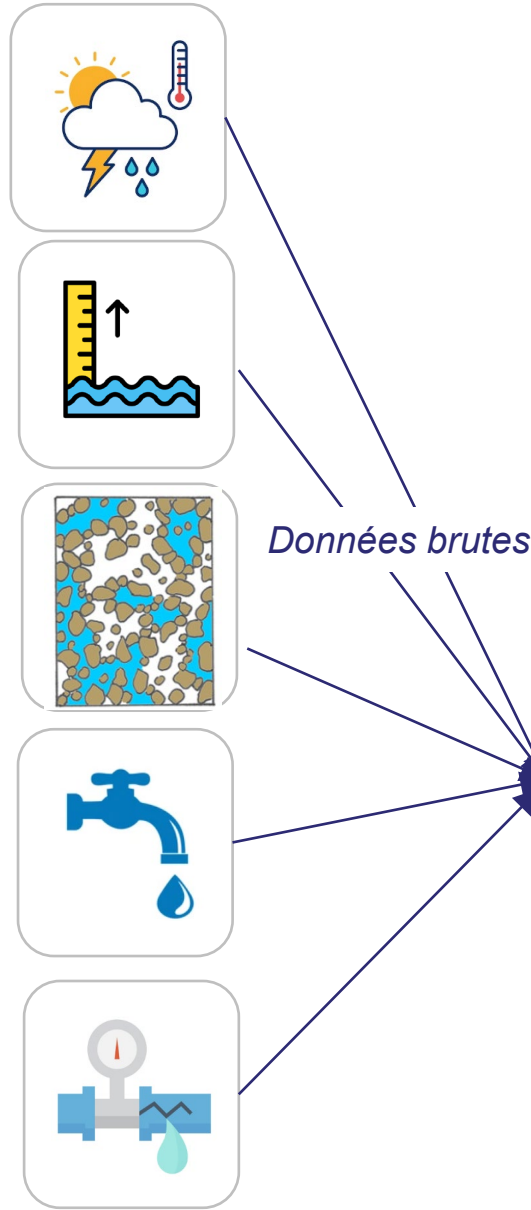
**PLATEFORME**  
Urbanisme pour la  
mobilité cycliste



**PLATEFORME**  
Valorisation des  
données énergie



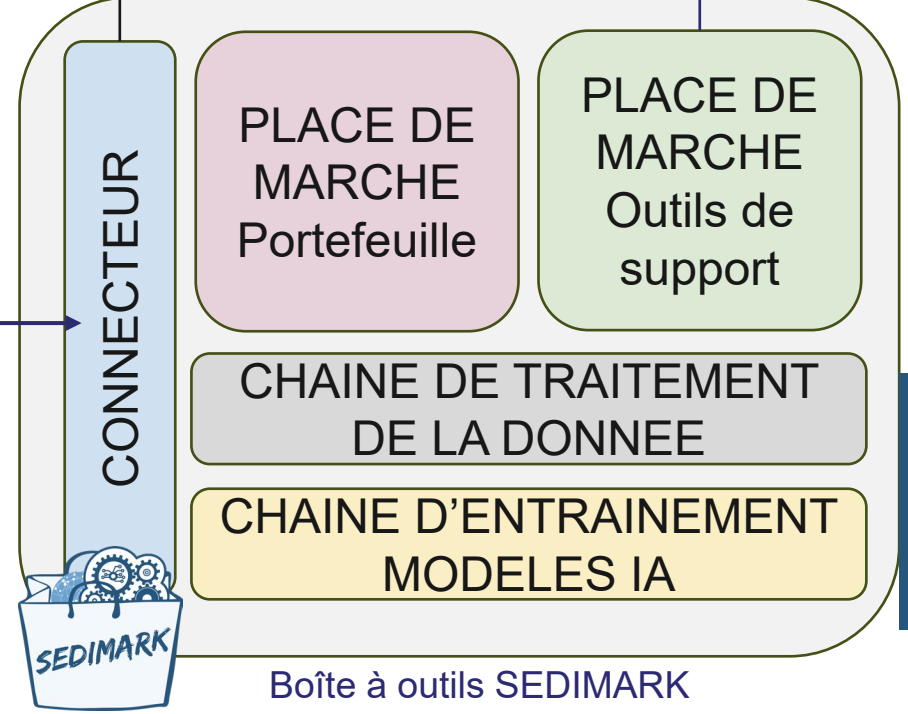
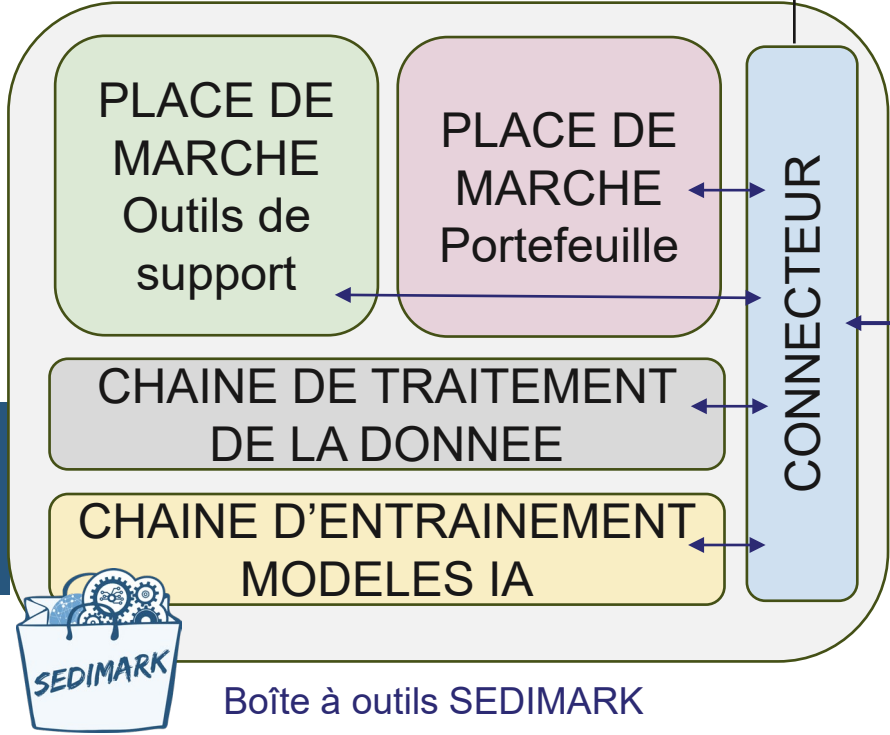
Infrastructure SEDIMARK de base



Données brutes



France  
**PLATEFORME**  
Valorisation des  
données eau



Helsinki

**PLATEFORME**  
Jumeau  
numérique de la  
mobilité urbaine

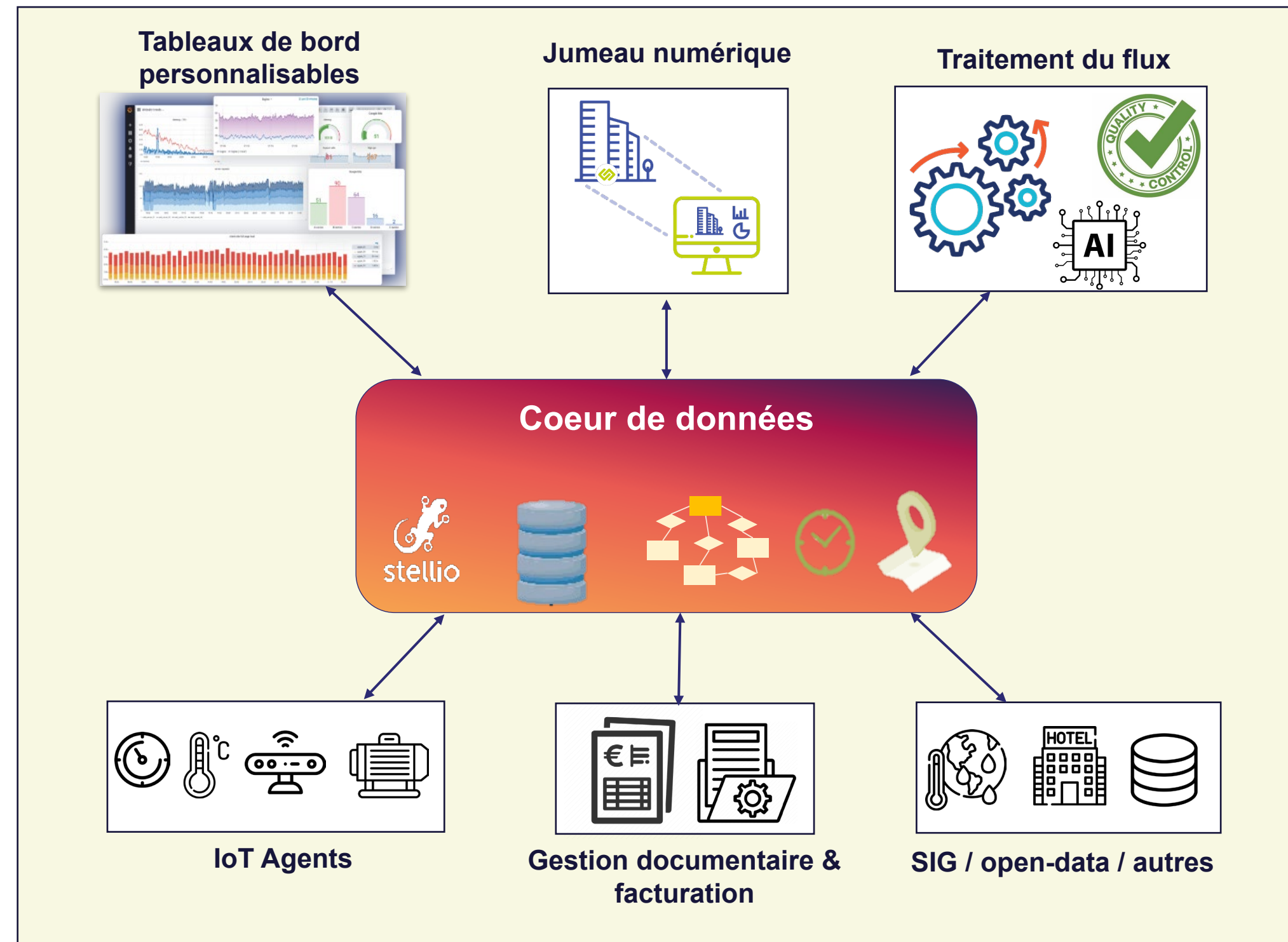
# PLATEFORME STELLIO

Offre EGM pour la Gestion et valorisation de l'eau



Nombreux cas d'usage : **Gestion + sauvegarde des fluides**

- **Smart irrigation** (réduction consommation + interventions)
  - Détection de **fuites** + **surfacturation**
  - **Risque submersion**
  - **Qualité de l'eau** (pollution, contamination)
- **Etat d'eau souterraine** (niveau, débit, qualité)
  - ... Et le vôtre ?



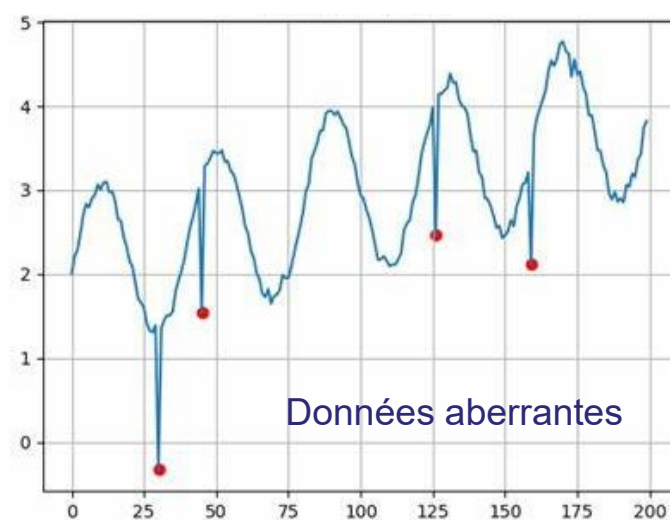
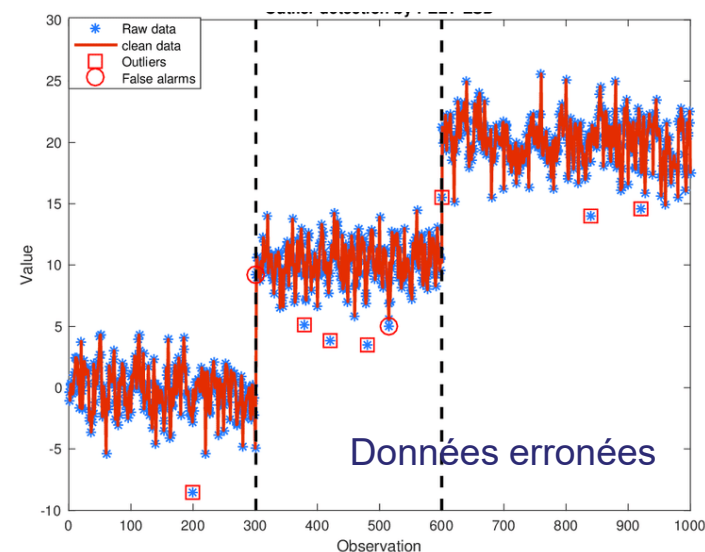


# BOÎTE À OUTILS [SEDIMARK / EGM]

## CONTRÔLE QUALITE

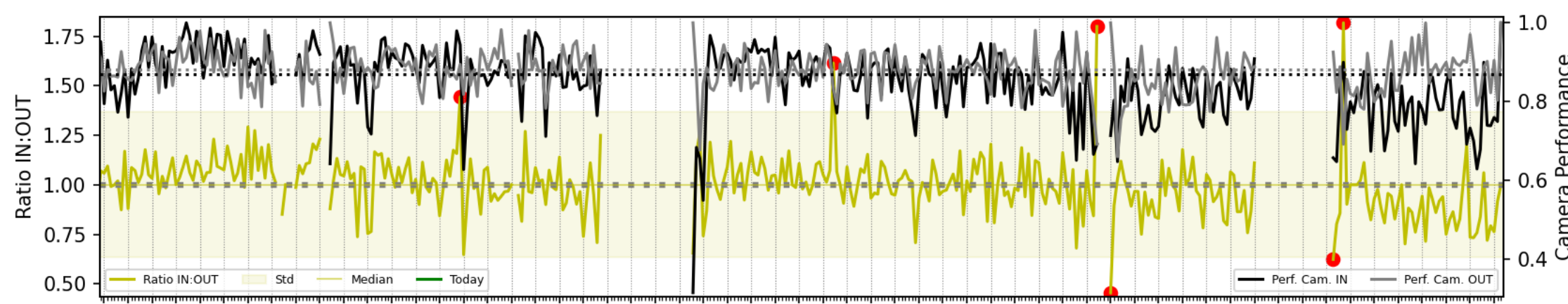
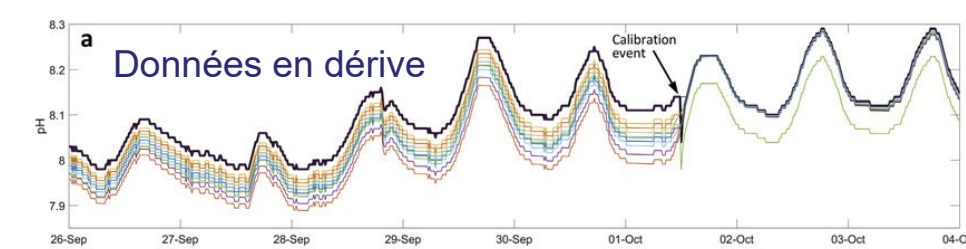
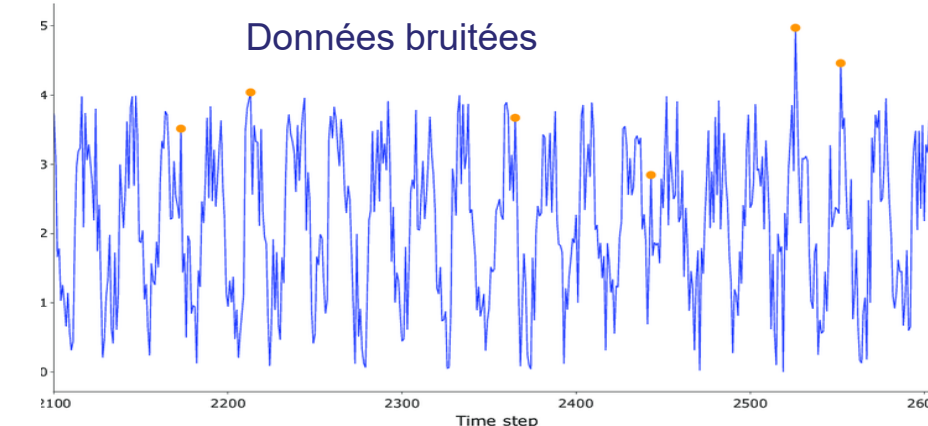
### Qualité des données

- Harmonisation
- Dédoublonnage
- Identification de :
  - Données manquantes ou aberrantes
  - Anomalies contextuelles
  - Dérives
- Réduction du bruit
- Complétion / remplacement de données manquantes ou aberrantes
- Augmentation & enrichissement des données
- Labélisation & annotation sémantique
- Métriques de qualité, validation & certification



Données dupliquées

	A	B	C	D
1	First name	Last name	Date of birth	E-mail address
2	Lucas	Bos	27/11/1980	lucas.bos@gmail.com
3	Finn	Meijer	15/02/1979	finmeijer@hotmail.com
4	Levi	Mulder	03/05/1980	lmulder@gmail.com
5	Sophie	Peters	06/07/1981	sophieeee1981@live.com
6	Liam	Brouwer	30/07/1982	l.brouwer@hotmail.com
7	Sam	De Wit	15/03/1980	sam.de.wit@gmail.com
8	Noah	Dijkstra	13/05/1983	dijkstra.noah@gmail.com
9	Julia	Smits	14/07/1982	julia.smits@gmail.com
10	Tess	Bakker	15/02/1979	tesssssb@hotmail.com
11	Luuk	van den Berg	03/10/1978	lvdb78@gmail.com
12	James	Schouten	27/06/1982	jameschouten@gmail.com
13	Nora	van Beek	09/09/1977	noravbeek@live.com
14	Zeger	Verbeek	13/10/1979	zegertie79@gmail.com
15	Milan	Willems	10/08/1983	millems83@outlook.com
16	Liv	van Vliet	12/04/1980	livvliet@outlook.com
17	Sam	Wit, de	15/03/1980	sam.de.wit@gmail.com
18	Sara	Prins	14/09/1983	prinssara@gmail.com
19	Lotte	Prins	17/11/1982	lotte.prins@gmail.com
20	Noud	Huisman	03/12/1984	n.huisman@gmail.com
21	Olivia	Kuiper	05/12/1981	o.liviakui@hotmail.com
22	Anna	Veenstra	18/10/1978	a.veenstra@gmail.com
23	Benjamin	Kramer	09/02/1984	benjamink@live.com
24	T.	Bakker	15/02/1979	tess.bakker@live.com
25	Lauren	Postma	03/10/1985	postma.lauren@gmail.com
26	Bram	Martens	14/07/1982	brartens@live.com
27	Adam	Hermans	28/09/1981	adam.hermans@gmail.com
28	Nina	van Dongen	11/06/1980	nina1980@hotmail.com
29	Max	Kuijpers	14/07/1981	maxkuijpers81@live.com
30	Henk	Mol	14/02/1983	henk.mol1983@gmail.com
31	Julia	Smits	14/07/1982	julia.smits@gmail.com
32	Agatha	Molenaar	06/02/1982	a.molenaar@kpn.nl
33	Luuk	Berg, van de	03/10/1978	luukvandeberg@hotmail.com
34	Maria	de Leeuw	01/08/1981	leeuwin1981@gmail.com
35	Zeger	Verbeek	13/10/1979	zegertie79@gmail.com
36	Marloes	van der Pol	20/04/1983	marloes.pol@outlook.com
37				



Données en dérive  
Données manquantes  
Données aberrantes  
Anomalie contextuelle

# BOÎTE À OUTILS [SEDIMARK / EGM]

## CONTRÔLE QUALITE + ANALYSE DE DONNEES

### Qualité des données

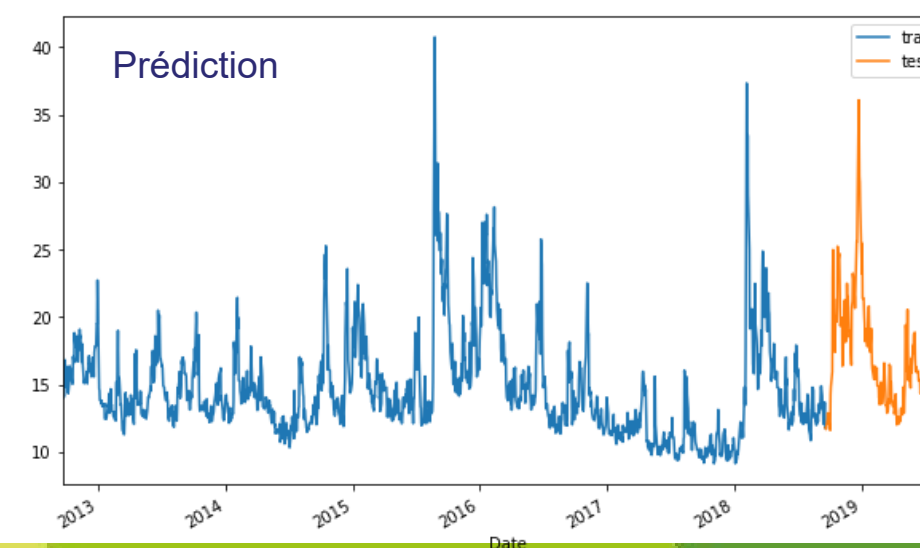
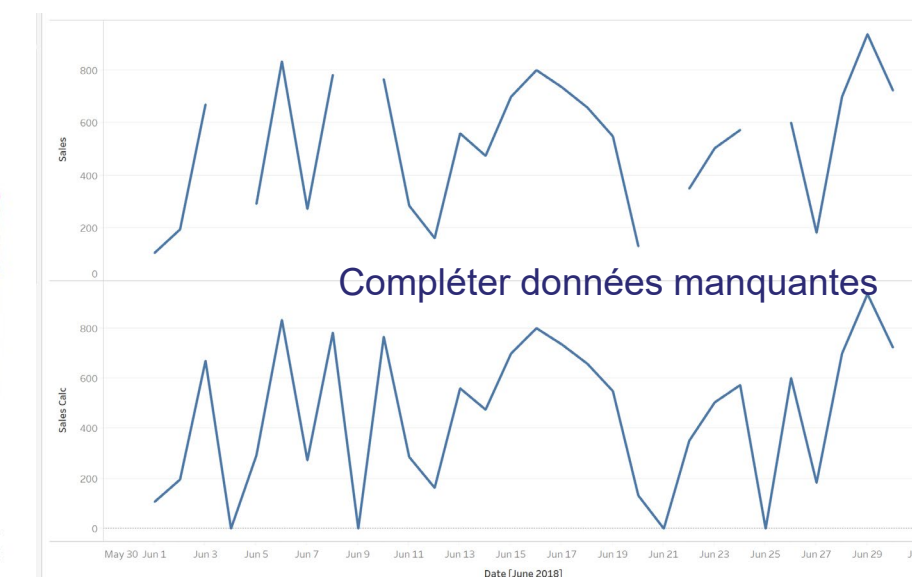
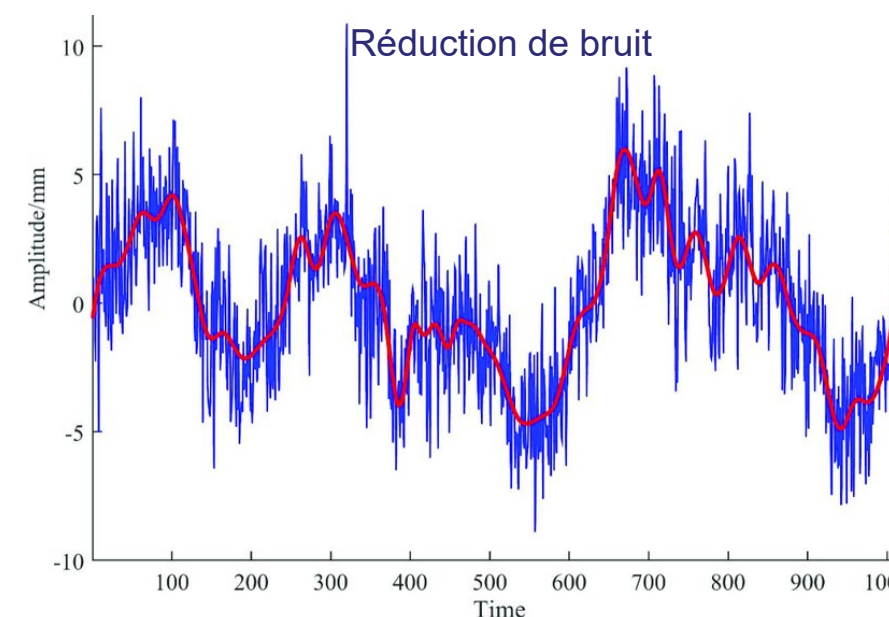
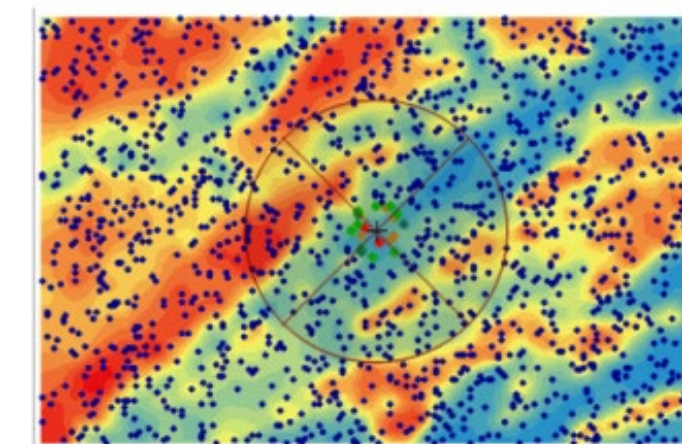
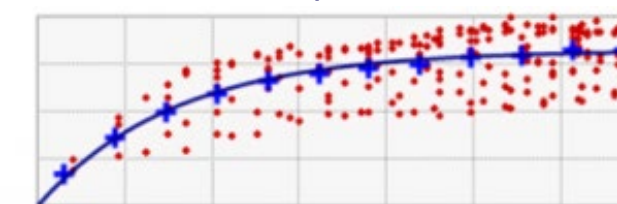
- Harmonisation
- Dédoublonnage
- Identification de :
  - Données manquantes ou aberrantes
  - Anomalies contextuelles
  - Dérives
- Réduction du bruit
- Complétion / remplacement de données manquantes ou aberrantes
- Augmentation & enrichissement des données
- Labélisation & annotation sémantique
- Métriques de qualité, validation & certification



### Analyse & Modélisation

- Exploration de données / analyse descriptive
- Analyse d'interpolation pour le remplacement / complétion des données
- Définition de :
  - KPIs
  - Anomalies contextuelles
  - Dérives
  - Critères de décision et différents niveaux d'alertes
- Analyse & modélisation prédictive pour la détection d'anomalies et l'anticipation d'évènements

Interpolation





# BOÎTE À OUTILS [EGM]

## SURVEILLANCE, ALERTES, REPORTING

### Qualité des données

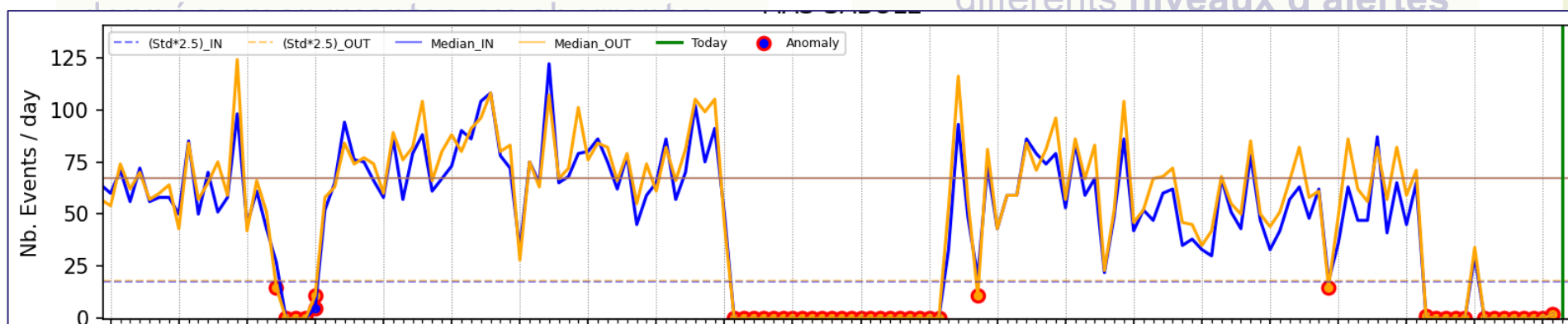
- Harmonisation
- Dédoublonnage
- Identification de :
  - Données manquantes ou aberrantes
  - Anomalies contextuelles
  - Dérives
- Réduction du bruit
- Complétion / remplacement de

### Analyse & Modélisation

- Exploration de données / analyse descriptive
- Analyse d'interpolation pour le remplacement / complétion des données
- Définition de :
  - KPIs
  - Anomalies contextuelles
  - Dérives
  - Critères de décision et différents niveaux d'alertes

### Surveillance & Alertes

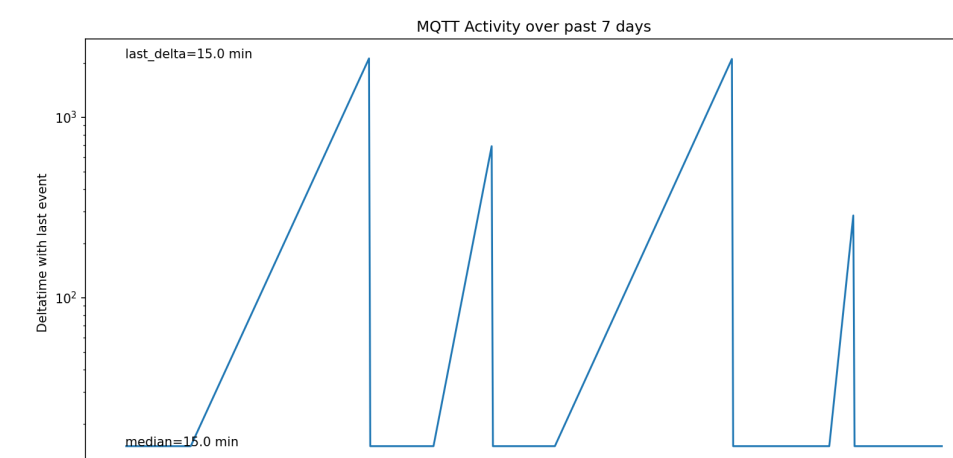
- Détection de:
  - Indisponibilité des capteurs
  - Dérives
  - Anomalies
  - Incohérences entres variables
- Création d'alertes + Rapports



- Métriques de qualité, validation & certification

stellio@egm.io  
À : Marianne MAROT (EGM); + 2 autres  
Lun 29

PKSUD ne reçoit plus de données depuis au moins 24h



Filtres | Parkings-CD66 | Publié | Éditer le tableau de bord

+ AJOUTER/MODIFIER LES FILTRES

Intervalle de temps par date: 2023-07-29T21:22:22 ≤ c...  
Nom du site: CD66\_PK\_MAS\_SABOLE  
Filtres hors du périmètre (0)

APPLIQUER LES FILTRES | EFFACER TOUT

Nombre de passages	Entrées	Sorties	Durées	Types de carburant
Total: 4,378	Jours ouvrés: 27 passages/jour	Weekends: 18 passages/jour		

Evolution des passages



# BOÎTE À OUTILS [EGM]

## CONFIDENTIALITE

### Qualité des données

- Harmonisation
- Dédoublonnage
- Identification de :
  - Données manquantes ou aberrantes
  - Anomalies contextuelles
  - Dérives
- Réduction du bruit
- Complétion / remplacement de données manquantes ou aberrantes
- Augmentation & enrichissement des données
- Labélisation & annotation
- Métriques de qualité, validation & certification

### Analyse & Modélisation

- Exploration de données / analyse descriptive
- Analyse d'interpolation pour le remplacement / complétion des données
- Définition de :
  - KPIs
  - Anomalies contextuelles
  - Dérives
  - Critères de détection à différents niveaux
- Analyse & modélisation prédictive pour la détection d'anomalies et l'anticipation d'évènements

### Surveillance & Alertes

- Détection de :
  - Indisponibilité des capteurs
  - Dérives
  - Anomalies
  - Incohérences entre variables

### Confidentialité

- Anonymisation & pseudo-anonymisation des données à caractère personnel / Conformité RGPD

#### Méthodes d'anonymisation



```
{  
  name: MAROT,  
  code_postal: 06370,  
  email: marianne.marot@egm.io,  
  location: {  
    lon: 44.7654,  
    lat: 7.7658 },  
  comments: Fuite dans l'appartement  
             n°44 de la Résidence Suzanne à  
             Mouans Sartoux,  
  pin: mypassword,  
}
```

+ Rapports

```
{  
  name: ***,  
  code_postal: 06***,  
  email: *****@***.**,  
  location: Nice,  
  comments: Fuite dans  
             l'appartement n°*** de la  
             Résidence ***  
             à ***,  
  pin: hp75jB5k5b6Hj6,  
}
```

# CONCLUSIONS



## PLATEFORME CENTRALISEE STELLIO EGM



## PLACE DE MARCHÉ DECENTRALISEE SEDIMARK

Open source  
Partage de données multi-domaines, FAIR, propres, sûres  
Control qualité, Analyse, Prédiction  
Outils + services intelligents  
Conformité EU RGPD  
Sécurité + Confiance

Collecte de données multi-domaines  
Surveillance, Alertes, Reporting  
Jumeau numérique

REJOIGNEZ-  
NOUS

ON RECRUTE

Smart  
Mountain  
for  
tomorrow



20<sup>th</sup> OCOVA FORUM

**MERCI POUR VOTRE ATTENTION**  
**THANK YOU FOR YOUR ATTENTION**



**Marianne MAROT**

Ingénieur R&D / Data scientist

[marianne.marot@egm.io](mailto:marianne.marot@egm.io) / [contact@egm.io](mailto:contact@egm.io)





# ANNEXE

SEDIMARK includes participants that cover the whole data value chain<sup>62</sup>:



**Data acquisition:** MYT, SDR and FV are data providers, generating huge amounts of data from sensors and users and will feed parts of them both raw and processed to SEDIMARK. Other partners also have direct access to data sources too (EGM, WINGS, UC, AtosIT).

**Data analysis:** AtosIT, SIE, MYT, EGM, WINGS, UC, SURREY, and NUID UCD are working on data analysis, with different focuses, i.e. on semantics, linked data, data discovery, data analytics and knowledge extraction.

**Data Curation:** a large number of partners are working in this area, AtosSP (interoperability, security), SIE (Automation), EGM (data validation, Interoperability, Automation), WINGS (Automation, Data quality), IOTA (data validation, trust/governance, incentivisation), INRIA (interoperability, annotation), UC (annotation, interoperability) and NUID UCD (data quality, annotation, trust).

**Data Storage:** SEDIMARK involves partners that have long track of expertise in data storage, i.e. IOTA (security/privacy, standardisation, distributed storage, scalability), SIE (data models, security, standardisation), EGM (scalability, availability, standardisation), UC (data models, scalability).

**Data Usage:** the following members of the SEDIMARK consortium cover the area of data usage: AtosIT (Prediction, Visualisation, Modelling, applications), SIE (prediction, visualisation), MYT (prediction, visualisation, applications, analytics), WINGS (prediction, exploration, modelling), FV (visualisation, control, applications), SDR (visualisation, control, applications), NUID UCD (prediction, modelling, simulation)

**End users:** SEDIMARK also includes links to end users of a data marketplace, through MYT (internal and external stakeholders), SDR (city authorities and citizens), FV (city authorities and citizens).

## SYSTEM VIEW

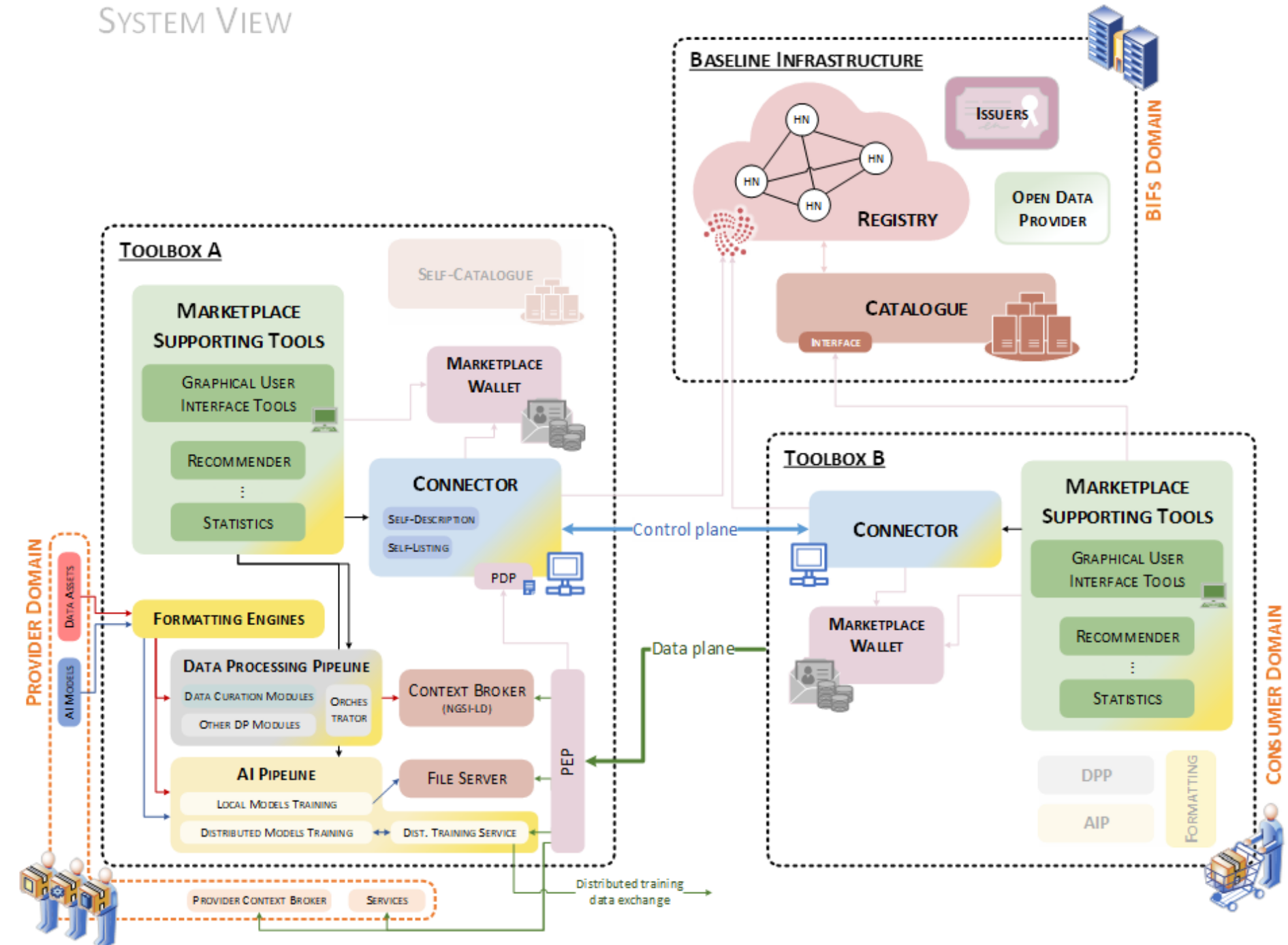


Figure 48: SEDIMARK platform system view