

# CEBA, un data lake dédié à l'observation des écosystèmes environnementaux

Francis Ogereau, Vincent Breton, Alexandre  
Claude, Gilles Mailhot, Jérémy Mezhoud,  
Christine Plumejeaud, Estelle Théveniaud, and  
David Sarramia

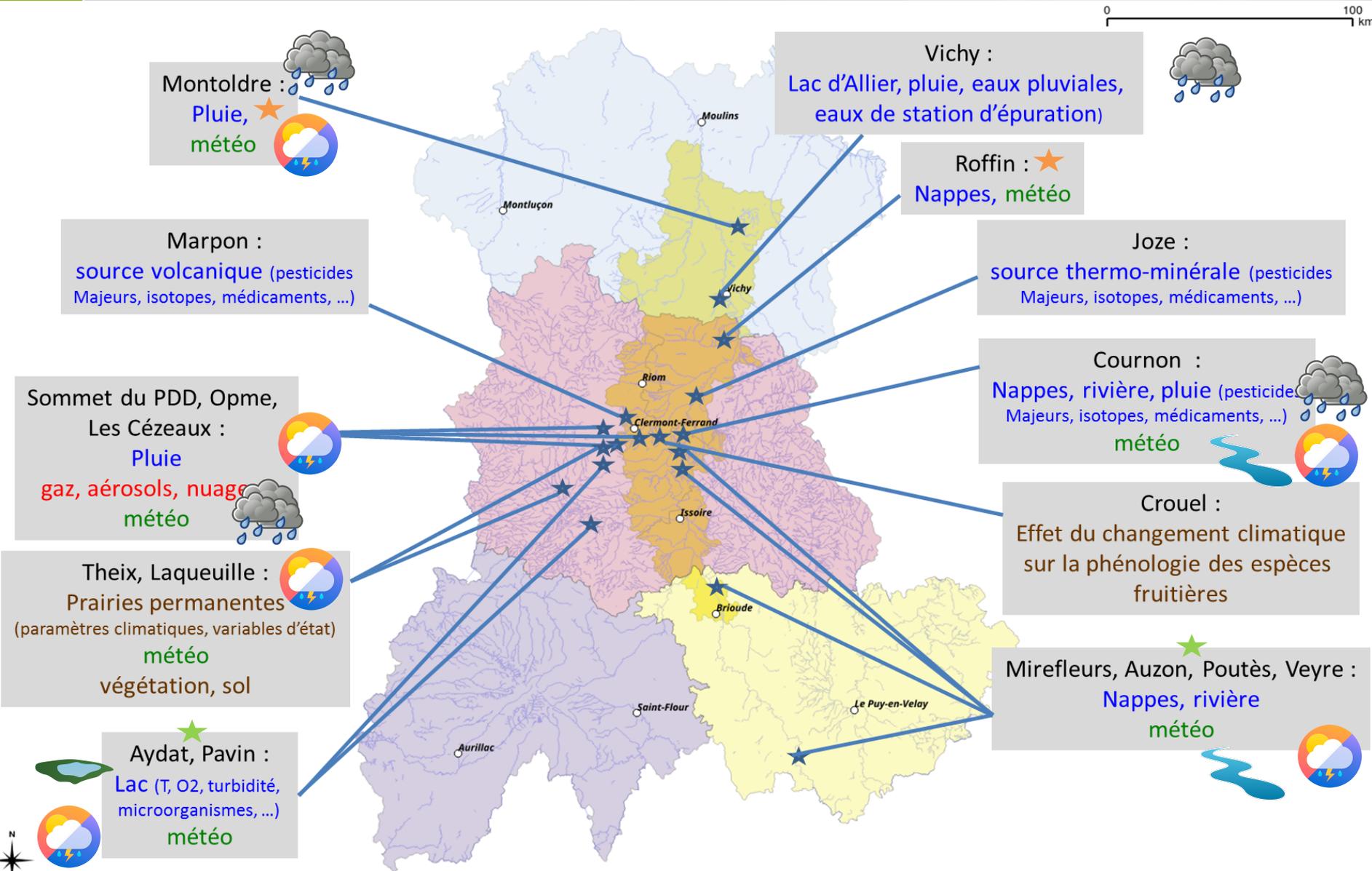


- Délivrable du Challenge 1 de l'I-Site clermontois : Agro-écosystèmes durables dans un contexte de changement global
- Une ambition importante :
  - Création d'un « grand » **observatoire de l'environnement** en Auvergne, unique en Europe.
- Des objectifs scientifiques à fort impact :
  - ➔ Concevoir un **environnement numérique** interconnecté valorisant les données environnementales existantes ou à acquérir.
  - ➔ **Automatiser le monitoring** de données à distance.
  - Accroître notre **compréhension des compartiments environnementaux** face au changement global (évolutions, interactions).
  - Comprendre **l'impact de ces compartiments sur les agroécosystèmes** et réciproquement.
  - Conceptualiser les **interrelations entre les compartiments** pour comprendre, **modéliser et prédire** le comportement des agroécosystèmes.
  - Accompagner **l'aide à la décision** (agriculteurs, entreprises, collectivités territoriales).



# CEBA - Zone Geographique

0 100 km



# Les données à (in)gérer

Capteurs  " temps réel "

Fichiers 

Base de données 

Données spatiales, GPS 

Image, audio, vidéo 



Quoi, quand, où



## Concept de Data Lake

- Site Web
- Outils d'ingestion
- Moteur d'indexation
- Catalogue de données
- Outils de visualisation
- Bases de données
- Système de fichiers



GeoNetwork



Grafana



PostgreSQL



beats

logstash



elasticsearch



kibana



**COLLECTER**

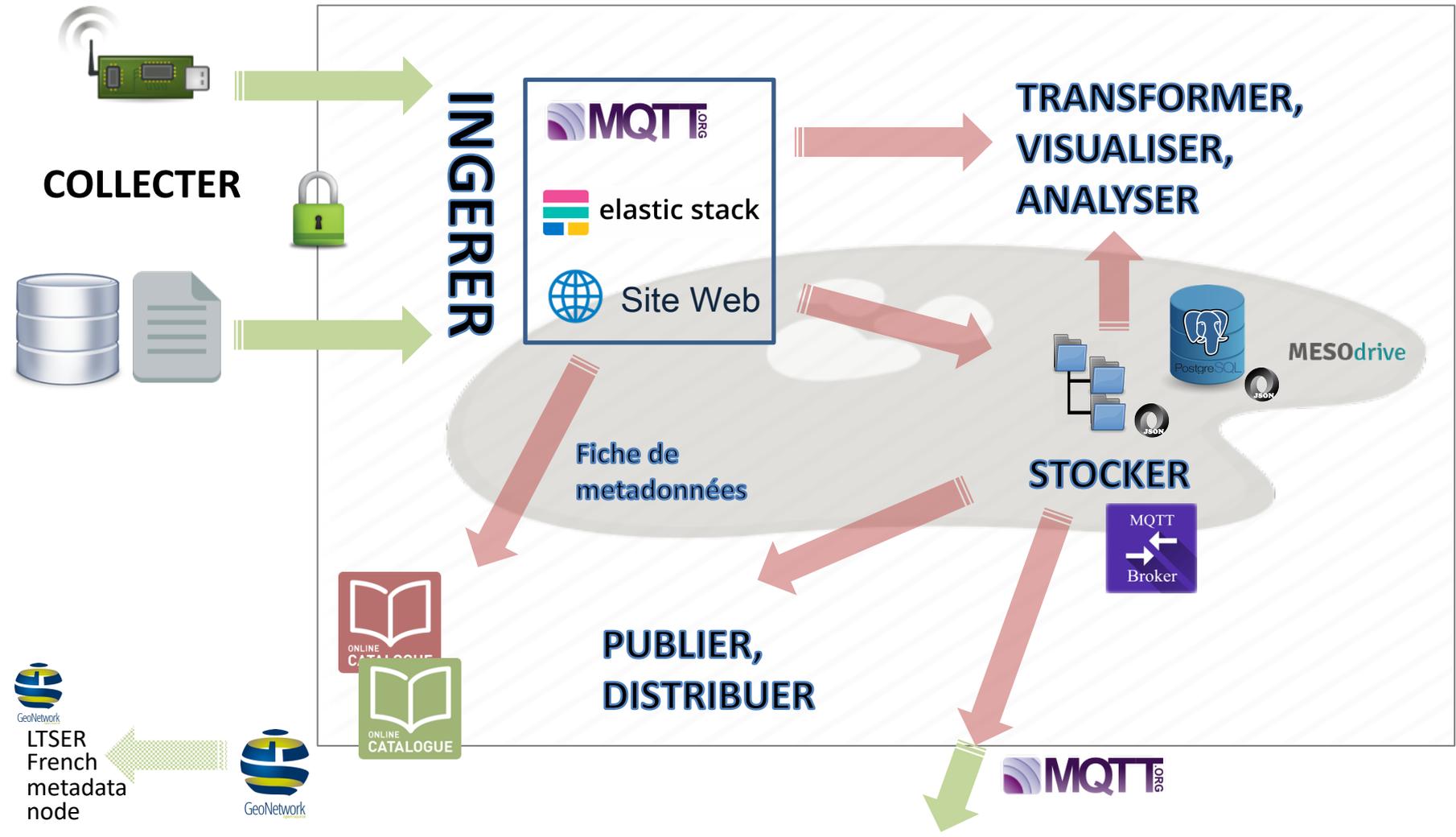
**TRANSFORMER, VISUALISER, ANALYSER**

**INGERER → STOCKER**

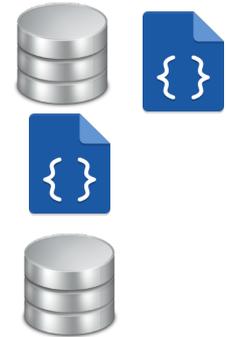
**PUBLIER, DISTRIBUER**



# CEBA – Composants / Big picture



- Réseau de capteurs sans fils
  - Réseau libelium (bd) (json)
  - Réseau ConnecSens (json)
- Base de données (bd)
- Site web
- Métadonnées
- DMP



### Description des données



#### Base de données MySQL

| Column        | Type        |
|---------------|-------------|
| ◇ id          | int(11)     |
| ◇ name        | varchar(45) |
| ◇ description | varchar(45) |
| ◇ id_ascii    | varchar(45) |
| ◇ units       | varchar(45) |
| ◇ value       | int(11)     |



| sensor  |
|---------|
| TCB     |
| HUMB    |
| PA      |
| TCC     |
| BAT     |
| PLV1    |
| PLV2    |
| PLV3    |
| ANE     |
| WV      |
| IN_TEMP |
| RAD     |
| TIME    |
| MAC     |
| ACC     |
| DATE    |
| GPS     |
| ALT     |



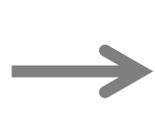
//



logstash

Table [JSON](#)

|              |                                  |
|--------------|----------------------------------|
| @timestamp   | December 10th 2018, 16:12:44.000 |
| @version     | 1                                |
| _id          | 325482                           |
| _index       | mtlsensors                       |
| _score       | 2                                |
| _type        | sensorDatawasp                   |
| frame_number | 31                               |
| id           | 325,482                          |
| id_secret    | 400578826                        |
| id_wasp      | RAD02                            |
| sensor       | BAT                              |
| timestamp    | 2018-12-10 16:12:44              |
| value        | 97                               |



elasticsearch



kibana

Création d'alias pour faciliter les requêtes/visualisation

```
POST /_aliases
{
  "actions": [
    {
      "add": {
        "index": "mtlsensors",
        "alias": "mtlsensorstcbalias",
        "filter": { "term": { "sensor.keyword": "TCB" } }
      }
    }
  ]
}
```



### Description des données

```

object {39}
  data-illuminance-unit : lux
  txInfo-codeRate : 4/5
  fPort : 2
  data-node-timestampUTC : 2018-10-02 07:09:43
  txInfo-adr : false
  servertimestampUTC : 2018-10-02T10:54:21.402434
  txInfo-dataRate-bandwidth : 125
  data-node-batteryVoltage-alarmIsLow : false
  txInfo-dataRate-modulation : LORA
  data-DataChannel : 1
  rxInfo-longitude : 0
  fCnt : 0
  deviceName : HE31294
  applicationName : ConneCSens-Nodes
  rxInfo-name : Kerlink01
  rxInfo-rssi : -26
  devEUI : 434e535301e31294
  data-illuminance : 352
  applicationID : 1
  rxInfo-mac : 7276ff00b030f4c
  data-node-batteryVoltage : 3.92
  data-node-batteryVoltage-unit : V
  rxInfo-latitude : 0
  rxInfo-loraSNR : 9
  txInfo-dataRate-spreadFactor : 7
  rxInfo-altitude : 0
  txInfo-frequency : 868300000
  data-temperature-alarmL : false
  data-atmosphericPressure-unit : hPa
  data-airHumidity-unit : %
  data-temperature-unit : °C
  data-atmosphericPressure : 967.76
  data-acceleration-unit : G
  data-acceleration-axisZ : -0.271
  data-temperature-alarmH : false
  data-airHumidity : 36
  data-temperature : 19.4
  data-acceleration-axisY : 0.021
  data-acceleration-axisX : -0.896
  
```



### Flux



elasticsearch



kibana

### Fichiers (au mesocentre)

#### Json Elasticsearch

```

1 {"index":{"_id":"1"}}
2 {"data-illuminance-unit": "lux", "txInfo-codeRate": "4/5", "fp
node-timestampUTC": "2018-10-02 07:10:13", "txInfo-adr": f
"servertimestampUTC": "2018-10-02T10:54:21.402434", "txInf
bandwidth": 125, "data-node-batteryVoltage-alarmIsLow": fa
dataRate-modulation": "LORA", "data-DataChannel": 1, "rxIn
fCnt": 0, "deviceName": "HE31294", "applicationName": "Co
rxInfo-name": "Kerlink01", "rxInfo-rssi": -26, "devEUI":
"434e535301e31294", "data-illuminance": 340, "applicationI
mac": "7276ff00b030f4c", "data-node-batteryVoltage": 3.92
batteryVoltage-unit": "V", "rxInfo-latitude": 0, "rxInfo-l
txInfo-dataRate-spreadFactor": 7, "rxInfo-altitude": 0, "
868300000}
3 {"index":{"_id":"2"}}
  
```

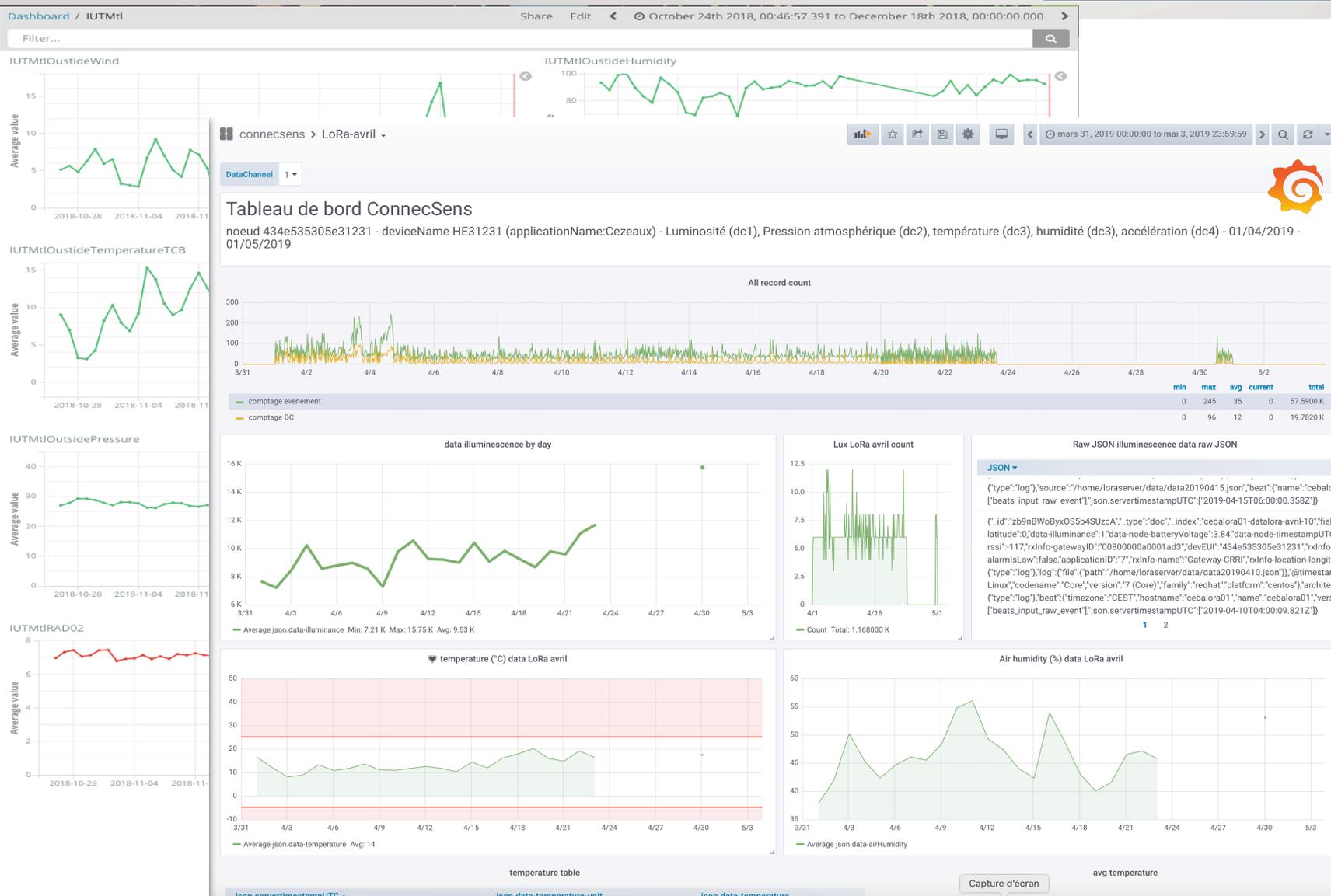
Table [JSON](#)

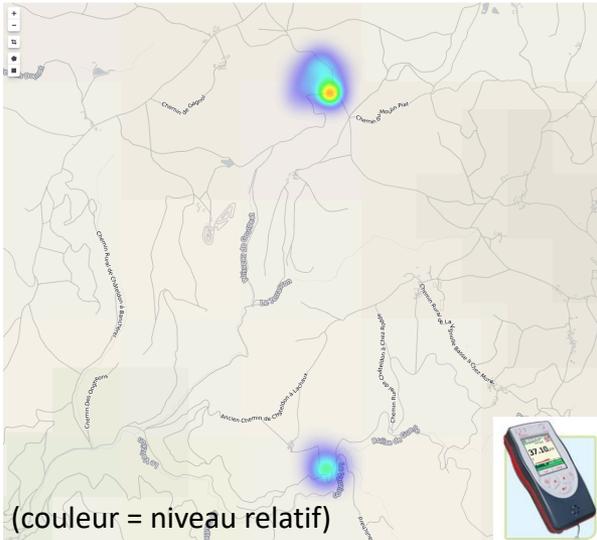
|                                 |                          |
|---------------------------------|--------------------------|
| t _id                           | Q Q * * 14               |
| t _index                        | Q Q * * flattened_4_1    |
| # _score                        | Q Q * * 1                |
| t _type                         | Q Q * * type4            |
| t applicationID                 | Q Q * * 1                |
| t applicationName               | Q Q * * ConneCSens-Nodes |
| # data-DataChannel              | Q Q * * 1                |
| # data-airHumidity              | Q Q * * 40.5             |
| t data-airHumidity-unit         | Q Q * * %                |
| # data-atmosphericPressure      | Q Q * * 982.14           |
| t data-atmosphericPressure-unit | Q Q * * hPa              |
| # data-illuminance              | Q Q * * 251              |
| t data-illuminance-unit         | Q Q * * lux              |



# Réseau de capteurs sans fils

## Monitoring/Dashboard avec Grafana / Kibana

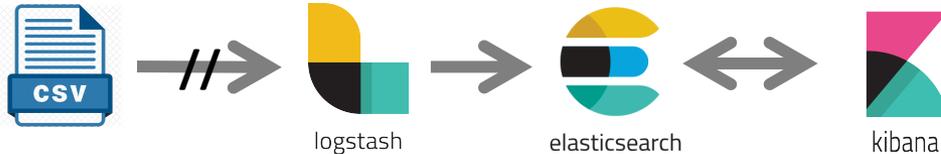





## Data description

Données d'un Camberra Colibri (csv) :

- GPS pour chaque mesure
- Mesure de radioactivité



## Requête sur les coordonnées GPS possibles

GET gamma\_colibri\_rophin/\_search

```
{
  "query": {
    "bool": {
      "must": {
        "match_all": {}
      },
      "filter": {
        "geo_polygon": {
          "location": {
            "points": [
              [3.5459891613159016, 46.01295812457846],
              [3.5592929180297688, 46.01295812457846],
              [3.5592929180297688, 46.002704419144706],
              [3.5459891613159016, 46.002704419144706],
              [3.5459891613159016, 46.01295812457846]
            ]
          }
        }
      }
    }
  }
}
```



# Réseau de capteurs sans fils

## Traitement de flux de données capteurs



json

Transport

[I]

Pipeline

[O]

Index

[F]



```
type: log
# Change to true to enable this input configuration.
enabled: true
# Paths that should be crawled and fetched. Glob based paths:
paths:
  - /home/l.../data/201904*
#json.keys_under_root: true
#json.add_error_key: true
fields:
  log_type: datalora-avril
#json.keys_under_root: true
#json.add_error_key: true
processors:
  - add_locale:
      format: abbreviation
```

```
input {
  beats {
    port => 5042
    ssl => true
    ssl_certificate => "/etc/logstash/.../logstash.crt"
    ssl_key => "/etc/logstash/.../logstash.key"
    ssl_verify_mode => "force_peer"
    ssl_certificate_authorities => [".../logstash-ca.crt"]
  }
}
filter {
  if [type] == "log" {
    grok {
      match => { "message" => "%{SYSLOG}" }
      add_field => [ "received_at", "%{date}" ]
      add_field => [ "received_from", "%{host}" ]
      remove_field => ["message"]
    }
    syslog_pri { }
    date {
      match => [ "syslog_timestamp", "%{SYSLOG_TIMESTAMP}" ]
    }
  }
}
output {
  if [fields][log_type] == "message" {
    elasticsearch {
      hosts => ["192.168.1.92:9200"]
      index => "cebalora01-message-%{date}"
    }
  }
  if [fields][log_type] == "test-data" {
    elasticsearch {
      hosts => ["192.168.1.92:9200"]
      index => "cebalora01-test-data-%{date}"
    }
  }
  if [fields][log_type] == "datalora-avril" {
    elasticsearch {
      hosts => ["192.168.1.92:9200"]
      index => "cebalora01-datalora-avril-%{date}"
    }
  }
}
```

cebalora01-data\*

Time Filter field name: json.servertimestampUTC

This page lists every field in the cebalora01-data\* index and the field's associated type as recorded by Elasticsearch. To change a field type, use the Elasticsearch Mapping API.

Fields (171) | Scripted fields (0) | Source filters (0)

Filter

| Name             | Type    | Format | Search... | Aggreg... |
|------------------|---------|--------|-----------|-----------|
| @timestamp       | date    |        | •         | •         |
| @version         | string  |        | •         |           |
| @version.keyword | string  |        | •         | •         |
| _id              | string  |        | •         | •         |
| _index           | string  |        | •         | •         |
| _score           | number  |        |           |           |
| _source          | _source |        |           |           |



- Base POSTGRES pour le stockage de données JSON
- Objectif : vérification cohérence, requêtes plus complexe
- Intégration (Automatisation en cours)

Récupérer la liste des sites par fichier JSON

```
select
  json_file.filename,
  json_row.data ->> 'applicationName' as
nom
from
  json_row, json_file
where
  json_row.json_file_id = json_file.id
group by
  json_file.filename, nom
;
```

|    | filename<br>text  | nom<br>text   |
|----|-------------------|---------------|
| 1  | data20190401.json | Cezeaux       |
| 2  | data20190401.json | Jardin-IRSTEA |
| 3  | data20190401.json | Montoldre     |
| 4  | data20190401.json | Radon-Sensors |
| 5  | data20190402.json | Cezeaux       |
| 6  | data20190402.json | Jardin-IRSTEA |
| 7  | data20190402.json | Montoldre     |
| 8  | data20190402.json | Radon-Sensors |
| 9  | data20190403.json | Cezeaux       |
| 10 | data20190403.json | Jardin-IRSTEA |
| 11 | data20190403.json | Montoldre     |
| 12 | data20190403.json | Radon-Sensors |
| 13 | data20190404.json | Cezeaux       |
| 14 | data20190404.json | Jardin-IRSTEA |
| 15 | data20190404.json | Montoldre     |
| 16 | data20190404.json | Radon-Sensors |

Récupérer la liste des Dev-EUI par application-name

|    | nom<br>text | deveui<br>text   |
|----|-------------|------------------|
| 1  | Cezeaux     | 434e535305e36210 |
| 2  | Cezeaux     | 434e535305e31297 |
| 3  | Cezeaux     | 434e535305e31273 |
| 4  | Cezeaux     | 434e535305e36250 |
| 5  | Cezeaux     | 434e535305e36226 |
| 6  | Cezeaux     | 434e535305e31233 |
| 7  | Cezeaux     | 434e535305e36244 |
| 8  | Cezeaux     | 434e535305507403 |
| 9  | Cezeaux     | 434e535305e31263 |
| 10 | Cezeaux     | 434e535305e31291 |
| 11 | Cezeaux     | 434e535305e31279 |
| 12 | Cezeaux     | 434e535305e31231 |
| 13 | Cezeaux     | 434e535305e31230 |
| 14 | Cezeaux     | 434e535305e31277 |
| 15 | ETNA        | 434e535302e36236 |
| 16 | ETNA        | 434e535302e36241 |

Afficher l'évolution du nombre de trames par devEUI par jour

|    | deveui<br>text   | dateserver<br>date | nbtrames<br>bigint |
|----|------------------|--------------------|--------------------|
| 1  | 434e535300e31208 | 2019-04-01         | 24                 |
| 2  | 434e535300e31216 | 2019-04-01         | 861                |
| 3  | 434e535300e31235 | 2019-04-01         | 240                |
| 4  | 434e535301e36231 | 2019-04-01         | 6                  |
| 5  | 434e535303e36199 | 2019-04-01         | 154                |
| 6  | 434e535303e36212 | 2019-04-01         | 154                |
| 7  | 434e535305e31230 | 2019-04-01         | 232                |
| 8  | 434e535305e31231 | 2019-04-01         | 230                |
| 9  | 434e535305e31233 | 2019-04-01         | 232                |
| 10 | 434e535305e31273 | 2019-04-01         | 232                |
| 11 | 434e535305e31277 | 2019-04-01         | 232                |
| 12 | 434e535305e31279 | 2019-04-01         | 232                |
| 13 | 434e535305e31291 | 2019-04-01         | 80                 |
| 14 | 434e535305e31297 | 2019-04-01         | 170                |
| 15 | 434e535300e31208 | 2019-04-02         | 24                 |
| 16 | 434e535300e31216 | 2019-04-02         | 1105               |



On peut :

- Voir projet/expérience
- Rechercher par mot clé
- Voir des zones géographiques

Un utilisateur peut créer:

- Un projet
- Un site
- Une expérience

## Éditer le Projet

...

### INFORMATIONS

|   |   |   |
|---|---|---|
| Nom du projet   | Statut  | Financement   |
| <input type="text" value="REseau de Suivi des Eaux en AUvergne"/> | <input type="text" value="Actif"/>                                      | <input type="text" value="Ville de Clermont-Ferrand"/><br><input type="text" value="Agence de l'Eau Loire-Bretagne"/><br><input type="text" value="Observatoire de Physique du Globe"/><br><input type="text" value="Fédérations des Recherches en Environnement"/> |
| Date de création du projet  | Organisme porteur   | Site web  |
| <input type="text" value="01 / 01 / 2014"/>                       | <input type="text" value="Observatoire de Physique du Globe de Clerm"/> | <input type="text" value="http://wwwobs.univ-bpclermont.fr/SO/auverwat"/>   |

### IMAGE

|  |
|--|
| Image  |
| <input type="text" value=""/><br><input type="button" value="Browse"/> |

### DESCRIPTION DU PROJET

Résumé

Ce projet, initialement appelé RESEAU, devenu AUVER-WATCH, est prévu pour un fonctionnement de 2014 à 2024 et s'attache à caractériser les masses

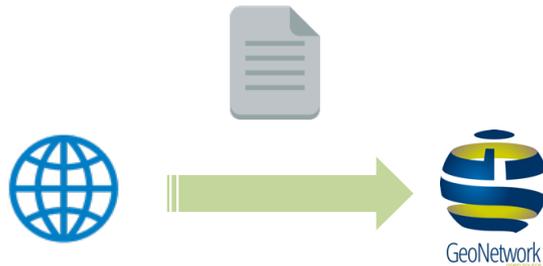
Description

**Auverwatch**

L'hydrologie au sens large est définie par ses acteurs comme la discipline consacrée à l'étude du cycle de l'eau pris sous les angles physiques, physico-chimiques et biologiques, et à l'étude des interactions entre le cycle de l'eau et les activités humaines dans ce milieu. Ses applications sont multiples tant du point de vue de la gestion de l'environnement que de son implication dans les activités humaines : production d'eau potable ou industrielle, géotechnique, production d'énergie... Les pressions qui pèsent ainsi sur les ressources en eau sont de plus en plus fortes, principalement en raison des activités humaines - notamment l'urbanisation, la croissance démographique, l'augmentation du niveau de vie, la concurrence croissante pour les ressources en eau et la pollution. De plus, le changement climatique et les variations des conditions naturelles viennent aggraver ces pressions. Dans ce contexte, l'enjeu majeur est de parvenir à une gestion du cycle de l'eau conforme aux exigences d'un

Un utilisateur peut :

- Créer un jeu de données
  - Metadonnée/ Thesaurus



Site Web CEBA

## Import de Métadonnées

### Informations obligatoires

Status

completed

Titre

Mesures d'activités radiologiques de source

Résumé

Mesure par spectrométrie de l'activité  
en radioéléments d'origine naturelle  
(chaînes de décroissance de

Ressource type

dataset

Type de représentation spatiale

vector

Date de début

31 / 12 / 2018

Date de fin

03 / 07 / 2019

Nom de l'étendue spatiale

Lachaux

Catégories de sujets

environment, geoscientificinformation

- Minimum: Quoi, Quand et Où
- Flux de données de capteurs
  - Potentiellement extraites automatiquement
  - Difficultés pour rassembler l'ensemble de données
  - Non auto-descriptive
- Jeu de données avec fichiers
  - Ajout manuel
- Fiche de métadonnées épurée du Réseau des Zones Atelier<sup>1</sup>
  - Utilisation des thésaurus Envthes, Gemet
- Saisie : site web
- Intégration : pipeline  du Réseau des Zones Atelier<sup>1</sup>

<sup>1</sup>Projet Bancariser les Données Ensemble



backTohome

Précédent

Suivant

Télécharger

Mode affichage

### Mesures d'activités radiologiques des sources minérales ZATU

Mesures par spectrométrie gamma de l'activité en radioéléments d'origine naturelle (chaînes de décroissance de l'Uranium238 et du Thorium 232) dans l'eau et les sédiments des sources minérales radioactives en Auvergne. Contributeurs : LPC

completed

#### À propos de cette ressource

##### Catégories

Jeux de données

Environnement

Informations géoscientifiques

GEMET - INSPIRE themes, version 1.0

- ressources minérales
- lieux de production et sites industriels
- installations de suivi environnemental

GEMET

- eau de source
- radioactivité naturelle
- rayonnement gamma
- dose équivalente

Langue

- fra

Identificateur de ressource unique

- ZATU\_a96c0982-d71f-452e-8c4c-fc41322f8ae8

Contraintes légales

This work is licensed under a Creative Commons Attribution 4.0 License (CC BY SA 4.0, <https://creativecommons.org/licenses/by-sa/4.0/>).

Aucune évaluation

Voir tous les commentaires

Ajouter votre commentaire

Étendue spatiale

Étendue spatiale

Étendue temporelle

Date de création

2018-12-31

Date de révision

2019-07-03

Fourni par



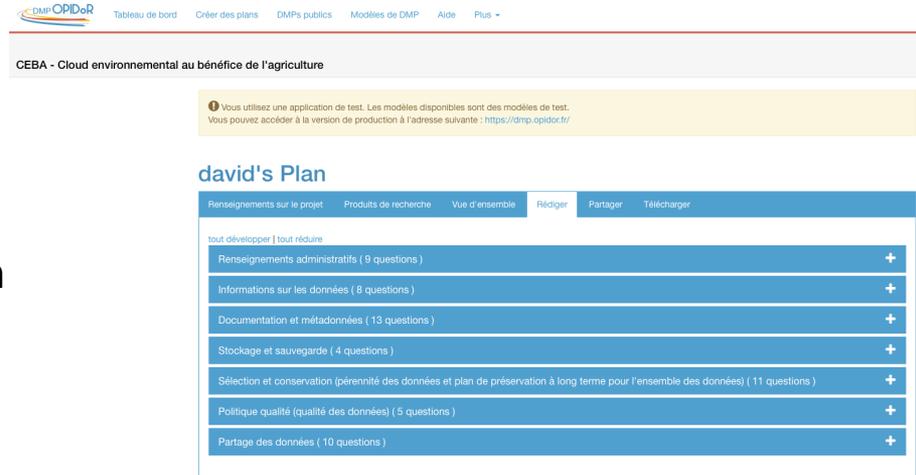
Mis à jour :

il y a 2 mois

Partager



- Conception d'un DMP pour le CEBA
  - Orienté jeu de données
  - Éléments principaux
    - Volumétrie
    - Durée/accès stockage
    - Partage
- Difficultés : très grande diversité d'interlocuteur
- Mise en œuvre
  - Initiale : fichier Word
  - En cours : à l'aide de 
    - Modèle créé en cours de validation



The screenshot shows the DMP OPIDoR web interface. At the top, there is a navigation bar with links: "Tableau de bord", "Créer des plans", "DMPs publics", "Modèles de DMP", "Aide", and "Plus". Below the navigation bar, the page title is "CEBA - Cloud environnemental au bénéfice de l'agriculture". A yellow warning box states: "Vous utilisez une application de test. Les modèles disponibles sont des modèles de test. Vous pouvez accéder à la version de production à l'adresse suivante : <https://dmp.opidor.fr/>". The main content area is titled "david's Plan" and has a sub-navigation bar with "Rédiger", "Partager", and "Télécharger". Below this, there is a list of questions to be answered, each with a plus sign to expand it:

| Question  | Statut |
|---|--------|
| Renseignements administratifs ( 9 questions )   | +      |
| Informations sur les données ( 8 questions )  | +      |
| Documentation et métadonnées ( 13 questions )   | +      |
| Stockage et sauvegarde ( 4 questions )  | +      |
| Sélection et conservation (pérennité des données et plan de préservation à long terme pour l'ensemble des données) ( 11 questions ) | +      |
| Politique qualité (qualité des données) ( 5 questions )   | +      |
| Partage des données ( 10 questions )  | +      |



# Premiers retours

| Positif   | Amélioration   | Perspectives  |
|---|--|---|
| Données brutes<br>Ingestion                     | Paramétrage du nœud/passerelle (nom expérience, typage de données...) et metadonnées associées | Automatisation de récupération d'informations des fichiers de configuration<br>Saisies site Web |
| « Temps réel » : assistance terrain, réactivité | Alertes, annotations   | Automatisation / description  |
| Prise en main Grafana                           | Formations   | Combinaison avec Kibana   |
| Indexation                                      |  |   |
| Requêtage                                       |  | Indexation multiple / usage   |
| Base de données : identifications de problèmes  | Intégration base de données de descriptions (par ex. capteurs)                                 | Automatisation  |
|   |  | Intégration du traitement des fichiers  |



- Architecture de préproduction
  - Projet /expérience pour créer des jeux de données → geonetwork
- La gestion des données :
  - Réseau de capteurs sans fils
    - 5 réseaux en ligne (Auzon, Aydat, Cézeaux, Etna, Montluçon)
    - Stockage dans une base de données JSON (travail en cours)
  - Fichiers (testé depuis geonetwork)
  - Base de données (travail en cours)
- Management des métadonnées (→ automatique)
  - Travail en cours / reseau de capteurs
- Solutions Open source



- Alexandre Claude Suite elastic & architecture
- Gilles Mailhot Coordination scientifique
- Jérémy Mezhoud Base de données
- Francis Ogereau Web, data-catalog
- David Sarramia Coordination & Suite elastic
- Estelle Théveniaud Geolab
- Antoine Mahul Mésocentre
- David Grimbichler Mésocentre



