

Mohamad ALI

# LA VULNÉRABILITÉ DU BÂTI

par l'Exploitation de l'OPEN DATA

15/05/2023



## ■ Sommaire :

- Introduction la vulnérabilité du bâti
- Base de données disponibles / Création de la base d'étude
- Validation de la qualité des données
- Rehaussement des données manquantes
- Identification des critères de vulnérabilité
- Création et opérationnalisation des critères de vulnérabilité

# INTRODUCTION

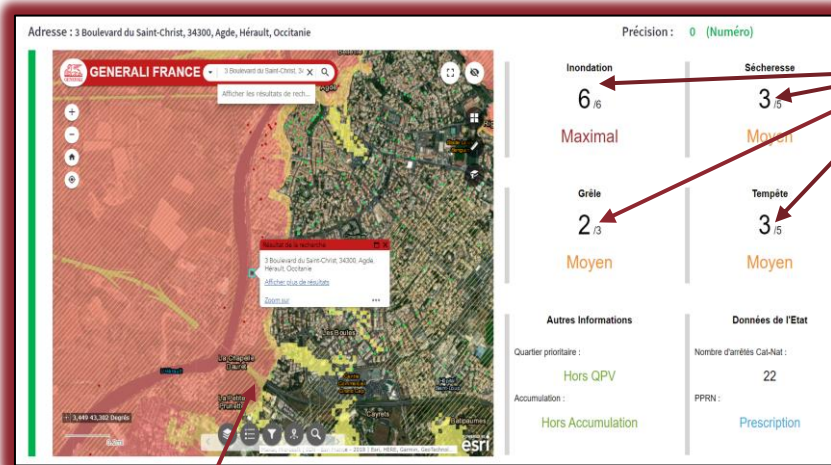
# Etude de la vulnérabilité du bâti

Ce que le GCL a accompli sur l'**aléa**...



Climate apps

La géointelligence et l'expertise climatique à votre portée.



Niveau d'exposition aux aléas climatiques (score)

Cartographie des zones exposées aux aléas climatiques (zonier)

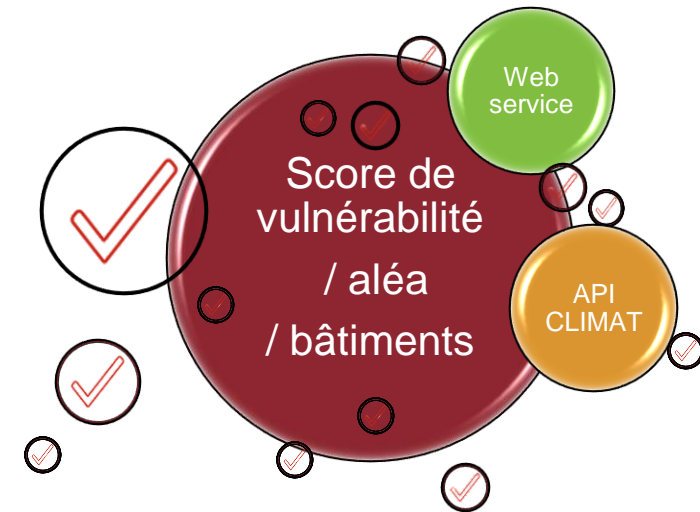
D'autres facteurs de sinistralité qui s'additionnent à l'aléa



**Grêle 2022, facteur de vulnérabilité : les toitures industrielles**

RI - Toitures en **fibrociment amianté**

..la prochaine frontière technique à franchir est la **vulnérabilité du bâti**



Solution et cas d'usage identique au score d'aléa

Surveillance/  
Renouvellement

Ciblage des contrats exposés :

- Renouvellement ajusté
- Derisking

Souscription  
/Tarif

Souscription avertie :

- Climate Apps

Tarification juste :

- ApiFue via WebService ArcGIS

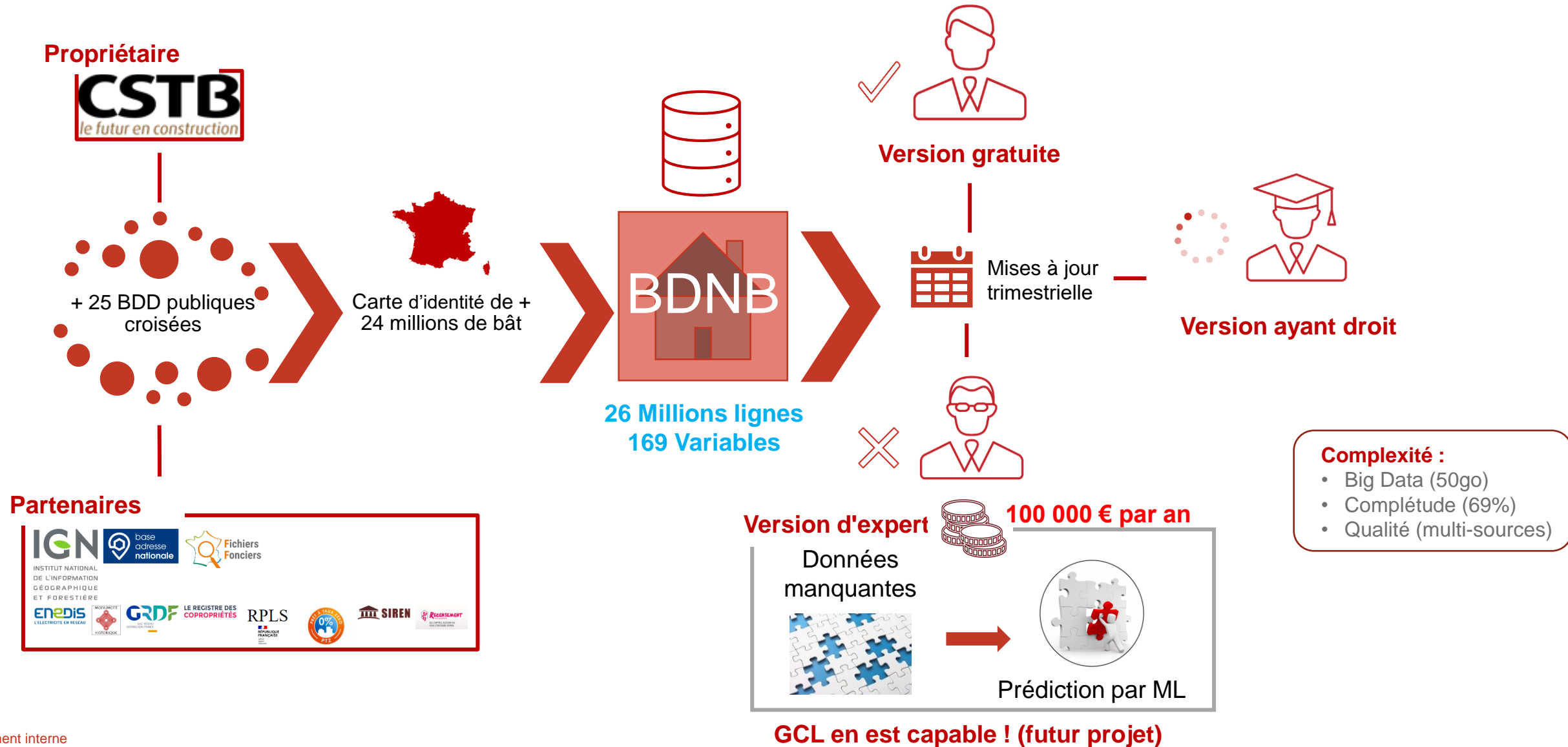
# BASE DE DONNÉES DISPONIBLES CRÉATION DE LA BASE D'ÉTUDE

Script :

- Shape (étudiant de Nantes) → python
- Orientation du toit → python
- Nature du toit → python
- BDNB, BD TOPO, Parcelle → R

# Etude de la vulnérabilité du bâti

Une **carte d'identité** pour plus de 24M de **bâtiments**...



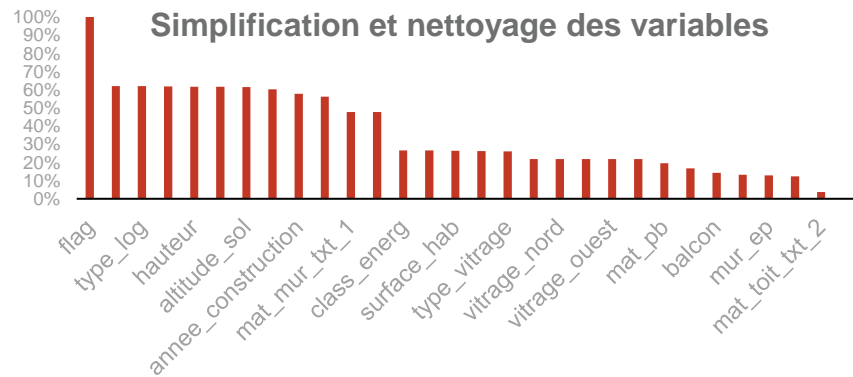
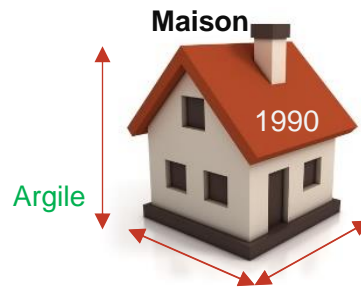
# Etude de la vulnérabilité du bâti

...et une multitude de **données à exploiter** conjointement...

## 1 Caractéristiques du bâti (BDNB)

### Classification & choix des variables

- Type de bâtiment
- Caractéristique du bâtiment
- Type de construction
- Sol/environnement



## 2 Données images aériennes (BD ORTHO)



- Orientation du toit
- Type de toit (industriel)
- Panneau solaire
- Piscine
- Proximité des arbres (en cours)



## 3 Données Marché (BD Adresses)

- 28 millions d'adresses dans notre géocodeur.



- Coordonnées GPS

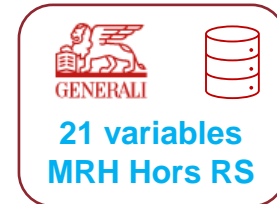


## 4 Données contractuelles (Generali)

- Contrats 2010 – 2022
- Sinistres Climatiques 2010 - 2022

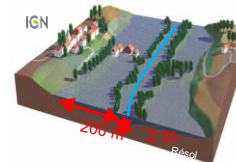


- Primes / Charges
- Type d'occupation / bâtiment
- Date souscription / résiliation / sinistre
- Coordonnées GPS



## 5 Données risques (GCL)

- Zoniers Climatiques
- Zoniers Hors Climatiques



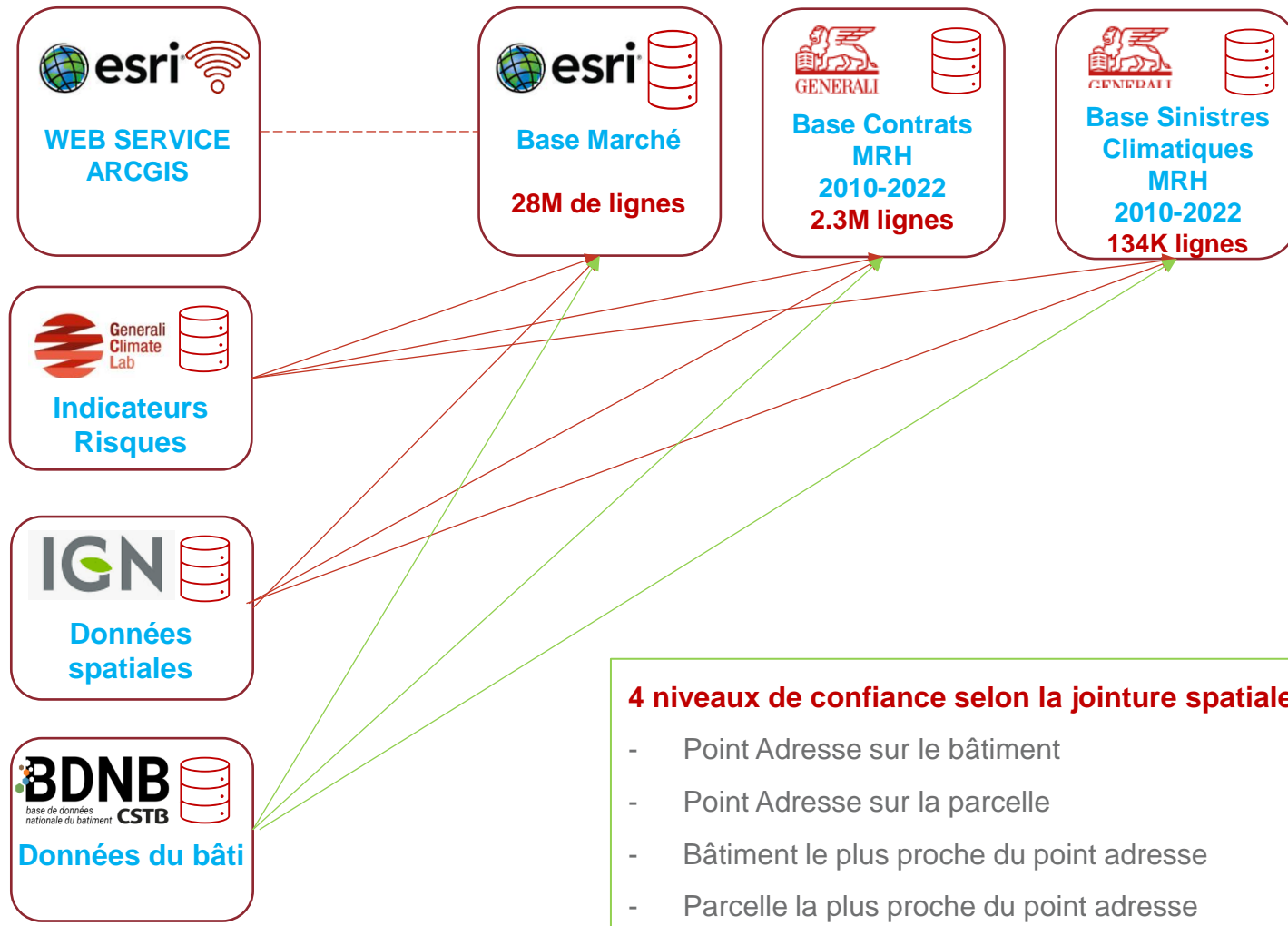
- V2 Inondation
- Sécheresse
- Grêle (bientôt)
- Zoniers DDE, VOL, INC





# Etude de la vulnérabilité du bâti

...pour en faire une **base de données GCL à 3 visions...**



## Cas d'usage :

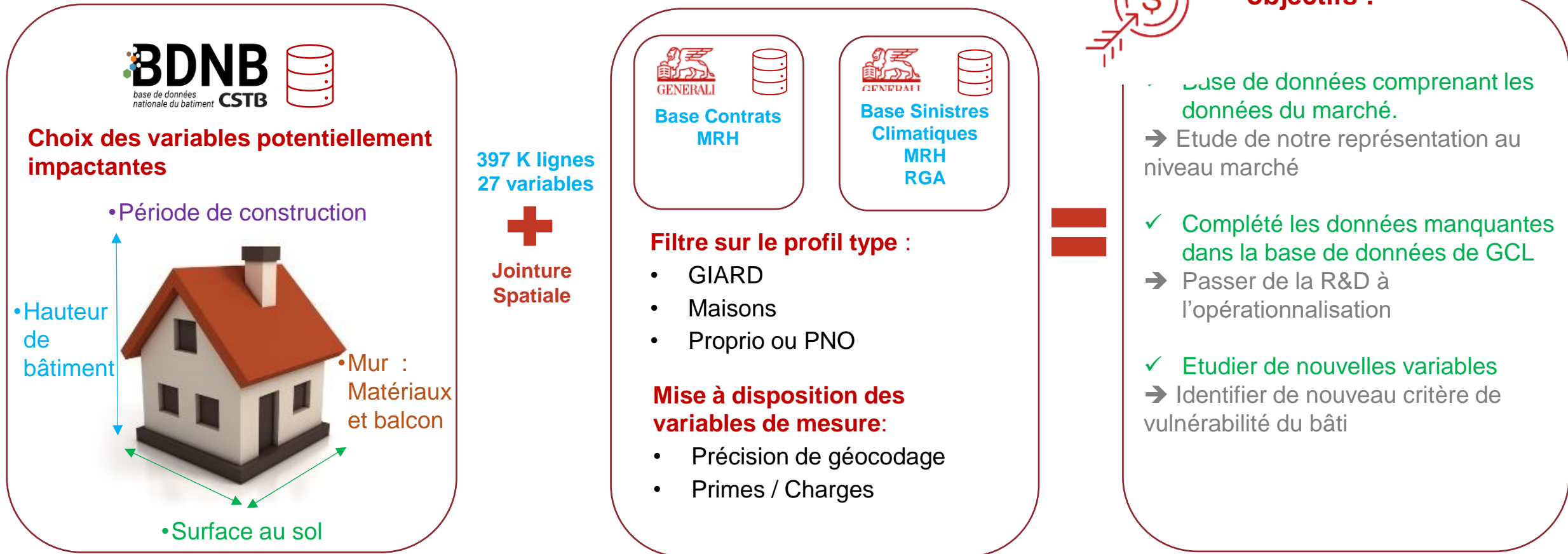
- **R&D** : Identification des critères de vulnérabilité
  - ➔ Machine Learning
  - ➔ Alimentation des travaux de **thèse**
- **Opérationnalisation** : requêtable dans nos systèmes via WebService ArcGIS
  - ➔ Tarification (Spread Tarifaire)
  - ➔ Souscription (Climate Apps)
- **Mesure** : gagner en précision sur les différents Kpis
  - ➔ S/P robuste (2010 à 2022)
  - ➔ Comparaison de notre exposition face au marché

# IDENTIFICATION DES CRITÈRES DE VULNRÉRABILITÉ DU BATÎ

# Etude de la vulnérabilité du bâti

Alimentation des travaux de thèse : RGA

**BDD** qui permet de faire de la **R&D** :



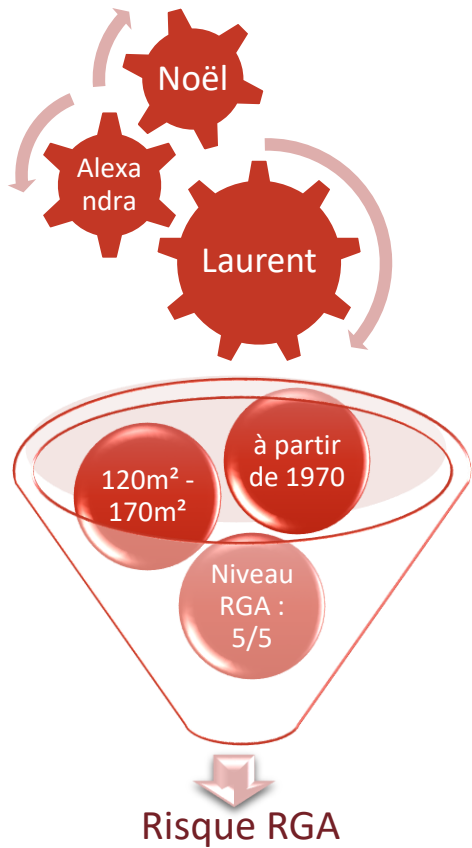
**Base de Données sur-mesure pour la thèse RGA**

# Etude de la vulnérabilité du bâti

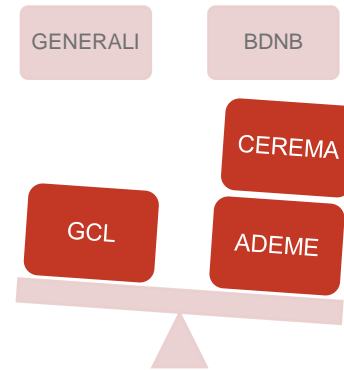
## Alimentation des travaux de thèse : RGA

**BDD** qui permet de passer de la **R&D** à l'**opérationnalisation**..

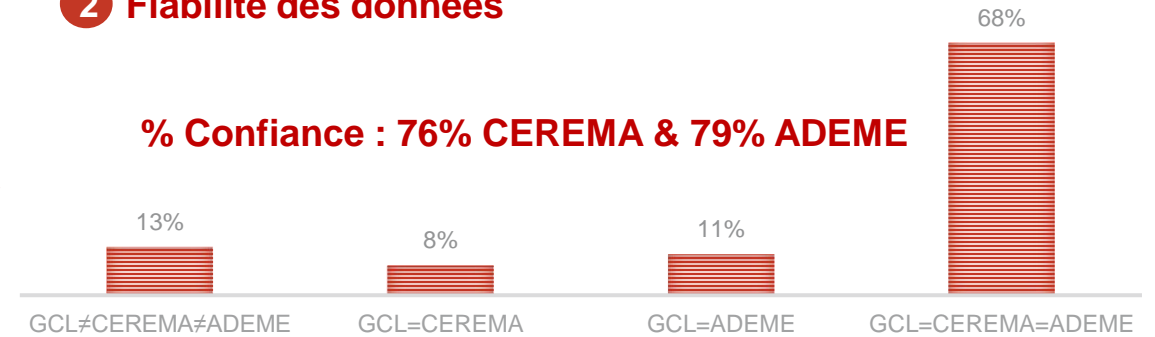
### Travaux de R&D sur le RGA



### 1 Identification des données



### 2 Fiabilité des données



### 3 Globalisation/systématisation

#### Solution Affaires en vigueur

Profil		1	2	3	4	5	Total
Année de construction	Surface						
Avant 1970	<120	18 491	19 724	26 653	3 884	3 464	72 216
		7%	7%	10%	1%	1%	27%
Avant 1970	120-170	5 644	6 043	8 168	1 106	1 251	22 212
		2%	2%	3%	0%	0%	8%
Avant 1970	>170	5 757	6 349	8 764	1 317	1 463	23 650
		2%	2%	3%	0%	1%	9%
A partir 1970	<120	14 665	17 665	20 169	3 187	2 714	58 400
		6%	7%	8%	1%	1%	22%
<b>A partir 1970</b>	<b>120-170</b>	<b>10 415</b>	<b>12 829</b>	<b>17 295</b>	<b>2 761</b>	<b>3 698</b>	<b>46 998</b>
		4%	5%	7%	1%	14%	18%
A partir 1970	>170	7 011	7 466	11 782	1 765	2 816	30 840
		3%	3%	4%	1%	1%	12%
NR	NR	2 531	3 037	4 048	540	614	10 770
		1%	1%	2%	0%	0%	4%
<b>Total</b>		<b>64 514</b>	<b>73 113</b>	<b>96 879</b>	<b>14 560</b>	<b>16 020</b>	<b>265 086</b>

Application du profil de risque sur tout le portefeuille MRH GIARD

#### Solution AN :



Climate Apps : GeoScan

#### Solution Prospection :



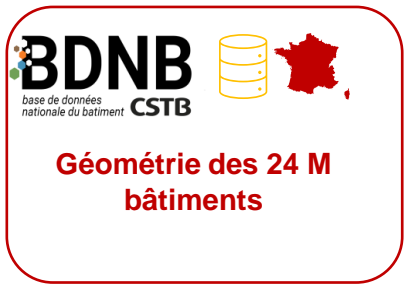
Ciblage des clients en zone 3 RGA non vulnérable



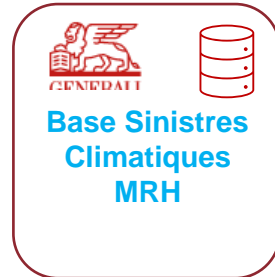
# Etude de la vulnérabilité du bâti

Alimentation des travaux de thèse : RGA

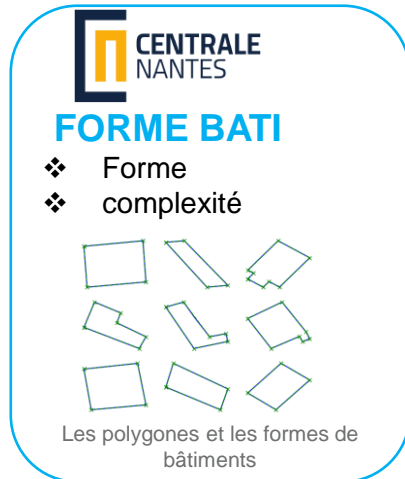
**BDD** qui permet d'identifier de **nouveau critère de vulnérabilité..**



10 995 Lignes  
49 Variables



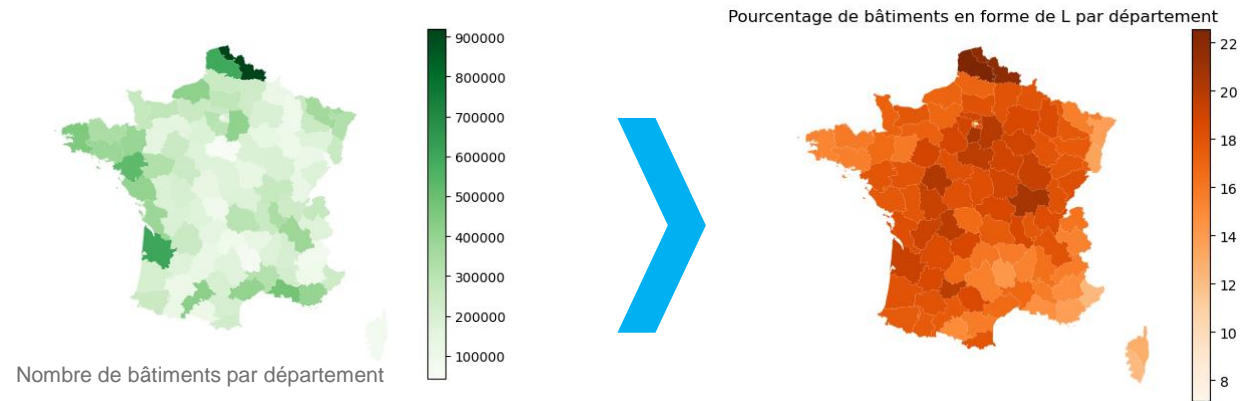
**Jointure Spatiale**



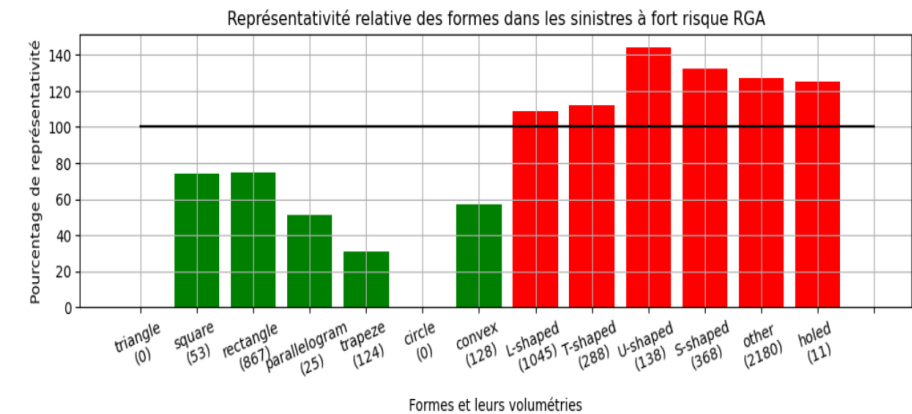
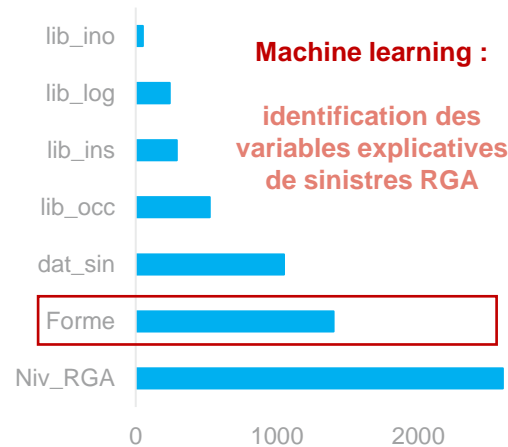
**Bureau d'étude :**

- Etudiant de Centrale Nantes
- Wiz'You
- GCL

## Etude Marché



## Etude de la sinistralité (Generali)



# Etude de la vulnérabilité du bâti

## Alimentation des travaux de thèse : Grêle



### BDD qui permet de faire de la R&D :



• Toiture : Matériaux, isolation...

• Période de construction...

• Mur : Matériaux, isolation et épaisseur...

• Baie : Matériaux, pourcentage, remplissage, orientation et vitrage...



• Surface au sol

398k Lignes  
54 Variables



Jointure  
Spatiale



Panneaux solaires  
Orientation du toit



Jointure  
Spatiale



Base Contrats  
MRH  
2018-2022  
(puis PRO)



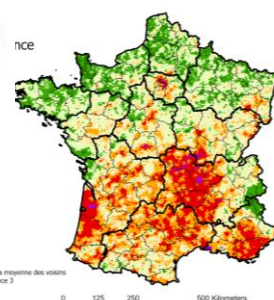
Base Sinistres  
Climatiques  
MRH



Jointure  
Spatiale



Generali  
Climate  
Lab



Aléa

Cartographie de l'aléa (Zonier grêle)



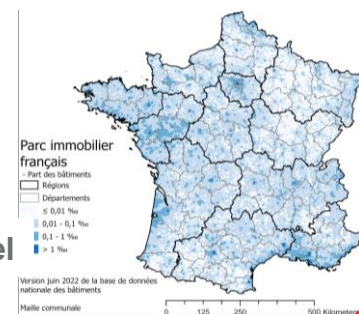
Identifier des critères de  
vulnérabilité du bâti face à l'aléa  
grêle

Exposition marché de l'aléa Grêle :

Exposition



Risque potentiel



# QUALITÉ DES DONNÉES



# CONTRÔLE COHÉRENCE DES DONNÉES

## BDNB...

### Base de Données Nationale des Bâtiments

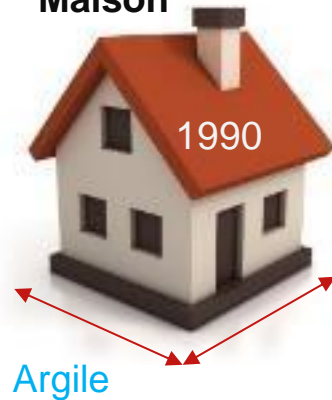


+ 25 BDD publiques croisées (Cerema, IGN, Ademe)

Carte d'identité de + 24 millions de bâtiments en France (dont 19 millions de maisons)

#### Maison

- Type de bâtiment
- **Caractéristique du bâtiment**
- Type de construction
- **Sol/environnement**



- Utilisation de :
- l'année de construction
  - La surface au sol

### Croisement des données **portefeuille** X **BDNB**



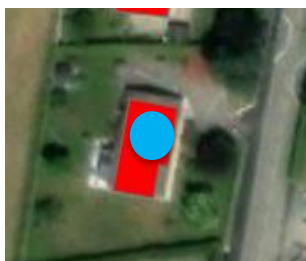
Empreinte  
géométrique



Bâtiments +  
Parcelle



#### Jointure spatiale : 4 niveaux de confiance



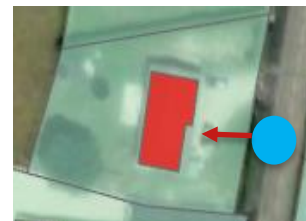
Niveau 1 : point sur le bâti



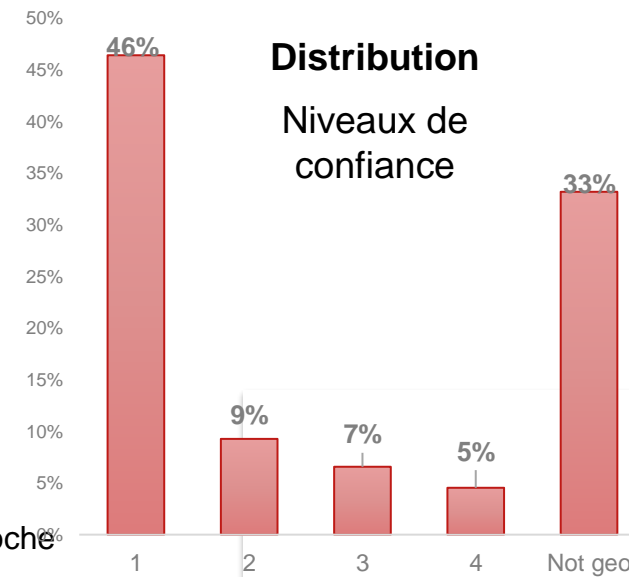
Niveau 3 : bâti le plus proche



Niveau 2 : point sur la parcelle



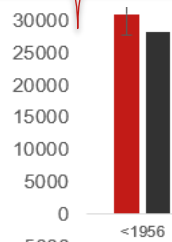
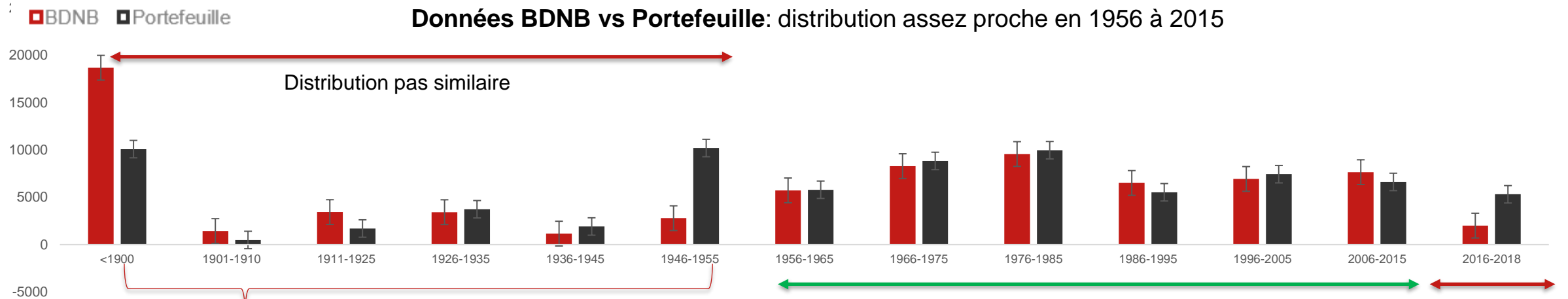
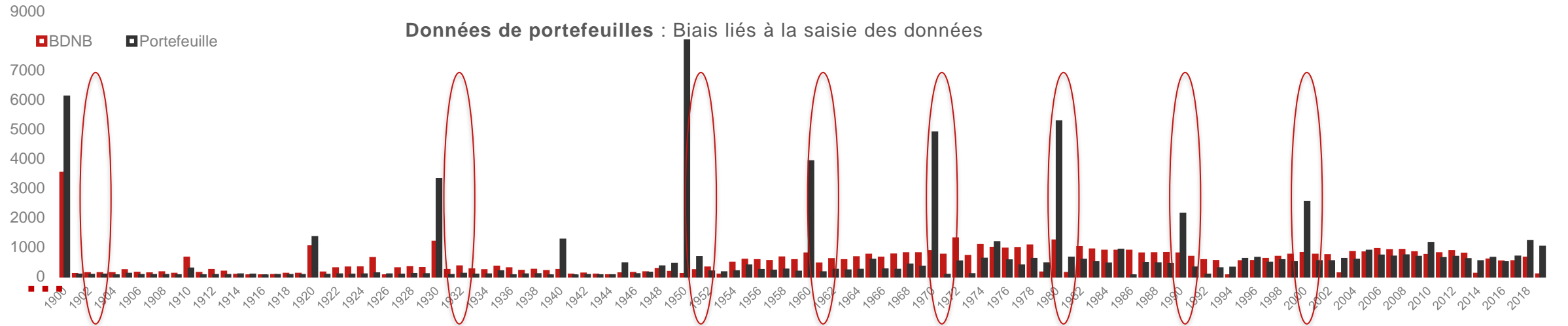
Niveau 4 : parcelle la plus proche





# CONTRÔLE COHÉRENCE DES DONNÉES

## ANNÉE DE CONSTRUCTION...

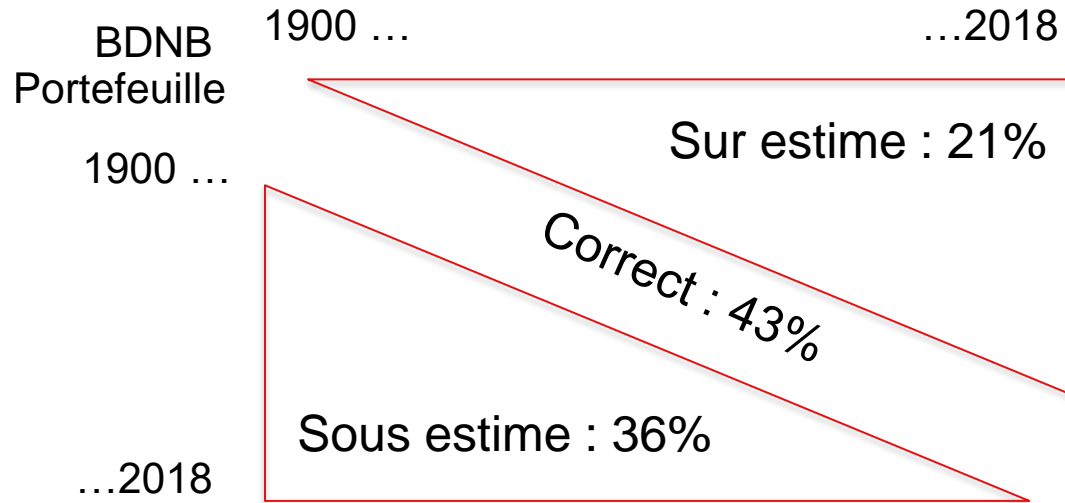


→ Distribution plus similaire après agrégation

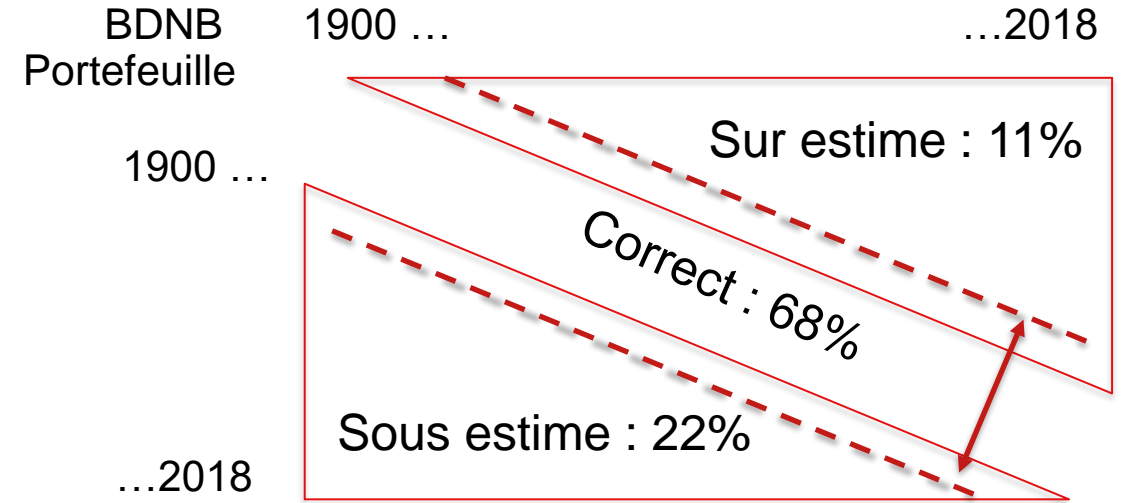
# CONTRÔLE COHÉRENCE DES DONNÉES

## ANNÉE DE CONSTRUCTION...

Données BDNB vs Portefeuille : simple



Données BDNB vs Portefeuille : +/- moyenne d'erreur (2 intervalles)



Données BDNB vs Portefeuille : métriques

**43%** taux de matching

**2** intervalles d'erreur moyen

**43%** avec taux de confiance 1, 2, 3

**44%** avec taux de confiance 1, 2

**45%** avec taux de confiance 1

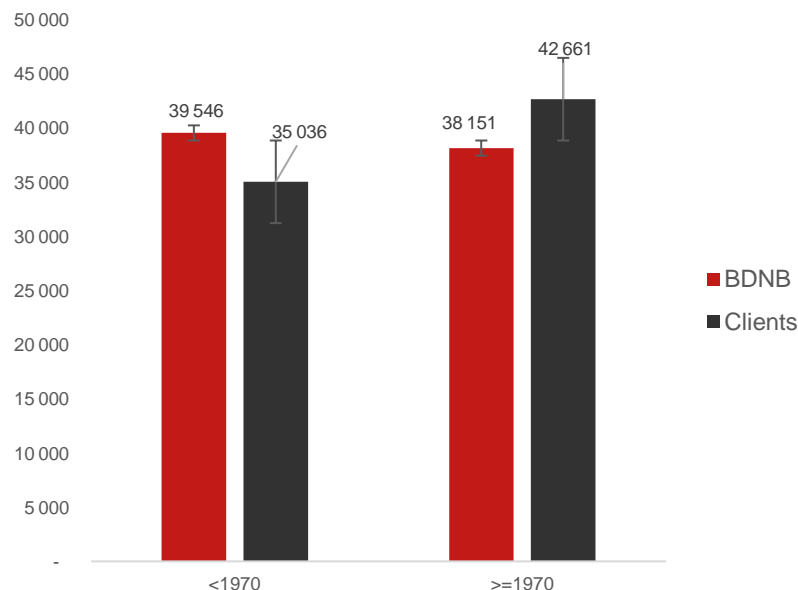
Données BDNB sont...

- Meilleures pour les années de construction récentes
- Sous estime l'année de construction

# CONTRÔLE COHÉRENCE DES DONNÉES

## ANNÉE DE CONSTRUCTION...

### Vision Tarif : deux intervalles



### Données BDNB :

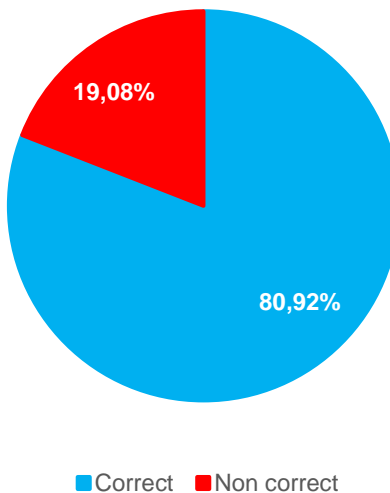
- Sur estimation des années anciennes
- Sous estimation des années récentes

**81%** taux de matching

**82%** avec taux de confiance 1

**82%** avec taux de confiance 1, 2

**81%** avec taux de confiance 1, 2, 3



### Année de construction BDNB : origine des données

Source : CEREMA

- « jannat » dans BD du **CEREMA** ;

Variable sur l'année d'achèvement du local présente dans les Fichiers fonciers bruts (comportant des erreurs de déclaration ou d'enregistrement).

- « **jannath** » une version **nettoyée** de « jannat » par CEREMA (certaines valeurs ont été attribuées par défaut).

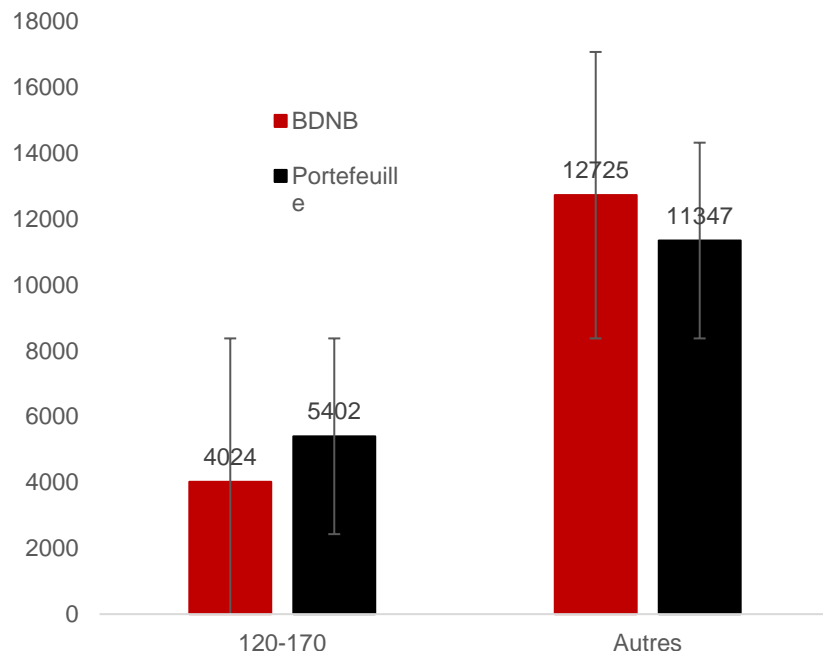
Date des données Cerema : 2020

Echantillon de validation : 77 697 contrats MRH

# CONTRÔLE COHÉRENCE DES DONNÉES

## SURFACE HABITABLE...

### Vision Tarif : deux intervalles



### Données BDNB :

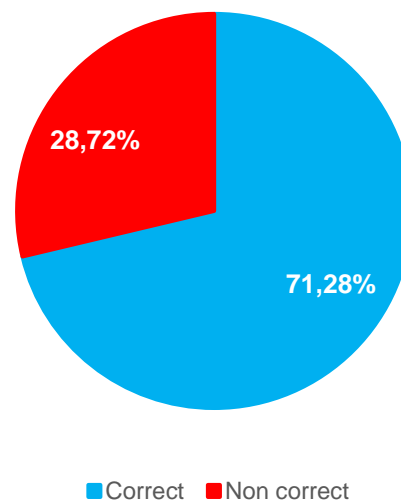
- Sur estimation des années anciennes
- Sous estimation des années récentes

**71.23%** taux de matching

**71.93%** avec taux de confiance 1

**71.79%** avec taux de confiance 1, 2

**71.43%** avec taux de confiance 1, 2, 3



### Année de construction BDNB : origine des données

**Source : CEREMA**

- « jannat » dans BD du **CEREMA** ;

Variable sur l'année d'achèvement du local présente dans les Fichiers fonciers bruts (comportant des erreurs de déclaration ou d'enregistrement).

- « jannath » une version **nettoyée** de « jannat » par CEREMA (certaines valeurs ont été attribuées par défaut).

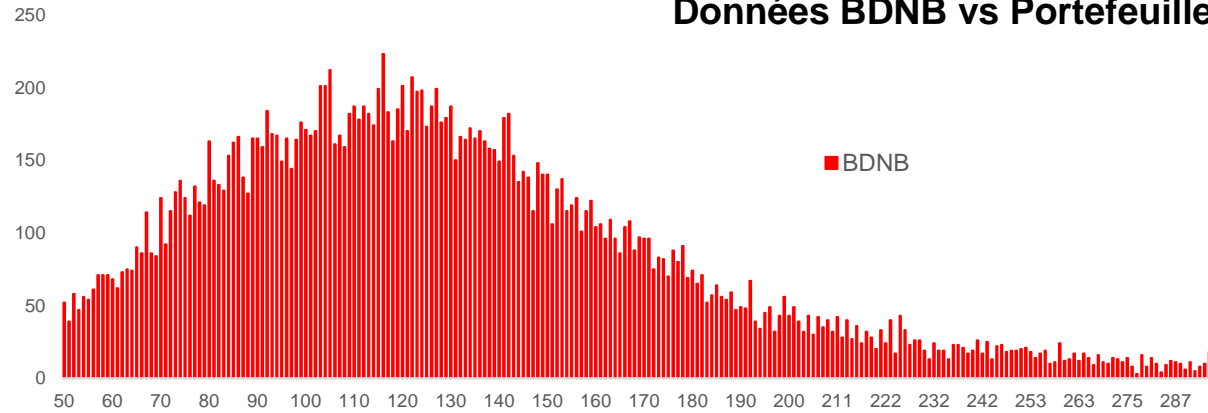
**Date des données Cerema : 2020**

**Echantillon de validation : 77 697 contrats MRH**

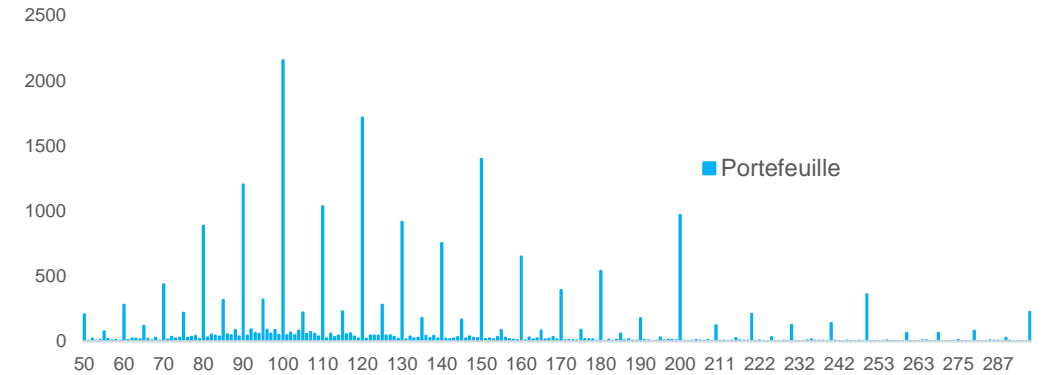
# CONTRÔLE COHÉRENCE DES DONNÉES

## SURFACE AU SOL...

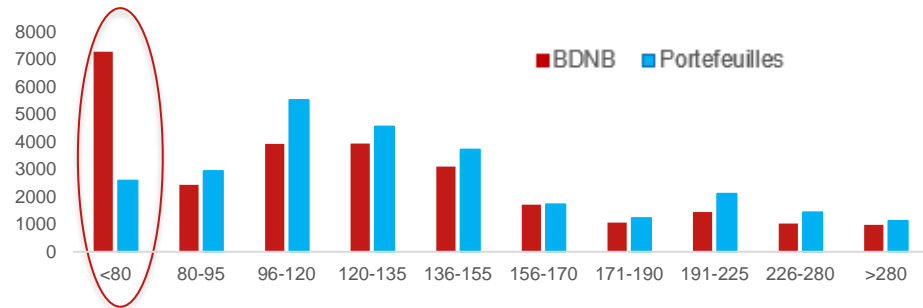
Données BDNB vs Portefeuille: distribution proche



Données BDNB : surface au sol (bâtiment de 1 niveau)



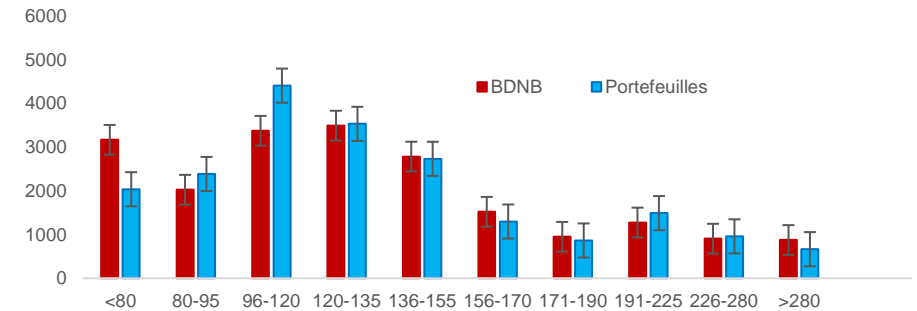
Données Portefeuille : surface habitable (calcul à partir du nombre de pièces)



**Forte différence :** BDNB sur estime les petits bâtiments

Correction ►

Sans la **jointure**  
**parcelle**





## Annexes

**BDNB****INPUT**

- BDNB (169V)  
(Choisir, nettoyage .....)

**OUTPUT**

- **BDNB\_INFO\_CLEAN(59V)**

**Base tarif****INPUT**

- base\_tarif\_mrh\_2021\_v2
- base\_tarif\_mrh\_2021\_v2\_bis
- portfolio\_predictions
- Contrats\_MRH\_absent\_2021\_gcl\_v2\_corr\_gcl  
(Merge, Rbind, Filtres, Corriger .....)

**OUTPUT**

- **base\_tarif\_mrh\_giard\_2010\_2022\_v2**

**Sinistres****INPUT**

- sinclim.csv  
(Merge, Rbind, Filtres, Corriger .....)

**OUTPUT**

- **sinclim\_mrh\_giard\_1998\_2022**

**Sinistres & contrats****INPUT**

- base\_tarif\_mrh\_giard\_2010\_2022\_v2
- sinclim\_mrh\_giard\_1998\_2022
- batiments\_parcelles\_shp\_dept  
(spatial join, Rbind, Filtres, Corriger .....)

**OUTPUT**

- **sinistres\_mrh\_giard\_1998\_2022**
- **contrats\_mrh\_giard\_2010\_2022**

**Sinistres & contrats & market****INPUT**

- Sinistres\_mrh\_giard\_1998\_2022
- Contrats\_mrh\_giard\_2010\_2022
- BDNB\_INFO\_CLEAN
- market.csv  
(spatial join, Rbind, Filtres )

**OUTPUT**

- **contrats\_mrh\_giard\_2010\_2022\_bdnb**
- **sinistres\_mrh\_giard\_2010\_2022\_bdnb**
- **market\_bdnb**

**CALCUL DU TAUX DE REMPLISSAGE****INPUT**

- Sinistres\_mrh\_giard\_1998\_2022
- Contrats\_mrh\_giard\_2010\_2022
- market.csv  
(spatial join, Rbind, Filtres, Corriger .....)

**OUTPUT**

- **contrats\_toML**
- **market\_bdnb\_taux**

# Etude de la vulnérabilité du bâti

## Feuille de route

### Kick Off :

#### Simplification des données :

- Sélection des variables caractéristiques du bâti
- Nettoyage des variables
- Rehaussement des données manquantes

#### RGA (Travail conjoint Noël X Laurent) :

- Fiabilisation des variables explicatives (année de construction, surface..)
- Analyse des données de sinistralité X variables
- Application du profil de risque sur tout le portefeuille MRH GIARD

### Sprint 1 :

#### Données :

- Base de données (simplifiée, nettoyée, rehaussée, fiabilisé) mise à disposition du GCL via une API

#### RGA :

- Proposer un score de vulnérabilité (forme, année et surface)
- Application de la vulnérabilité du bâti dans la tarification

#### Grêle :

- Mise à disposition d'une base de données bâti pour les travaux de grêle

### Sprint 3 :

#### Données :

- Rehaussement des données
- Application d'un score de vulnérabilité pour chaque bâtiment de France selon chaque aléa.

#### Opérationnalisation :

- Mise à disposition du score dans la Climate Apps

### 2024 :

#### Idées :

- Calcul d'un coût de reconstruction pour le Cat Modeling ?

Kick Off : 12/2022

Sprint 1 : 06/2023

Sprint 2 : 09/2023

1<sup>ère</sup> année d'alternance

Start : 09/2022

2023

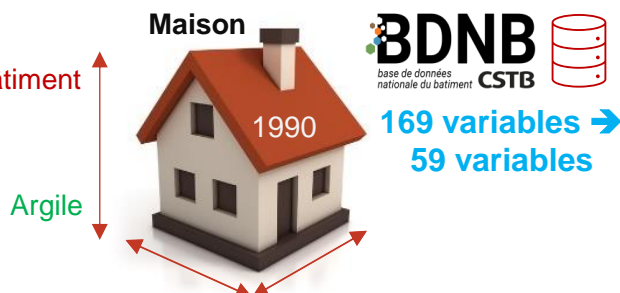
**Un alternant (24 mois) : Data Analyst**



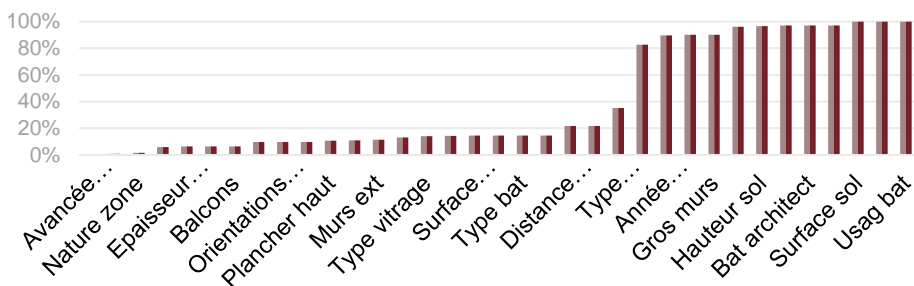
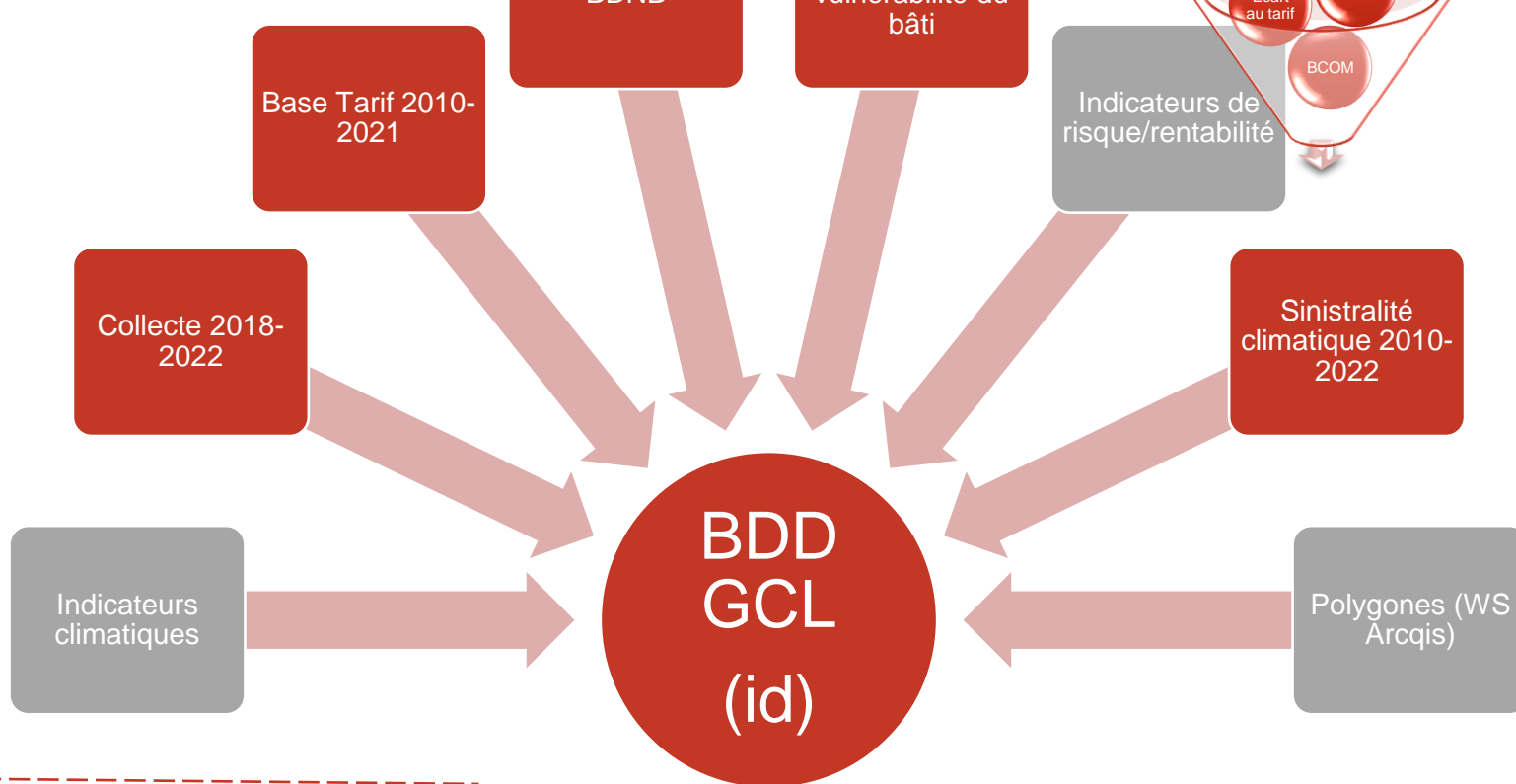
## Nettoyage et retraitement

### 1 Identification des variables pertinentes :

- Type de bâtiment
- Caractéristique du bâtiment
- Type de construction
- Sol/environnement



### 3 Base de données GCL



% Complétude

Source 1 : ADEME

Mur (11%)  
Année (14%)

Retraitement

Nettoyage

Choix &  
niveau de  
confiance

Source 2 : CEREMA

Mur (90%)  
Année (89%)



### Objectif d'application :

Une BDD 

- Accessible → API
- Exploitable → Simple
- Fiable → Niveau de confiance
- Generali → jointure sinistre/exposition

### Cas d'usage :

Machine Learning (exposition, critères de vulnérabilité)  
Diagnostic climatique pour chaque bâtiment de France  
→ Externalisation du Lab !

# Base de données bâtiments

## BDNB : retraitement

### 1 Identification des variables pertinentes :

- Type de bâtiment
- Caractéristique du bâtiment
- Type de construction
- Sol/environnement



### 3 Simplification des données

EX: murs & année de construction

Source 1 : ADEME

Source 2 : CEREMA

% Complétude

Mur (11%)  
Année (14%)

Retraitement

Nettoyage

Choix & niveau de confiance

Mur (90%)  
Année (89%)

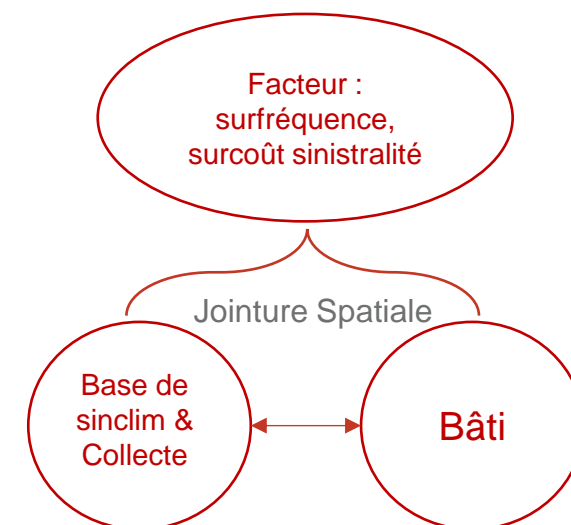
Retraitement

Nettoyage

### 2 Taux de complétude & rehaussement



### 4 Affectation aux données Generali



Objectif d'application :

Une BDD



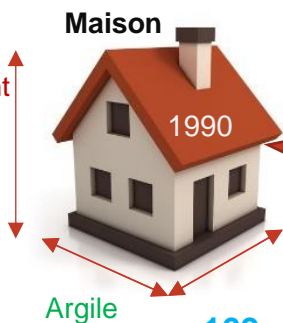
- 1 Accessible → API
- 2 Exploitable → Simple
- 3 Fiable → Niveau de confiance
- 4 Generali → jointure sinistre/exposition

# Base de données bâtiments

## BDNB : retraitement

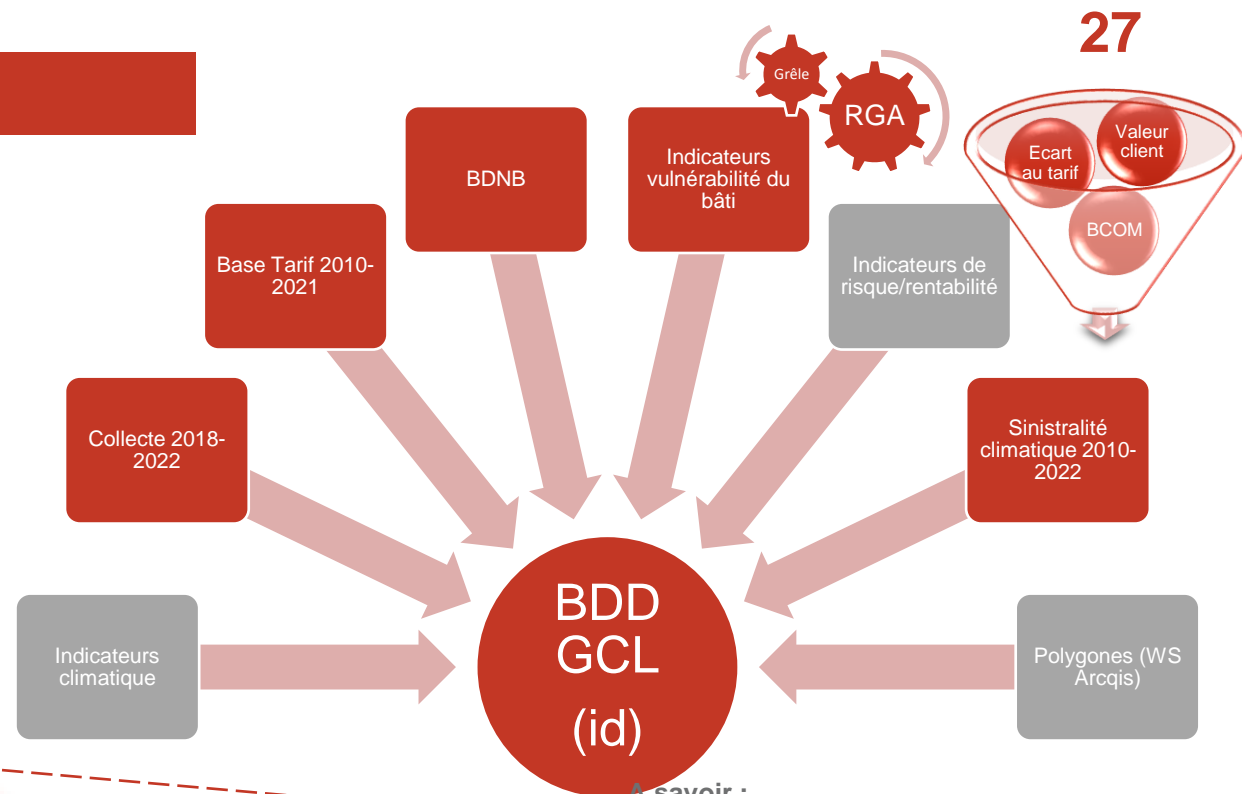
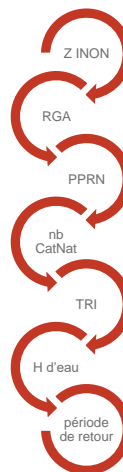
### 1 Identification des variables pertinentes :

- Type de bâtiment
- Caractéristique du bâtiment
- Type de construction
- Sol/environnement

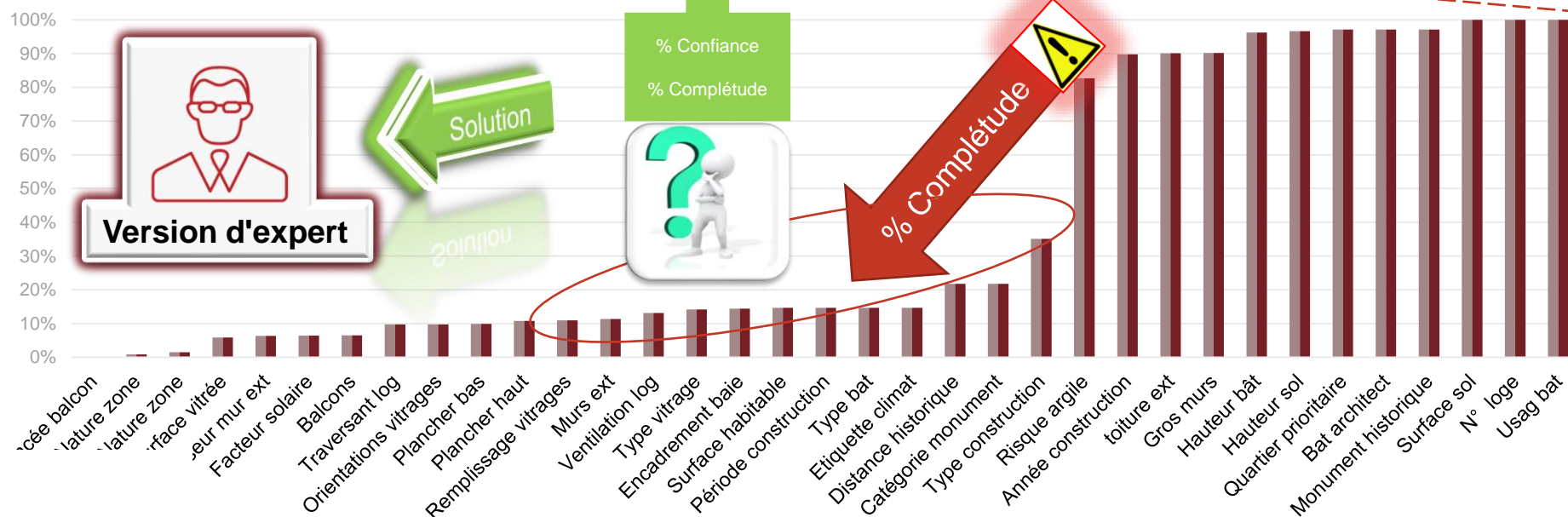


Génie civil et matériaux

169 variables → 59 variables



### 2 Taux de complétude & rehaussement



A savoir :

stocké dans Base de données **postgresql** requêttable via la **Climate-API** (url, Climate Apps)

Cas d'usage :

Machine Learning (exposition, critères de vulnérabilité) Diagnostique climatique pour chaque bâtiment de France

=> Externalisation du Lab !

**Objectif d'application :**

Une **BDD**

- 1 Accessible → API
- 2 Exploitable → Simple
- 3 Fiable → Niveau de confiance
- 4 Generali → jointure sinistre/exposition