



Journée Scientifique et Technique du CFMS du 11 décembre 2019
« *Machine Learning et Big Data en Géotechnique* »

La standardisation des données géotechniques : les projets MINnD Géotechnique et IDBE Geotech

Mickaël BEAUFILS

BRGM



Plan de la présentation

Quelques mots sur le BRGM

Interopérabilité, Machine Learning et Big Data : des sujets liés

Les activités de standardisation des données géotechniques pour le BIM en France et à l'international

Machine Learning et Big Data au BRGM

BRGM : Bureau de Recherches Géologiques et Minières

Un organisme pluridisciplinaire

Plus de 1000 salariés

Plusieurs type de mission

Etablissement Public à caractère Industriel et Commercial (EPIC)

Appui aux politiques publiques (BSS, Géorisques, GéoCatalogue, ...)

Etablissement de recherche : Label CARNOT

Le digital au cœur des activités du BRGM

Une expérience solide autour de l'interopérabilité des systèmes d'informations

Des activités de recherche appliquée autour du Web Sémantique, du Big Data et du Machine Learning

Supporté par des infrastructures informatiques reconnues : BRGM = Data Center Recherche de la Région Centre



L'interopérabilité des données : un enjeu majeur pour le Machine Learning et le Big Data



72,024 views | Mar 23, 2016, 09:33am

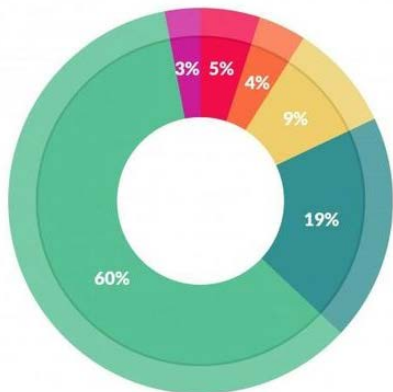
Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says



Gil Press Contributor

Big Data

I write about technology, entrepreneurs and innovation.



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

A glimpse inside the mind of a data scientist

Must-have skills? Daily challenges? Find out what actual data scientists really think about their critical role in data science



What tasks consume most of your time?

“Data preparation tends to be one of my most time-intensive activities.

“Gaining access to data and getting it into the proper format can be extremely time-consuming.

Quelques données produites et maintenues par le BRGM

Données géophysiques



Géothermie

Forages



Observations géologiques

RESSOURCES MINÉRALES

GÉOTHERMIE

Modèles géologiques



GÉOLOGIE

DGR

STOCKAGE GÉOLOGIQUE DU CO₂

Analyses / tests



LABORATOIRES & EXPÉRIMENTATION



Géosciences pour une Terre durable
brgm

DISN



SYSTÈMES D'INFORMATION

Modèles hydrogéologiques

DEPA

DRP



Etudes de risques

Pollution des sols



ENVIRONNEMENT & ÉCOTECHNOLOGIES

APRÈS-MINE

Cartes de risques

Niveau des eaux souterraines



EAU



RISQUES

Installations classées

Qualité des eaux souterraines

Qualité des eaux de surface

Cavités

Mesures sismiques

L'interopérabilité et le FAIR data au cœur de la stratégie du BRGM



Savoir que des données existent

Pouvoir les récupérer

Pouvoir les comprendre

Pouvoir les réutiliser

Source: <https://www.ands.org.au/working-with-data/fairdata/training>

Des données structurées pour des usages étendus

Modélisation

BIM / Jumeau
numérique

Pour pouvoir répondre plus efficacement à nos missions

Utilisation de standards = mutualisation des ressources =
partage des frais = cout global diminué

Big Data



Web
Sémantique

Une nécessité pour les grandes structures

Un socle commun pour éviter les cloisonnements et faciliter
le pluridisciplinaire et le collaboratif

Machine
Learning

Une extension de la démarche scientifique

La possibilité de partager les résultats, mais aussi les
méthodes, les interprétations et les données sources

Le projet MINnD



Modélisation des Infrastructures Interopérables pour les Infrastructures Durables

Deux saisons : Saison 1 : 2014-2018 ; Saison 2 : 2018-2021

Projet national collaboratif (71 partenaires en 2018)

Objectif : proposition de standards / outils / méthodes pour étendre les capacités du Building Information Modeling (BIM) aux infrastructures (Bridge, Rail, Road, Tunnel) + accompagnement pour la mise en place du BIM

Organisation : groupes de travail thématiques

En lien avec buildingSmart International (bSI) et l'Open Geospatial Consortium (OGC)

Participation du BRGM

Coordinateur du Thème 2 : Expérimentations (Saison 1)

Coordinateur du groupe de travail sur la standardisation des données géotechniques (Saison 1 et 2)

Apporte sa connaissance des Systèmes d'Information

Contribue à assurer le lien avec l'OGC



- **Sujet** : Appliquer la philosophie de l'open BIM aux données géotechniques
- **Besoin** : Permettre aux acteurs du projet d'accéder aux données relatives à l'environnement de l'infrastructure pendant toute sa durée de vie
- **Equipe** : 40 personnes issus de 13 organisations

- Recherche / organisme public



- Maitrise d'ouvrage (ou assimilé)



- Ingénierie



- Constructeurs



MINnD Géotechnique : la proposition

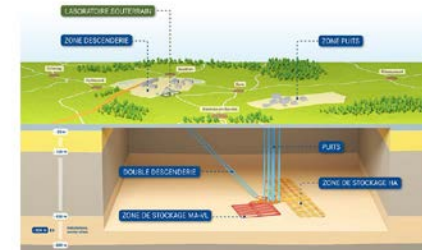
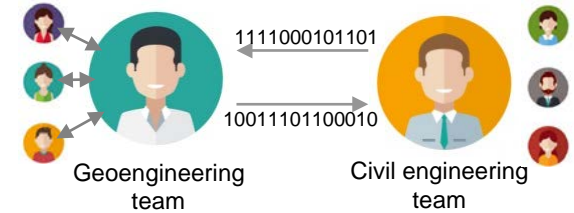
Better Information Management (BIM?)

Au sein des équipes géotechniques, organisation des données allant des observations et mesures brutes jusqu'aux interprétations et préconisations

Entre les équipes géotechniques et génie civil, meilleure collaboration à l'ère du BIM

Proposition de standards pour l'expression des données géotechniques et leur échange

Basé sur la norme NFP 94-500 et les guides AFTES, l'expérience des membres et des projets en cours



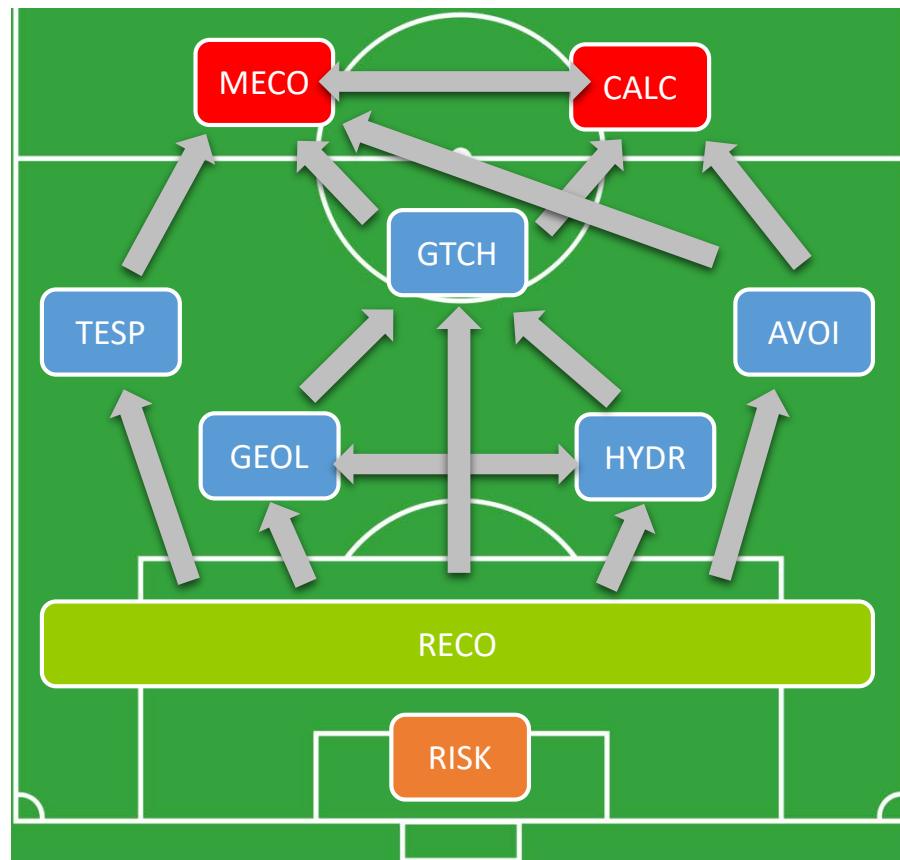
Les missions géotechniques étudiées dans MINnD Géotechnique

Designation	Id
Structure sizing and definition of the geotechnical influence zone	CALC
Construction methods	MECO
Geological modeling	GEOL
Hydrogeological modeling	HYDR
Geotechnical modeling	GTCH
Anthropic environment modeling	AVOI
Excavations and Site Pollution modeling	TESP
Observations and Measurements	RECO
Risk and uncertainty assessment	RISK

Book C:
Draft
Conception Plan
(including RISK)

Book B:
Environment
Modeling
(including
uncertainties)

Book A:
Factual data
collection



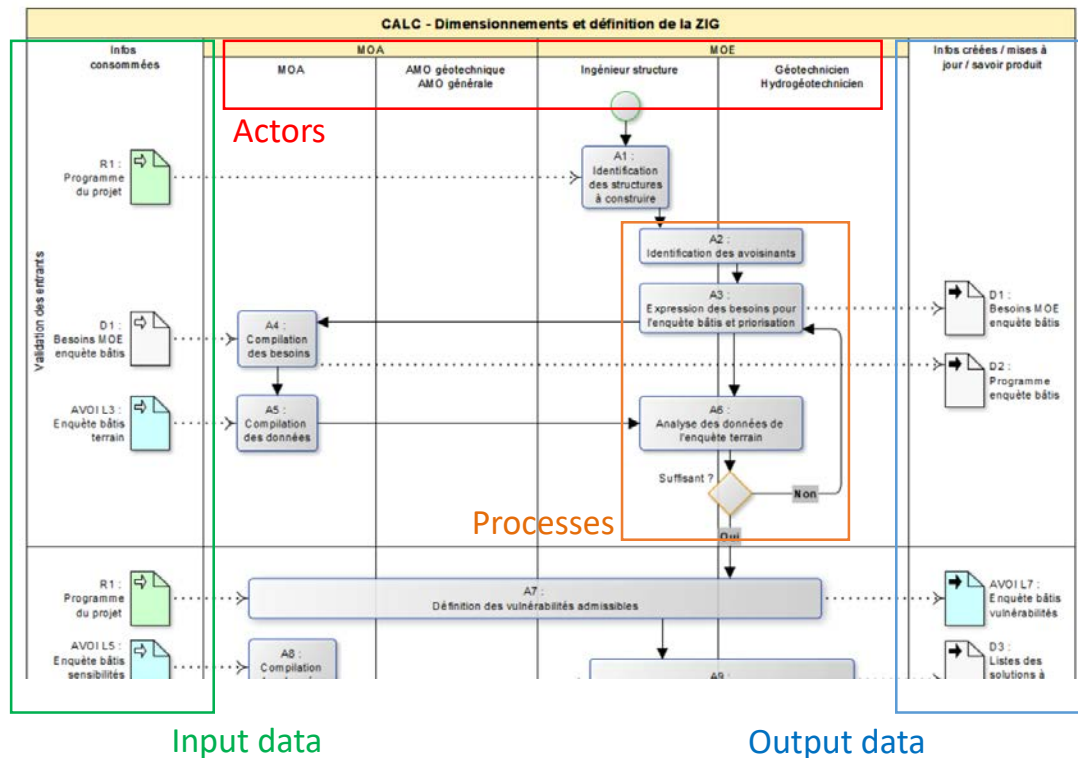
Décrire les missions géotechniques

Information Delivery Manual (IDM)

- Description de l'enchaînement des processus
- ISO 29481-1:2016
- 1 par sujet

Informations

- Qui fait quoi
- Données utilisées
- Données produites



Bien décrire les données collectées et produites pour le présent et le futur

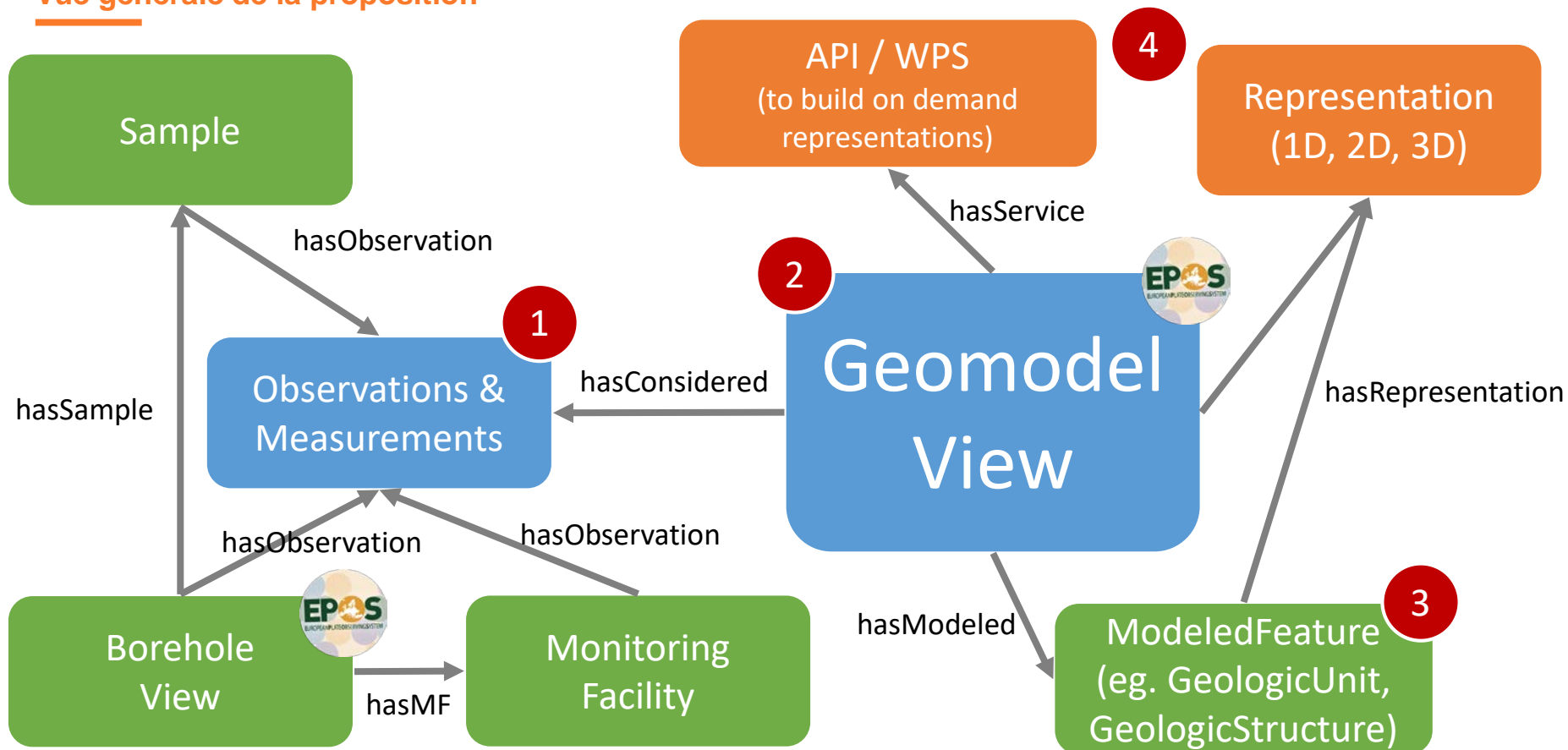
Description des données de reconnaissance

- Description des observations et mesures en géotechnique
 - ✓ Valeur, incertitude, propriétés observées, méthodes d'obtention
 - ✓ Relations observations / interprétations
- Description des supports d'observations
 - ✓ Forages, échantillons, point

Description des données interprétées

- Définition des objets modélisés
 - ✓ Unités géologiques, unités hydrogéologiques, unités géotechniques
 - ✓ Failles, cavités
 - ✓ Masses d'eau
- Définition des propriétés associées et de leur association
 - ✓ Attribut vs interprétation

Vue générale de la proposition



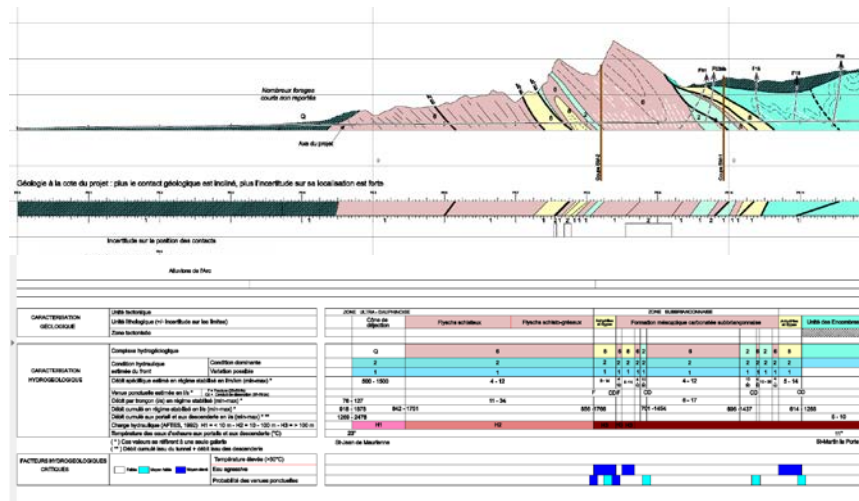
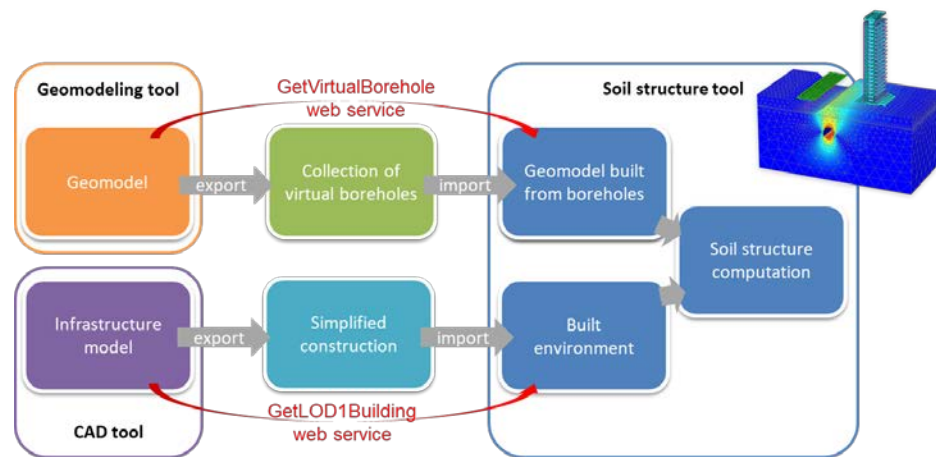
Explorer de nouvelles façons d'échanger / manipuler des données

Outils de co-visualisation / accès aux données BIM et geosciences

- Prototype basé sur l'outil eveBIM du CSTB

Services de création de coupes à la volée, forages virtuels

- Pour calcul sol-structure
- Pour production de livrable



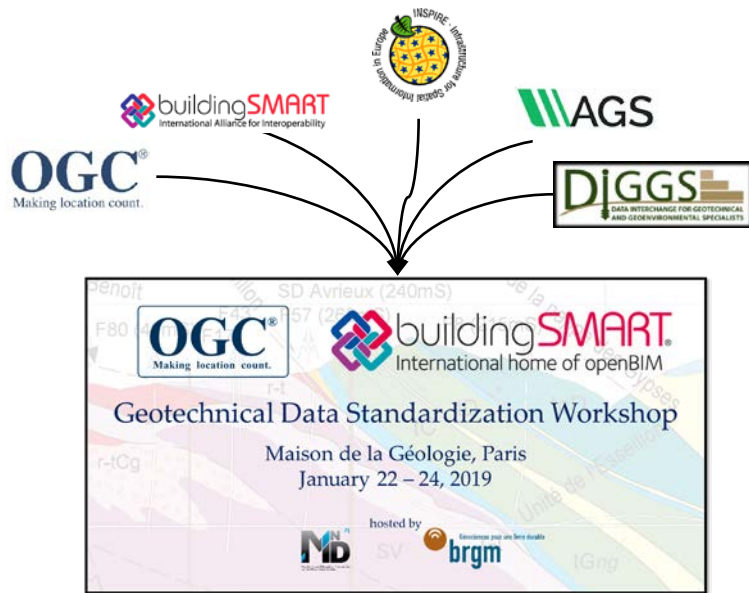
Quelques initiatives en cours pour la standardisation des données géoscientifiques



Earth, Space, and Environmental Sciences
Interest Group



Une approche poussée à l'internationale






IDBE GEOTECH

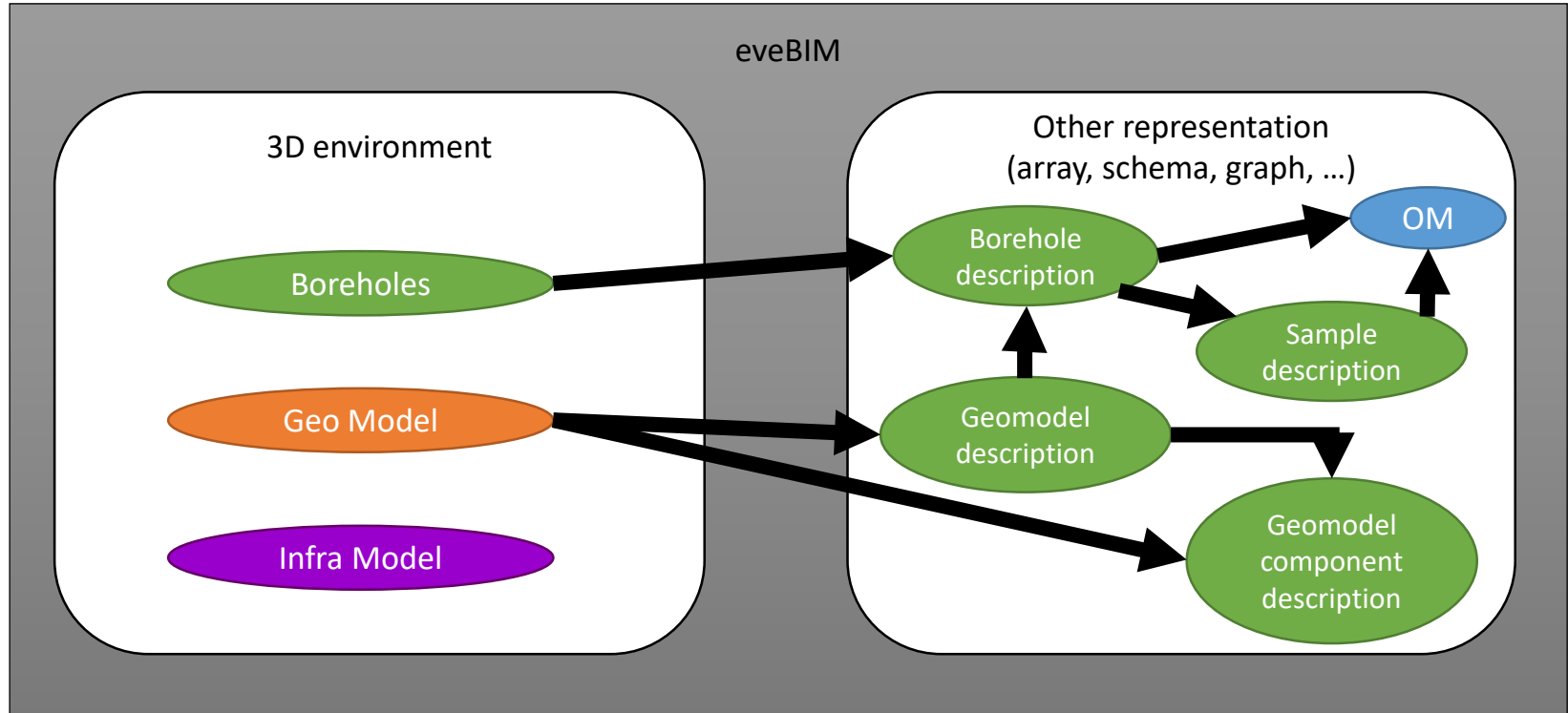
Geotechnical Data Standardization Workshop

- 60 participants du monde entier
- Présentations disponibles sur <https://github.com/opengeospatial/IDBE-Geotech>
- Résolutions:
 - ✓ Collaboration officielle entre OGC et bSI sur la standardisation des données géotechniques pour le BIM via le groupe **IDBE Geotech**
 - ✓ **IDBE Geotech** s'appuie fortement sur les travaux de MINnD pour les étendre à l'international

IDBE Geotech

- Animateurs:
 - ✓ Mickaël Beaufiles (BRGM, France) 
 - ✓ Jonas Weil (IC-Group, Autriche) 
 - ✓ Rie Wada (Oyo corporation, Japon) 









Coaccès données géoscience / infra : prototype avec OYO / CSTB

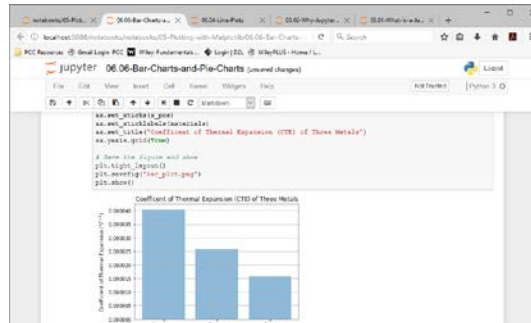


Le Big Data au BRGM

- Clusters Big Data
 - Serveurs répartis
 - 170 To utiles
 - Hadoop
- Data lab
 - Infrastructure distribuée de stockage et calcul
 - Deux valorisations
 - ✓ Données centralisées dans un Data Lake
 - ✓ Support data science

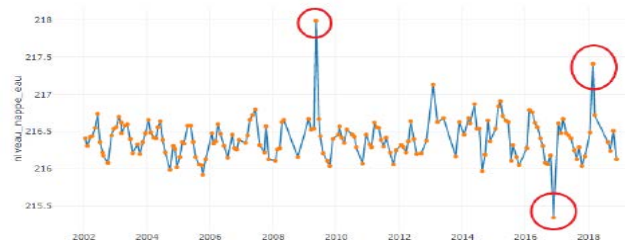
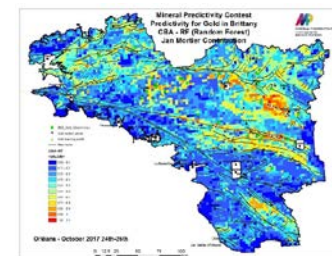


Donnée	Volume
	> 270 000 données
	> 240 000 références
	> 125 millions d'analyses
	> 67 millions d'analyses
	> 14 millions de mesures
	> 21 millions de mesures
	> 27 millions de mesures
	> 620 000 données



Intelligence Artificielle et géosciences au BRGM

- Promesses scientifiques
 - Augmentation des capacités prédictives (en qualité et/ou quantité)
 - Nouvelles perspectives de découvertes scientifiques (modèles agnostiques de toute thématique)
 - Valorisation du patrimoine numérique (rapports, notices de cartes géologiques, spectrographie, ...)
- Premières réalisations au BRGM
 - Cartographie prédictive appliquée (Machine Learning)
 - ✓ Aux ressources minérales
 - ✓ À la pollution
 - Prédiction des niveaux de nappes phréatiques
 - ✓ Détection automatique d'anomalies
- Thèse sur l'explicabilité des données



En conclusion

L'interopérabilité est un enjeu majeur pour la géotechnique

Projets MINnD et IDBE Geotech en cours pour la standardisation des données géotechniques

Le BRGM est présent et actif

Sur les sujets interopérabilité

Sur les sujets Machine Learning et Big Data

Opportunités de collaboration sur ces sujets

Merci pour votre attention

- **Contacts :**

- Sujet interopérabilité : Mickaël Beaufils - m.beaufils@brgm.fr
- Sujet Big Data et Machine Learning : Vincent Labbé – v.labbe@brgm.fr

