

Du capteur au supercalculateur, pour une meilleure compréhension de la qualité de l'air

François Bodin Université de Rennes 1

Introduction



- La maîtrise de la qualité de l'air nécessite une meilleure compréhension de la répartition des polluants sur le territoire
- La mesure ne suffit pas à l'explication
 - La simulation numérique permet d'aider à la compréhension
- Le projet AQMO (Air Quality & MObility)
 - Développe une plateforme urbaine de bout en bout
 - Complète les outils actuels de mesure de la qualité de l'air
 - Se fonde sur les technologies de l'Internet des Objets, du Cloud et du HPC
- Contexte d'intégration de l'*Edge* et du HPC
 - Avec l'idée de l'«IA everywhere »
 - Big Data & Extreme-scale Computing exascale.org

Objectif du projet

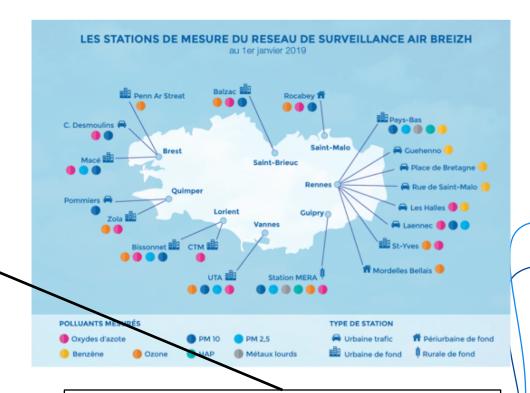


- Etendre de la mesure de la qualité de l'air
 - Temporelle
 - Spatiale
- Analyser la dispersion des polluants par la simulation numérique
 - Comprendre et identifier les causes
 - Poser des questions « et-si »
- Mettre à disposition des données auprès des citoyens
 - Via le service public métropolitain de la données

Qualité de l'air



- Principaux polluants
 - Les oxydes d'azote
 - · le trafic routier
 - L'ozone polluant
 - Généré à partir du trafic routier sous certaines conditions atmosphériques
 - · Les particules fines
 - · le trafic routier
 - · le chauffage résidentiel
 - l'activité agricole mais aussi des sources naturelles (poussières désertiques, incendies, volcans, écumes, ...)
 - Les métaux lourds
 - industrie + combustion carbonée
 - Les HAP (Hydrocarbures Aromatiques Poly- cycliques)
 - combustion carbonée (bois, charbon, hydrocarbures)
 - Les COV (Composés Organiques Volatils)
 - · carburants, peintures, solvants, colles, ...
 - Le Monoxyde de carbone
 - · combustion carbonée
- Indice global de qualité de l'air
 - 3 polluants définissent l'indice
 - particules fines (PM 10)
 - dioxyde d'azote (NO2)
 - ozone (O3)



Polluants qui sont pris en compte dans AQMO dans sa première phase

Partenaires - interdisciplinarité



- Air Breizh
 - surveillance de la qualité de l'air en Bretagne
- AmpliSIM
 - startup simulation numérique météo
- Genci
 - coordination des moyens de calculs FR
- CNRS/Idris
 - centre de calcul
- Irisa/Université de Rennes 1
 - coordinateur

- Keolis
 - opérateur des bus rennais
- Neovia
 - PME spécialisée en projet HPC
- Rennes Métropole
- Ryax-Technology
 - startup gestion de ressources IoT
- Ucit
 - SaaS HPC















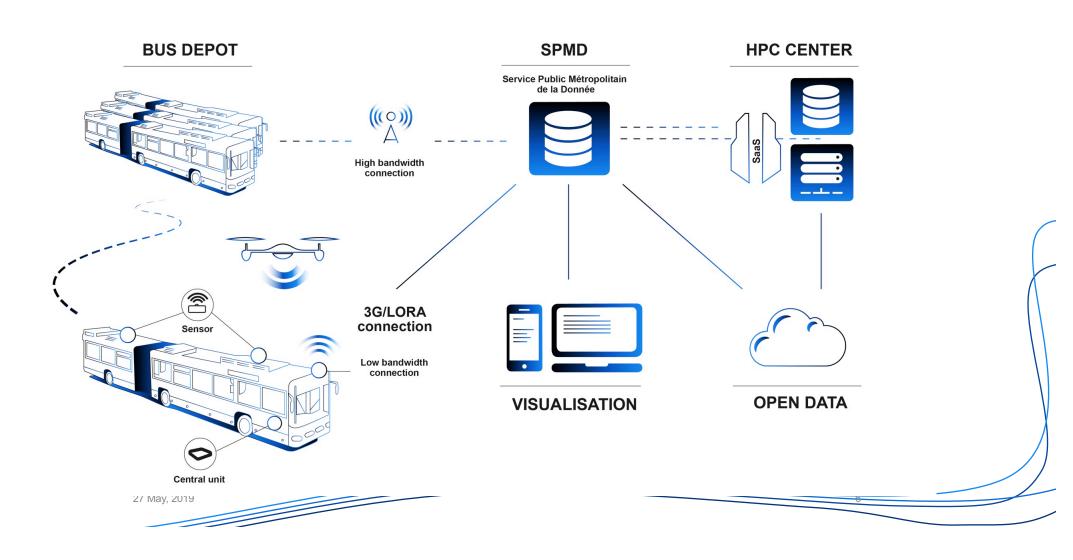






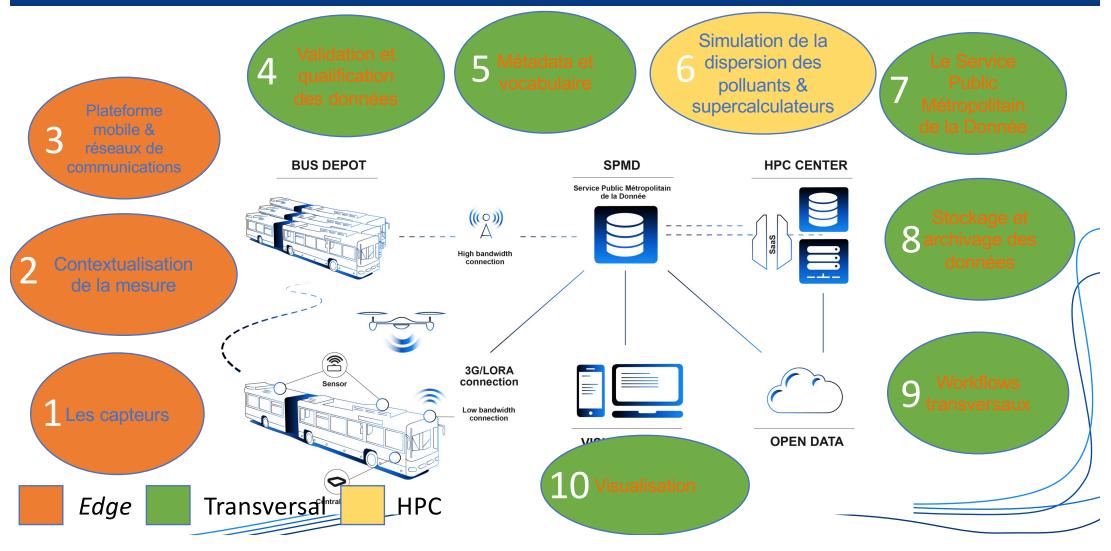
Vue d'ensemble du projet





Plan de la présentation





Les capteurs



- De nombreuses technologies sur un large spectre
 - Instruments scientifiques pointus
 - Capteurs faibles coûts
- Quel compromis
 - Peu de capteurs coûteux ou beaucoup de capteurs peu coûteux ?
- Attention capteurs ≠ mesures
 - Comment calibrer?
 - Comment valider?

Exemple de capteurs



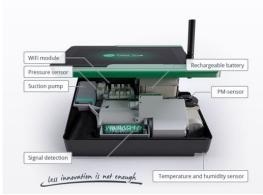
- SDS011
 - ~30 euros



- Instrument portable, Fidas Frog / Palas
 - ~20k€ -



- Dizaines de milliers d'euros
- Tests, contrôles et agréments délivré par le Laboratoire Central de Surveillance de la Qualité de l'Air



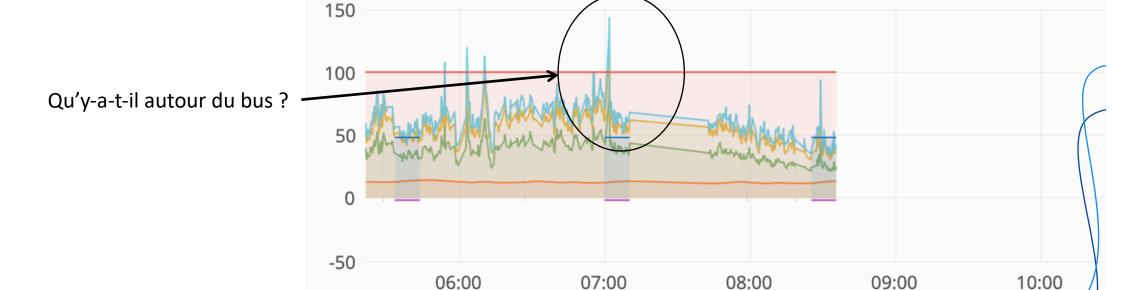
Contextualisation de la mesure



Coordonnées GPS

27 May, 2019

• Environnement autour de la mesure



100200061

— Pm 1 — Pm 2.5 — Pm 10 — Temperature — Humidity — latitude — løngiti

Calcul à la périphérie (edge computing)



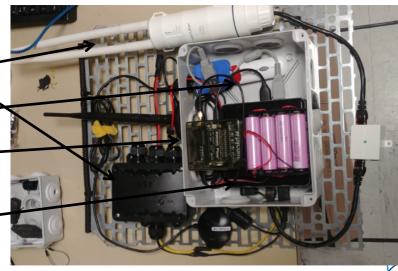
- La contextualisation de la mesure peut nécessiter des traitements de données lourds
 - Analyse d'images
 - Réduction de la transmission de données
 - Préservation de la vie privée
- Utilisation des technologies de l'IA
 - Image de contrôle (floutée)
 - Seul le comptage est remonté
 - A la demande, en fonction des mesures



Plateforme mobile



- Un gestionnaire d'alimentation (UPS)
- Des unités de communications
 - Une communication WiFi-
 - Une connexion 4G—
 - Une connexion LoRa-
- Une unité de calculs et stockage
 - Intel NUC avec GPU___
- Des capteurs
 - PM
 - Une caméra
 - Avec analyse d'images (DL)
- Extensible à d'autres capteurs et utilisateurs





Réseaux de communications



- Utilisation de multiples canaux
 - Wifi Bus au dépôt
 - Téléversement des volumes de données importants
 - LoRa pour le monitoring et le contrôle de l'unité centrale
 - Réseau d'Acklio
 - Réseau Rennes Métropole
 - 3/4G pour la maintenance et le prototypage
 - Coût souvent trop important pour un déploiement en production
- Gestion de l'intermittence des communications nécessaire
 - Pas de stabilité des canaux, dépend de la position du bus
- Isolation & sécurisation via un SDN / VPN / Firewall

Validation de la mesure



- Comprendre les limitations et la signification des mesures
 - Les mesures sont ultimement la base d'actions politiques
- Comparer les capteurs utilisés avec les instruments de référence
 - Recherche de corrélations
 - Caractérisation des erreurs
 - Détermination des conditions de bon fonctionnement
- Dans AQMO
 - Station Laennec d'AirBreizh qui sert de référence fixe
 - Utilisation d'un Fidas Frog en mobilité

Qualification des données



- Chaque mesure doit être qualifiée avant stockage
 - Chaque mesure possède une étiquette
 - Utilisation d'une échelle internationale
- Echelle du Natural Environment Research Council*
 - 0: Aucune procédure de contrôle de la qualité n'a été appliquée à la valeur des données
 - 1: Des données de bonne qualité qui ne font partie d'aucun dysfonctionnement identifié et qui ont été vérifiées comme étant cohérentes avec les phénomènes réels pendant le processus de contrôle de la qualité
 - 2: Valeur de données qui correspond probablement à des phénomènes réels, mais qui n'est pas confirmée, ou valeur de données faisant partie d'un dysfonctionnement jugé trop faible pour affecter la qualité globale de l'objet de données dont il fait partie
 - 3: Valeur des données reconnue comme **inhabituelle** lors d'un contrôle de qualité faisant partie d'une caractéristique qui est probablement incompatible avec des phénomènes réels
 - 4: Une valeur de données manifestement erronée
 - 5: Valeur des données **ajustée pendant le contrôle de qualité**. La meilleure pratique recommande fortement que la valeur avant le changement soit préservée dans les données ou les métadonnées qui les accompagnent.

• ...

*http://vocab.nerc.ac.uk/collection/L20/current

27 May, 2019

15

Metadata et vocabulaire



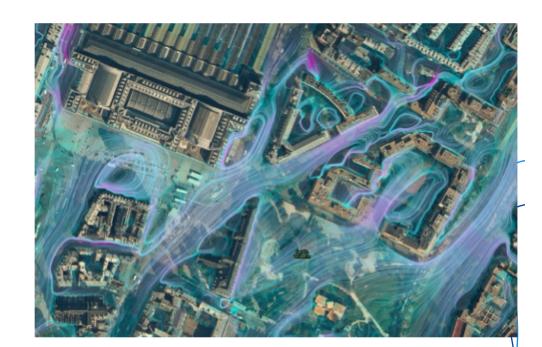
- Construction d'un vocabulaire commun
 - Utilisation des standards internationaux
 - Accord nécessaire sur la sémantique et la syntaxe
 - Gouvernance commune et coordonnée
 - Sur la base des standards utilisés par AirBreizh
- Par exemple sur la base de SKOS
 - Système simple d'organisation des connaissances

27 May, 2019

Simulation de la dispersion des polluants



- Reproduire sur un ordinateur comment les différents polluants se dispersent dans l'atmosphère
 - Sur la base des équations de la physique en fonction des vents, du cadastre des sources d'émissions, des bâtiments, du relief, des sols, ...
- Quelques heures / CPU pour les modèles grossiers (10 mètres)
- 30 000 heures x CPU par jour simulé (700 km²) pour les modèles à grille fine (3m)



Supercalculateurs dans AQMO – Un changement de pratique



- En charge du calcul du modèle numérique fin
- Mais nécessite d'étendre l'usage traditionnel (batch)
 - Accessibles aux métropoles
 - Mise en œuvre comme un Software as a Service (SaaS)
 - Calculs sur une base régulière (e.g. journalière)
 - Calculs lors d'événements catastrophiques (e.g. incendies)
- Et aussi de considérer le traitement des données en flux
 - Simulation numérique combinée à de l'assimilation de données pour le calcul de la dispersion des polluants



Le supercalculateur Turing de l'IDRIS

27 May, 2019

Urgent computing

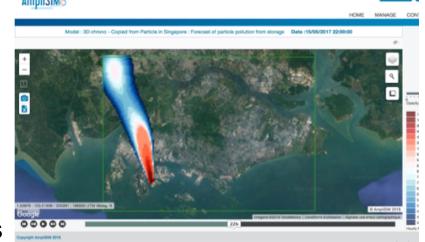


• Répondre rapidement en situations de crise (e.g. incendies)

• Identifier les zones pour lesquelles il faut mettre en place des mesures

spécifiques (e.g. évacuation)

- Un temps de réponse rapide est essentiel
 - Quelques dizaines de minutes
 - La réquisition des moyens doit être prédéfinie
 - Les profils d'exécution adaptés
 - Les données nécessaires doivent être présentes dans l'infrastructure de calcul
 - Pré-calculer ce qui peut l'être
 - Tests réguliers et systématiques



Le Service Métropolitain de la Donnée



- En charge de la diffusion des données ouvertes
 - Définie une forme de gouvernance et de fédération

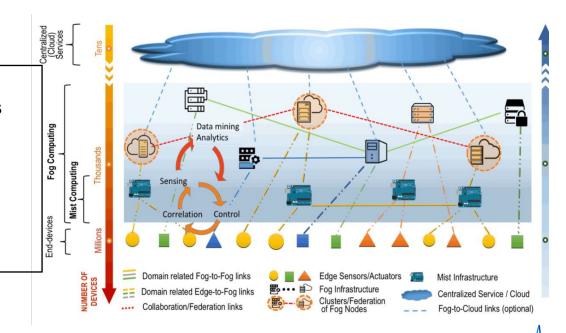
- Dans le cadre d'AQMO permet aussi la consolidation avec les sources citoyennes
 - Ambassad'Air à Rennes (capteurs Luftdaten)
 - Smart citizen à Barcelone

27 May, 2019

Workflows transversaux



- Déploiement de workflows sur de nombreux composants hétérogènes
 - Modèle proche du Fog par exemple
 - Modèle en couches pour permettre un accès généralisé à un continuum partagé de ressources informatiques évolutives
 - Un réseau décentralisé
 - Modèle de calcul distribué et fédéré



Src: NIST Special Publication 500-325, Fog Computing Conceptual Model Recommendations of the National Institute of Standards and Technology

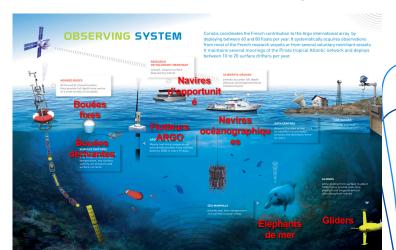
27 May. 2019

Stockage et archivages des données



- Support du cycle de vie des données par une fédération d'infrastructures
 - Périphérie
 - Serveur intermédiaire, stockage agile
 - Cloud, centres de calculs
- Les éléments à considérer
 - Référentiel communautaire de description des données
 - Modèle utilisateurs
 - Modèle de ressources transversal
 - Modèles de règles de gestion des données
 - Capacité de de monitoring
 - Transport de la donnée avec optimisation de l'accès
 - Processus de validation des jeux de données
 - ...
- Archivage souvent négligé
 - ~ 50k€ / an / petabyte

Un exemple proche Seadatanet https://www.seadatanet.org/



Visualisation des données



Visualisation des séries temporelles des données de capteurs

 Visualisation des cartes de dispersion et éventuellement en immersion 3D



- Visualisation des modèles numériques
 - En général réservée aux scientifiques car complexe et plutôt destiné à l'étude des comportements des modèles

27 May, 2019

23

Conclusion



- Le premier bus équipé est opérationnel depuis 5 mois
 - 20 bus équipés d'ici fin 2020 dont au moins 1 bus électrique
- Le projet aborde de nombreuses problématiques de l'Internet des objets et du HPC
 - Comment organiser la logistique des données ?
 - Comment valider et qualifier les données ?
 - Comment distribuer les workflows sur des infrastructures hétérogènes et distribuées ?
 - Comment s'assurer que les données seront encore pertinentes dans 20 ans ?
- Les infrastructures restent à inventer
 - Permettre un cycle des données entre de nombreux acteurs garantissant des données de qualité
 - AQMO propose une plateforme expérimentale
 - Maitrise des coûts sur le long terme

27 May, 2019



Co-financed by the Connecting Europe Facility of the European Union





















