

Titre de la thèse – PhD Title:

Filtrage contextuel local et traitement local opportuniste des données de santé

Contextual Filtering and Opportunistic Local Processing of Health Data

Collaboration- Partnership: *projet Premyom***IMT Atlantique : Campus** ☐ Brest ☐ Nantes ☒ Rennes**Laboratory :** IRISA**École doctorale :** ☒ SPIN ☐ 3MG**Financement - Funding:** **Projet Premyom** (<https://premyom.com>)**Résumé :**

Les objets connectés de santé ont des usages multiples : mesure de l'activité physique, détection de chute, surveillance de la pression artérielle, etc. Ces dispositifs portables, de plus en plus courants, génèrent en continu des volumes massifs de données qui peuvent être redondantes (mesures inchangées), voire non pertinentes ou inutiles selon le contexte d'utilisation. Dans ce cadre, cette thèse vise à concevoir une architecture distribuée permettant de filtrer et d'enrichir localement les données à l'aide d'informations contextuelles (présence d'objets, posture, activité, etc.) avant leur transmission vers un système central. En combinant traitement embarqué et infrastructure active, l'objectif est de réduire la charge de traitement centralisé, les coûts de transmission et les atteintes à la vie privée. Le travail s'appuiera sur des cas d'usage concrets, en lien avec le projet Premyom, pour définir les besoins des chercheurs et formaliser les arbitrages entre traitements locaux et centralisés.

Environnement académique :

IMT Atlantique, grande école d'ingénieurs généralistes, a pour ambition de conjuguer le numérique, l'énergie et l'environnement pour transformer la société et l'industrie par la formation, la recherche et l'innovation. L'établissement qui est présent sur trois campus (Brest, Nantes et Rennes) appartient à l'Institut Mines-Télécom et dépend du ministère l'Économie, des finances et de la souveraineté industrielle et numérique. Reconnue internationalement pour la qualité de sa recherche, les scientifiques d'IMT Atlantique accompagnent environ 300 doctorants. La thèse proposée s'inscrit dans les activités de recherche de l'équipe : E4SE et de du laboratoire IRISA et du département SRCD. Les activités scientifiques de ce département relèvent de la cyber-sécurité, l'informatique et les réseaux.

Contexte et état de l'art scientifique :

Dans les systèmes actuels de collecte de données de santé, les dispositifs portables (lunettes intelligentes, montres, capteurs biométriques, etc.) transmettent généralement toutes les données vers des serveurs centraux où elles sont stockées et analysées. Si cette méthode garantit la disponibilité de l'ensemble des données brutes, elle entraîne des coûts de traitement élevés, des risques pour la vie privée, et une charge significative de traitement préalable (souvent appelée « curation ») pour nettoyer, normaliser et contextualiser les données.

Dans de nombreux cas, une grande partie des données ainsi collectées est non pertinente, voire manquante ou au contraire inutile. Par exemple, les données issues d'un dispositif médical ne sont utiles que lorsqu'il est effectivement porté. Ces situations peuvent souvent être identifiées localement grâce à des informations contextuelles disponibles dans l'environnement immédiat (proximité d'autres objets connectés, niveau de signal, activité ambiante, etc.). Exploiter ces informations pour filtrer ou enrichir les mesures au plus près de leur source permettrait de réduire considérablement la charge du système central.

Cette thèse propose de développer une approche hybride, dans laquelle le dispositif lui-même et son environnement proche participent activement au tri, à l'agrégation et à l'enrichissement des données, tout en laissant au système central la gestion de la cohérence globale et des traitements plus lourds. Cette manipulation décentralisée des données — réalisée directement par ou à proximité des dispositifs qui les génèrent — correspond à une approche de type *Edge computing*, qui constitue le cœur de la thématique de recherche.

Objectifs de la thèse ?

- Identifier les besoins de curation et les critères de qualité des données dans des cas d'usage réels (ex. suivis cliniques, dispositifs à visée comportementale).
- Définir une méthodologie pour caractériser localement les données pertinentes, en s'appuyant sur les informations contextuelles accessibles in situ.
- Concevoir une architecture de traitement local reposant sur des microservices embarqués pour filtrer, enrichir, ou annoter les données avant leur transmission.
- Explorer l'automatisation de la chaîne de traitement à partir de profils d'usage ou de contraintes exprimées par les chercheurs et cliniciens.

Profil du candidat :

- Bonnes connaissances en traitement de données capteurs et systèmes distribués.
- Expérience en développement et déploiement de services dans des architectures distribuées (Edge computing, microservices).
- Intérêt pour les applications en santé connectée et en IoT médical.
- Capacité à interagir avec des chercheurs en santé ou en sciences humaines.
- Esprit d'analyse, autonomie et esprit de synthèse.
- Bon niveau d'anglais scientifique.

Références :

- L. Chen et al., "HARMONI: Context-aware Filtering of Sensor Data for Continuous Remote Health Monitoring," in Proc. of IEEE HealthCom, 2009.
- A. Bouachir et al., "SELF-CARE: Selective Fusion with Context-Aware Low-Power Edge Computing for Wearable Health Monitoring," IEEE Access, 2022.
- M. Vayena et al., "Data Acquisition, Curation, and Use For a Continuously Learning Health System," NAM Perspectives, 2018.
- J. Schmid et al., "Unpacking Real-World Data Curation: Principles and Best Practices," Duke-Margolis Center for Health Policy, 2020.

Plan de travail :

1. **Analyse exploratoire des cas d'usage**
 - Collaboration avec les partenaires du projet Premyom pour identifier des scénarios concrets.
 - Détermination des données utiles / inutiles dans les chaînes de collecte actuelles.
2. **Prototype de traitement local distribué**
 - Implémentation de microservices de filtrage / annotation embarqués.
 - Expérimentations sur des dispositifs portables avec des capteurs contextuels.
 - Étude comparative des traitements pouvant être réalisés localement (avec peu de données) versus ceux nécessitant une agrégation centrale (grandes quantités de données).
3. **Automatisation de la chaîne de traitement et analyse de l'équilibre local / central**
 - Formalisation des besoins scientifiques (qualité, délais, granularité).
 - Génération automatique de chaînes de traitement adaptées aux cas d'usage.

Candidature:

Renseignements complémentaires:

Ce sujet de thèse se déroulera au sein de l'équipe E4SE du laboratoire IRISA hébergée à IMT Atlantique, (campus de Rennes). Elle est en lien avec le projet PREMYOM (<https://premyom.com>) et les collaborations existantes dans le domaine e-santé.

Les candidats doivent fournir les documents suivants :

- Un CV détaillé incluant leur parcours académique et les expériences pertinentes.
- Une lettre de motivation expliquant leur intérêt pour le sujet et en quoi leur profil correspond au poste.

Les candidats sont libres d'ajouter tout élément susceptible de venir en appui de leur candidature, comme des lettres de recommandation, des relevés de notes détaillés, des prix ou distinctions académiques.

Les candidatures doivent être envoyées par courriel à l'adresse suivante : jean-marie.bonnin@irisa.fr

Date de fin de candidature : 15/09/2025

Date de démarrage de la thèse : fin 2025
Durée du contrat : 36 mois
Localisation : IMT Atlantique, campus de Rennes
Contact : jean-marie.bonnin@irisa.fr

Abstract:

Connected health devices are used for multiple purposes: tracking physical activity, fall detection, blood pressure monitoring, etc. These increasingly common wearable devices continuously generate massive amounts of data, which can often be redundant (e.g., unchanged measurements), irrelevant, or unnecessary depending on the usage context. In this setting, the goal of the thesis is to design a distributed architecture capable of filtering and enriching data locally using contextual information (e.g., presence of objects, posture, activity) before transmitting it to a central system. By combining embedded processing with active infrastructure, the objective is to reduce centralized processing load, transmission costs, and privacy risks. The work will be grounded in real-world use cases, in connection with the Premyom project (<https://premyom.com>), to define researcher needs and formalize the trade-offs between local and centralized data processing mechanisms.

Context:

IMT Atlantique, internationally recognised for the quality of its research, is a leading general engineering school under the aegis of the Ministry of Industry and Digital Technology, ranked in the three main international rankings (THE, SHANGHAI, QS). Located on three campuses, Brest, Nantes and Rennes, IMT Atlantique aims to combine digital technology and energy to transform society and industry through training, research and innovation. It aims to be the leading French higher education and research institution in this field on an international scale. With 290 researchers and permanent lecturers, 1000 publications and 18 M€ of contracts, it supervises 2300 students each year and its training courses are based on cutting-edge research carried out within 6 joint research units: GEPEA, IRISA, LATIM, LABSTICC, LS2N and SUBATECH.

The proposed thesis is part of the research activities of the team: E4SE and of the laboratory IRISA and the department SRCD. The scientific activities of this department are related to cyber-security, computer science and networks.

Scientific context and Motivation:

In current health data collection systems, wearable devices (smart glasses, watches, biometric sensors, etc.) typically transmit all collected data to centralized servers where it is stored and analyzed. While this method ensures the availability of raw data, it results in high processing costs, privacy risks, and a heavy burden of preprocessing and standardization (often referred to as data curation) to clean, normalize, and contextualize the information.

In many cases, a large portion of the data collected in this way is irrelevant, missing or unnecessary. For example, data from a medical device is only relevant when the device is actually worn. Such situations can often be detected locally using contextual information from the immediate environment (e.g., proximity to other connected objects, signal strength, ambient activity). Leveraging such local information to filter or enrich measurements could significantly reduce the central processing load and associated costs.

This thesis proposes to develop a hybrid approach, in which the device itself and its nearby environment actively participate in sorting, aggregating, and enriching data, while leaving the system-wide coherence and heavier processing tasks to the central infrastructure. The objective is to define a methodology and an adaptive architecture to intelligently distribute preprocessing responsibilities across different layers of the system. This decentralized manipulation of data—performed directly by or near the data-generating devices—corresponds to an “Edge computing” approach, which will be central to the thesis.

Require skills:

- Solid knowledge in sensor data processing and distributed systems.
- Experience in developing and deploying services in a distributed architecture (Edge computing, microservices).
- Interest in connected health and medical IoT applications.
- Ability to interact with health or social science researchers.
- Ability to work independently and as part of a team; critical thinking.
- Proficiency in scientific English, both written and spoken.

References:

- L. Chen et al., "HARMONI: Context-aware Filtering of Sensor Data for Continuous Remote Health Monitoring," Proc. IEEE HealthCom, 2009.
- A. Bouachir et al., "SELF-CARE: Selective Fusion with Context-Aware Low-Power Edge Computing for Wearable Health Monitoring," IEEE Access, 2022.
- M. Vayena et al., "Data Acquisition, Curation, and Use For a Continuously Learning Health System," NAM Perspectives, 2018.

- J. Schmid et al., "Unpacking Real-World Data Curation: Principles and Best Practices," Duke-Margolis Center for Health Policy, 2020.

Work Plan

1. **Exploratory Analysis of Use Cases**
 - Collaborate with the Premyom project partners to identify concrete scenarios.
 - Identify useful vs. unnecessary data in current data collection workflows.
2. **Prototype for Distributed Local Processing**
 - Implement local services for filtering and annotation.
 - Run experiments on wearable devices equipped with contextual sensors.
 - Conduct a comparative study of processing tasks that can be performed locally (with limited data) vs. those requiring central aggregation (large-scale data).
3. **Automation of Processing Chain and Local/Central Trade-off Analysis**
 - Formalize scientific requirements (quality, delay, granularity).
 - Automatically generate tailored data processing chains for identified use cases.
 - Analyze the trade-offs between edge-level and centralized processing to determine optimal system partitioning.

Application:

Additional Informations:

This PhD project will take place within the E4SE team at the IRISA laboratory, hosted by IMT Atlantique (Rennes campus), and is linked to the Premyom project (<https://premyom.com>) and ongoing collaborations in the field of e-health.

Candidates should submit the following documents:

- A detailed CV including academic background and relevant experience.
- A cover letter explaining their motivation for applying and how their profile fits the position.

Applicants are welcome to include any additional materials that may support their application, such as recommendation letters, detailed transcripts, awards, or other distinctions.

Applications must be sent by email to: jean-marie.bonnin@irisa.fr

Application deadline: 15/09/2025

PhD starting date: end 2025

Contract duration: 36 months

Location: IMT Atlantique, Rennes campus

Contact: jean-marie.bonnin@irisa.fr