

## **Sujet de thèse**

# **“ Architectures réparties et conteneurs logiciels sécurisés pour coopérations médicales multi-sites ”**

## **Contexte**

La médecine de précision a pour objectif d'optimiser les soins médicaux d'un patient, de façon individualisée, en accompagnant les médecins dans la prise de décision grâce à des applications informatiques intégrant des algorithmes de précision. Ces applications vont permettre de représenter et de comparer systématiquement une information individuelle dans le contexte d'une population avec des caractéristiques similaires. Le développement de la médecine de précision dans le cadre de maladies chroniques comme la transplantation rénale pourrait permettre une meilleure compréhension et une meilleure prise en charge des malades [SS17]. En effet, ces pathologies sont étiologiquement hétérogènes incluant de multiples causes et menant à des symptômes et des traitements différents [ABIO15]. Par conséquent, il est actuellement impossible, même pour un clinicien expérimenté, de prédire l'évolution clinique d'un malade et d'anticiper la sécurité et l'efficacité des traitements [DC+17].

L'échange de données dans les projets de recherche impliquant plusieurs centres et/ou plusieurs partenaires a conduit à un changement de paradigme dans le système de partage de données. Classiquement, des infrastructures centralisées sont créées pour stocker, traiter ou archiver des informations [MA+19].

À l'ère de la RP GD, ces structures pourraient ne plus être adaptées aux projets de santé collaboratifs en raison de la réglementation sur la confidentialité des données sensibles. Pour relever le défi d'accéder à ces données et de les utiliser tout en assurant la sécurité des données, il devient indispensable de créer des bases de données distribuables sur site, reliées à un intégrateur de calcul, où chaque centre intègre ce module local (base de données + intégrateur). Ensuite, les centres collectent, stockent et contrôlent les données de leurs propres patients. Le principe fondateur de l'architecture est qu'aucune donnée individuelle ne circule en dehors des centres.

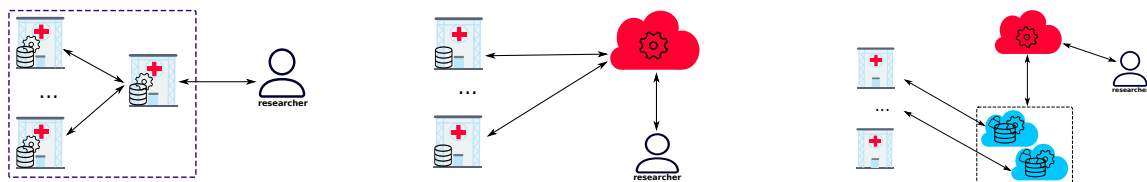
## **Objectifs**

Notre objectif sera de développer ce principe dans les projets multicentriques KTD-INNOV (INNOVation pour le diagnostic de la transplantation rénale) et EU-TRAIN (EUr TRANsp-INnov). Les deux projets sont conçus pour intégrer des données cliniques et biologiques systématiques à grande échelle de patients transplantés d'un rein afin de développer et de valider une application de médecine de précision. Cette application sera donc connectée à toutes les bases de données sans passer par une base de données centralisée, et

uniquement pour collecter des résultats populationnels résumés. Les nouveaux défis consistent ensuite à appliquer cette approche à des calculs ou à des algorithmes plus complexes. La solution de base de données distribuée consolide la sécurité des données et facilite la collaboration sur des projets de recherche multicentriques, dans lesquels chaque centre peut contrôler et rendre compte de l'utilisation des données par leurs propres patients.

Dans le cadre de cette thèse, nous visons le développement de deux nouveaux concepts qui sont cruciaux pour la conception et l'implémentation d'analyses médicales coopératives réparties :

1. **Une méthode guidée par des architectures réparties** pour la conception d'analyses biomédicales. Les trois illustrations suivantes montrent, à titre d'exemple, trois instances avec calcul et stockage de données localement auprès de hôpitaux (à gauche), dans des nuages communautaires (en rouge, au milieu et à droite) ou aussi à l'aide de nuages publics (en bleu, à droite)



Dans un premier temps, les partenaires choisissent une architecture appropriée pour la coopération entre chercheurs. Cette architecture fixera ses caractéristiques principales : localisation de données et des calculs, leur mise à l'échelle et les politiques globales de sécurité et de la protection des données.

2. **Une notion de bio-conteneurs sécurisés** pour leur implémentation et gestion. Ces bio-conteneur permettent l'exécution d'algorithmes à différents endroits tout en encapsulant et protégeant des données critiques contribuées par ex. par un site pour une analyse exécutée sur un autre site.

## Travaux

L'étudiant.e sera en charge de :

- 1) Développer une méthode pour la définition d'une *coopération médicale multi-sites à l'aide d'architectures réparties* en étendant nos résultats préliminaires à ce sujet [BGS+19].
- 2) Définir la notion de *bio-conteneur logiciels répartis sécurisés* sur la base de composants Docker. Le modèle de sécurité pour bio-conteneurs ainsi que de la

protection des calculs et des données sera construit en étendant notre modèle pour la sécurisation d'analyses biomédicales [BS18].

- 3) Afin de montrer l'applicabilité de la méthode, l'étudiant.e prendra en main le prototype KiTapp développé par l'équipe ATIP-Avenir du Pr. Gourraud du CRTI. Cet outil permet une mise en contexte d'un patient transplanté de DIVAT au sein d'une population de patient qui lui ressemble.

Il faudra d'abord tester la faisabilité de l'algorithme sur des bases de données distribuées d'abord en local puis dans des configurations réparties. Ensuite, une évaluation du niveau de sécurité au niveau des centres et au moments des transferts de données est nécessaire. Finalement, le choix d'une architecture appropriée et une implémentation à l'aide de bio-conteneurs seront livrés.

- 4) Intégration des architectures et des bio-conteneurs avec les outils de workflow bio-médicaux utilisés dans les projets KTD-INNOV et EU-TRAIN.

## Références

[ABIO15] Agence de la Biomédecine. **REIN 2015 annual report** -Réseau Epidémiologie et Information en Néphrologie

[BGS+19] F-z. Boujdad, A. Gaignard, M. Südholt, W. Garzon, L. Benavides, R. Redon : “**On distributed collaboration for biomedical analyses**”, accepté pour publication à l'atelier international CCGrid-Life, May 2019.

[BS18] F-z. Boujdad, M. Südholt: **“Constructive Privacy for Shared Genetic Data”**, Proc. 8th Int. Conf. on Cloud Computing and Services Science (CLOSER), Feb. 2018.  
<https://hal.archives-ouvertes.fr/hal-01692620v2>

[DC+17] R. Danger, M. Chesneau *et al.* : **“A composite score associated with spontaneous operational tolerance in kidney transplant recipients”**; Kidney Int. ; 91(6):1473-1481, June 2017  
<https://www.ncbi.nlm.nih.gov/pubmed/28242033>

[MA+19] M. Mendu, S. Ahmed *et al.* : **“Development of an electronic health record-based chronic kidney disease registry to promote population health management”**, BMC Nephrol ; 20(1):72, Mar. 2019  
<https://www.ncbi.nlm.nih.gov/pubmed/30823871>

[NL18] A. Negrouk, D. Lacombe : **“Does GDPR harm or benefit research participants? An EORTC point of view”**; Lancet Oncol. ; 19(10):1278-1280, Oct. 2018  
<https://www.ncbi.nlm.nih.gov/pubmed/30303112>

[SS17] M. Sirota, M. Sarwal : **“Transplantomics: Toward Precision Medicine in Transplantation Research”**, Transplantation. 101(8):1777–1782, Aug. 2017  
<https://www.ncbi.nlm.nih.gov/pubmed/28121910>