

Unification des données par les connaissances

Méthode de construction d'un entrepôt de données historisé de santé guidée par les ontologies

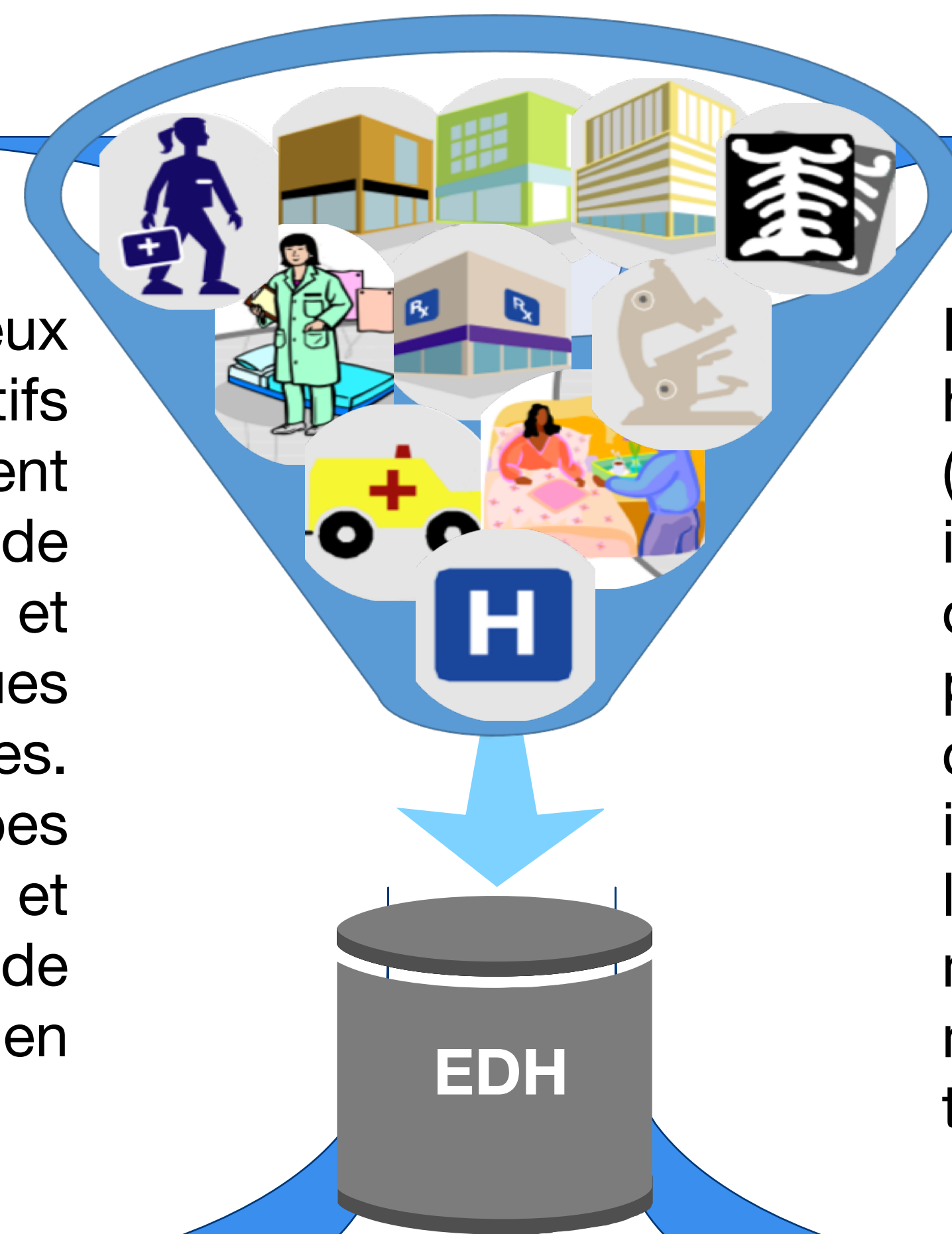
Christina Khnaisser^{1,3}, Luc Lavoie¹, Hassan Diab², Anita Burgun³, Jean-François Ethier^{1,2,3,4}

1 Groupe de recherche interdisciplinaire en informatique de la santé (GRIIS), Département d'informatique, Faculté des sciences, Université de Sherbrooke
2 Centre intégré universitaire de santé et de services sociaux de l'Estrie Centre hospitalier universitaire de Sherbrooke (CIUSSS-CHUS)
3 Sciences de l'information au service de la médecine personnalisée, INSERM UMR 1138 Équipe 22 Centre de Recherche des Cordeliers, Université Paris Descartes
4 Département de médecine, Faculté de médecine et des sciences de la santé Université de Sherbrooke

1 {christina.khnaisser, luc.lavoie}@usherbrooke.ca
2 hdiab.chus@ssss.gouv.qc.ca
2 anita.burgun@aphp.fr
4 jf.ethier@usherbrooke.ca

Mise en contexte

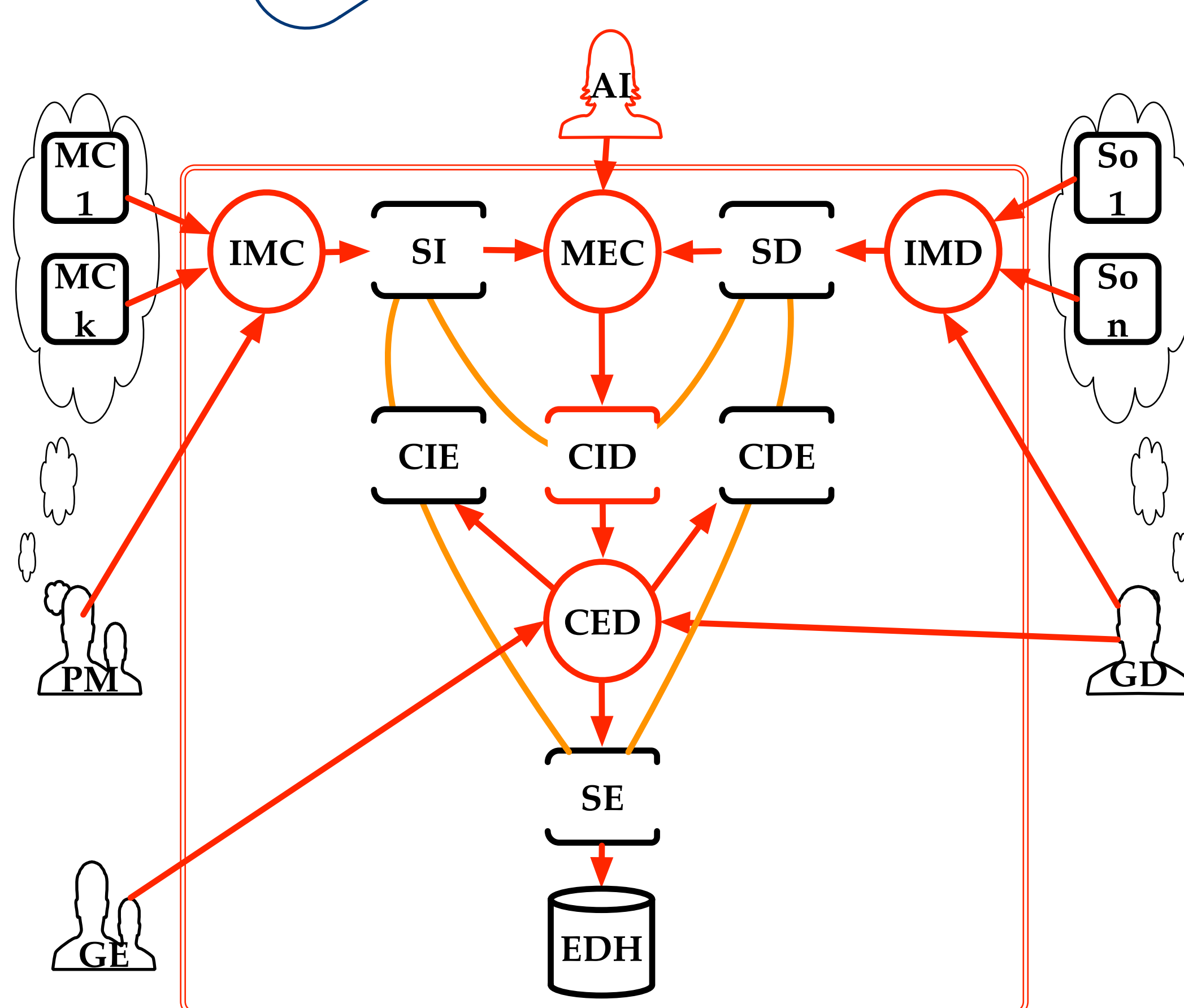
Les systèmes informationnels sont désormais au centre de nombreux secteurs d'activité. Ainsi, dans le secteur de la santé, les dispositifs médicaux, les dossiers cliniques et sociaux, la recherche, etc. génèrent un très grand volume de données chaque jour. Les établissements de santé désirent réutiliser ces données pour différents types d'analyse et les partager d'une façon exploitable. L'efficacité des études cliniques dépendant de la quantité et de la qualité des données recueillies. L'extraction des informations d'intérêt correspondantes à certains types d'études demeure un grand défi vu l'hétérogénéité structurelle et sémantique des sources. Une bonne structuration d'entrepôt de données est nécessaire pour faciliter l'expression des requêtes en faisant abstraction de cette complexité.



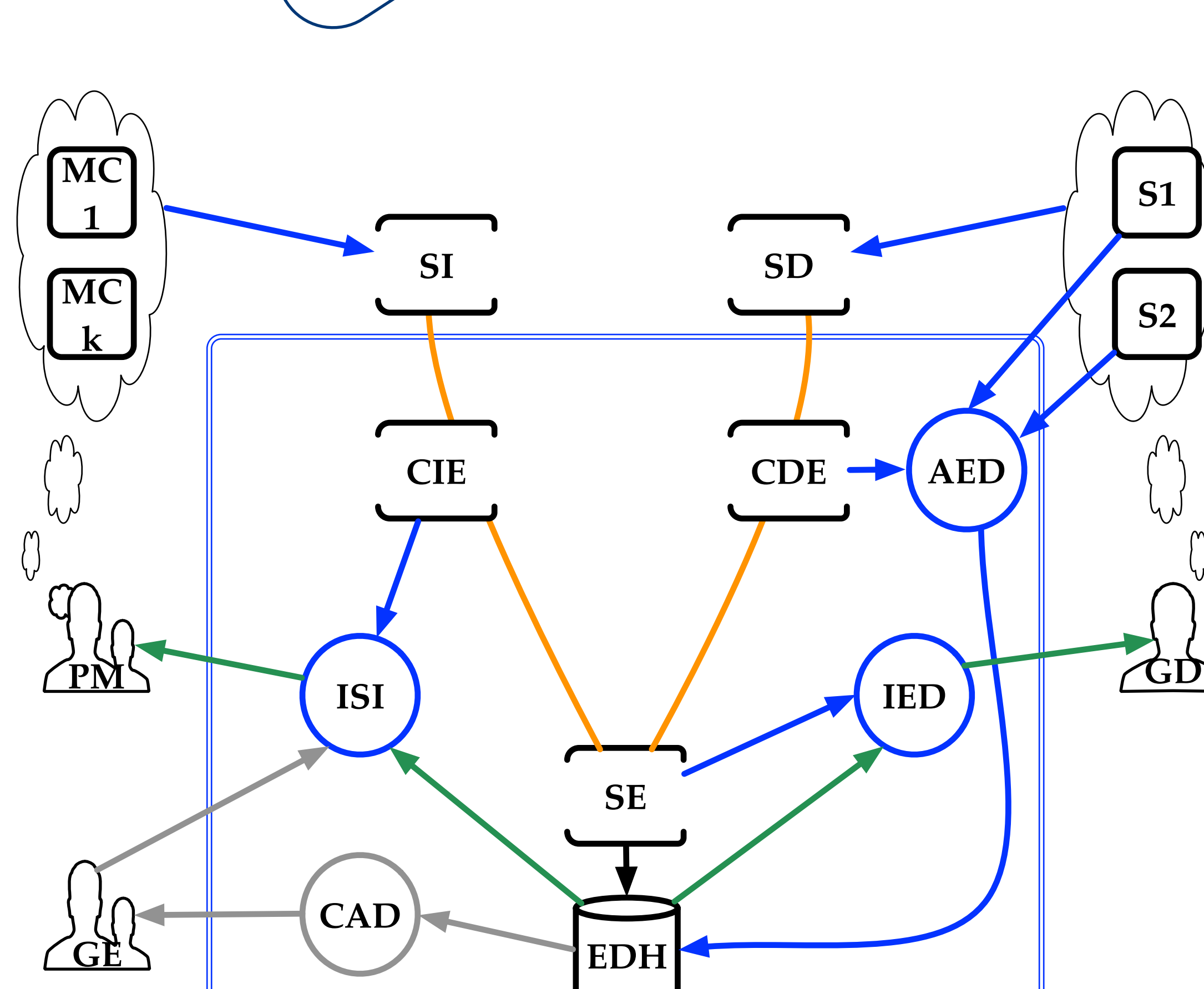
Problématique

L'approche préconisée consiste à construire un entrepôt de données historisé (EDH) en se basant sur un modèle de connaissance (ontologie) pour faciliter l'intégration des sources, l'extraction des informations et l'expression des requêtes. Cependant, la mise en place d'un EDH demeure en grande partie manuelle et coûteuse. Plus précisément, les différents artefacts nécessaires pour obtenir un EDH complet et fonctionnel sont obtenus par diverses méthodes non intégrées qui nécessitent un effort d'intégration considérable. En plus, les méthodes existantes sont basées sur des règles ad hoc avec un minimum d'historisation et de gestion des données manquantes. Ces méthodes limitent l'expressivité des requêtes et l'interopérabilité temporelle et donc à l'exploitation de données historiques.

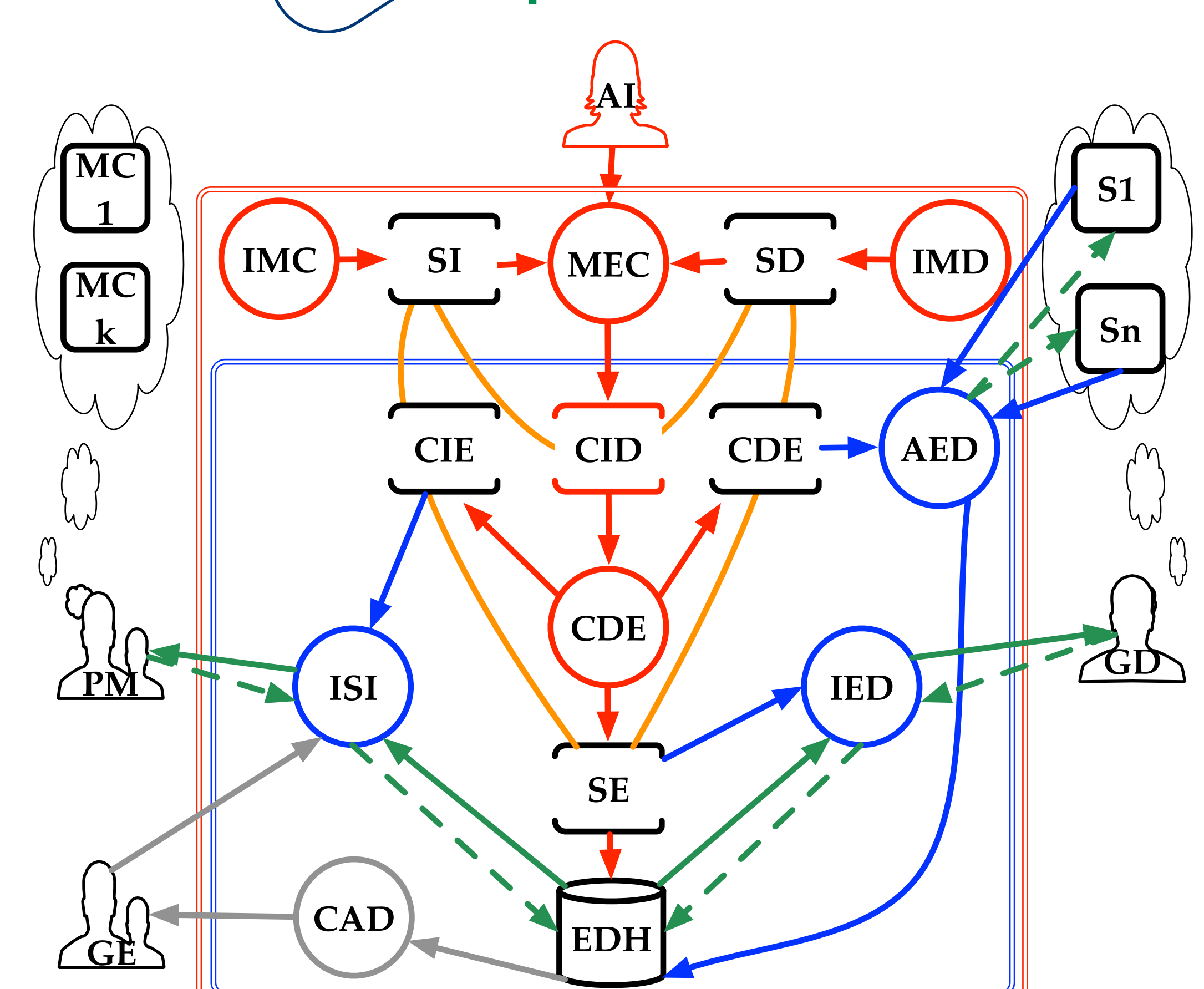
1 Construction



2 Consultation



3 Exploitation



Processus

AED : alimentation de l'entrepôt de données
CAD : contrôle d'accès aux données
CED : construction de l'entrepôt de données
IMC : importation du modèle de connaissances
IMD : importation du modèle de données
MEC : mise en correspondance
ISI : interaction système informationnel
IE(D) : interaction entrepôt de données

Dépôts

CID : correspondant information-données
CIE : correspondant information-entrepôt
CDE : correspondant données-entrepôt
MC : modèle de connaissances
SD : schéma de données
SI : schéma d'information
SO : source de données

Agent

AI : Analyste informatique
PM : personnel médical
GD : gestionnaire des données sources
GE : gestionnaire de l'entrepôt

Références

Khnaisser, C. 2016. Méthode de construction d'entrepôt de données temporalisé pour un système informationnel de santé. <http://savoirs.usherbrooke.ca/handle/11143/8386>.

Khnaisser, C., Lavoie, L., Diab, H., and Ethier, J.-F. 2015. Data Warehouse Design Methods Review: Trends, Challenges and Future Directions for the Healthcare Domain. In: T. Morzy, P. Valduriez and L. Bellatreche, eds., *New Trends in Databases and Information Systems*. Springer International Publishing, 76-87.

Khnaisser, C., Lavoie, L., Burgun, A., and Ethier, J.-F. 2017. *Unified Bitemporal Historization Framework*. Université de Sherbrooke (GRIIS), Sherbrooke, Québec, Canada.

Objectifs

Le but est d'élaborer une méthode semi-automatisée de construction d'un EDH guidée par les ontologies en respectant les contraintes découlant de son intégration à un système informationnel. Plus précisément, il s'agit :

- de proposer un modèle ontologique temporel qui permettra l'arrimage optimal entre l'approche relationnelle et l'approche ontologique réaliste;
- de proposer une méthode et un atelier de construction d'un entrepôt de données historisé à partir d'une ontologie de référence;
- de prouver l'efficacité de la méthode en l'appliquant à un cas réel dans le domaine de la santé.

Conclusion

Le système devrait permettre d'automatiser certaines décisions fondées sur des critères fondamentaux (théorie relationnelle), des critères de conception (normalisation, bitemporalité, etc.), de critères technologiques (choix du SGBDR d'hébergement, des primitives de temporalisation, du dialecte SQL, etc.). L'approche semi-automatisée rendue possible par les ateliers est innovante dans le sens où, jusqu'à présent, les EDH du domaine de la santé ont été développés manuellement selon des règles ad hoc. Elle le sera également par la prise en compte d'un modèle temporel unifié et par l'intégration des contraintes du domaine provenant du modèle de connaissances.