



Développement d'un outil informatique pour la collecte rapide de spécimens biologiques.

Gauvreau, Danny^{1,2}; Djiana, Rose^{3,4}; Thibeault, Denis^{3,4}.

¹ Département de biochimie, Centre hospitalier universitaire de Montréal (CHUM), Montréal, QC.

² Service de biochimie, Hôpital de Verdun, CIUSS du Centre-Est de l'île de Montréal, Montréal, QC.

³ Division de biochimie médicale, Centre universitaire de santé McGill (CUSM), Montréal, QC.

⁴ Département de médecine, Université McGill, Montréal, QC.

Résumé

Objectif : Proposer une procédure efficace pour repérer rapidement des échantillons de patients présélectionnés, sur une chaîne analytique à petite, moyenne ou grande échelle afin de réaliser des études de corrélation. **Méthodes :** À l'aide de notions de fouille de données (« Data Mining »), du système informatisé de laboratoire (*Cerner Millenium*), du système de gestion des spécimens (*Power Express Cennexus* de la compagnie Beckman Coulter Inc.) et d'un logiciel de base de données, nous proposons un modèle schématique de procédures à effectuer pour retrouver efficacement des spécimens sur une chaîne analytique. De plus, à l'aide du logiciel Microsoft® Access, nous avons créé un outil informatique afin de retrouver efficacement les spécimens voulus pour effectuer des études de corrélation. Des critères sont à déterminer au préalable selon la situation locale qui se présente et peuvent comprendre : les analyses touchées, le nombre de spécimens voulus, la gamme de valeurs attendues, etc. **Résultats :** La procédure proposée est constituée de cinq étapes distinctes : 1- Extraction des données du SIL et du logiciel de gestion des échantillons; 2- Importation des données de chaque système dans une table indépendante du logiciel Microsoft® Access; 3- Épuration des données recueillies à l'aide de requêtes selon les critères déterminés au préalable et création d'une table des données épurées pour chaque système; 4- Liaison des tables à l'aide d'un/des champs commun(s) pour l'identification des échantillons biologiques; 5- Création d'une liste d'étiquettes des échantillons sélectionnés et retenus pour le ramassage de spécimens. Une macro peut être créée afin d'automatiser le processus. **Discussion :** Lors d'un processus de corrélation à petite échelle, le temps technique épargné est moins significatif comparé à une procédure manuelle de récupération de spécimens. Cependant, à moyenne et grande échelle, le temps technique épargné corrèle de façon négative avec l'ampleur du nombre de spécimen à rechercher. **Conclusion :** Avec la centralisation des laboratoires, beaucoup d'études de corrélations à grande échelle devront être réalisées. La procédure décrite de l'outil informatique développé et présenté ici s'applique à toutes les échelles d'étude et peut générer des économies significatives en temps techniques.

Introduction

Le laboratoire diagnostique génère une énorme quantité de données patients. Plusieurs problèmes qui se présentent dans un laboratoire peuvent être résolus à l'aide de ces données. Le **problème exposé ici**, un exemple parmi tant d'autres, est de retracer des échantillons de patients afin d'effectuer une **étude de corrélation** lors d'un changement d'appareils et/ou de méthodes. Selon l'ampleur de l'étude à réaliser, cette dernière peut être d'envergure colossale et prendre un temps technique très considérable.

Le concept de fouille ou forage de données (« Data Mining ») consiste à utiliser de grandes bases de données afin de découvrir de nouvelles connaissances (*Knowledge discovery in database - KDD*)¹. La découverte de nouvelles connaissances à l'aide de base de données est représentée par un processus en plusieurs étapes qui consiste en :

1. L'analyse du problème d'application;
2. La sélection des données;
3. Le prétraitement des données;
4. L'apprentissage (fouille de données);
5. L'évaluation et l'interprétation des résultats.

Méthode

Nous avons utilisé une approche qui s'apparente au concept de KDD et de son processus afin de créer un schéma de procédure pour répondre à la problématique de la recherche de spécimens biologiques afin de réaliser simultanément une ou plusieurs études de corrélation au CUSM.

Les données utilisées proviennent du système informatisé de laboratoire (SIL), *Cerner Millenium*, et du système de gestion des spécimens, *Power Express Cennexus* de la compagnie Beckman Coulter. À l'aide du logiciel Microsoft® Access, qui nous sert à la fois d'entrepôt de données et nous permet le prétraitement et la sélection des données. Des critères de sélection sont à déterminer au préalable selon la situation qui se présente et ces derniers fourniront les indicateurs pour la création des requêtes de sélection et tableaux croisés pour le prétraitement des données :

- Les analyses touchées (Ex: TSH, fT₃ et fT₄);
- Le nombre d'échantillons désirés (Ex: minimum 40, en accord avec le guide du CLSI EP-9A2);
- La gamme de valeurs attendues (à l'intérieur et à l'extérieur des valeurs de référence (V.R.) attendues, en dehors de la linéarité de la méthode et le seuil de décision médicale).

Résultats

La **procédure proposée** est constituée de **cinq étapes** distinctes :

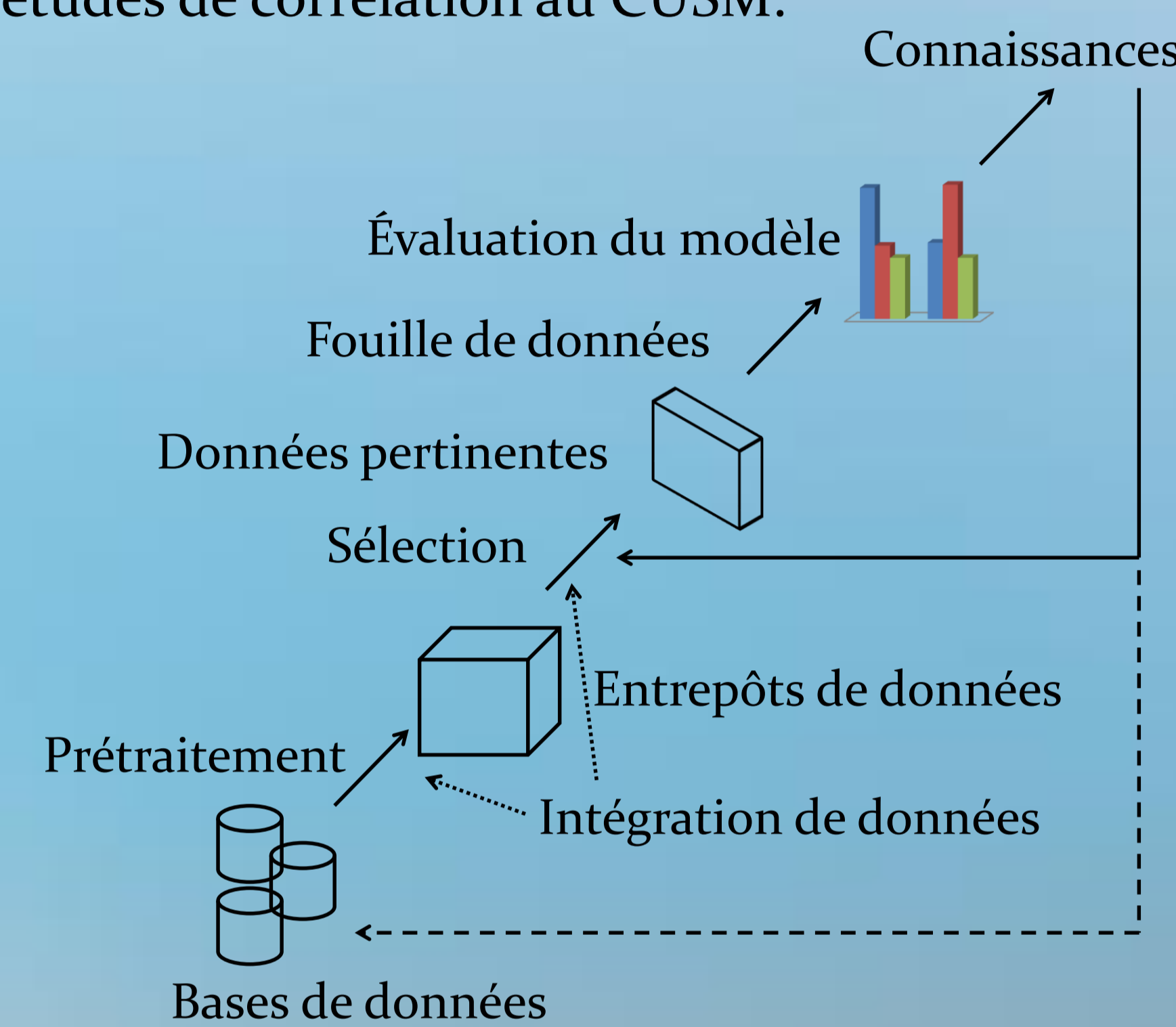
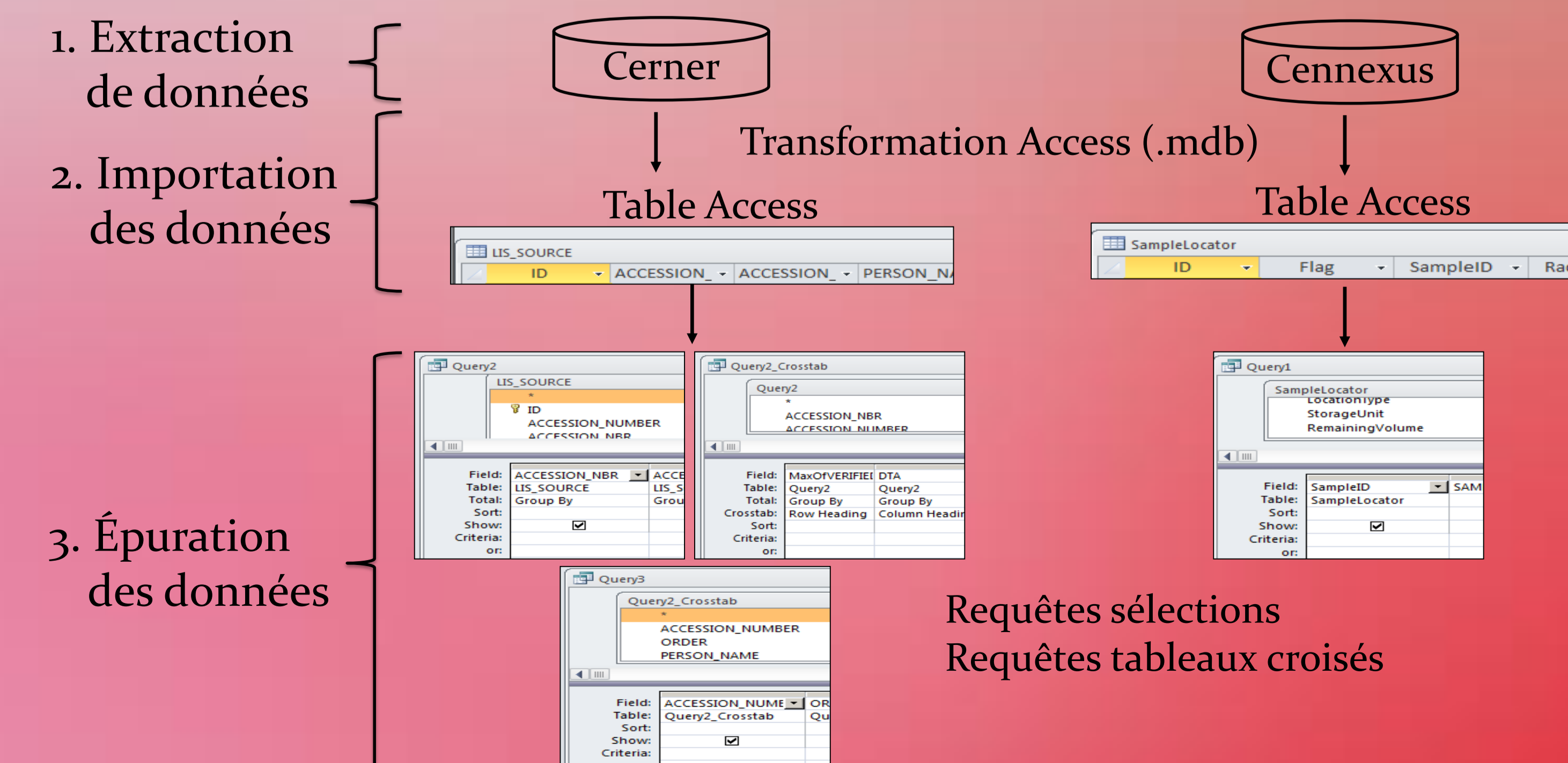
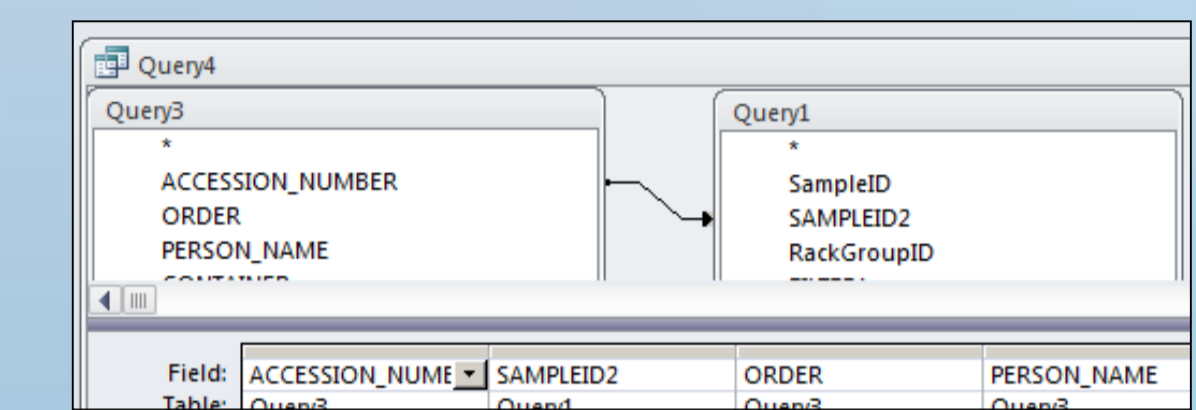


Figure 1: Schéma des étapes qui composent le principe de KDD¹.

Résultats (suite)

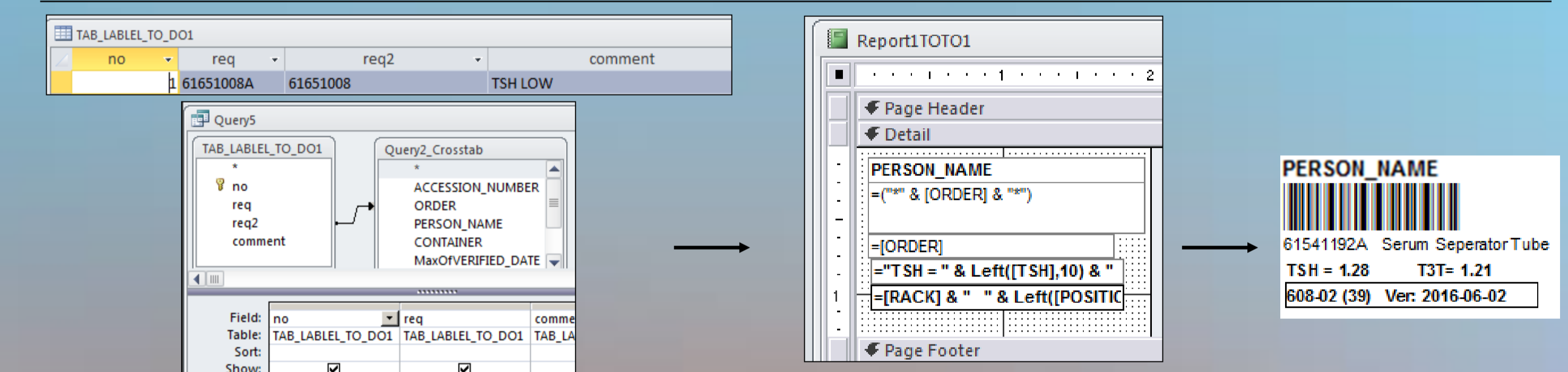
4. Liaison des tables



5. Création d'une table de spécimens retenus et encore disponibles dans le laboratoire, avec leur emplacement et création d'une liste d'étiquettes pour aliquotage. Un tri d'impression des étiquettes ordonné selon les racks immenses de tubes est appliqué pour une collecte des tubes plus efficace.

Tableau 1: Nombres de tubes sélectionnés à rechercher dans le frigo.

Abréviations	ACE [1 (8 - 52) 120]	ACETA [17 (66 - 135) 1324]	AFI [0 (1.2 - 1.53) 200]	AFP [0,5 (0 - 500) 3000]	...	VANCO [2,5 (5 - 40) 100]	Total (84 analyses)
Linéarité haute	4	2			...		229
> V.R.	19	5	2	4	...	3	1226
< V.R.	22		25		...	8	565
Normale	40		1		...	44	475
Total d'échantillons	85	7	28	4	...	55	2495



* Une macro peut-être créée afin d'automatiser le processus.

Discussion

Lors d'une étude de corrélation à petite échelle, le temps technique épargné est moins significatif comparé à une procédure manuelle de récupération des spécimens. Par contre, à moyenne et grande échelle, le temps technique épargné corrèle négativement avec l'ampleur du nombre de spécimens à rechercher. Compte tenu des ressources humaines de plus en plus limitées au laboratoire, l'efficacité à identifier et à collecter les tubes nécessaires à une étude devient essentiel. La découverte de connaissance à partir de base de données (KDD) peut être appliqué à plusieurs autres problématiques du laboratoire de diagnostique autre que les études de corrélation.

Conclusion

Avec la centralisation des laboratoires, beaucoup d'études de corrélations vont devoir être réalisées et la procédure décrite ci-dessus et l'outil informatique créé fonctionne à toute les échelles et peut générer des économies de temps techniques très significatif.

Références

1. Fayyad, U. M.; Piatetsky-Shapiro G. and Smyth P. 1996. From Data mining to knowledge discovery in databases. *AI Magazine* 17(3): 37-54