

Constitution d'un catalogue commun à plusieurs CRLCC

Un projet de la FNCLCC

Emmanuel Reyrat, Jean Genève

► La Fédération des Centres de Lutte Contre le Cancer (FNCLCC) a entrepris, il y a 2 ans, la mise en place d'un système d'information permettant le partage de données issues des tumorothèques gérées par les vingt Centres Régionaux de Lutte Contre le Cancer (CRLCC). Cette application se fonde sur la transmission sécurisée d'un ensemble de renseignements de base sur les échantillons conservés dans les tumorothèques. ◀



E. Reyrat : Responsable informatique de la FNCLCC.
J. Genève : Directeur du Bureau d'Études cliniques et thérapeutiques de la FNCLCC.
Fédération des Centres de Lutte Contre le Cancer (FNCLCC),
101, rue de Tolbiac, 75654 Paris
Cedex 13, France.
reyrat@fnclcc.fr

Une volonté stratégique

La Tumorothèque Virtuelle Fédérale répond, avant tout, à un besoin qu'éprouvent les scientifiques des CRLCC (Centres Régionaux de Lutte Contre le Cancer), d'être en mesure de connaître, à tout moment, le contenu des tumorothèques de l'ensemble des Centres afin de pouvoir réaliser des études prospectives sur les prélèvements cliniquement annotés.

En outre, les Centres profitant de leur regroupement au sein d'une Fédération structurée et de leur habitude d'échange de données, souhaitaient être force de proposition dans ce type de structure, notamment vis-à-vis de projets similaires. C'est ainsi que la FNCLCC (Fédération des Centres de Lutte Contre le Cancer) a été amenée à participer à divers groupes de travail de l'Institut National du Cancer, sur la mise en place d'un catalogue national des tumorothèques s'appuyant, au niveau régional, sur les Cancéropôles.

De plus, en des temps où l'on parle beaucoup d'échanges de données médicales (dossier médical personnel, dossier communiquant en cancérologie...), il était important de prouver qu'il était possible de partager des informations et d'échanger des données structurées.

Les diverses étapes à franchir...

La première étape à franchir pour lancer un tel projet était de dresser la liste des informations à partager et de déterminer le format d'échange.

Pour répondre à ce besoin, nous avons bénéficié des travaux d'un groupe de travail piloté par la DHOS (Direction de l'Hospitalisation et de l'Organisation des Soins), dans le cadre des actions liées au lancement des appels d'offres « Tumorothèques » au niveau national. Ce groupe de travail a été chargé de définir une liste d'items, qui constitue le plus petit dénominateur commun permettant d'identifier, sans ambiguïté, un échantillon provenant d'un patient donné par rapport à un autre échantillon. Cette liste n'est limitée que dans le cadre d'échanges d'informations. Chaque gestionnaire de tumorothèque peut néanmoins recueillir des informations complémentaires qu'il juge nécessaires et utiles. Mais l'expérience montre qu'il est déjà difficile de collecter l'ensemble des informations constituant cette liste minimale.

Dans un deuxième temps, un groupe de travail constitué de professionnels des Centres (informaticiens et biologistes...) a rédigé un manuel de codage permettant de

préciser aux correspondants le type de données attendues à partir de ces 40 items et le format de ces données. Il a également fallu mettre en place un processus de contrôle-qualité nécessaire au moment de l'intégration des informations dans la base centrale.

Puis, un format d'échanges qui se veut « universel » a été choisi. Les correspondants doivent créer un fichier-texte avec séparateur. Ceci permet totalement de s'affranchir des systèmes locaux de gestion des tumorothèques. En effet, n'importe quel outil, du simple tableur au système le plus sophistiqué, est en mesure de permettre une extraction de données dans un tel format. Ainsi, chaque gestionnaire de tumorothèque reste (et doit rester) maître du choix de l'application qu'il veut utiliser.

En effet, si dans le cadre de la gestion d'une tumorothèque, il faut insister sur l'aspect « qualité » du prélèvement, il ne faut surtout pas oublier également la notion de qualité de l'information qui définit ce prélèvement.

Si la qualité du prélèvement ne dépend pas de notre système de recueil de données et doit être contrôlée au niveau du service qui gère ces échantillons, en amont de la transmission de données, nous avons dû nous préoccuper de la qualité des données transmises.

Pour ce faire, un processus de contrôle - avant intégration dans la base de données - a été défini et sa mise en place doit progressivement évoluer et s'enrichir de nouveaux paramètres. Ce processus permet déjà de vérifier la cohérence d'un certain nombre d'items existant dans les classifications internationales : code topographique, code histologique, TNM, date de prélèvement pertinente par rapport à la date de consultation des données, etc.

De simples contrôles comme ceux-là, s'ils risquent de jouer un rôle néfaste sur l'exhaustivité des données (une fiche rejetée risque de ne pas être corrigée, ni renvoyée, donc de ne pas être intégrée) permettent de garantir que les données intégrées sont d'une qualité suffisante pour permettre leur exploitation.

Au final, à partir de ces fichiers et au travers de ces contrôles, il a été possible de constituer une base de données sur un serveur spécifique installé dans les locaux de la FNCLCC.

Comment est structurée l'application ?

Pour accéder à l'application, un Centre doit s'identifier par un couple « login-mot de passe ». Cette identification joue un double rôle : la sécurité des échanges et le lien entre la personne « identifiée » et les données auxquelles elle peut avoir accès.

Il est à noter que la sécurité est également assurée par un protocole de cryptage des échanges et par un système de contrôle de l'adresse IP du poste accédant à l'application.

Une fois identifié, l'utilisateur affiche un menu lui offrant quatre choix :

• Téléchargement du fichier

La personne qui transmet les données est amenée à sélectionner, sur son poste de travail, le fichier-texte contenant les données à transmettre.

Ce fichier fait, ensuite, l'objet d'un double contrôle en local :

- validité des enregistrements (l'ensemble des champs requis est bien présent pour chaque enregistrement transmis) ;

- lien entre la personne qui transmet les dossiers et l'établissement précisé dans chaque enregistrement et auquel seront rattachées les données transmises (pour éviter qu'une simple erreur d'identification du Centre puisse endommager l'ensemble des données d'un autre établissement).

Lorsque les fiches ont passé ce premier contrôle local, elles sont envoyées vers le serveur par une liaison cryptée et font l'objet d'un ensemble de contrôles de contenu (code diagnostique, TNM, morphologie, validité des dates...). À l'issue de ces contrôles, les fiches sont effectivement intégrées dans la base de données, ou rejetées. Dans ce dernier cas, le motif du rejet doit être précisé.

• *Visualisation des données caractérisant un prélèvement*
Un Centre peut consulter le contenu de chacun des enregistrements qu'il a envoyés (et qui ont été intégrés) pour valider l'information. Il ne peut en aucun cas visualiser les données d'un autre Centre.

• *Visualisation et téléchargement de la liste des fiches rejetées accompagnée du motif de chaque rejet*

Cette fonctionnalité permet de faciliter la correction de l'information au niveau de la tumorothèque elle-même, correction qui doit être suivie d'une nouvelle extraction et d'un renvoi des fiches pour une nouvelle intégration.

• *Visualisation de tableaux de synthèse dénombrant, selon des critères prédéfinis (localisation des tumeurs...), les échantillons contenus dans la base*

Cette fonctionnalité est complétée par la possibilité d'extraire les informations dans un tableur ou un traitement de texte, de façon à pouvoir les exploiter très facilement.

... Et les premières difficultés à surmonter

Outre la difficulté de collecter des données fiables, il est d'ores et déjà possible de dresser une liste, non exhaustive, des problèmes rencontrés lors de la mise en œuvre d'une telle application :

• Extraire d'une base de données un fichier-texte demande, dans certains cas, l'intervention d'un professionnel de l'informatique qui devra, si nécessaire, créer un petit programme d'extraction des données de la tumorothèque et, éventuellement, du dossier médical (pour les données cliniques). Or, des impératifs plus urgents (CCAM, T2A) peuvent les contraindre à différer leur intervention.

• Certains établissements ont mis en place très récemment une gestion locale structurée de leur tumorothèque. Ils n'ont donc pas forcément de données à transmettre dans l'immédiat.



• L'ensemble des acteurs n'est pas forcément en mesure, aujourd'hui, de transmettre ce minimum requis d'items définis par le groupe de travail. Bien sûr, cette difficulté n'est que passagère et pourra être résolue dans un proche avenir dans la mesure où chacun devra être en mesure de transmettre l'ensemble de ces items aux cancérôpôles dans le cadre d'autres projets.

• Une autre difficulté potentielle est le « partage » d'une tumorothèque par deux ou plusieurs structures diverses (CLCC et CHU par exemple) entraînant la nécessité de déterminer, à plusieurs, des règles du jeu pour l'extraction de données et leur transmission à une tierce entité. La difficulté est ici liée à l'identification du propriétaire de l'information.

• Les notions éthiques et juridiques (information des patients, utilisation non-commerciale de telles données...) doivent être prises en compte dans la mise en œuvre de ce type de partage. Cette difficulté devrait être très vite levée, quand un système de contrat entre les acteurs sera mis en place.

• Enfin, un tel système n'est possible et viable que s'il existe une réelle volonté de partage des données, et donc un réel intérêt à partager ces données. Il faut donc, rapidement, définir des règles d'accès et d'utilisation de ces données et des échantillons auxquels elles sont associées.

D'autres projets existent

La FNCLCC n'est pas la seule à travailler sur un tel système. L'Institut National du Cancer a constitué des groupes de travail pour travailler, entre autres, sur la mise en place d'un catalogue national des tumorothèques, système de recueil de données s'appuyant sur les cancérôpôles.

Ce futur catalogue national des tumorothèques s'appuiera sur les différents cancérôpôles dont l'organisation peut se décliner selon trois modèles possibles :

• Un cancérôpôle n'a pas encore mis en place un système de recueil régional centralisé des données des tumorothèques. Chaque établissement gérant sa propre tumorothèque envoie, de façon directe, ses données vers la base centrale de l'Institut.

• Un cancérôpôle a mis en place une base régionale de centralisation des données. Chaque établissement gérant sa propre tumorothèque transmet au cancérôpôle, au travers d'un système de recueil des données, ces fiches descriptives d'échantillons dans une base centrale régionale. C'est le cancérôpôle qui, dans un deuxième temps, envoie les données à l'Institut du Cancer.

• Un cancérôpôle (ou une autre entité régionale) centralise les données des tumorothèques des établissements. Il n'existe plus de données gérées localement par les établissements. Ce système transmet ensuite les informations vers l'Institut du Cancer.

Ce schéma de recueil de données reste totalement compatible avec le système de la FNCLCC et, bien que les finalités soient proches, chaque application sera en mesure de répondre aux besoins propres de chacun (Figure 1). Notre fierté est d'avoir montré que le modèle fédéral était viable, et que, malgré les difficultés, un tel système pouvait être mis en œuvre.

Conclusion

Comme nous venons de le voir, toutes les difficultés n'ont bien sûr pas été levées.

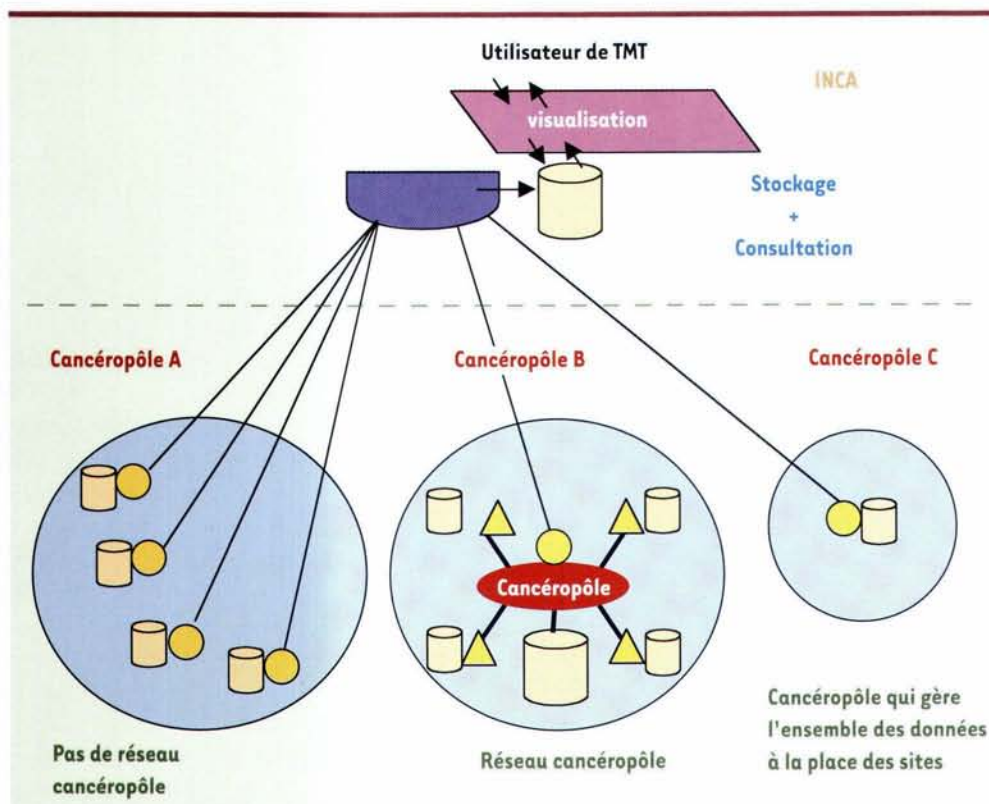


Figure 1. Recueil de données.

Mais nous pouvons considérer que nous avons rempli au moins notre premier objectif : montrer qu'un tel système de recueil de données des tumorothèques est réalisable.

Il ne nous reste plus qu'à réussir le pari d'intégrer prochainement les données de l'ensemble des vingt Centres.

Nous pourrions alors débiter l'application directe de cette base, à savoir son utilisation à des fins scientifiques, conduites d'études génomiques adossées à des données cliniques dans le cadre d'essais rétrospectifs et prospectifs de cohortes ou de recherche clinique chez des patients ayant une même pathologie. Cela met l'accent sur l'importance de la « flexibilité » du contenu de la base qui doit être prête à accueillir des données supplémentaires propres à chaque pathologie (typage Her2 pour le cancer du sein par exemple).

Ces études clinico-génomiques pourraient être élaborées par un ou plusieurs CRLCC en réponse à des appels d'offre animés par un bureau scientifique indépendant. ♦

SUMMARY

Production of a common catalogue for several « Centres Régionaux de Lutte Contre le Cancer » : a « Fédération des Centres de Lutte Contre le Cancer » project

The « Fédération des Centres de Lutte Contre le Cancer » (FNCLCC) has been undertaken 2 years ago, the set-up of an information system allowing the partition of data from tumour libraries managed by the twenty « Centres Régionaux de Lutte Contre le Cancer » (CRLCC). This organization was established on a secure transmission of basis informations about samples kept in tumour libraries. ♦

TIRÉS À PART

E. Reyrat