Avis sur l’externalisation d’une étape de la phase analytique des examens s’appuyant sur le séquençage de l’ADN

Avant propos

L'externalisation d'une étape de la phase analytique des examens s'appuyant sur le séquençage, et en particulier de la partie bioinformatique, est un sujet dont les sociétés savantes se sont emparées récemment. Au regard de la complexité du sujet, nous avons constitué un panel d'experts (bioinformaticiens et biologistes) impliqués dans les analyses de biologie médicale en France. Il s'agissait de construire une réflexion argumentée sur i) la maîtrise des risques liés aux étapes impliquant des analyses bioinformatiques, ii) les différents modèles d'organisations qui pourraient exister pour la réalisation de ces examens de séquençage. L'objectif de ce document est de poser un avis consensuel validé par ce panel d'experts.

Avis

L'analyse biologique par séquençage haut débit appliqué au diagnostic réalisé par les laboratoires de biologie médicale (LBM) peut être décomposée en plusieurs étapes regroupées conceptuellement en 2 spécialités imbriquées : la partie "*wet*" (e.g. extraction des acides nucléiques, création des librairies, séquençage, validation d'une mutation identifiée) et la partie "*dry*" (e.g. élaboration du design et du panel de gènes, préparation de la feuille de travail, génération des données brutes, validation des métriques, traitement et analyse des données), définissant la "phase analytique". Un courrier de la Direction Générale de la Santé du 22 mai 2023 précise que "la sous-traitance n'est possible que pour l'ensemble de la phase analytique et non pour une partie seulement", et rend ainsi cette phase analytique indivisible.

De plus, ce même courrier précise que "la sous-traitance à un prestataire qui n'est pas un LBM contrevient à la loi". Autrement dit, seul un LBM est en capacité de garantir la maîtrise de la phase analytique, et ceci pour toutes les étapes. En effet, les étapes "*wet*" et "*dry*" présentent des risques similaires qui ne peuvent être maîtrisés que dans le cadre d'un LBM : il convient donc de considérer strictement en miroir un prélèvement et une donnée, un automate et un serveur de calcul, un matériel de laboratoire et un logiciel, un réactif et une base de données, et enfin un personnel habilité par le LBM (par exemple un ingénieur en biologie et en bioinformatique). De par sa nature dématérialisée, la donnée numérique apparaît *de facto* plus vulnérable que le prélèvement biologique, augmentant ainsi le risque inhérent à son mésusage.

Dans le cadre de certains modèles d'organisation, des risques critiques spécifiques à la partie "*dry*" peuvent alors être identifiés comme étant impérativement à maîtriser par un personnel habilité par le LBM : i) risque de non maîtrise du cycle de vie du traitement bioinformatique (mise à jour non validée des logiciels, des bases de données, des serveurs de calculs), ii) risque de non maîtrise de l'espace des données (accès, sécurisation, politique de sauvegarde et d'archivage, maîtrise du choix de l'infrastructure), iii) risque majeur de détournement des données de santé des patients (modification de la finalité initiale déclarée, notamment en dehors du champ du diagnostic).

Ainsi, il est fondamental de garder la maîtrise intellectuelle (maîtrise des risques et de l'expertise) de l'ensemble des étapes de la phase analytique, et ceci dans le cadre d'un

LBM. Il en découle que seuls les modèles d'organisation internes au LBM sont garants de cette maîtrise intellectuelle. L'organisation coopérative est envisageable avec un cadre légal et des conditions de maîtrise adéquates. De plus, la légalisation de l'externalisation de la partie "*dry*" de la phase analytique entraînerait par contagion la possibilité de l'externalisation de la partie "*wet*" et provoquerait de facto la disparition de la notion de LBM.

Dans ce contexte, les solutions acceptables pour une mise en conformité sont : i) l'internalisation de la maîtrise des risques liées aux solutions de traitement bioinformatique, ii) la coopération entre LBM (la maîtrise intellectuelle reste sous l'autorité des LBM), sous réserve de se conformer aux exigences légales. Ces solutions nécessitent : i) un délai de mise en conformité, ii) un renforcement des compétences et ressources en bioinformatique au sein des LBM.

Cet engagement est un prérequis à un diagnostic de qualité pour le patient.

Définitions

Externalisation
L'externalisation consiste en la délégation de responsabilité de la maîtrise d'une partie d'un processus analytique (e.g. séquençage, gestion des prélèvements ou des données, gestion des automates ou des traitements bioinformatiques) par un personnel non habilité par le LBM.

Internalisation
L'internalisation consiste en la maîtrise de l'ensemble des processus analytiques par un personnel habilité par le LBM.

Données sensibles
CNIL (https://www.cnil.fr/fr/definition/donnee-sensible)
"Les données sensibles forment une catégorie particulière des données personnelles. Ce sont des informations qui révèlent la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique."

Maîtrise intellectuelle
Union de la maîtrise des risques et de l'expertise. La maîtrise des risques est la conception des systèmes de contrôle interne de qualité permettant de vérifier que la qualité prévue des résultats est bien obtenue. Il est important que ce système de maîtrise permette aux membres du personnel d'obtenir des informations claires et faciles à comprendre sur lesquelles baser leurs décisions techniques et médicales (selon la norme NF EN ISO 15189 (§ 5.6.1)).

Traitement bioinformatique
Opération informatique utilisée pour l'analyse de données biologiques numériques, telles que le séquençage de l'ADN, de l'ARN ou des protéines. Un pipeline est une succession de traitements automatisés à l'aide d'une chaîne de logiciels et d'algorithmes, qui comprend tout, ou une partie, de l'analyse informatique des données.

Le personnel habilité par un LBM

Seul le personnel habilité par un LBM, dont les tâches sont encadrées strictement par un système unique de management de la qualité répondant aux exigences de la norme ISO 15189, est en capacité de garantir la maîtrise du processus analytique et celle de l'ensemble de ses composantes.

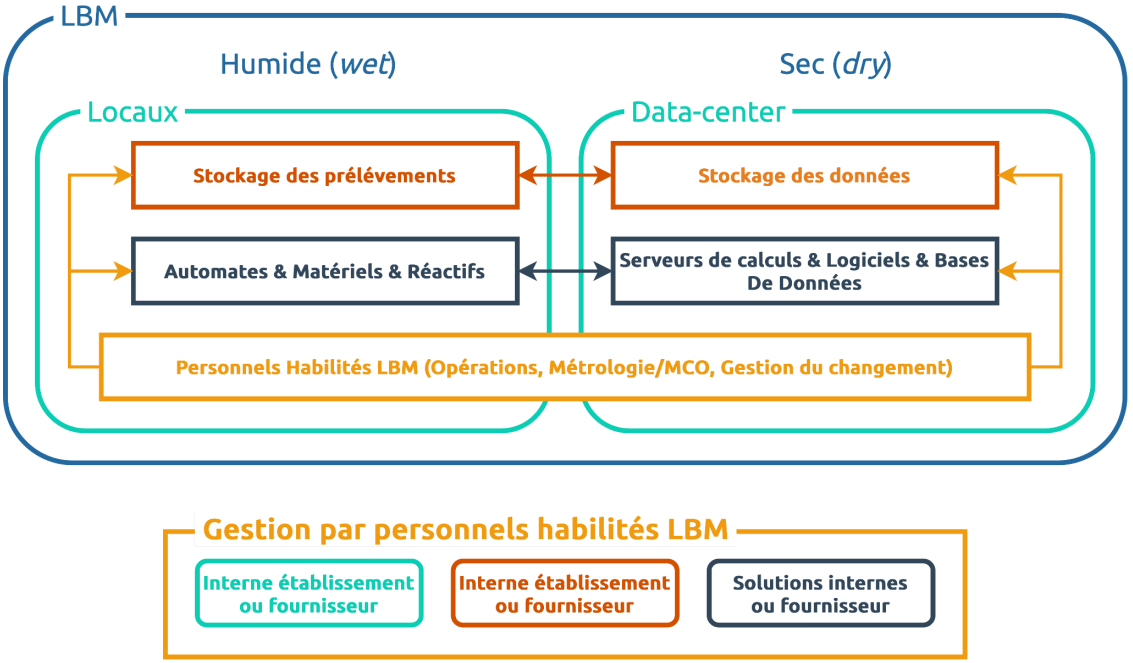
Parmi les analyses de biologie médicale, les méthodes de séquençage à haut débit utilisées en génétique comportent une phase analytique complexe, divisée en plusieurs étapes cruciales. **Cette phase analytique peut uniquement être menée à bien que par la maîtrise de deux compétences complémentaires : la biologie et la bioinformatique.**

Il est essentiel de souligner que la réussite de cette phase analytique dépend non seulement de l'expertise en biologie et en bioinformatique, mais également de la **capacité à maîtriser les risques à chaque étape du processus** sous la responsabilité du biologiste médical responsable du LBM qui supervise l'ensemble du processus.

En effet, **la maîtrise du processus complet**, depuis la réception de la demande jusqu'à la remise des résultats, **est indivisible d'un point de vue de la responsabilité et de la qualification réglementaire d'un LBM.** Cela implique que pour chaque étape, de la préparation de l'échantillon au traitement des données de séquençage, seul le personnel habilité par un LBM est en capacité de garantir la maîtrise du processus analytique et celle de l'ensemble de ses composantes, et ainsi la fiabilité et la validité des résultats obtenus.

Risques similaires entre partie “*wet*” et partie “*dry*”

Par analogie, les éléments de la partie “*wet*” et de la partie “*dry*” doivent être strictement considérés en miroir, particulièrement au regard de la similarité des risques.



Si la partie “*wet*” est définie comme l’ensemble des étapes analytiques correspondant au traitement de l’échantillon, la partie “*dry*” est définie comme les étapes analytiques correspondant aux traitements de la donnée numérique par des moyens informatiques ou bioinformatiques au sein d’un examen de biologie médicale. **Par analogie, cette partie “*dry*” regroupe des éléments équivalents aux éléments devant être maîtrisés par la partie “*wet*” d’une analyse de biologie médicale** :

- les locaux correspondent à l’infrastructure bioinformatique (data-center),
- le stockage d’un prélèvement correspond à un stockage et un archivage des données,
- les automates, matériels et réactifs correspondent aux serveurs de calculs, logiciels et bases de données.

Au regard de la réglementation en vigueur, **la maîtrise des locaux, du stockage des prélèvements, des séquenceurs et autres automates doivent être maîtrisés par du personnel habilité par le LBM. Il en va de même avec tous les éléments de la partie “*dry*”**, comme le datacenter, les serveurs de données, ou les pipelines, par exemple.

Il est également à signaler qu’**une autorisation de maîtrise d’un seul élément de la partie “*dry*” par un organisme non LBM** (sociétés externes, nationales et européennes) pourrait conduire par analogie à l’autorisation de maîtrise de l’ensemble des autres éléments de la phase analytique, y compris la partie “*wet*”. Par contagion, cette pratique **pourrait à terme concerner d’autres examens de biologie médicale, et ainsi faire disparaître le principe de LBM.**

Maîtrise des risques critiques de la partie “*dry*”

Des risques critiques spécifiques à la partie “*dry*” doivent être identifiés comme étant impérativement à maîtriser par un personnel habilité par le laboratoire de biologie médicale.

Traitement bioinformatique des données

La maîtrise du traitement bioinformatique des données de santé est fondamentale. Chaque étape du processus de traitement joue un rôle déterminant dans la fiabilité et la validité des résultats obtenus.

Par conséquent, **le laboratoire de biologie médicale (LBM) doit assurer une maîtrise totale de l'ensemble de la chaîne de traitement**. Cela implique la vérification d’un ensemble d’indicateurs permettant de garantir le bon déroulement du processus.

Plus spécifiquement, **le LBM doit accorder une attention particulière à la maîtrise du cycle de vie du traitement bioinformatique. Tout changement significatif dans une étape de ce processus doit être soumis à une validation rigoureuse par le LBM**. Un changement est considéré comme majeur dès lors qu'il nécessite une adaptation de la part de l'utilisateur par rapport aux résultats du traitement. Il est important de noter qu'un simple test de non-régression ne constitue pas une preuve suffisante de l'absence de changement majeur.

Ainsi, en veillant à la maîtrise constante du traitement des données de santé, le LBM peut garantir la fiabilité et l'intégrité des informations produites, renforçant ainsi la confiance des parties prenantes et assurant la qualité des services rendus.

Exemples :

- Changement de version d’un traitement bioinformatique sans maîtrise de la validation par le LBM
- Changement des composants, y compris des versions, d’un traitement bioinformatique sans maîtrise de la validation par le LBM
- Mise à jour des bases de données sans maîtrise de la validation par le LBM
- Modification des formats d’entrée et de sortie d’un traitement bioinformatique

Espace de données

Dans le cadre d'un LBM, la sécurité des données de santé est fondamentale. L'espace de données, qu'il soit interne ou externalisé dans un cloud, doit être maîtrisé par un personnel habilité par le LBM.

L'externalisation vers un cloud présente des risques en matière de sécurité et de confidentialité des données : confier ces données sensibles à un prestataire externe

implique une perte potentielle de contrôle direct sur la maîtrise de l'infrastructure et les mécanismes de sécurité. Malgré les assurances fournies par certains fournisseurs de services cloud, la supervision et **la maîtrise directe du personnel habilité par le LBM est cruciale pour garantir la conformité aux normes de sécurité**, telles que celles imposées par l'Hébergeur de Données de Santé (HDS), la solution SecNumCloud, ou encore les clauses de sortie en cas de résiliation du contrat, ou simplement en cas d'utilisation complémentaire des données. Par ailleurs, **l'empilement des couches de différents prestataires au sein d'un processus augmente les risques de perte de maîtrise**.

Autant que pour une solution interne, il est nécessaire de maintenir une maîtrise totale de la sécurité des données, en assurant la formation et la sensibilisation du personnel, ainsi que d'assurer la mise en place des mesures de protection adaptées aux spécificités de l'activité médicale.

Exemples :

- Le serveur de données ne garantit pas un haut niveau d'exigence tant du point de vue technique, qu'opérationnel ou juridique (qualification ANSSI, référentiel SecNumCloud)
- Le serveur de données ne permet pas d'accéder à l'intégralité données générées pour le LBM
- Le serveur de données ne permet pas la gestion des droits d'accès et de la sécurité par le LBM

Détournement de données de santé

Le détournement de données de santé représente une violation flagrante des réglementations telles que le RGPD et les directives de la CNIL, engendrant ainsi un risque critique de non-maîtrise des données.

Ce détournement se caractérise par une **altération de la finalité initiale** pour laquelle les données ont été collectées. Il peut survenir lors de l'implication d'un prestataire externe au LBM, qui pourrait utiliser les données à d'autres fins que celles pour lesquelles elles ont été fournies, **mettant ainsi en péril la confidentialité et l'intégrité des informations médicales**. Notons que l'utilisation des données à des fins de recherche et de développement peut constituer un détournement lorsqu'elles étaient initialement destinées à des traitements diagnostiques.

De plus, les lois extraterritoriales telles que le Cloud Act et la loi FISA augmentent le risque de détournement en permettant l'accès des autorités étrangères aux données stockées dans le cloud, sans nécessairement en informer le propriétaire initial des données.

Il est donc impératif pour le LBM de prendre des mesures strictes pour prévenir tout détournement potentiel de données, en choisissant soigneusement ses prestataires externes et en surveillant étroitement l'utilisation qui est faite des informations médicales confiées.

Exemples :

- Utilisation des données pour l'amélioration d'un algorithme (recherche et développement)
- Utilisation des données pour le traitement d'autres analyses (annotations biologiques et cliniques) n'appartenant pas au LBM
- Mise à disposition des données à des fins de recherche clinique ou de recherche translationnelle
- Mise à disposition des données à une autre entité (e.g. UK Biobank partage ses données avec des assureurs)
- Infrastructure cloud HDS basée sur un prestataire sujet aux lois extraterritoriales (e.g. Health Data Hub (HDH) et Microsoft/Azure)

Modèles d'organisation pour la partie "*dry*"

Il est fondamental de conserver la maîtrise de toutes les étapes de la phase analytique au sein d'un LBM, et cela ne peut être assuré que par des modèles d'organisation internes et coopératifs.

Une mise en conformité peut être effectuée en s'adossant sur l'un des modèles d'organisation conforme proposés ci-après. Notons que les modèles d'internalisation sont actuellement conformes vis-à-vis de la loi, et que les modèles de coopération pourraient le devenir si la maîtrise des risques critiques est assurée. **Un délai de mise en conformité est à prévoir.**

Internalisation au sein du LBM

L'ensemble des étapes, incluant les traitements bioinformatiques (développements et cycles de vie) et l'administration des données (stockage, archivage et sécurisation), **sont de la responsabilité directe du LBM, ou d'un service support appartenant à la même entité juridique** que le LBM (Direction des Services Numériques, Service BioInformatique). **L'internalisation consiste donc à maîtriser en interne les éléments et les risques associés par le LBM et les services supports**.

Exemples :

- Service de bioinformatique au sein de l'établissement
- Unité fonctionnelle de bioinformatique au sein du LBM
- Plateforme de séquençage au sein d'un établissement

Une ou plusieurs étapes peuvent être réalisées par une solution bioinformatique proposée par un prestataire externe, et déployée en interne au sein du LBM (ou d'un service support), avec ou sans contrat de maintenance. Dans le cadre de l'étape concernée, la compétence est aux mains du prestataire, mais les éléments (e.g. serveurs de calculs, serveurs de stockages, versions des solutions, données) et les risques associés sont maîtrisés en interne par le LBM ou les services supports.

Exemples :

- Déploiement en interne d'un pipeline d'analyse complet (e.g. installation d'un pipeline, image virtuelle du pipeline, serveur incluant le pipeline).
- Déploiement en interne d'un traitement spécifique (e.g. installation d'un logiciel spécifique à une étape de l'analyse).
- Déploiement en interne d'un outil spécifique (e.g. installation d'un outil additionnel en périphérie du pipeline, comme une interface de visualisation).

Coopération avec un autre LBM

Une ou plusieurs étapes de la phase analytique (y compris un traitement bioinformatique spécifique) **sont réalisées en partenariat avec un autre LBM** qui possède les compétences et la maîtrise des risques. Cette mutualisation peut être réalisée par deux

LBM, décrite sous la forme d'un contrat de coopération permettant de définir la mise en commun des ressources et des moyens, ou possiblement entre 2 entités légales dans une structuration de type GCS (Groupement de Coopération Sanitaire).

Il apparaît intéressant de définir et de référencer des LBM pour lesquels des expertises bioinformatiques spécifiques sont avérées. Une extension du périmètre des LBM de référence pourrait ainsi inclure la dimension bioinformatique.

Exemples :

- Un LBM réalise un traitement bioinformatique spécifique (e.g. analyse RNASeq) pour un autre LBM
- Dans un GCS, 3 LBM se partagent la phase analytique (e.g. extraction du matériel biologique, séquençage de l'ADN et traitement bioinformatique)

Solutions de mise en conformité

En cas de non-conformité d'un laboratoire de biologie médicale (LBM) en matière de traitement bioinformatique et de gestion des données de santé, des solutions de mise en conformité sont possibles.

Dans le contexte de garantir la préservation de la maîtrise intellectuelle au sein des laboratoires de biologie médicale (LBM), plusieurs solutions organisationnelles peuvent être envisagées pour une mise en conformité d'un laboratoire.

Tout d'abord, **lorsqu'un LBM externalise le traitement bioinformatique sans maîtriser les risques, il peut choisir entre deux approches principales**.

La première consiste en l'internalisation, où l'ensemble du processus de traitement des données reste sous son contrôle direct, permettant une supervision étroite et une adaptation rapide aux besoins spécifiques du laboratoire. **Cette internalisation peut faire intervenir des solutions provenant de prestataires externes**. Cela garantit la fiabilité et la confidentialité des informations médicales tout en minimisant les risques liés aux traitements bioinformatiques des données de santé.

La seconde option est la coopération entre les LBM, où ces derniers peuvent aller jusqu'à former un groupement (type GCS) pour partager les coûts et les ressources nécessaires à la mise en place d'infrastructures informatiques robustes et sécurisées. De plus, les LBM qui ont internalisé certaines solutions provenant de prestataires externes, afin de mieux maîtriser les risques associés, peuvent tirer profit d'une coopération renforcée entre eux. En s'unissant, **les LBM peuvent bénéficier d'une expertise collective** tout **en conservant un certain degré d'indépendance** dans la gestion de leurs données et de leurs processus.

Que ce soit par le biais de l'internalisation ou de la coopération entre LBM, y compris par la mise en place de LBM de référence labellisés en bioinformatique, l'objectif demeure le même : **garantir la sécurité, la fiabilité et la confidentialité des données de santé, tout en assurant une gestion efficace des risques et une préservation de la maîtrise intellectuelle** au sein des laboratoires de biologie médicale, au service des patients.

Références

Code de la Santé Publique :
https://codes.droit.org/PDF/Code%20de%20la%20sant%C3%A9%20publique.pdf

CNIL :
https://www.cnil.fr/

RGPD :
https://www.cnil.fr/fr/reglement-europeen-protection-donnees
https://www.economie.gouv.fr/entreprises/reglement-general-protection-donnees-rgpd

Coopération (Ministère du travail de la santé et des solidarités) :
https://sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/biologie-medicale/biologie-medicale-questions-reponses-pour-les-professionnels/article/definitions-et-principes-generaux

GCS :
https://sante.gouv.fr/professionnels/gerer-un-etablissement-de-sante-medico-social/cooperations/article/le-groupement-de-cooperation-sanitaire-gcs

Laboratoire de biologie médicale de référence :
https://sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/biologie-medicale/article/laboratoires-de-biologie-medicale-de-reference

Contributions

Claire BARDEL (Lyon)
David BAUX (Montpellier)
Pierre BLANC (Paris)
Laurent CASTERA (Caen)
Marie DE TAYRAC (Rennes)
Anne-Sophie DENOMMÉ-PICHON (Dijon)
Christophe HABIB (Toulouse)
Antony LE BECHEC (Strasbourg)
Alban LERMINE (Paris)
Jean MULLER (Strasbourg)
Charles VAN GOETHEM (Montpellier)