

Guide pour la rédaction d'un cahier des charges en cas d'externalisation de la partie bioinformatique

GT PAX

Table des matières

Préambule.....	1
Distinction entre LBM accrédité et prestataire non LBM.....	2
Eléments du cahier des charges	3
Présentation générale.....	3
Spécification des données d'entrée et de sortie.....	4
Spécifications du pipeline bioinformatique	5
Qualification et validation de méthode	7
Modifications et requalifications	8
Informatique et données	9
Maintenance des équipements et logiciels.....	10
Gestion des non-conformités	11
Qualification du personnel du prestataire.....	12
Visites et audits.....	13
Formation initiale et continue du personnel du LBM.....	14
Support et continuité d'activité	15
Responsabilités et engagements	16
Références.....	17
Glossaire et dictionnaire	18
Organisations et réglementation	18
Formats de fichiers	18
Pipelines et analyse.....	19
Bioinformatique et séquençage.....	19
Informatique et données	19
Liste des participants.....	21

Préambule

Lors de la phase analytique d'un examen de biologie médicale de type séquençage massif en parallèle (communément désigné sous l'acronyme de NGS), l'externalisation de la partie bioinformatique comporte de nombreux risques nécessitant la maîtrise de la part du laboratoire. Nous proposons ce guide fournissant une aide à la rédaction d'un cahier des charges en vue de **l'externalisation de la partie bioinformatique d'un Laboratoire de Biologie Médicale (LBM) réalisant des analyses de génétique humaine à visée diagnostic**.

L'objectif est de proposer des éléments de maîtrise des risques associés à l'externalisation de la chaîne de traitement (pipeline bioinformatique) et un cadre pour garantir la qualité et la sécurité du traitement des données nécessaire à une mise en conformité réglementaire.

Pour une préparation optimale du cahier des charges, il est recommandé que le LBM réalise, au préalable, une analyse des risques. Cette démarche permettra de mettre en évidence les éléments essentiels et d'établir les responsabilités attendues du prestataire.

Concernant les divers documents fournis par le prestataire, il convient de préciser dans le cahier des charges les délais à respecter.

Ce document s'appuie sur des documents de référence existants et diffusés par les sociétés savantes ou institutions telles que BioInfoDiag, NGS-Diag, l'ANPGM et l'INCa.

Le cahier des charges devrait être rédigé conjointement par au moins un biologiste médical, en collaboration avec la Direction des Services de l'Information (DSI) et le Délégué à la Protection des Données (DPO). **L'avis d'un bioinformaticien est fortement recommandé pour garantir la pertinence technique du projet.** Si la structure ne dispose pas de cette expertise en interne, il est conseillé de faire appel à un avis externe pour compléter cette étape.

Nous tenons à préciser que l'externalisation de toute ou partie de la bioinformatique d'une analyse d'un LBM **n'est préconisée ni par BioInfoDiag¹, ni par l'ANPGM², comme une solution principale et pérenne**. Cependant, dans le cas où l'externalisation est tout de même utilisée comme solution principale, une attention particulière doit être portée à la solution dégradée de cette solution externalisée et doit être maîtrisée. Cette dernière doit être envisagée pour pallier les risques majeurs liés aux pannes réseaux, à une panne majeure du prestataire, à la nullité du contrat, à une attaque informatique ou toute autre raison qui empêcherait le recours à la solution externe. **En aucun cas le cahier des charges ne se substitue au dossier de qualification de la solution bioinformatique.**

Nous tenons également à rappeler que **ce document n'est pas exhaustif**, y compris dans les listes d'items proposés, se veut uniquement être une **aide à la rédaction d'un cahier des**

¹ <https://bioinfo-diaq.fr/blog/2024/avis-externalisation/>

² https://anpgm.fr/toutes-les-actualit%C3%A9s/veille-l%C3%A9gislative/courrier-anpgm_cofrac-externalisation-bioinformatique_f%C3%A9vrier-2024/

charges, et que **seul le biologiste médical référent est responsable de son activité** et des risques associés.

Distinction entre LBM accrédité et prestataire non LBM

Un **prestataire non LBM doit démontrer sa compétence** par des éléments objectifs (références, certifications, maîtrise des processus, documentation qualité, etc.) définis dans un **contrat de prestation**. Un **laboratoire de biologie médicale accrédité** selon la norme ISO 15189 dispose d'un **cadre réglementaire** et qualité reconnu, couvrant l'ensemble des exigences de compétence technique, d'organisation, de traçabilité et de maîtrise des risques. Cette distinction peut orienter le niveau d'audit, de vérification ou de supervision attendu par le LBM lorsqu'il évalue un prestataire.

Dans le cadre d'une **collaboration entre plusieurs LBM** (hors Groupement de Coopération Sanitaire ou Groupement Hospitalier de Territoire), les échanges peuvent s'appuyer sur les exigences communes de la norme ISO 15189 et sur la reconnaissance mutuelle de leurs systèmes qualité, et ainsi faire l'objet d'une **convention inter-laboratoires** (à l'instar d'une convention préleveur externe ou d'une convention de coordination). Cette convention encadre le transfert sécurisé des données de séquençage, la répartition des tâches de la phase analytique (pipelines bioinformatiques, annotation, interprétation assistée), les responsabilités respectives, la traçabilité, la cybersécurité, la confidentialité et la conformité aux exigences de la norme ISO 15189 pour cette étape de l'examen. Elle garantit que, même si certaines étapes de traitement bioinformatique sont mutualisées ou réalisées par un autre LBM accrédité, **chaque laboratoire demeure responsable et accrédité pour la partie analytique qu'il effectue**, notamment pour le pipeline utilisé et les conclusions rendues dans le compte-rendu.

Eléments du cahier des charges

Présentation générale

Le LBM définit ici le **contexte global** de la méthode et décrit les éléments principaux externalisés, conformément aux recommandations de l'ANPGM³ :

- Pathologies étudiées : préciser les indications cliniques ciblées.
- Volume d'échantillons : nombre d'échantillons attendus par an ou sur une période définie.
- Fréquences des analyses : périodicité prévue pour la réalisation des examens.
- Délais analytiques requis : temps maximal acceptable entre la réception de l'échantillon et la délivrance des résultats.

Le LBM précise les **paramètres techniques** permettant l'analyse des données issues du séquençage génétique, notamment :

- Type de Séquençage : long-read, short-read, UMI, DNA-seq, RNA-seq, amplicon, etc.
- Fournisseur et instrument : marque, génération du séquenceur (ex : 2ème ou 3ème génération), modèle exact.
- Enrichissement ciblé : panel de gènes, WES (Whole Exome Sequencing), WGS (Whole Genome Sequencing) ou autre méthode appliquée.

Le LBM définit les exigences pour l'**étape bioinformatique**, considérée comme un élément critique de la phase analytique :

- Étapes principales de la phase analytique à réaliser par le prestataire : démultiplexage des lectures ; alignement sur le génome de référence ; appel de variants ; annotation des variants ; filtrage et priorisation des résultats ; génération des métriques qualités ; log d'exécution.
- Types d'anomalies génétiques recherchées : SNVs, indels, SVs, CNVs, fusions, etc.
- Performances et délais requis : sensibilité, spécificité, limites de détection, temps de traitement maximal par lot ou par échantillon.
- Résultats et livrables attendus : description des données de sortie (brutes, intermédiaires, finales), formats de fichiers, rapports et tableaux de métriques.

³ https://anpgm.fr/media/documents/BP-GT4_Val_methode_NGS.pdf

Spécification des données d'entrée et de sortie

Le LBM définit les données d'entrée et de sortie nécessaires au fonctionnement de la solution bioinformatique externalisée. Le prestataire devrait être en mesure de prendre en charge différents formats d'entrées, de produire les sorties attendues et d'assurer la cohérence, l'intégrité et la traçabilité de l'ensemble des fichiers manipulés. Le LBM peut exiger que la solution accepte plusieurs types d'entrées, qu'elle génère différents jeux de sorties et que les formats utilisés respectent les standards internationaux en vigueur.

Les types de données manipulées sont rappelés conformément au référentiel ANPGM-GT1⁴ :

- **données brutes** : issues directement du séquenceur ;
- **données primaires** : séquences et valeurs de qualité ;
- **données secondaires** : résultats issus des traitements analytiques (alignements, variants, métriques, etc.).

Les types de données d'entrée et de sortie incluent, a minima :

- **Données brutes ou primaires** : données issues du séquençage et transmises par le LBM (ex. BCL, FAST5, FASTQ, uBAM). Le prestataire devrait garantir leur intégrité (contrôles de checksum, absence de modification) et leur conservation conformément aux exigences contractuelles.
- **Données secondaires** : données intermédiaires générées par la chaîne de traitement (ex. BAM/CRAM alignés, fichiers VCF bruts, métriques qualité, logs d'exécution). Ces données doivent être traçables, versionnées, associées aux paramètres du pipeline et restituées au LBM sur demande.
- **Données tertiaires** : résultats finaux issus de l'analyse bioinformatique (ex. VCF annoté, fichiers TSV de variants filtrés, rapports automatisés, métadonnées analytiques). Les formats doivent être exploitables par le LBM et compatibles avec ses outils d'interprétation et de validation des résultats.

En complément, le LBM peut exiger la mise à disposition d'**interfaces de visualisation et d'exploration des données**, notamment pour faciliter la vérification de la qualité et l'interprétation des résultats :

- Visualisation des lectures alignées (BAM/CRAM) via une interface intuitive intégrant par exemple IGV, avec gestion sécurisée des accès.
- Exploration, filtrage et interrogation des variants via un tableau interactif en ligne ou un portail dédié, avec possibilité d'export (TSV, Excel, JSON).
- Interface dédiée aux métriques et aux logs du pipeline, permettant la consultation des indicateurs qualité (taux de duplication, couverture, ratios, profondeur, contamination, etc.), l'accès aux logs d'exécution, l'historisation des traitements et la visualisation des alertes ou erreurs.

Le prestataire devrait documenter précisément les formats gérés, les métadonnées associées, les mécanismes de contrôle d'intégrité, ainsi que les modalités de restitution et de transfert sécurisé des données (cf. § *Informatique & Données*). Cette structuration garantit l'interopérabilité, la reproductibilité des analyses et la conformité avec les exigences du LBM.

⁴ https://anpgm.fr/media/documents/BP-GT1_Recommandations_NGS_Info_fzWYgEK.pdf

Spécifications du pipeline bioinformatique

Le prestataire devra se conformer aux exigences du LBM, de la norme ISO 15189 et de la réglementation en vigueur en communiquant clairement sur les méthodes, outils et paramètres, paramètres et bases de données utilisés. Il lui incombe de fournir une documentation complète, lisible et à jour permettant au personnel habilité du LBM de maîtriser l'utilisation de la solution, son paramétrage, ses évolutions et l'accès aux données générées. Les recommandations publiées par l'ANPGM⁵ et le réseau NGS-Diag⁶ constituent un cadre de référence pour la qualification et le suivi des pipelines bioinformatiques.

Un pipeline, ou chaîne de traitement de données, désigne toute solution logicielle permettant la transformation et l'analyse des données de séquençage.

Le prestataire devra notamment :

- fournir la **documentation technique**, incluant :
 - une description détaillée des logiciels et autres composants utilisés, y compris les versions et les algorithmes implémentés.
 - les différentes étapes du pipeline bioinformatique, des points d'entrée (fichiers FASTQ, BAM, etc.) aux points de sortie (fichiers BAM, VCF, TSV, etc.), en précisant les traitements des étapes d'alignement, d'appel et d'annotation des variants.
- assurer la **traçabilité complète des outils, versions, bases de données, paramètres**, environnements (conteneurs, images), ainsi que les journaux d'exécution ; fournir la description détaillée de chaque modification (cf. § *Modifications Mineures et Majeures*) et son impact attendu.
- fournir la **documentation de validation analytique** (sensibilité, spécificité, justesse, répétabilité, limites de détection, périmètre fonctionnel), incluant les données de qualification du pipeline et les tests de non-régression.
- proposer une **formation adaptée** au personnel habilité du LBM concernant l'utilisation, la personnalisation des paramètres et la compréhension des métriques qualité.
- formaliser une **procédure de configuration** permettant au personnel du LBM de modifier, ou de demander la modification, des paramètres ou composants du pipeline selon les besoins du laboratoire.
- documenter le **périmètre de détection** de la solution, incluant les seuils (fréquence allélique, profondeur minimale, qualité), les types et tailles d'événements détectés (SNV, indels, CNV, SV, fusions...), ainsi que les zones non couvertes ou non analysables.
- garantir la **reproductibilité** des analyses (gestion des versions, gel des environnements, contrôle des dépendances).
- définir un **processus de gestion des incidents**, incluant la détection, la qualification, la correction, la traçabilité et la communication au LBM.
- décrire les **besoins et contraintes techniques** (ressources de calcul, stockage, durée d'exécution, dépendances externes) ainsi que les exigences de cybersécurité associées (authentification, journalisation, chiffrement, ségrégation des données).

⁵ https://anpgm.fr/media/documents/BP-GT4_Val_methode_NGS.pdf

⁶ https://anpgm.fr/media/documents/BP_NGSDiag_005_Qualification_Bioinfo_Notice_technique_Final.pdf

Le prestataire devra garantir au LBM un accès complet, sécurisé et pérenne aux données générées par la solution (cf. §*Informatique et données*) , ainsi qu'à l'ensemble des métriques, journaux, rapports et métadonnées nécessaires pour permettre au LBM de comprendre l'objectif du pipeline, d'évaluer ses limites et d'interpréter les résultats en toute autonomie.

Qualification et validation de méthode

La qualification de la solution bioinformatique fournie par le prestataire doit être réalisée avant sa mise en production. Comme pour toute conception de logiciel pour le diagnostic clinique⁷, la procédure de qualification et de validation⁸ doit être documentée, inclure une évaluation approfondie des performances et les éléments de preuve conservés.

Liste des éléments de preuve :

- la **documentation**, concernant l'infrastructure technique utilisée pour l'analyse, incluant le matériel, (serveurs, équipements) et les systèmes d'exploitation, ainsi que les configurations logicielles spécifiques.
- un **jeu de données de validation** ou celui fourni par le LBM : sur demande du LBM, le prestataire devrait fournir a minima une description du jeu de données de validation utilisé, permettant de reproduire et vérifier les analyses.
- l'ensemble des éléments permettant d'évaluer les performances de la solution, tels que :
 - mesures statistiques :
 - sensibilité (rappel) : Capacité de la solution à détecter correctement les vrais variants.
 - précision : Capacité de la solution à ne pas rapporter de faux variants.
 - F-mesure : Moyenne harmonique des précisions et rappels.
 - durée analytique, ou temps d'exécution d'une analyse en fonction :
 - de la taille du jeu de données.
 - des versions des outils et des bases de données utilisées.
 - du dimensionnement de l'équipement informatique.
 - des points d'entrée (FASTQ, BAM...) et de sortie (BAM, VCF, TSV...) de la solution.
- la documentation concernant la conception des logiciels (ex : cycle de développement, habilitation du personnel...), comme indiqué dans le document de conception de logiciel pour le diagnostic clinique⁷

⁷ https://www.e-cancer.fr/content/download/224959/3070300/file/Conception_de_logiciels_pour_le_diagnostic_clinique_par_sequencage_haut_debit_mel_20180221.pdf

⁸ https://anpgm.fr/media/documents/BP_NGSDiag_005_Qualification_Bioinfo_Notice_technique_Final.pdf

Modifications et requalifications

Conformément au document de validation de méthode de l'ANPGM⁹, *“Toute modification du pipeline doit être tracée. Une validation par des tests appropriés est nécessaire pour toute modification impactant le fonctionnement global du pipeline”*.

Le prestataire doit documenter **toutes les modifications apportées au pipeline** et mettre en œuvre des tests adaptés pour garantir la continuité, la reproductibilité et la fiabilité des analyses.

Modification majeure : toute modification pouvant influencer le fonctionnement global ou les performances du pipeline.

Modification mineure : toute modification n'ayant pas d'impact significatif sur le fonctionnement global ou les performances du pipeline.

Le LBM peut considérer comme **modification majeure** notamment :

- un changement de **version majeure** ou l'ajout d'un **outil** ou de **base de données** influant sur les étapes critiques du pipeline (e.g. alignement, appel de variants, annotation des variants, priorisation des variants).
- un changement d'un **composant du système informatique**, incluant le matériel (nouveau serveur) et le logiciel (e.g. gestionnaire de workflow, ordonnanceur, système d'exploitation).
- **l'ajout d'une nouvelle fonctionnalité** ayant un impact sur les analyses ou la production des résultats.

Une modification majeure doit faire l'objet d'une **nouvelle qualification complète**, incluant :

- un test de non-régression du pipeline bioinformatique ;
- une évaluation des performances globales de la solution ;
- une validation formelle par le LBM avant mise en production.

De plus, toute modification majeure doit être **notifiée aux utilisateurs et au LBM**, avec une description claire de sa nature et de son impact. Le LBM peut exiger un **délai de notification** ou une **planification spécifique** avant mise en production (par exemple en cas d'interruption de service ou d'impact critique).

Une modification mineure devrait faire l'objet d'un **test de non-régression** réalisé par le prestataire, qui en fournit la preuve au LBM.

⁹ https://anpgm.fr/media/documents/BP-GT4_Val_methode_NGS.pdf

Informatique et données

La solution proposée par le prestataire doit être conforme aux exigences de sécurité informatique et de gestion des données, validée par les autorités compétentes (LBM, DSI, DPO), et en accord avec les recommandations des sociétés savantes et la législation en vigueur, notamment les règlements RGPD, les certifications idoines (ISO 27001, SecNumCloud), ainsi que les exigences relatives aux dispositifs médicaux (CE-IVD, IVDR, HDS, ISO 15189).

Le prestataire devrait fournir :

- un **système de chiffrement** robuste pour sécuriser la transmission, le stockage et la récupération des données sensibles
- les **preuves d'accès à un système sécurisé**, exclusivement réservé au personnel habilité du LBM, permettant le stockage et le traitement des données, conformément aux normes de sécurité et aux exigences légales (notamment RGPD, certifications HDS, SecNumCloud et lois extraterritoriales). Il est préconisé d'inclure dans les preuves un rapport de couverture des risques d'exposition et une description des méthodes utilisées pour réduire la surface d'attaque.
- une **traçabilité complète des mises à jour** des équipements et des logiciels (versions utilisées, modifications appliquées) et de l'infrastructure hébergeant la solution (ex : versions des systèmes d'exploitation, matériel employé).
- les informations relatives au **cycle de vie des données** : description détaillée du cycle de vie de toutes les données, y compris les fichiers intermédiaires générés au cours du processus¹⁰.
- un **accès aux données** pendant la durée du contrat. Le LBM doit pouvoir accéder aux données générées par le pipeline via des protocoles standards sécurisés, tels que SSH et HTTP via TLS (HTTPS) ou équivalents, en fonction du type et de la durée de vie des données¹⁰ (cf. §*Entrées et Sorties*).

Toutes les données générées ou fournies par le LBM, ainsi que celles créées dans le cadre de la solution du prestataire, appartiennent exclusivement aux patients, sous la responsabilité du LBM. Ainsi, le LBM doit s'assurer qu'à l'issue du contrat, les données soient restituées et que le prestataire aura supprimé les données du LBM (**clause de restitution**). Il est conseillé de définir dans le cahier des charges les modalités de restitution (format, procédure...). Le prestataire s'engage à détruire les données dans un délai défini après vérification effective de la restitution complète des données par le LBM.

Le prestataire ne peut utiliser ces données en dehors de l'usage prévu par le cahier des charges, la prescription et la réglementation en vigueur (RGPD). Ainsi, la réutilisation des données par le prestataire, en dehors de l'objectif initial de l'analyse de biologie médicale à visée diagnostique, est totalement interdite (conformément à la RGPD). L'accès aux données est limité aux membres habilités du LBM, et ces données ne peuvent être utilisées, consultées ou modifiées par d'autres parties, y compris par le prestataire.

¹⁰ https://anpgm.fr/media/documents/BP-GT1_Recommandations_NGS_Info_fzWYgEK.pdf

Maintenance des équipements et logiciels

Le prestataire est tenu d'assurer une **maintenance complète, proactive et documentée** des équipements et logiciels fournis, afin de garantir leur **fonctionnement optimal**, leur fiabilité et leur conformité continue avec les exigences de qualité et réglementaires.

Le prestataire s'engage notamment à fournir :

- Protocoles et procédures détaillées :
 - modes opératoires pour la maintenance préventive et corrective.
 - procédures de mise à jour logicielle et matérielle.
 - contrôles qualité et tests après maintenance.
 - mises à jour de sécurité pour garantir la conformité aux normes en vigueur.
- Traçabilité et suivi :
 - enregistrement de toutes les interventions, mises à jour et actions de maintenance.
 - suivi détaillé des versions logicielles et matérielles utilisées.
 - accès en temps réel à l'historique des maintenances pour le LBM, permettant vérification et audit à tout moment.
- Respect des délais et continuité de service :
 - planification et réalisation des interventions dans les délais définis contractuellement.
 - minimisation des risques d'interruption des services ou d'impact sur la qualité et la fiabilité des résultats.
 - gestion des incidents critiques et mise en place de solutions temporaires si nécessaire pour assurer la continuité des analyses.

Cette approche garantit que les équipements et logiciels utilisés dans le cadre des analyses bioinformatiques restent **fiables, conformes et traçables**, tout en permettant au LBM de maîtriser les risques liés à l'externalisation.

Gestion des non-conformités

Une **non-conformité** désigne tout manquement ou écart par rapport aux normes, exigences ou spécifications définies, susceptible d'affecter la **qualité, la sécurité ou la performance** des produits, services, processus ou systèmes. Ces écarts doivent être **identifiés, analysés et corrigés rapidement** afin de limiter leur impact et prévenir leur réapparition.

Les non-conformités peuvent être classées en deux catégories :

- **Non-conformités mineures** : impact limité sur la qualité, la sécurité ou le fonctionnement, présentant un faible risque pour l'activité.
- **Non-conformités majeures** : défaillances significatives nécessitant des corrections immédiates en raison de leur impact potentiel élevé sur la qualité, la sécurité ou la continuité des services.

Le prestataire doit :

- Assurer la **traçabilité complète** des non-conformités liées à sa solution, via un système permettant le suivi et l'audit des incidents.
- **Notifier immédiatement** le LBM en cas de malfunction, panne ou tout autre écart significatif
- Mettre en œuvre une **surveillance continue** des analyses afin de détecter et signaler les anomalies dès leur apparition.
- Proposer et mettre en place des **actions correctives adaptées** pour résoudre les non-conformités identifiées.

Dans les cas où une non-conformité est attribuée au prestataire et selon l'analyse de son impact, la reprise ou la correction des analyses concernées sera réalisée par le prestataire, à ses frais, afin de garantir la qualité et la fiabilité des analyses sans entraîner de surcoût pour le LBM.

Une gestion rigoureuse, transparente et documentée des non-conformités constitue un élément fondamental pour assurer la conformité aux exigences réglementaires, maintenir la qualité et la sécurité des données, préserver la confiance et la collaboration entre le LBM et le prestataire.

Qualification du personnel du prestataire

La maîtrise des risques par le LBM repose sur la **compétence et l'organisation du prestataire**, qui doivent être clairement démontrées et vérifiables. Il est préconisé de demander au prestataire des garanties précises et documentées concernant les qualifications, compétences et l'organisation de son personnel.

Il est conseillé au laboratoire de demander au prestataire de :

- justifier de la qualification de son personnel en transmettant les preuves de compétences pour toutes les personnes impliquées dans les prestations pour le LBM.
- fournir un organigramme à jour précisant les rôles et responsabilités de chacun au sein de l'équipe impliquée.
- mettre à disposition les fiches de poste détaillant les missions, compétences requises et expériences attendues pour chaque fonction.

Il est à préciser que, même si le **personnel du prestataire est qualifié**, seul le **personnel du LBM est habilité** et autorisé à réaliser des analyses au sein de son laboratoire. Aucune intervention du personnel du prestataire n'est permise lors de la réalisation des examens, et toute modification de la solution externalisée doit être effectuée sous la responsabilité exclusive du personnel habilité du LBM.

Étant donné la nature sensible des **données médicales** traitées, une attention particulière doit être portée à ces informations et au personnel qui y accède, celles-ci **relevant du secret médical**.

Cette approche garantit la **sécurité, la conformité réglementaire et la qualité des analyses**, tout en maintenant une responsabilité claire entre le LBM et le prestataire.

Visites et audits

Le LBM se réserve le droit de réaliser **des visites sur site et des audits auprès du prestataire**, conformément aux exigences de la norme ISO 15189 et aux principes de maîtrise des processus externalisés. Ces évaluations ont pour objectif de :

- vérifier la conformité des **pratiques bioinformatiques**.
- contrôler l'intégrité des **chaînes de traitement** (pipeline, versionnage, validation).
- évaluer l'adéquation des **ressources humaines, matérielles et logicielles** avec les exigences du présent cahier des charges.

Le prestataire s'engage à **faciliter pleinement ces contrôles**, en garantissant :

- un accès transparent à ses installations et infrastructures informatiques.
- la disponibilité des procédures opératoires, systèmes de gestion des versions, journaux de traçabilité.
- l'accès à la documentation technique et aux personnes impliquées dans la prestation.

Cet audit contribue à :

- assurer la traçabilité complète des opérations.
- maîtriser les risques et la cybersécurité.
- garantir la robustesse analytique des traitements bioinformatiques.
- vérifier le respect des engagements contractuels.
- soutenir une démarche de conformité réglementaire et d'amélioration continue.

Formation initiale et continue du personnel du LBM

Afin de garantir une utilisation optimale et sécurisée des solutions proposées par le prestataire, ce dernier devrait assurer la **formation initiale et continue permettant l'habilitation des membres du personnel du LBM à l'usage de l'outil**. Cette exigence vise à permettre au LBM de maîtriser pleinement en interne la solution fournie, tant sur les aspects fonctionnels (utilisation quotidienne, interprétation des résultats) que techniques (paramétrage, maintenance de premier niveau). La formation constitue un élément clé pour réduire les risques opérationnels, assurer une continuité de service, et répondre aux exigences de qualité et de conformité réglementaire.

Le prestataire devrait :

- mettre en place une **formation initiale adaptée**, dispensée à l'ensemble des utilisateurs concernés au sein du laboratoire, avant la mise en production de la solution. Cette formation devrait couvrir les aspects théoriques et pratiques.
- proposer un programme de **formation continue** permettant de maintenir et d'actualiser les compétences des utilisateurs, notamment en cas de mise à jour ou d'évolution de la solution.
- fournir une **attestation de formation** pour chaque membre du personnel formé, laquelle devra être intégrée aux fiches d'habilitation du laboratoire conformément aux exigences qualité.

Le prestataire devrait également transmettre un plan de formation détaillé, précisant les objectifs pédagogiques, les supports fournis (documentation technique, guides d'utilisation, tutoriels), ainsi que les modalités et calendriers des sessions de formation.

Support et continuité d'activité

Pour garantir la pérennité et la disponibilité des prestations externalisées, le LBM doit pouvoir compter sur **un système de support fiable ainsi que sur des mécanismes robustes de continuité et de reprise d'activité en cas d'incident**. Ces dispositions sont essentielles pour prévenir les interruptions de service, limiter les impacts opérationnels et assurer la qualité des analyses réalisées. Une communication efficace entre le LBM et le prestataire est également nécessaire pour optimiser la gestion des incidents et la traçabilité des actions correctives.

Le prestataire devrait :

- mettre à disposition un **système de support adapté**, incluant :
 - Un système de gestion de tickets pour le signalement, le suivi et la résolution des incidents.
 - Une documentation technique détaillée pour guider les utilisateurs dans la résolution autonome des problèmes simples.
 - Un système de gestion de la traçabilité garantissant le suivi des actions réalisées et la documentation associée.
- respecter les **délais de prise en charge** et d'intervention définis en accord avec le LBM dans le cadre du cahier des charges.
- fournir un **Plan de Continuité d'Activité (PCA)** détaillant les mesures à mettre en œuvre pour assurer la continuité des services en cas de perturbation.
- fournir un **Plan de Reprise d'Activité (PRA)** précisant les procédures de restauration rapide des activités après une interruption majeure.
- fournir un **Contrat de Niveau de Service (SLA : *service-level agreement*)** précisant les objectifs attendus et le niveau de service souhaité.
- **anticiper la cessation d'activité** en proposant :
 - un plan de transition détaillé, incluant la migration des données et des processus vers une solution alternative, avec un calendrier et des modalités claires.
 - la garantie d'un accès prolongé aux données et aux outils pendant la période de transition, afin d'éviter toute interruption brutale du service.
 - un accompagnement technique et documentaire pour faciliter la reprise en interne ou par un nouveau prestataire.

Pour assurer une gestion efficace des interactions et des incidents, les deux parties désignent des référents dédiés :

- Le LBM nomme un ou plusieurs référents internes chargés de coordonner les échanges avec le prestataire, de superviser les demandes de support et de valider les actions correctives.
- Le prestataire désigne un ou plusieurs référents dédiés, responsables de garantir la disponibilité et la réactivité du support, de suivre les incidents signalés et d'assurer une communication transparente avec le LBM, notamment en cas d'escalade liée à un incident critique.

Cette organisation vise à renforcer la collaboration et à minimiser les risques liés à des interruptions ou des défaillances des prestations externalisées.

Responsabilités et engagements

Afin d'assurer une compréhension partagée des responsabilités et de garantir une coordination optimale entre le LBM et le prestataire, une **matrice RACI (Responsable, Approbateur, Consulté, Informé)** devrait être définie conjointement avant la mise en production de la solution. Cette matrice permettra de clarifier les rôles et responsabilités pour chaque étape du processus, de l'installation à l'exploitation quotidienne, en passant par la gestion des incidents et la mise en œuvre des plans de continuité ou de reprise d'activité. L'élaboration de la matrice RACI, accompagnée d'un suivi rigoureux des indicateurs de performance, constitue une étape clé pour sécuriser la collaboration et prévenir les litiges liés aux responsabilités partagées.

En cas de non-respect des engagements de services ou des délais contractualisés par le prestataire, le LBM définit des pénalités applicables, formalisées dans le contrat. Ces pénalités visent à protéger l'activité du LBM et à garantir le respect des engagements liés à la qualité, à la réactivité et à la continuité des services.

Le biologiste médical reste pleinement responsable des résultats transmis aux prescripteurs. La validation finale des analyses et des rapports relève exclusivement du biologiste médical chargé de l'activité, conformément à la réglementation en vigueur. En aucun cas, cette responsabilité ne peut être transférée au prestataire ou à toute autre partie externe.

Références

- **ANPGM/COFRAC 2024** : Avis sur l'externalisation d'une étape de la phase analytique des examens s'appuyant sur le séquençage de l'ADN
<https://anpgm.fr/toutes-les-actualit%C3%A9s/veille-l%C3%A9gislative/courrier-anpgm-cofrac-externalisation-bioinformatique-f%C3%A9vrier-2024/>
- **ANPGM_GT1** : Recommandations générales pour la gestion informatique des données et des analyses de séquençage à haut débit pour les laboratoires de diagnostic moléculaire de maladies génétiques (ANPGM).
https://anpgm.fr/media/documents/BP-GT1_Recommandations_NGS_Info_fzWYgEK.pdf
- **ANPGM_GT4** : Validation de méthode : Validation initiale approfondie de la méthode de séquençage haut débit pour la recherche de mutations constitutionnelles (analyse qualitative) (ANPGM)
https://anpgm.fr/media/documents/BP-GT4_Val_methode_NGS.pdf
- **BioInfoDiag/COFRAC 2024** : Avis sur l'externalisation d'une étape de la phase analytique des examens s'appuyant sur le séquençage de l'ADN
<https://bioinfo-diag.fr/blog/2024/avis-externalisation/>
- **BP_NGSDiag_005** : Qualification des solutions bioinformatiques: note technique (NGS-DIAG)
https://anpgm.fr/media/documents/BP_NGSDiag_005_Qualification_Bioinfo_Note_technique_Final.pdf
- **HDS** : Certification obligatoire en France pour les Hébergeurs de Données de Santé (HDS), garantissant la sécurité, la confidentialité et la traçabilité des données
https://esante.gouv.fr/sites/default/files/media_entity/documents/referentiel_certification_hds--fr--v2.pdf
- **INCa** : Conception de logiciels pour le diagnostic clinique par séquençage haut-débit (INCA)
https://www.e-cancer.fr/content/download/224959/3070300/file/Conception_de_logiciels_pour_le_diagnostic_clinique_par_sequencage_haut_debit_mel_20180221.pdf
- **ISO 27001** : Sécurité de l'information, cybersécurité et protection de la vie privée - Systèmes de management de la sécurité de l'information – Exigences
<https://www.iso.org/fr/standard/27001>
- **ISO 15189** : Exigences concernant la qualité et la compétence
<https://www.iso.org/fr/standard/76677.html>
- **IVDR** : Réglementation européenne qui renforce les exigences de sécurité, de classification et d'évaluation pour les dispositifs de diagnostic in vitro.
<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32017R0746>
- **RGPD** : Règlement général sur la protection des données (CNIL)
<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>
- **SecNumCloud** : Label ANSSI pour des services cloud sécurisés et souverains, adaptés aux données sensibles.

<https://cyber.gouv.fr/offre-de-service/solutions-certifiees-et-qualifiees/comprendre-evaluation-de-securite/qualification-de-produit-et-services/referentiels-qualification/>

Glossaire et dictionnaire

Organisations et réglementation

- **LBM** : Laboratoire de Biologie Médicale
- **DSI** : Direction des Systèmes d'Information
- **DPO** : Délégué à la Protection des Données
- **ISO** : *International Organization for Standardization* - Organisation internationale de normalisation ; dans le contexte des LBM, se réfère notamment à la norme **ISO 15189**, définissant les exigences de qualité et de compétence pour les laboratoires de biologie médicale.
- **IVDR** : *In Vitro Diagnostic Regulation* - base réglementaire actuelle pour la mise sur le marché, la mise à disposition et la mise en service des dispositifs médicaux de diagnostic in vitro sur le marché européen
- **PCA** : Plan de Continuité de l'Activité - ensemble des mesures garantissant la continuité des activités en cas de perturbation.
- **PRA** : Plan de Reprise de l'Activité - ensemble des procédures permettant de rétablir les services critiques après un incident majeur.
- **RACI** : Responsable, Approbateur, Consulté, Informé ; type de document ou matrice précisant les responsabilités et rôles lors de la gestion d'un projet.
- **SLA** : *Service Level Agreement* - Accord de niveau de service
- **Traçabilité** : capacité à suivre l'origine et l'historique d'un échantillon, d'un fichier ou d'un résultat.

Formats de fichiers

- **BCL** : *Base Call Format* - format produit par les séquenceurs Illumina, contenant les informations brutes de lecture et de qualité avant conversion en FASTQ.
- **BAM** : *Binary Alignment/Map* - version binaire et compressée du format SAM, utilisée pour le stockage et le traitement efficace des lectures alignées.
- **BED** : format texte pour décrire des régions génomiques (chromosome, start, end, annotation).
- **CRAM** : format compressé alternatif à BAM, permettant de stocker les alignements avec réduction de taille disque en utilisant le génome de référence.
- **FAST5** : fichier brut généré par les séquenceurs Oxford Nanopore, contenant les signaux électriques et métadonnées associées.
- **FASTQ** : Format de fichier standard pour stocker les lectures brutes issues du séquençage, incluant les séquences et les valeurs de qualité.
- **GTF / GFF** : formats de fichiers pour annoter les génomes et les transcrits.
- **LOG** : fichiers de journalisation générés par les pipelines, contenant l'historique d'exécution et les erreurs.
- **Metrics QC / QC report** : fichiers de métriques qualité, indicateurs sur la couverture, la duplication, la profondeur, etc...

- **SAM** : *Sequence Alignment/Map* - format texte standard pour représenter les lectures alignées sur un génome de référence, lisible directement par un humain ou un logiciel.
- **TSV** : *Tab-Separated Values* - format de fichier texte où les données sont séparées par des tabulations, utilisé pour exporter des résultats analytiques ou des tables de variants.
- **VCF** : *Variant Call Format* - format texte standard pour décrire les variants génétiques détectés à partir d'alignements.

Pipelines et analyse

- **Alignement** : positionnement des lectures sur un génome de référence.
- **Appel de variants** (*Variant Calling*) : identification des variations génétiques à partir des lectures alignées.
- **Annotation** : ajout d'informations biologiques et cliniques aux variants détectés.
- **Filtrage** : sélection des variants d'intérêt selon des critères de qualité et pertinence biologique.
- **Pipeline** : enchaînement structuré d'outils et de scripts permettant de transformer des données brutes en résultats interprétables.
- **Reproductibilité** : capacité à obtenir les mêmes résultats sur les mêmes données avec le même pipeline.

Bioinformatique et séquençage

- **Amplicon** : séquençage ciblé de fragments spécifiques d'ADN amplifiés par PCR.
- **CNVs** : *Copy Number Variants* - variations du nombre de copies de segments d'ADN.
- **DNA-seq** : séquençage de l'ADN génomique, utilisé pour identifier variants, mutations et structures génomiques.
- **Fusions** : réarrangements génétiques entraînant la fusion de deux gènes ou segments génomiques.
- **Indels** : Insertions ou délétions de quelques bases dans le génome.
- **PCR** : *Polymerase Chain Reaction* - technique d'amplification enzymatique d'un fragment spécifique d'ADN, utilisée pour enrichir des régions ciblées avant séquençage.
- **RNA-seq** : séquençage de l'ARN, utilisé pour quantifier l'expression des gènes et détecter des variantes transcriptionnelles.
- **SNVs** : *Single Nucleotide Variants* - variants ponctuels affectant une seule base.
- **SVs** : *Structural Variants* - variations structurales du génome (insertions, délétions, inversions, duplications).
- **UMI** : *Unique Molecular Identifier* - identifiant Moléculaire Unique, séquence courte ajoutée aux fragments d'ADN pour distinguer les molécules originales et corriger les biais de PCR.
- **WES** : *Whole Exome Sequencing* - séquençage de l'ensemble des exons codants du génome.
- **WGS** : *Whole Genome Sequencing* - séquençage complet du génome.

Informatique et données

- **Checksum** : somme de contrôle permettant de vérifier l'intégrité d'un fichier (ex : MD5, SHA256).
- **Container / Docker / Singularity** : environnement isolé permettant d'exécuter un pipeline avec toutes ses dépendances.
- **Logs** : fichiers enregistrant l'exécution du pipeline, utiles pour l'audit et le diagnostic d'erreurs.
- **Versioning** : suivi des versions des outils, bases de données et pipelines pour garantir traçabilité et reproductibilité.

Liste des participants (ordre alphabétique)

- *David Baux*
- *Claire Bardel Danjean*
- *Pierre Blanc*
- *Laetitia Bourgeade*
- *Laurent Castera*
- *Marie De Tayrac*
- *Anne-Sophie Denomme-Pichon*
- *Antony Le Behec*
- *Alban Lermine*
- *Jean Muller*
- *Aurelien Trimouille*
- *Charles Van Goethem*