

Proposition de création d'un réseau métier CNRS en bioinformatique

MERIT : réseau MetiER en bioinformaTique

Groupe de travail :

- BIHOUEE Audrey (Université de Nantes)
- CHIAPELLO Hélène (INRAE)
- CORRE Erwan (CNRS – INEE)
- LEFORT Vincent (CNRS – INS2I)
- LOUIS Alexandra (CNRS – INSB)
- VANDEL Jimmy (CNRS – INSB)

Constat :

L'évolution des technologies d'acquisition de données en sciences du vivant génère une avalanche de données. Les métiers permettant de gérer, traiter et analyser ces données représentent aujourd'hui un enjeu primordial. Le métier de bioinformaticien est intrinsèquement interdisciplinaire. Les ingénieurs du domaine dépendent de la Branche d'Activité Professionnelle des sciences du vivant, de la terre et de l'environnement (BAP A) ou de celle de l'informatique, statistiques et calcul scientifique (BAP E). Au sein de ces deux BAP, ils sont souvent affectés aux emploi-types suivants, selon leur grade (IE ou IR) et leurs activités :

1. BAP A
 - Ingénieur biologiste en analyse de données (IR - [A1A41](#)).
 - Ingénieur biologiste en traitement de données (IE - [A2A41](#)).
2. BAP E
 - Chef de projet ou expert en Ingénierie logicielle (IR - [E1C43](#)).
 - Ingénieur en ingénierie logicielle (IE - [E2C45](#)).

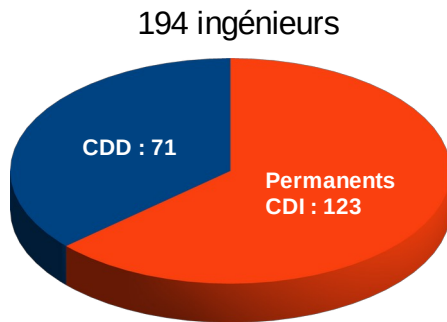
Cependant, les activités des bioinformaticiens peuvent emprunter à d'autres emploi-types :

- Chef de projet ou expert en infrastructures (IR - [E1B42](#)) et administrateur de bases de données / intégrateur d'applications (IE – [E2B44](#)) pour les responsables de plateformes.
- Expert en calcul scientifique (IR - [E1E45](#)) et ingénieur en calcul scientifique (IE – [E2E47](#)) pour les développeurs d'applications et les bio-analystes.
- Expert en information statistique (IR – [E1D44](#)) et ingénieur statisticien (IE – [E2D46](#)) pour les bioinformaticiens réalisant des analyses biostatistiques.

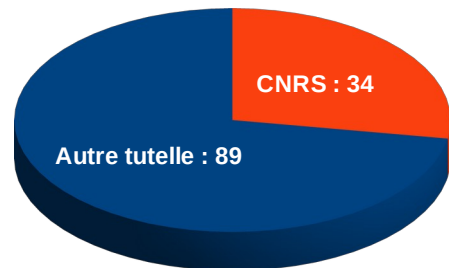
Ces divers éléments composant le métier de bioinformaticien sont regroupés et décrits dans une [fiche métier](#) éditée en 2020 par la Société Française de Bioinformatique ([SFBI](#)). Il y est mentionné le contexte interdisciplinaire dans lequel le métier s'exerce et les nombreuses thématiques scientifiques abordées. Ainsi, les laboratoires accueillant des bioinformaticiens sont partagés entre différents instituts du CNRS. De plus, les ingénieurs bioinformaticiens peuvent parfois exercer leurs activités au sein d'équipes ou de laboratoires thématiquement éloignés de la bioinformatique.

Public cible:

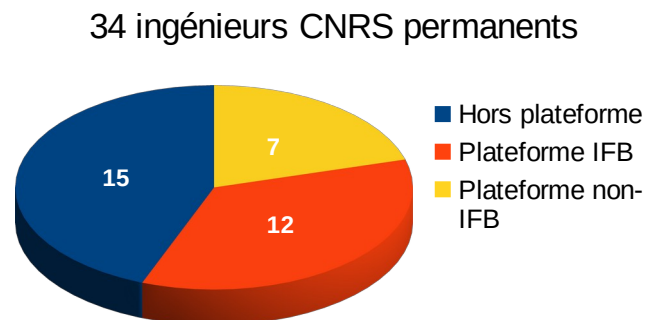
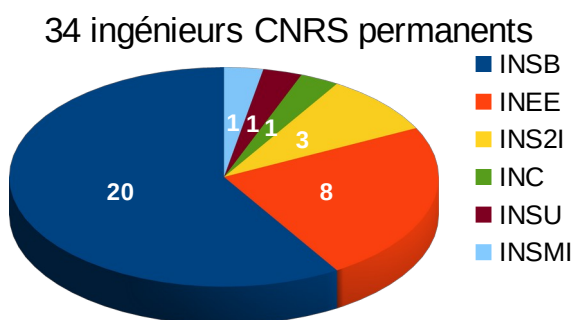
Afin de caractériser la population d'ingénieurs bioinformaticiens qui pourraient être concernés par un réseau métier en bioinformatique, nous nous sommes appuyés sur les réseaux locaux existants dans le Grand-Ouest (Bretagne et Nantes), à Lille et à Montpellier. Ces réseaux regroupent 194 ingénieurs bioinformaticiens dont 34 sont des agents titulaires du CNRS :



123 ingénieurs permanents - CDI



Les 34 ingénieurs du CNRS sont rattachés à différents instituts. Ils exercent leurs activités au sein des laboratoires ou sur des plateformes :



Si l'on considère que les données de la population d'ingénieurs du Grand-Ouest, de Lille et de Montpellier constituent un échantillon représentatif de la population nationale d'ingénieurs bioinformaticiens, nous pouvons extrapoler à l'échelle nationale. Pour cela, nous avons utilisé les informations des participants à JOBIM en 2019 et 2020.

- JOBIM 2019 : 150 ingénieurs sur 434 inscrits, dont 41 ingénieurs venant du Grand-Ouest, de Lille et de Montpellier (41 / 150 ~ 27%).
- JOBIM 2020 : 209 ingénieurs sur 680 inscrits, dont 46 ingénieurs venant du Grand-Ouest, de Lille et de Montpellier (46 / 209 ~ 22%).

Les ingénieurs du CNRS

Nous estimons que la population des ingénieurs du Grand-Ouest, de Lille et de Montpellier représente entre 22% et 27 % des ingénieurs bioinformaticiens qui participent à JOBIM. Nous pouvons donc considérer que le CNRS compte entre 126 et 154 ingénieurs bioinformaticiens titulaires qui participent à JOBIM. Comme tous les ingénieurs bioinformaticiens ne vont pas systématiquement à JOBIM, ces chiffres sous-estiment la population d'ingénieurs bioinformaticiens titulaires au CNRS. Par ailleurs, les ingénieurs bioinformaticiens

ne souhaiteront pas tous participer au réseau métier. Nous pensons donc que ce réseau pourrait concerner de l'ordre de **150 ingénieurs titulaires du CNRS**.

Les ingénieurs non-CNRS

Parmi les 89 ingénieurs permanents du Grand-Ouest, de Lille et de Montpellier ayant un employeur autre que le CNRS, 19 exercent leurs activités au sein d'UMR ou d'UAR ayant le CNRS comme tutelle. Les 70 autres ingénieurs sont affectés à des unités qui ne sont pas liées au CNRS.

Besoins :

- Pour faire face à l'évolution technologique rapide, les ingénieurs bioinformaticiens ont régulièrement besoin de se former.
- Pour rompre leur isolement thématique, les ingénieurs bioinformaticiens ont besoin de lieux d'échanges leur permettant de se rencontrer et de partager leurs expertises.

Structuration de la communauté :

1. Institut Français de Bioinformatique ([IFB](#))

L'IFB fédère les activités de service en bioinformatique portées par les plateformes et les équipes associées de bioinformatique. Bien que les plateformes concentrent souvent les ingénieurs en bioinformatique, nombre d'entre eux exercent leurs activités dans les laboratoires de recherche n'étant pas directement rattachés à une plateforme. Le périmètre de l'IFB ne couvre donc pas l'ensemble de la communauté des ingénieurs bioinformaticiens. Un réseau métier en bioinformatique pourra donc s'adresser à l'ensemble des ingénieurs, indépendamment de leur rattachement à une plateforme ou un laboratoire.

2. Groupement de Recherche en BioInformatique Moléculaire ([GdR BIM](#))

Le GdR BIM a pour vocation essentielle d'animer la communauté bioinformatique pour les activités de recherche. Il s'adresse en priorité aux chercheurs. Les ingénieurs peuvent participer aux activités d'animation scientifique financées par le GdR, mais celui-ci ne couvre pas les besoins essentiels de formations pour les ingénieurs. Le GdR n'a pas pour vocation de structurer la communauté des ingénieurs en bioinformatique. Un réseau métier en bioinformatique s'adressera spécifiquement aux ingénieurs. Il sera donc complémentaire du GdR BIM et il pourra servir de relais pour consolider les échanges entre ingénieurs et chercheurs.

3. Société Française de BioInformatique (SFBI)

La SFBI est la Société savante Française de la Bioinformatique. Elle a pour but de promouvoir, au travers de rencontres et d'échanges scientifiques, la recherche interdisciplinaire à l'interface de la biologie, de l'informatique, des mathématiques, des statistiques et de la physique et de rassembler la communauté francophone de la bioinformatique. Elle s'adresse à un vaste public de chercheurs et d'ingénieurs, sa mission est généraliste et n'est pas spécifiquement tournée vers les problématiques techniques auxquelles sont confrontés les ingénieurs.

4. Exemple de structuration à l'INRAE

- Au sein de l'INRAE, la communauté des bioinformaticiens est structurée avec 5 Centres Automatisés de Traitement de l'Information (CATI) qui regroupent environ 120 ingénieurs. Les CATI organisent des actions communes, comme par exemple des hackathons interCATIomiques dédiés à des développements ou à de la résolution de problèmes liés aux données omiques. En amont d'un hackathon, les CATI organisent des formations de mise à niveau pour les participants.

- Le [PEPI IBIS](#) est un réseau métier des bioinformaticiens pour les données omiques.

Positionnement et réseaux existants :

L'IFB fédère déjà les activités des plateformes de bioinformatique. Le réseau métier aura pour vocation de s'adresser à l'ensemble des bioinformaticiens quelque soit leur contexte de travail (plateformes, équipes de recherche, plateaux techniques, ...). Bien évidemment, les activités du réseau seront envisagées en concertation avec les autres acteurs du territoire national (IFB, AVIESAN,...). Des actions communes en partenariat seront privilégiées au démarrage du réseau.

Quelques réseaux d'échange en bioinformatique ont vu le jour. Ils se sont formés spontanément au sein des communautés locales (liste non exhaustive) :

- [Réseau métier des Ingénieurs en Bioinformatique de Lille](#).
- Réseau montpelliérain des ingénieurs en bioinformatique ([MBI](#)).
- Réseau informel en bioinformatique porté par la plateforme BiRD de Nantes (p'tits dej bioinfo, tutos bioinfo, rencontres mensuelles).
- Structuration en cours dans le Grand-Ouest dans le cadre de l'axe bioinformatique du GIP BioGenOuest.
- Réseau informel d'échange en bioinformatique entre Nice et Sophia-Antipolis.

De part leurs activités interdisciplinaires, les bioinformaticiens peuvent se reconnaître dans des réseaux métiers existants :

- [Calcul](#) pour le calcul scientifique.
- [DevLOG](#) pour le développement logiciel.
- [Resinfo](#) pour les infrastructures de calcul et de stockage.
- [RIS](#) pour les biostatistiques.
- [Bioinfo Diag](#) pour le diagnostic et la prise en charge des pathologies humaines.

Formations :

L'IFB se positionne comme un acteur majeur des activités de formation en bioinformatique. Il vise en priorité l'organisation de formations à destination des biologistes et bioinformaticiens. L'IFB vise également à créer de nouvelles ressources de formations et à les diffuser rapidement, notamment sur des sujets à forte demande comme les principes FAIR de science ouverte en bioinformatique pour lesquels les ressources de formation sont encore assez peu développées. Pour diffuser efficacement ces nouvelles formations et former une masse critique de formateurs, l'IFB se positionne comme maître d'ouvrage pour (i) créer des nouvelles formations, (ii) les consolider à travers des partenariats et la formation de formateurs, puis (iii) diffuser largement ces formations sur le territoire national.

L'AVIESAN organise depuis près de 10 ans des formations sur le traitement de données NGS et nouvelles technologies omiques pour des niveaux débutant ou avancé. Au delà des enseignements qui ont profité aux ingénieurs intégrant une part de plus en plus importante de leurs activités dans l'analyse de données, ces formations ont contribué à la structuration d'une communauté de formateurs pour beaucoup d'entre eux ingénieurs dans des équipes de recherche ou des plateformes.

Le réseau métier en bioinformatique pourra être un des partenaires qui pourra bénéficier des formations créées par l'IFB ou soutenues par l'AVIESAN afin de créer des Actions Nationales de Formation (ANF).

Proposition :

La création d'un réseau métier CNRS en bioinformatique permettra de fédérer les ingénieurs bioinformaticiens à l'échelle nationale et de contribuer à structurer cette communauté. Un tel réseau concernera plusieurs instituts du CNRS :

- En premier lieu l'[INSB](#), qui dénombre le plus gros effectif de bioinformaticiens.
- L'[INS2I](#) qui occupe une place particulièrement importante de part ses équipes de recherche en bioinformatique et leurs plateformes associées.
- L'[INEE](#) pour la bioinformatique étudiant la biodiversité (révolution en cours sur l'ADN environnemental).
- L'[INSMI](#) pour la bioinformatique en lien avec les biostatistiques (par exemple l'IMAG à Montpellier).
- L'[INSU](#) pour la bioinformatique en lien avec l'océanologie (par exemple l'Institut Méditerranéen d'Océanologie à Marseille, impliqué dans le projet TARA Océan).
- L'[INC](#) impliqué dans l'infrastructure MetaboHUB par exemple.

Si le réseau regroupe effectivement des personnels travaillant dans des laboratoires issus de plusieurs instituts du CNRS, il pourrait être pertinent de demander son intégration à la Mission pour les Initiatives Transverses et Interdisciplinaires ([MITI](#)).

Actions envisagées :

En 2022 :

- **Comité de pilotage** : constituer le comité de pilotage du réseau. Le choix des personnes impliquées devra respecter des critères de parité, de représentation des tutelles et instituts CNRS, de représentation géographique et de variété thématique.
- **Réseaux locaux** : recenser les éventuels réseaux locaux et acteurs existants, les informer de l'initiative et instaurer un dialogue avec ces partenaires potentiels. Pour certains sites académiques disposant d'une communauté reconnue en bioinformatique mais n'ayant pas de réseau local, favoriser l'émergence d'un tel réseau.
- **Communication** : créer les outils de communication du réseau (site web, liste de diffusion, messagerie instantanée,...).

A partir de 2023 :

- **Kick-off**: nous envisageons d'organiser au printemps (entre le 6 mars et le 8 avril pour éviter les vacances solaires) une rencontre de 2 jours en présentiel pour le lancement du réseau. Cette rencontre comprendra des ateliers de réflexion ayant plusieurs objectifs : (i) définir l'organisation et la structuration du réseau, (ii) identifier des thématiques prioritaires pour la communauté des ingénieurs en bioinformatique afin d'établir des groupes de travail et (iii) définir les actions de formation à mener. Il s'agira d'envisager une ANF pour 2024 et de co-construire les formations avec l'IFB, l'INRAE et AVIESAN.
- **Formation** : préparer le dossier pour une ANF à organiser en 2024.
- **Groupes de travail** : mettre en place des groupes de travail au sein du réseau pour assurer la veille technologique et proposer des axes thématiques à développer.

- **Métier** : travailler à la reconnaissance du métier de bioinformaticien et des problématiques liées à l'interdisciplinarité, en lien avec l'Observatoire des Métiers et de l'Emploi Scientifique ([OMES](#)) et la SFBI.

En 2024 :

- **ANF** : organiser une formation dont la thématique aura été définie lors du kick-off, en lien avec l'IFB, l'INRAE et/ou AVIESAN.

Actions récurrentes à partir de 2024 :

- **Animation** : organiser une rencontre annuelle des membres du réseau, sous la forme de journées thématiques et/ou de hackathons.
- **Groupes de travail** : prendre en compte les propositions des groupes de travail, les faire évoluer (identification de nouvelles thématiques).

Financement :

Le fonctionnement du réseau sera essentiellement basé sur la bonne volonté des membres du comité de pilotage. Certaines actions nécessiteront néanmoins la prise en charge (en totalité ou en partie) de frais de missions. Les sources de financement envisageables sont :

- Les instituts du CNRS volontaires.
- L'IFB selon la pertinence de l'action à financer par rapport à sa stratégie.
- La SFBI.
- Les Labex.

Plus spécifiquement, pour le **kick-off** qui sera organisé sur 2 jours au printemps 2023, nous envisageons de rassembler de 50 à 100 personnes. Voici une estimation des dépenses selon le nombre de participants:

	50 pers.	80 pers.	100 pers.
3 pauses café (12€)	600	960	1 200
3 repas (50€)	2 500	4 000	5 000
Hébergement (1 nuit 120€)	5 100	8 160	10 200
Transport (A-R 120€)	5 100	8 160	10 200
Total	13 300 €	21 280 €	26 600 €

Le kick-off se déroulera certainement sur Paris, ce qui économisera la part du budget dédiée au transport et à l'hébergement pour les ingénieurs de la région parisienne. Dans le tableau ci-dessus, nous considérons 15 % de participants parisiens. Nous souhaitons que le réseau puisse prendre en charge l'ensemble des dépenses, soit un budget de l'ordre de 26 k€ pour 100 participants. Il est important de pouvoir assurer pour chaque participant une prise en charge complète (repas, hébergement et transport). En effet, les laboratoires acceptent rarement de prendre en charge des missions pour leurs ingénieurs. Afin d'assurer le financement du kick-off et donc la participation du maximum d'ingénieurs, nous envisageons de solliciter les contributions de la SFBI et de l'IFB.

Thématiques :

Voici une liste non-exhaustive de thématiques et de technologies qui pourraient être suivies par les groupes de travail afin de proposer des actions à mener par le réseau :

- Analyse : Whole Genome, Microbiote 16S/Shotgun, transcriptomics (Single Cell et bulk), Méta-Génomique/Transcriptomique, RNAseq et SRP
- Workflows et containers : Snakemake, NextFlow, Conda/PIP, Docker, Singularity/Apptainer
- Visualisation : R, R Markdown, Shiny, bibliothèques python (matplotlib, seaborn,...)
- Calcul : distribué sur un cluster, utilisation du Cloud Biosphère
- FAIR et Open Science
- Apprentissage statistique, machine learning et deep learning
- Intégration de données omiques
- Ontologies et web sémantique (RDF, SPARQL) et bases de données NoSQL
- Assemblage et annotation de génomes

Laboratoires :

Voici une liste indicative et non-exhaustive de laboratoires (regroupés par institut) ayant des agents potentiellement concernés par le réseau :

- INSB : IRL3614 (EBEA), UAR2014 (PLBS), UAR3760 (IBCP), UMR5048 (CBS), UMR5095 (IBGC), UMR5203 (IGF), UMR5235 (LPHI), UMR5237 (CRBM), UMR5535 (IGMM), UMR6075 (CRCI2NA), UMR6286 (US2B), UMR6291 (ITX-lab), UMR7104 (IGBMC), UMR7256 (IGS), UMR7257 (AFMB), UMR7275 (IPMC), UMR7280 (CIML), UMR7288 (IBDM), UMR1283/8199 (EGENODIA), UMR8227 (LBI2M), UMR8576 (UGSF), UMR9002 (IGH), UMR9017 (CIIL), UMR8197 (IBENS), UMR8030 (GENOSCOPE), UMR7592 (IJM), UMR7238 (LCQB), UMR9198 (I2BC), UMR8120 (GQE-LeMoulon), UMR 9213 (IPS2), UMR3525 (Génétique des génomes), UMR3348 (Intégrité du génome, ARN et cancer), UMR9020 (CANTHER)
- INEE : FR2424 (SBR), UMR5175 (CEFE), UMR5290 (MIVEGEC), UMR5554 (ISEM), UMR5558 (LBBE), UMR6197 (BEEP), UMR7144 (AD2M), UMR8198 (EEP), UMR9190 (MARBEC), UMR7205 (ISYEB), UMR8079 (ESE), UMR7206 (EAE)
- INS2I : UMR5506 (LIRMM), UMR5800 (LaBRI), UMR 6004 (LS2N), UMR6074 (IRISA), UMR9189 (CRISAL), UMR9015 (LISN)
- INSIS : UMR7357 (ICube)
- INSMI : UMR5149 (IMAG), UMR8071 (LaMME), UMR8524 (LPP)
- INSU : UAR3282 (OREME), UAR3470 (PYTHEAS)
- INC : UAR3290 (MSAP)