



ANNEE 2018-2019

Le métier de Bio- informaticien

PREPROFESSIONNALISATION

DAVID EKCHAJZER
MIASHS PARIS 1 PANTHEON-SORBONNE



Table des matières

Le métier de Bio-informaticien	0
Présentation du métier	3
Origines	3
Prérogatives.....	4
Le bio-informaticien dans le marché de l'emploi	5
Etudes.....	5
Etat de l'emploi	6
Le profil du bio-informaticien	7
Avenir du métier	8
Evolution du métier dans les années à venir	8
Le bio-informaticien face aux enjeux éthiques et politiques.....	9
ANNEXES	11
Bibliographie	12
Figures	13
Interview Nicolas Ballois	17
Interview Jérôme Hedin	21

Le 14 avril 2003, le premier séquençage complet de l'ADN s'achève. Il aura fallu 10 ans et 2,7 milliards de dollars pour y parvenir¹. Aujourd'hui, en 2018, vous pouvez séquencer en quelques heures votre ADN pour une cinquantaine d'euros grâce à des plateformes comme myHeritage². Cette avancée exponentielle n'est pas sans rappeler celle des ordinateurs, caractérisée par la loi de Moore³ qui peut être vulgarisée par : « tous les 18 mois, la puissance de calcul des processeurs est doublée à prix constant ». Cette analogie⁴ est pertinente, car c'est en partie grâce à l'avènement de l'informatique et à la croissance de la puissance de calcul que les biologistes ont pu accomplir cette prouesse. L'éclosion d'un nouveau domaine des biotechnologies a été à l'origine de la réunion de l'informatique et de la biologie : la bio-informatique. Cette science est à mi-chemin entre la biologie et l'informatique, plus précisément entre les biotechnologies⁵ et le big data⁶. Le métier de bio-informaticien qui en découle consiste en la gestion (traitement, simulation, synthétisation) de données biologiques à travers l'informatique. Le bio-informaticien se doit donc d'avoir une double compétence dans ces domaines. L'étude de ce métier est intéressante, car il est peu connu du grand public. Il est pourtant appelé à jouer un rôle déterminant dans les années à venir dans les avancées biotechnologiques et médicales, voire à transformer la vision que l'on a du vivant. De plus, il permet d'associer le numérique et la biologie, deux secteurs en pleine révolution. C'est pourquoi j'ai choisi d'aborder ce métier pour ce travail de recherche. Je l'ai découvert en 2015, par le biais des conférences sur les nouvelles technologies liées à la biologie. Passionné par la biologie et l'informatique, je me suis tout de suite intéressé au métier de bio-informaticien et j'ai suivi son évolution ces dernières années. Ce travail de recherche a été une aubaine pour en apprendre davantage, notamment pour comprendre certains aspects du métier avec lesquels j'étais peu familier. Comme vous pourrez le voir, je me suis appuyé sur deux interviews de bio-informaticiens, chacun abordant un aspect de leur métier. J'ai pu poser mes questions à Nicolas Ballois, bio-informaticien dans une grande société pharmaceutique, UCB bio-pharma que j'ai contacté par le biais de twitter. J'ai également pu interviewer Jérôme Henin, enseignant-chercheur au CNRS⁷ en biologie moléculaire. Celui-ci considère « [qu'il n'est] pas un bio-informaticien au sens étroit du terme, mais au sens large, [car il] utilise l'informatique pour faire de la biologie. ». Je l'ai contacté grâce à l'une de ses étudiantes. Mes recherches m'ont également amené à solliciter la communauté des bio-informaticiens français à travers twitter, mais également à travers leur site web de référence Bioinfo-fr⁸.net qui regorge d'informations statistiques, techniques et plus générale sur la bio-informatique. Pour le reste, mon travail s'appuie sur la bibliographie que vous trouverez en annexe.

¹ *Projet génome humain — Wikipédia. 2018. Projet génome humain — Wikipédia. [EN LIGNE] Disponible sur » :* https://fr.wikipedia.org/wiki/Projet_g%C3%A9nome_humain. [Accédé le 04 Décembre 2018].

² *MyHeritage propose d'identifier vos origines à travers le séquençage de votre ADN pour 49€99*

³ *Loi de Moore — Wikipédia. 2018. Loi de Moore — Wikipédia. [EN LIGNE] Disponible sur :* https://fr.wikipedia.org/wiki/Loi_de_Moore. [Accédé le 06 Décembre 2018]

⁴ *Voir figures en Annexe : Référence 4 | Coût de séquençage de l'ADN comparé à la loi de Moore | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>*

⁵ *L'OCDE définit la biotechnologie comme « l'application à des organismes vivants des principes scientifiques et de l'ingénierie à la transformation de matériaux vivants ou non-vivants ¹aux fins de la production de connaissances, de biens et de services. »*

⁶ *Le BigData désigne un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler. (lebigdata.fr)*

⁷ *Centre Nationale de la Recherche Scientifique*

⁸ <https://bioinfo-fr.net/>

Présentation du métier

Pour présenter le métier peu connu de bio-informaticien, je me suis appuyé sur les définitions de l'ONISEP⁹ et du CIDJ¹⁰, que j'ai enrichi par les entretiens que j'ai effectués. En effet, l'interview de deux bio-informaticiens travaillant dans des secteurs distincts (pharmaceutique et recherche publique) et ayant des prérogatives différentes m'a permis de comprendre que le métier de bio-informaticien a plusieurs aspects qu'il convient de dissocier.

Origines

La première mention de l'informatique appliquée à la biologie peut être trouvée dans une lettre écrite par le père de l'informatique, Allan Turing à l'un de ses confrères Mike Woodger « *Notre nouvelle machine [le Ferranti Mark I] va arriver lundi. J'espère faire dans les premiers travaux quelque chose en rapport avec "l'embryologie chimique". En particulier, j'espère pouvoir expliquer l'apparition de suite de Fibonacci dans les cônes de sapin.* »¹¹ »¹². Allan Turing en 1951 alors que l'ordinateur n'en est qu'à ses débuts, voit déjà dans son invention un moyen de comprendre le vivant. Il faudra attendre quelques années pour que le monde de la biologie s'empare de cette pratique. Selon Jérôme Henin, enseignant-chercheur de recherches au CNRS, on peut donner deux origines au métier de bio-informaticien. Tout d'abord, « l'informatique est dans la biologie parce qu'elle est partout ». Dans les années 60, les biologistes et les chimistes moléculaires commencent à séquencer des protéines de plus en plus précisément. Ils produisent alors de plus en plus de données qu'ils n'arrivent plus à traiter.¹³ L'avènement parallèle de l'ordinateur est une aubaine pour les biologistes, qui voient dans la puissance de calcul des machines un moyen de traiter leurs données. Nicolas Ballois, se souvient avoir vu l'explosion de ce nouveau domaine quand il était étudiant, alors même que ce métier était d'abord sous-traité dans des pays en voie de développement (comme l'Inde). S'en est suivi une croissance exponentielle du secteur. En France, le métier de bio-informaticien n'a pas évolué comme dans les pays anglo-saxons, peut-être à cause d'une simple différence linguistique. Pour Jérôme Hedin, il y a dans le métier de bio-informaticien « l'idée qu'on traite l'information biologique, qui a à voir avec le mot anglais *bio-informatics*¹⁴ et il y a en France cette dualité entre [...] traiter l'information biologique ou alors utiliser l'informatique dans la biologie. ». En anglais, deux métiers lient l'informatique à la biologie : celui de « *bio-informaticien* » qui traite l'information biologique, et celui de « *computed biologist* » qui utilise l'informatique dans la biologie. En France, pourtant, ces deux métiers ont évolué sans jamais se dissocier. On observe cette différence lors de la création des différentes sociétés savantes de bio-informatique. Celle, internationale, a été nommée « International Society for Computational Biology and Bioinformatics »¹⁵, alors que la société française a été

⁹ Office Nationale d'Informations sur les Enseignements et les professions | Onisep. 2018. bio-informaticien / bio-informaticienne - Onisep. [EN LIGNE] Disponible sur : <http://www.onisep.fr/Ressources/Univers-Metier/Metiers/bio-informaticien-bio-informaticienne>. [Accédé le 01 Décembre 2018].

¹⁰ Centre d'Information et de Documentation Jeunesse

¹¹ Il s'agissait pour Turing de définir un modèle statistique dans la formation du cône d'un sapin

¹² turing_woodger février 1951. [EN LIGNE] Disponible sur : http://www.alanturing.net/turing_woodger_feb51/. [Accédé le 02 Décembre 2018].

¹³ Séquençage de l'ADN — Wikipédia. 2018. Séquençage de l'ADN — Wikipédia. [EN LIGNE] Disponible sur : https://fr.wikipedia.org/wiki/S%C3%A9quen%C3%A7age_de_l'ADN. [Accédé le 28 Novembre 2018].

¹⁴ « *informatic* » en anglais étant utilisé dans le sens du traitement de l'information

¹⁵ Leur site internet : <https://www.iscb.org/> (ISCB)

nommée « Société Française de Bio-informatique »¹⁶. Julien Fumey¹⁷, docteur en bio-informatique et membre du conseil d'administration de la Société Française de Bio-informatique, dans un article¹⁸ pour la communauté des bio-informaticiens français, explique cette différence ainsi : « En français, le mot « informatique » désigne à la fois la science¹⁹ et la technique, ce qui n'est pas le cas en anglais. D'ailleurs, la frontière entre science et technique est très mince ; et la science a besoin de la technique qui, elle-même, a besoin de la science. Dans la pratique, on est rarement dans la science pure ou dans la technique pure, en tout cas en bio-informatique. ». C'est pourquoi je traiterai du métier de Bio-informaticien dans l'acceptation française du terme.

Prérogatives

Selon le CIDJ, « le bio-informaticien organise les données biologiques [...] pour les rendre exploitables grâce au développement et à l'utilisation de puissantes bases de données. C'est à lui de concevoir ou d'acheter les logiciels spécifiques au secteur d'activité et à l'entreprise. »²⁰. Ses prérogatives vont varier en fonction de ses compétences et de l'entreprise dans laquelle il évolue. Une grande partie du travail est donc de recevoir et de mettre en forme une énorme quantité de données. Nicolas Ballois, travaillant dans une unité récemment créée au sein d'une entreprise pharmaceutique, « reçoit des données parfois [difficiles] à regarder [...], les formate pour les mettre dans une base de données [...], branche sur cette base de données [...] les outils qui permettent de traiter les données en faisant des statistiques ». Mais toutes les entreprises n'ayant pas les mêmes moyens, l'analyse peut être plus ou moins poussée. Certains bio-informaticiens vont se contenter de mettre en forme leurs données. « Il y a [...] des sociétés qui font tout avec Excel et qui encodent tout à la main », quand d'autres vont développer leurs propres logiciels d'exploitation de données. « Il y a des sociétés qui vont beaucoup plus loin, en faisant des analyses multivariées qui créent des modèles à partir de ça ». La mise en forme et l'exploitation de ces données a plusieurs vocations. Elle permet d'abord de répondre aux besoins quotidiens des biologistes pour mettre en place des plans d'expérience²¹, exploiter les données récoltées, les mettre dans une forme compréhensible pour le biologiste. Chez UCB-biopharma, l'unité de Nicolas Ballois a ouvert une *hotline*²² pour être en contact permanent avec tous les biologistes de l'entreprise dans le monde. « Il y a une partie de la *hotline* pour répondre aux scientifiques, par exemple s'ils ont besoin de données demandées par les autorités sanitaires ». On voit donc qu'une grande partie du travail du bio-informaticien consiste en une assistance du biologiste. Mais le traitement des données peut également être fait au profit de la gestion de l'entreprise, « voir par exemple si le rendement de production d'un produit peut être augmenté de 20% » dans le cas de UCB-biopharma. En effet, les entreprises pharmaceutiques ou des biotechnologies engendrent beaucoup de données biologiques. Le traitement de ces données peut leur permettre de réduire leur coût de production, leur coût de gestion... Une autre prérogative du bio-informaticien est la gestion des outils dont il dispose. Cela passe par plusieurs tâches qui lui incombent. Tout d'abord, l'acquisition (payante ou gratuite) de logiciels. « On a le choix des outils que l'on utilise. Il existe

¹⁶ Leur site internet : <https://www.sfbf.fr/> (SFBF)

¹⁷ Son site internet : <https://www.julienfumey.fr/>

¹⁸ m4rsu, Julien Fumey, 2015. Aux origines de la bioinformatique. Bioinfo-Fr, [EN LIGNE]. Disponible sur : <https://bioinfo-fr.net/aux-origines-de-la-bioinformatique?highlight=%27origine%27> [Accédé le 6 Décembre 2018].

¹⁹ On nomme la science en anglais « *computer science* »

²⁰ Josée Lesparre, 2018. Aux origines de la bioinformatique. CIDJ, [EN LIGNE]. Disponible sur : <https://www.cidj.com/metiers/bio-informaticien-bio-informaticienne> [Accédé le 27 Novembre 2018].

²¹ On nomme plan d'expérience la suite ordonnée d'essais d'une expérimentation, chacune permettant d'acquérir de nouvelles connaissances en contrôlant un ou plusieurs paramètres d'entrée (le Parisien)

²² Assistance téléphonique

beaucoup de logiciels. Il faut être un peu expert pour pouvoir les utiliser. ». C'est pourquoi une partie importante du travail consiste dans l'apprentissage et l'enseignement des différentes technologies. « On aide les gens à se former aux logiciels d'analyse de données ». De plus, ces outils peuvent être développés par les bio-informaticiens eux-mêmes. C'est le cas de Jérôme Henin, qui « contribue à ces logiciels au sein d'une communauté de développeurs appartenant à [des] labos différents ». Enfin, le bio-informaticien peut faire de la recherche expérimentale. C'est d'ailleurs la principale activité de Jérôme Henin, qui « fait [...] des simulations [...] qu'on appelle des expériences *in silico* ». Ce terme latin, qu'on pourrait traduire par « en silice », est une analogie aux expressions latines *in vivo* (« en vie »), *in vitro* (« en éprouvette »)... et fait référence au silicium, l'un des composants des processeurs. Cette nouvelle méthode expérimentale consiste en la simulation d'organismes vivants ou de molécules dans un ordinateur pour mieux les observer ou pouvoir répéter un nombre quasi infini de fois des expériences.²³ Cette prérogative du bio-informaticien se retrouve en France, surtout dans les laboratoires publics de recherche qui doivent disposer de centres de calcul très puissants (les institutions publiques françaises en possèdent trois). Si les prérogatives du bio-informaticien sont multiples, elles ont néanmoins un dénominateur commun : elles nécessitent une bonne connaissance de l'outil informatique et des phénomènes biologiques.

Le bio-informaticien dans le marché de l'emploi

Pour effectuer une analyse pertinente du métier de bio-informaticien dans le marché de l'emploi, j'ai pu m'appuyer sur le témoignage des personnes que j'ai interviewées, que j'ai pu confirmer grâce à une étude²⁴ très complète, menée en juin 2018 auprès de 400 bio-informaticiens français. C'est l'étude la plus récente et la plus complète qui existe sur le métier de bio-informaticien et qui apporte un éclairage statistique sur celui-ci. L'étude permet de se faire une idée du profil du bio-informaticien français, mais également de confronter les secteurs public et privé. Au regard de cette confrontation, mon analyse du marché de l'emploi a pu être affiné.

Etudes

D'après l'étude « Enquête Bioinfo-fr 2018 », les bio-informaticiens sont des personnes très formées. En effet, 35% d'entre eux sont doctorants 60% sont Bac+5²⁵. On peut expliquer cela par la nécessité d'une double compétence dans des domaines de pointe. Sur le contenu des études, il y a une très grande diversité des profils, à l'image des deux entretiens que j'ai pu effectuer. Nicolas Ballois a « intégré Agro ENSAIA. Là-bas, [il a] fait [son] cycle ingénieur. [Ensuite, il a] fait la filière pharmacie industrielle. » De son côté, Jérôme Henin a « fait un master de Chimie Physique [...], puis une thèse officiellement [en] Chimie théorique [...] sur des thèmes biologiques grâce à l'outil informatique ». Ils ont néanmoins en commun une prépa

²³ *In silico* — Wikipédia. 2018. In silico — Wikipédia. [EN LIGNE] Disponible sur : https://fr.wikipedia.org/wiki/In_silico. [Accédé le 03 Décembre 2018].

²⁴ Kumquat, Gwenaëlle Lemoine, 2018. Enquête Bioinfo-fr 2018 : un portrait de la bioinfo. Bioinfo-Fr, [EN LIGNE]. Disponible sur : <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo> [Accédé le 24 Novembre 2018].

²⁵ Voir figures en Annexe : Référence 25 | Type de diplôme | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

BCPST, connue pour son côté généraliste. Plus généralement, les bio-informaticiens ont des profils très différents, venant d'études de biologie, d'informatique, de pharmaceutique, de chimie moléculaire, de mathématique, de statistiques... 40% des personnes interrogées ont une dominance dans leur travail plutôt équilibrée, alors que 30% sont plus tournées vers l'informatique, et 30% vers la biologie.²⁶ C'est justement cette pluralité des profils qui font fonctionner les équipes de bio-informaticiens qui doivent répondre à des problèmes aux frontières de toutes ces disciplines. Nicolas Ballois insiste sur l'importance du travail en équipe dans son domaine. Dans son équipe, « [ils sont] tous à la base des ingénieurs plutôt généralistes qui viennent de la biologie mais [ils] aimerait bien avoir quelqu'un avec des connaissances informatiques un peu plus solides ». On voit bien la nécessité d'avoir une équipe ayant des compétences diverses. Si les équipes de bio-informaticiens sont hétéroclites, c'est aussi parce que l'offre de formation dans ce domaine spécifique est assez faible, bien qu'elle tende à se densifier. L'association des Jeunes Bio-informaticiens de France (JeBiF) tient une liste²⁷ des formations spécialisées en France sur son site internet. On dénombre 1 DUT, 4 licences, 17 masters, 2 spécialités en école d'ingénieurs, et un diplôme universitaire. Soit 25 formations dans toute la France, ce qui est un nombre assez faible. La majorité des bio-informaticiens viennent donc de domaines différents ou se spécialisent lors de leur doctorat (d'où le grand nombre de doctorants, 35%, face au 1% que compte la France). Comme dans la majorité des métiers liés aux nouvelles technologies, la formation en continue prend une place extrêmement importante. Les logiciels évoluant sans cesse, une grande partie du travail consiste à apprendre à s'en servir. C'est pourquoi Nicolas Ballois conseille « de ne pas trop se former sur un outil spécifique. » pour ne pas être tributaire de celui-ci. Cette formation passe par internet, notamment par la communauté de bio-informaticiens français (Bioinfo-fr) qui publie régulièrement des formations sur des domaines²⁸ précis et offre un *chat* en ligne²⁹ permettant de connecter tous les bio-informaticiens de la communauté. La formation est également faite au sein même des équipes de bio-informaticiens comme Nicolas Ballois nous l'indique « Ça m'arrive de me mettre à côté d'un collègue pendant une heure pour essayer de l'aider à régler un problème ».

Etat de l'emploi

On observe une réelle différence dans le marché de l'emploi entre les secteurs privé et public. Tout d'abord, selon « Enquête Bioinfo-fr 2018 », le public est le premier secteur d'emploi en bio-informatique, à hauteur de 75%³⁰. « Les moyens de diffusion de cette enquête ont pu biaiser les résultats en faveur du secteur public » explique l'autrice de l'étude. Selon Jérôme Henin, « les recrutements [sont] en stagnation ou en léger déclin côté public. ». De plus, les recrutements dans le public reposent beaucoup sur des contrats plutôt précaires. 10% des contrats sont en stage, 55% en CDD ou contrat thèse contre 30% de CDI.³¹ Concernant les rémunérations, le salaire médian dans le public se situe dans la tranche

²⁶ Voir figures en Annexe : Référence 26 | Dominance dans le travail | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

²⁷ JeBiF - RSG France. 2018. Les formations | JeBiF – RSG France. [EN LIGNE] Disponible sur : <https://jebif.fr/fr/bioinformatique/les-formations/>. [Accédé le 06 Décembre 2018].

²⁸ La partie didacticiel de la communauté : <https://bioinfo-fr.net/categorie/didacticiel>

²⁹ Lien du chat en ligne : <https://webchat.freenode.net/?channels=bioinfo-fr>

³⁰ Voir figures en Annexe : Référence 30 | type d'emplois | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³¹ Voir figures en Annexe : Référence 31/34 | type de contrat selon public/privée/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

25000€-30000€³², contre 20520€³³ pour tous les secteurs confondus en France. Le secteur privé, quant à lui, se porte bien. Selon Nicolas Ballois, « il y a juste à claquer des doigts pour trouver du boulot. Et les salaires sont bons ! ». Les entreprises qui embauchent appartiennent au secteur pharmaceutique ou biotechnologique, voire agroalimentaire, trois secteurs en pleine croissance. Le type de contrat prédominant est le CDI, qui représente 65% des contrats dans le privé. Les autres contrats se partagent entre 15% de CDD et 10% de stages ou d'alternances.³⁴ Le salaire médian dans le privé se trouve dans la tranche 30000€-35000€³⁵. On voit donc que les contrats sont plus stables dans le privé et plus rémunérateurs. On doit tout de même nuancer, car 45% des contrats concernent des 18-30 ans dans le public³⁶ ; ce qui signifie que le public embauche beaucoup plus de jeunes que le privé. Les salaires et la stabilité de l'emploi tendant à évoluer avec l'âge, on peut ainsi expliquer en partie la différence entre public et privé. Le métier de bio-informaticien est tourné vers l'international, comme l'illustre l'exemple de Nicolas Ballois qui travaille en Belgique. D'après lui « les équipes de data ne sont pas énormes, ce qui veut dire qu'il vaut mieux être un peu mobile, au moins au niveau de la France, sinon au niveau de l'Europe [ou] du monde ». De plus en fonction des domaines de spécialité de chacun, les laboratoires ou entreprises qui travaillent dans ces domaines peuvent amener les bio-informaticiens à s'expatrier. Comme le remarque Jérôme Henin, dans son domaine (simulation moléculaire) « il y a deux, trois labos qui se débrouillent ; mais globalement, la France est un peu à la traîne, les voisins européens, en particulier les Britanniques et les Allemands étant très forts ». Près de 20% des répondants à l'« Enquête Bioinfo-fr 2018 » travaillent à l'international³⁷ alors même que cette enquête, comme le rappellent ses créateurs, « a principalement été diffusée via les lecteurs de [leur] blog ainsi que la liste de la SFBI (qui tout deux touchent une population majoritairement française) ».

Le profil du bio-informaticien

Nicolas Ballois nous explique que « les profils de bio-informaticiens sont très recherchés parce qu'il y a beaucoup de profils très bons en informatiques [...] à l'inverse il va y avoir des biologistes [...] mais dès qu'ils vont être confrontés à une multitude de données ils ne sauront rien en faire [...]. Nous sommes le chaînon entre les deux. Nous ne sommes pas de super-bons informaticiens [...] mais on comprend ce que les biologistes nous demandent et de l'autre côté on comprend comment les logiciels et le traitement des données fonctionne. ». Le bio-informaticien se démarque donc par sa compréhension des deux domaines dans lequel il évolue bien qu'il ne soit généralement pas un expert dans chacun de ces domaines. Au niveau du genre, il y a une petite disparité : 59% d'hommes pour 40% de femmes et 1% d'autres³⁸ qu'il faut comparer au domaine de l'informatique qui compte 93%

³² Voir figures en Annexe : Référence 32/35 | Salaire en K€ selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³³ Jullien Sellier, 2018. Le salaire médian en France est de 1.710 euros nets par mois. RTL, [EN LIGNE]. Disponible sur : <https://www.rtl.fr/actu/conso/le-salaire-median-en-france-est-de-1-710-euros-net-par-mois-7794751410> [Accédé le 1 Décembre 2018].

³⁴ Voir figures en Annexe : Référence 31/34 | type de contrat selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³⁵ Voir figures en Annexe : Référence 32/35 | Salaire en K€ selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³⁶ Voir figures en Annexe : Référence 36 | Répartition des âges selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³⁷ Voir figures en Annexe : Référence 37 | Localisation du travail actuelle | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

³⁸ Voir figures en Annexe : Référence 38 | Genre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

d'hommes pour 7% de femmes³⁹. Le domaine de la bio-informatique est donc plutôt mixte et tend à le devenir au vu de la mixité des bio-informaticiens en stage (5% des contrats sont des stages exercés par des femmes et 5% par des hommes)⁴⁰. Les bio-informaticiens sont plutôt jeunes. L'âge médian tend vers la trentaine, 30% des bio-informaticiens ayant entre 26 et 30 ans⁴¹. On peut expliquer cette jeunesse par la nouveauté du domaine et la création récente de formations spécifiques. Enfin, du point de vue des technologies, les bio-informaticiens sont près de 25% à utiliser Python (langage le plus utilisé en science et traitement de données)⁴², 20% à utiliser R et 20% à utiliser Bash. Nicolas Ballois le confirme : « Python est très à la mode en ce moment, comme R, un *freeware* assez puissant. Les doctorants ont fait leurs thèses avec ça. ». Finalement, si l'on devait dresser le profil type du bio-informaticien français, on pourrait dire que c'est un homme, d'un âge compris entre 26 et 30 ans, en CDD dans une institution publique après un Bac+5 gagnant entre 30000 et 35000€ par an et qui utilise quotidiennement Python.

Avenir du métier

Les entretiens que j'ai pu réaliser dans le cadre de ce travail m'ont permis de dégager un lourd questionnement sur l'avenir du métier de bio-informaticien, tant au niveau des technologies et des débouchés qu'au niveau éthique. J'ai également perçu une potentielle désillusion des bio-informaticiens face à la complexité du vivant en lisant une tribune de Laurent Alexandre sur ce qu'il appelle « l'hiver biotechnologique ». J'ai donc décidé de traiter dans la dernière partie de mon travail l'avenir du métier de bio-informaticien à travers un questionnement sur l'évolution de la technique, mais également sur les questions que cette évolution engendre.

Evolution du métier dans les années à venir

L'augmentation de la puissance de calcul dans les années à venir va permettre aux bio-informaticiens de traiter de plus grandes quantités de données et faire des analyses plus poussées. Dans la société de Nicolas Ballois, ils vont « aller vers beaucoup plus de puissance de calcul pour réduire les temps et les coûts de production avec des [processus] de fabrication de produit [et] rendre [la production] plus écolo ». De là à parler d'intelligence artificielle, il n'y a qu'un pas que Jérôme Henin ne voudrait pas franchir : « C'est une question qu'on est tous en train de se poser. Je pense qu'on est actuellement dans une bulle de l'intelligence artificielle. [...] On essaie de se convaincre que ça peut tout résoudre. [...] Il y a des travaux qui ont l'air prometteur mais, en y regardant de plus près, ce n'est pas forcément plus

³⁹ Stack Overflow. 2018. Stack Overflow Developer Survey 2018. [EN LIGNE] Disponible sur : <https://insights.stackoverflow.com/survey/2018/#demographics>. [Accédé le 05 Décembre 2018].

⁴⁰ Voir figures en Annexe : Référence 40 | type de contrat selon le genre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

⁴¹ Voir figures en Annexe : Référence 41 | Age | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>

⁴² Voir figures en Annexe : Référence 42 | Langage de programmation | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo> | Python (langage) - Wikimonde. 2018. Python (langage) - Wikimonde. [EN LIGNE] Disponible sur : https://wikimonde.com/article/Python_%28langage%29. [Accédé le 06 Décembre 2018].

intéressant que les outils dont on disposait auparavant. ». Les bio-informaticiens ne sont pas convaincus par l'hégémonie de l'IA sur leur travail. Selon Nicolas Ballois, « il y aura toujours besoin de gestionnaires de projets dans les entreprises pour implémenter ces solutions-là⁴³. Si je regarde les équipes aujourd'hui, [elles] ne sont plus du tout techniques [elles] ne savent pas programmer [...] ce n'est plus que des gestionnaires de portefeuille de données ou de projets ». On peut donc penser que c'est l'augmentation de la puissance de calcul qui apportera une évolution au métier, plutôt que l'IA. La « bulle d'intelligence artificielle » dont parle Jérôme Henin fait références aux « hivers de l'IA ». En 1965 puis en 1980, les investissements dans le secteur se sont effondrés après qu'on se soit aperçu que les chercheurs avaient fait des promesses irréalisables⁴⁴ à leurs sponsors. S'en est suivie une période de désillusion, avec un ralentissement des recherches⁴⁵ qualifiée d'hiver. Laurent Alexandre, chirurgien-urologue, chef d'entreprise, auteur et polémiste sur les questions de technologie, écrit le 4 juillet 2018, dans *Le Monde*, une tribune titrée « L'hiver biotechnologique »⁴⁶. Selon lui, « l'ingénierie du vivant – cellules-souches, modifications génétiques, organes artificiels – est supposée pouvoir traiter les maladies dégénératives et accélérer le recul de la mort. Mais, entre les fantasmes technologiques et la commercialisation d'un traitement bien évalué, il y a un immense fossé que certains chercheurs et les start-ups ignorent. ». Il fait ici une analogie avec « l'hiver de l'IA » et prédit une désillusion et une baisse des investissements dans le milieu, notamment sur les recherches bio-informatiques. Selon Nicolas Ballois, « cet hiver a déjà commencé. Les grandes compagnies pharmaceutiques [commencent] petit à petit à abandonner les solutions biotechnologiques [...]. Après, c'est peut-être une deuxième génération des biotech qui est en train d'arriver. ». Jérôme Henin, quant à lui, parle d'« un printemps frisquet qui n'en finit pas. Dans [son] domaine, on promettait qu'on pourrait simuler des molécules de plus en plus grosses, remplacer les essais cliniques par des simulations. Et puis, [...] comme le tassement de la loi de Moore⁴⁷, on arrive à une certaine maturité et on retombe un peu. On se rend compte que beaucoup de choses qu'on a promises ne pourront pas se faire à court terme. ». Les biotechnologistes se rendent compte que le vivant est plus complexe que l'on n'e l'avait imaginé, à l'instar des informaticiens qui se rendent compte de la complexité du cerveau humain. L'évolution du métier de bio-informaticien est donc incertaine, tant au niveau des technologies que des investissements.

Le bio-informaticien face aux enjeux éthiques et politiques

Le nombre de sociétés qui s'installent dans le secteur de la bio-informatique est en pleine expansion, notamment celles qui s'intéressent au génome⁴⁸. On peut nommer l'américaine 23andme⁴⁹ et l'israélienne myHeritage⁵⁰, qui proposent de retracer l'histoire génétique de vos ancêtres à partir de votre ADN. La première proposait il y a quelques années de vous envoyer une analyse complète de vos gènes, de vous indiquer les avantages

⁴³ Logiciels ayant recours à l'intelligence artificielle

⁴⁴ Notamment l'éminence de la création d'un cerveau numérique avec une conscience égale à l'homme

⁴⁵ AI winter - Wikipedia. 2018. AI winter - Wikipedia. [EN LIGNE] Disponible le : https://en.wikipedia.org/wiki/AI_winter. [Accédé le 03 Décembre 2018].

⁴⁶ Le Monde.fr. 2018. Biotechnologies : « winter is coming ». [EN LIGNE] Disponible sur : https://www.lemonde.fr/sciences/article/2018/07/04/biotechnologies-winter-is-coming_5325807_1650684.html. [Accédé le 04 Décembre 2018].

⁴⁷ Correspond à la fin de la loi de Moore. Les processeurs ne doublent plus de puissance à prix constant tous les 18 mois. L'évolution devient de moins en moins exponentielle.

⁴⁸ Ensemble du matériel génétique, c'est-à-dire des molécules d'A.D.N., d'une cellule. (LAROUSSE)

⁴⁹ Leur site internet : <https://www.23andme.com>

⁵⁰ Leur site internet : <https://www.myheritage.fr>

généétiques dont vous disposez et les risques que vous encourez (diabète, cancer, maladie chronique...). Cette disposition a été interdite par les Etats-Unis.⁵¹ Mais les possibilités qu'a ouvertes l'analyse des données biologiques individuelles sont énormes et les risques encourus encore plus importants. Tout d'abord, au niveau de la conservation des données personnelles. Nicolas Ballois se pose la question des données biologiques personnelles « Pour tout ce qui est maladie chronique, on développe de plus en plus d'applications pour suivre [les] traitements, quelles sont les doses de médicaments, prévoir les [...] crises. C'est très bien pour les patients, mais ça veut dire qu'on garde des données qui leur appartiennent. » L'ADN, par exemple, est l'une des seules traces qu'un humain ne peut pas changer. Pourtant, de nombreuses personnes ont donné leur code ADN à des sociétés privées comme 23andme, qui ne sont aujourd'hui pas ou peu contrôlées. Il y a également la problématique du vol de ces données. « Dans les conférences où je vais, on parle [...] de données personnelles notamment à cause du scandale du vol de données avec Facebook et Google. ». Il y a également les problématiques de sélection consciente de l'ADN. La société GenePartner⁵² propose de choisir un partenaire idéal par rapport à vos gènes. On peut également citer la création d'armes létales spécifiques à un type de génome. Dans un article destiné à la communauté des bio-informaticiens français, Yoann Muscaz explique que « le cas le plus connu est celui du "Docteur La Mort", Wouter Basson de son vrai nom, qui a essayé en plein Apartheid de développer une bactérie capable de tuer uniquement la population de couleur noire. »⁵³. Cet inventaire non exhaustif de problématiques n'est aujourd'hui pas traité au niveau (inter)national. Ce doit être le rôle du bio-informaticien d'apporter ces questionnements dans la sphère publique. C'est pourquoi Jérôme Henin tient à nous rappeler que « les politiques de recherches [qu'il] mène sont constitutives des politiques de recherche publique. Donc il y a une vraie citoyenneté à avoir quand on est chercheur. ». Les bio-informaticiens demandent l'ouverture d'espaces de discussion et de prises de décisions sur ces questions : « Nous ne faisons plus de bioéthique en France, déplore Nicolas Ballois. Ces sujets ne sont pas du tout à la mode. Je pense qu'il faudrait à nouveau des états de la bioéthique pour les traiter, d'autant plus que l'on a de nouvelles technologies pour lesquelles on n'a jamais légiféré au minimum au niveau national et bien évidemment au niveau européen. ». Le bio-informaticien a donc également un rôle éthique, voire politique à jouer pour sensibiliser la population aux questions de bioéthiques et participer aux légiférations aux côtés des législateurs.

Nous avons pu voir que le métier de bio-informaticien est apparu en réponse à l'apport massif en donnée apporté par la biologie. Ce métier lié au tri et à la gestion de données biologique a d'abord été un emploi peu qualifié et délocalisé à l'étranger il est devenu en quelques années un métier central des biotechnologies exercé par des personnes aux profils variés ayant pour dénominateur commun une forte connaissance en informatique et en biologie. Il s'agit donc d'un métier très qualifié qui s'exerce aussi bien dans le privé que dans le public et donne lieu à une rémunération plus importante que la moyenne française. Enfin, les bio-informaticiens sont au cœur de questions bioéthiques dont ils doivent aujourd'hui s'emparer pour faire évoluer la société au regard des avancées apportées par leur nouvelle matière.

⁵¹ Yoann M., Yoann Muscaz, 2018. *De l'explosion du tout séquençage à la médecine personnalisée*. Bioinfo-Fr, [EN LIGNE]. Disponible sur : <https://bioinfo-fr.net/de-lexplosion-du-tout-sequencage-a-la-medecine-personnalisee?hlite=%27%C3%A9thique%27> [Accédé le 6 Décembre 2018].

⁵² Leur site internet : <https://www.genepartner.com>

⁵³ Yoann M., Yoann Muscaz, 2018. *De l'explosion du tout séquençage à la médecine personnalisée*. Bioinfo-Fr, [EN LIGNE]. Disponible sur : <https://bioinfo-fr.net/de-lexplosion-du-tout-sequencage-a-la-medecine-personnalisee?hlite=%27%C3%A9thique%27> [Accédé le 6 Décembre 2018].

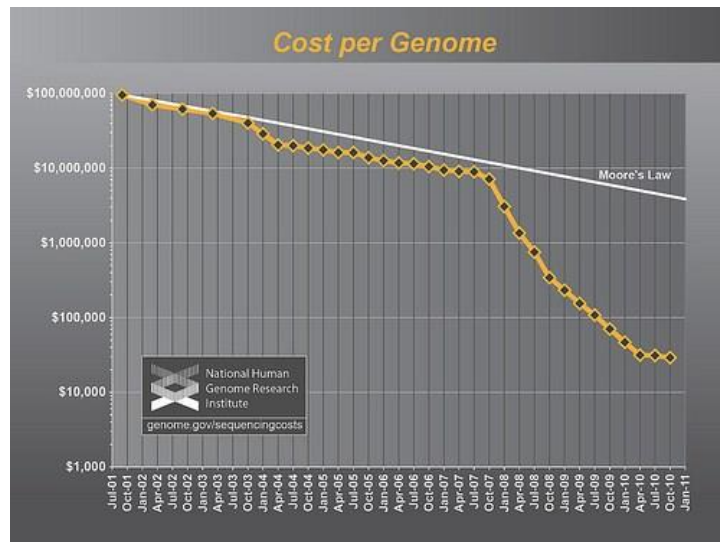
ANNEXES

Bibliographie

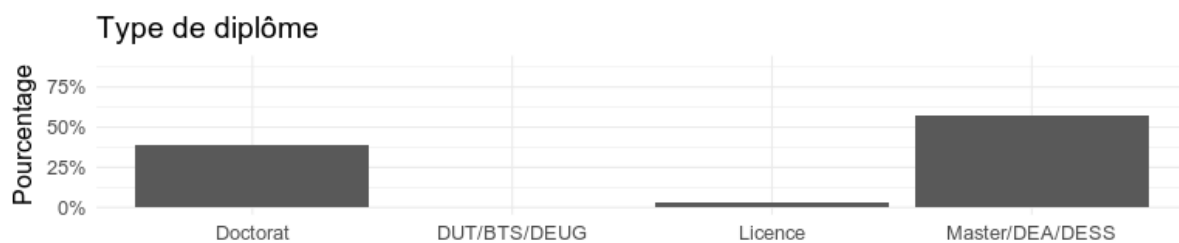
Nom de la source	Type de document	Liens	Auteur	Date
TEDX Paris Salon – Programmer le vivant	Conférence	https://www.youtube.com/watch?time_continue=30&v=S0OFfuUOX5U	Franck Delaplace	Octobre 2012
Site de présentation Franck Delaplace	CV	https://ibisc.univ-evry.fr/~delapla/	Franck Delaplace	inconnue
HuffPost : Programmer les organismes vivants	Article	https://www.huffingtonpost.fr/franck-delaplace/programmer-organismes-ted_b_1891002.html	Franck Delaplace	19/09/2012
Site du projet « Synbiotic »	Site	http://synbiotic.spatial-computing.org/start	Communautaire	inconnue
Académie d'Oxford : Bioinformatique	Livre	https://academic.oup.com/bioinformatics/	Joana Pereira et Andrei N Lupas	1er décembre 2018
État de l'emploi bioinformatique en France : analyse des offres de la SFBI	Article	https://bioinfo-fr.net/etat-de-lemploi-bioinformatique-en-france-analyse-des-offres-de-la-sfbi	Chargée de recherche : Nils Giordano	30 mars 2016

Figures

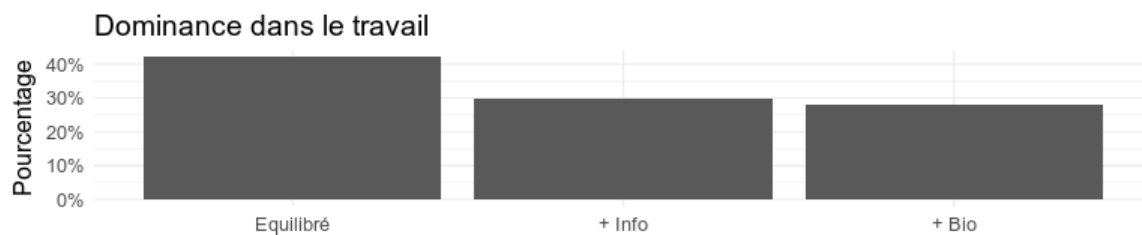
Référence 4 | coût de séquençage d'un génome humain comparé à la loi de Moore | <https://bioinfo-fr.net/de-lexplosion-du-tout-sequençage-a-la-medecine-personnalisee>



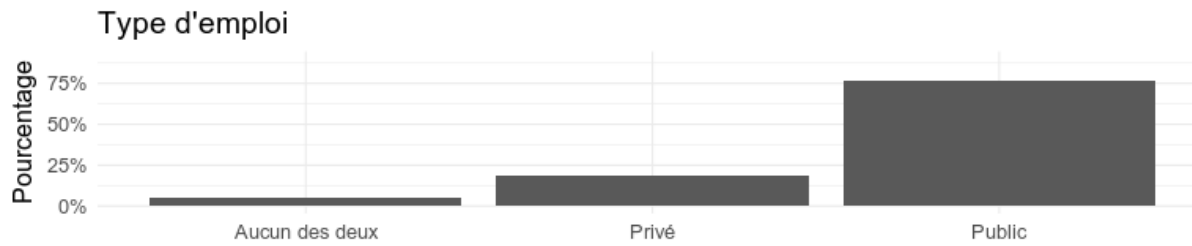
Référence 25 | Type de diplôme | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



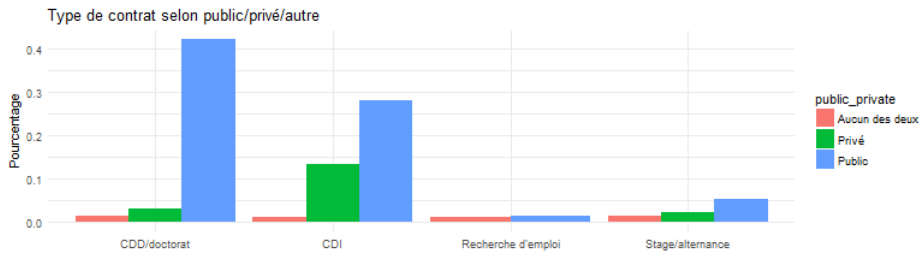
Référence 26 | Dominance dans le travail | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



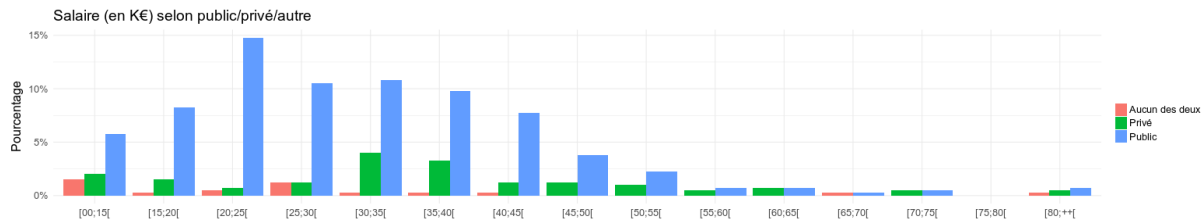
Référence 30 | Type d'emplois | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



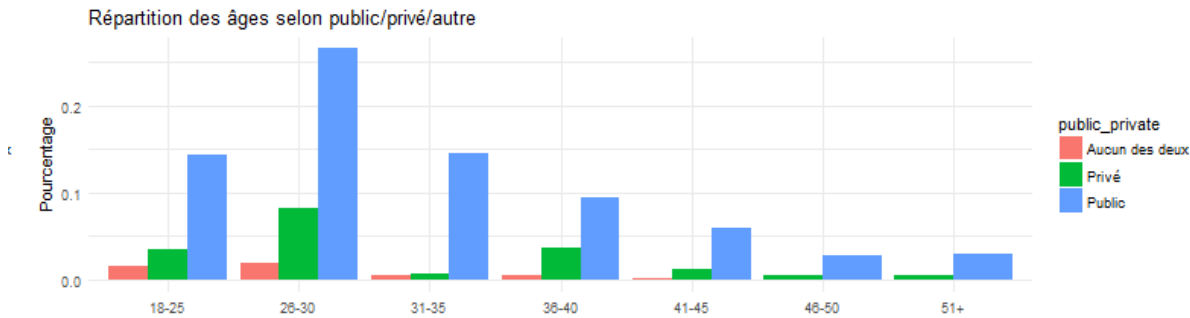
Référence 31/34 | type de contrat selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



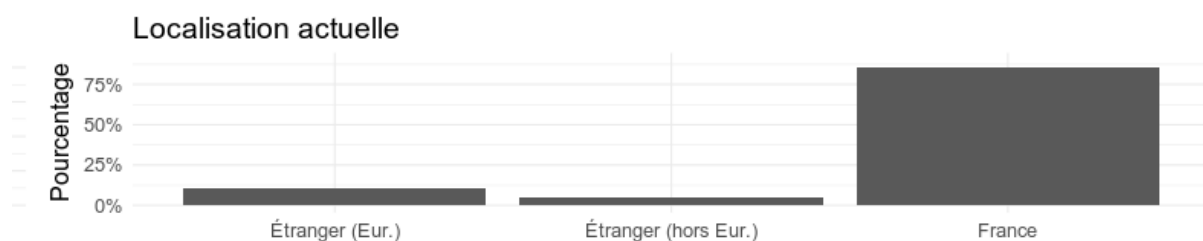
Référence 32/35 | Salaire en K€ selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



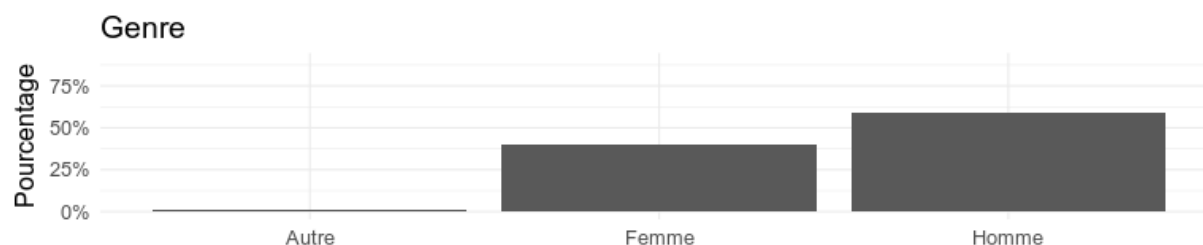
Référence 36 | Répartition des âges selon public/privé/autre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



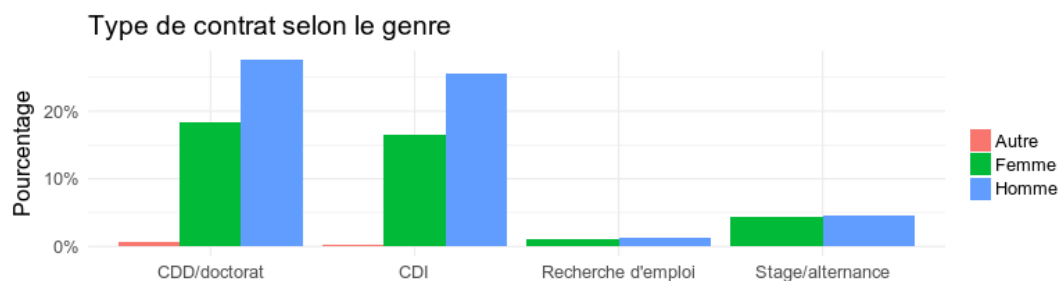
Référence 37 | Localisation du travail actuelle | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



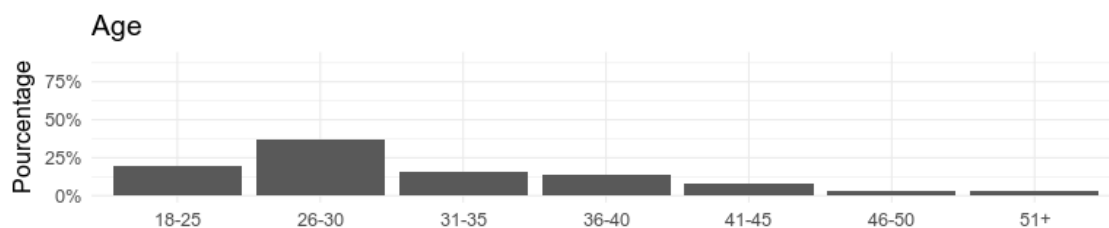
Référence 38 | Genre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



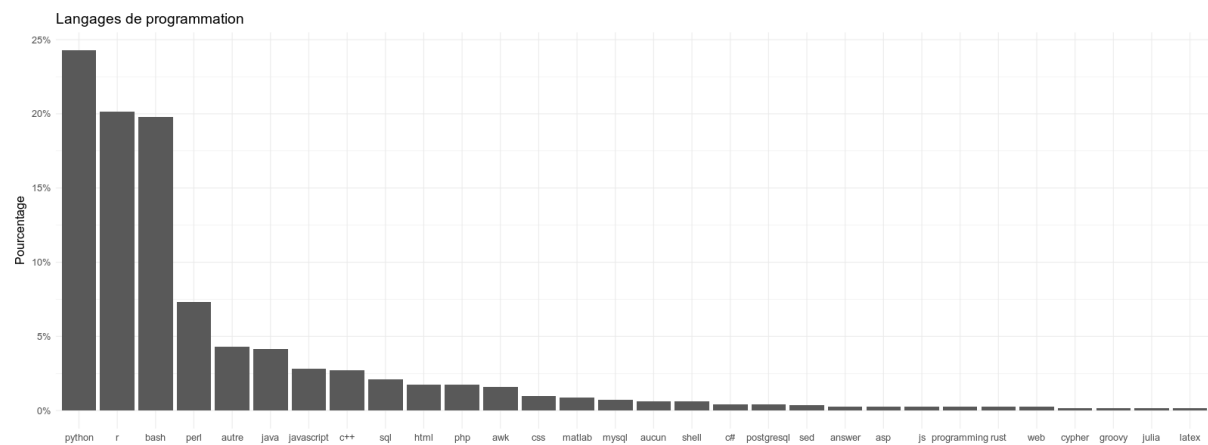
Référence 40 | Type de contrat selon le genre | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



Référence 41 | Age | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



Référence 42 | Langage de programmation | <https://bioinfo-fr.net/enquete-bioinfo-fr-2018-portrait-de-bioinfo>



Interview Nicolas Ballois

Quelles études avez-vous faites ?

J'ai intégré une prépa scientifique générale BCPST à la prépa Saint-Louis boulevard Saint-Michel. A la suite des concours, j'ai intégré une école agro, ENSAIA, où j'ai fait mon cycle ingénieur spécialisé dans les phytopathologies (maladie des plantes). Par la suite, je sentais que ça ne me suffisait pas. J'ai hésité entre un Phd et des études de pharmaceutique, ce que j'ai fait. J'ai fait la filière pharmacien industriel, concernant tous les métiers liés à la production de médicaments. Cela va de la recherche clinique au marketing, à la production... En France, l'industrie pharmaceutique est remplie de pharmaciens.

Vous avez donc un langage commun.

Oui, mais cela va être très différent selon les pays. Par exemple, en Belgique, il y a peu de pharma mais c'est bondé de PhD. Je me suis rendu compte que, dans l'industrie, le data analyste était le roi du pétrole, donc j'ai fait un stage de fin d'études dans ce secteur. Ils m'ont gardé pendant 4 ans. J'ai travaillé à UCB BioPharma, qui a des pôles dans le monde entier avec 9000 employés. Le plus gros site de fabrication est en Belgique, près de Bruxelles. On travaille avec tous les sites européens et asiatiques. Mon rôle, c'est d'aider toute la production pharmaceutique à gérer ses données. Globalement, dans l'industrie pharmaceutique, il y a des données qui viennent de plein de logiciels différents, des cuves, des tests en laboratoire ; et, à chaque étape, des bases de données dédiées qui agrègent de l'information.

Ces données doivent-elles être traitées en direct ?

Oui, c'est possible. Tout dépend des moyens et des technologies que l'on souhaite. Parce que c'est très bien d'enregistrer des données, mais il faut pouvoir les traiter derrière.

Quelles sont vos prérogatives ?

On reçoit des données qui, parfois, sont vraiment pas jolies à regarder. On les formate pour les mettre dans une base de données Oracle et, ensuite, on branche sur cette base de données les outils qui permettent de traiter les données en faisant des statistiques. Nous, ces des statistiques très simples : le suivi d'un pH au cours du temps par exemple ; mais il y a des sociétés qui vont beaucoup plus loin, en faisant des analyses multivariées, qui créent des modèles à partir de ça. Il y a aussi des sociétés qui font tout avec Excel et qui encodent tout à la main.

Les outils qui traitent les données, vous les développez ou vous les achetez ?

Cela dépend. Pour tout ce qui est sondes reliées aux usines, on les achète. Il y a de gros logiciels qui savent faire ça (Delta V, Orchestra, Rockwell). Mais une fois qu'on obtient des données avec ces outils-là, on n'a encore rien fait. Nous, ce qu'on fait, c'est transformer les données d'une base de données vers l'autre grâce à des consultants SAS qui font le boulot. Une fois qu'on a fait ça, on peut traiter les données. Le service dans lequel je travaille a été créé pour analyser les données. On a le choix des outils qu'on utilise. Là encore, il existe beaucoup de logiciels. Il faut être un peu expert pour pouvoir les utiliser.

Quelles sont vos prérogatives ?

Nous ne sommes clairement pas IT (tech). Il y a un autre service informatique pour faire ça. Nous, on est vraiment spécifique de l'outil d'analyse. Il y a une part de hotline, pour répondre aux scientifiques par exemple s'ils ont besoin de données demandées par les autorités sanitaires ; là on partage l'écran et on les aide à le faire. Il y a une part de training où l'on aide les gens à se former aux logiciels d'analyse de données. Il y a la part purement data management, où l'on récupère des données qu'on doit nettoyer et mettre en forme, et puis

des projets à plus long terme, voir par exemple si le rendement de production d'un produit peut être augmenté de 20%. Il faut à la fois être biologiste pour comprendre la question, et informaticien (data analyste) pour pouvoir mettre en forme les données. Nous faisons également de l'aide aux développements de nouveaux procédés, par exemple des plans d'expériences.

Quels sont les types de données que vous manipulez ?

On a surtout des données biologiques, sous forme numérique bien sûr. On commence à faire du textMining, de l'IA et aussi du boolean.

Quel langage de programmation utilisez-vous au quotidien ?

Je ne savais pas du tout utiliser de langage script ; au bout d'un moment il a fallu s'y mettre. J'utilise beaucoup du MySQL, ça va de soi. Mais les sociétés ont des langages internes qui dérivent du C++ ou du Python. On s'est beaucoup formé sur du Web pour présenter les données. On peut vraiment utiliser tous les langages de programmation. Nous, au lieu d'apprendre le Python, on a acheté un logiciel qui permet de coder visuellement avec des boîtes mais, évidemment, c'est très cher.

Est-ce que vous travaillez en équipe ?

Nous avons chacun nos projets précis, mais nous travaillons en OpenSpace et on interagit. Ça m'arrive de me mettre à côté d'un collègue pendant 1h pour essayer de régler un problème. L'idée, comme on gère une hotline, c'est qu'on ne veut pas que les gens aillent voir quelqu'un parce qu'il est plus expert, mais on veut partager toutes les informations.

D'où le métier de bio-informaticien tire son origine ?

Quand j'ai commencé à rentrer dans le milieu, c'était l'avènement du pyroséquençage (séquençage de l'ADN haute performance). On arrivait à faire en une journée ce qu'on faisait en un mois dix ans avant. Je me souviens avoir bossé à l'INRA. Des personnes venaient de nous présenter un nouveau pyroséquenceur, mais il fallait pouvoir traiter des données qui sortaient, donc ils mettaient à disposition un service 24/24 d'analyse des séquences d'ADN. Il suffisait d'envoyer les données en Inde pour savoir de quoi il s'agissait. C'est vraiment là que j'ai entendu parler du métier de bio-informaticien. Aujourd'hui, je vois très clairement à quel point ces compétences qui sont très recherchées et très intéressantes.

Quelles sont vos relations avec les biologistes

On travaille très étroitement avec eux, notamment à travers la hotline que nous avons mise en place. Les profils de bio-informaticien sont très recherchés parce qu'il y a beaucoup de profils très bons en informatique qui vont pouvoir travailler dans tous les milieux comme l'aéronautique. A l'inverse, il va y avoir des biologistes qui seront très bons pour travailler sur de la recherche ; mais, dès qu'ils vont être confrontés à une multitude de données, ils ne savent rien en faire. Nous sommes le chaînon entre les deux. C'est vraiment là où nous nous positionnons et où nous sommes appréciés. Nous ne sommes pas de supers bons informaticiens. On ne serait pas capables de mettre en place un serveur ; mais on comprend ce que les biologistes nous demandent. De l'autre côté, on comprend comment les logiciels et le traitement des données fonctionnent. On a vraiment un rôle à jouer là-dessus, c'est clair.

Quels sont les différents profils que vous avez dans votre équipe ?

Nous sommes tous à la base des ingénieurs plutôt généralistes, qui viennent de la biologie ; mais on aimerait bien avoir quelqu'un avec des connaissances en informatique un peu plus solides. Pas pour faire du travail de routine, mais pour nous montrer des choses qu'on ne connaît pas.

Est-ce qu'il y a une singularité française dans le métier de Bio-informaticien ?

Moi je n'ai jamais travaillé en France, mais une particularité française qui n'est pas une bonne chose, c'est qu'on parle français dans le boulot, ce qui est un vrai frein car, en conférence, les

Français ne parlent pas très bien anglais. En Belgique, tout le monde parle anglais au travail. C'est vraiment fondamental de parler anglais. L'autre particularité est qu'on recrute beaucoup sur diplôme en France, ce qui est une erreur. On devrait plutôt recruter sur compétence. Python est très à la mode en ce moment, comme R, un freeware assez puissant. Toutes les doctorantes et tous les doctorants ont fait leur thèse avec ça.

Comment se porte le marché de l'emploi dans ce domaine ?

Très bien ! Attention, parce que les équipes de data ne sont pas énormes, ce qui veut dire qu'il vaut mieux être un peu mobile au moins au niveau de la France, idéalement au niveau de l'Europe. Si tu es mobile au niveau du monde, alors il y a juste à claquer des doigts pour trouver du travail. Et les salaires sont bons !

On parle beaucoup d'analyse et de traitement des données biologiques grâce à l'informatique. Est-ce que l'utilisation de l'informatique pour maîtriser ou synthétiser le vivant (Synbiotic) peut faire partie du métier de bio-informaticien ?

Oui évidemment. Ce sont plutôt les universitaires et le public qui y travaillent, bien que le privé doive investir là-dedans. Ce sont sûrement des mouvances au niveau mondial et on entend beaucoup parler ça dans les conférences de R&D.

Comment voyez-vous votre métier dans dix ans ?

Aujourd'hui, on est un peu à l'âge de pierre du traitement des données de la bio-informatique. On va aller vers beaucoup plus de puissance de calcul pour réduire les temps et les coûts de production avec des processus de fabrication de produits plus écolos. On va implémenter ce genre de technologies. Je crains qu'on implémente des solutions « clefs en main », qui seront vendues par les sociétés informatiques, et que le bio-informaticien n'ait aucune idée de l'algorithme qu'il y a derrière. C'est déjà un peu le cas dans certaines sociétés, ce que je trouve dommage.

Est-ce que l'IA peut supprimer le métier de bio-informaticien ?

Non, il y a un tel apport. De toute manière, il y aura toujours besoin de gestionnaires de projet dans les entreprises pour implémenter ces solutions-là. Si je regarde les équipes aujourd'hui, les personnes ne sont plus du tout techniques ; ils ne savent pas programmer ou plus programmer. Ce ne sont plus que des gestionnaires de portefeuille de données ou de projets. On est l'une des rares équipes à avoir une petite connaissance technique dans notre métier.

Quels sont les questionnements éthiques, voire politiques qui découlent de votre métier ?

Je n'engage bien entendu pas les sociétés avec lesquelles j'ai travaillé. Ce que je pense, c'est que l'industrie médicale et l'industrie aéronautique, sont les plus réglementées au monde. Rien ne peut être fait sans que ça se sache. Le gros avantage avec l'industrie pharmaceutique, c'est que ça fait plus d'un siècle qu'on engrange des données sur les essais cliniques et qu'on les anonymise. Après, l'industrie pharma commence à être très gourmande en données personnelles. Pour tout ce qui est maladie chronique, on développe de plus en plus d'applications pour suivre les traitements : quelles sont les doses de médicaments à utiliser pour les asthmatiques, pour les patients épileptiques, pour prévoir leurs prochaines crises... C'est très bien pour les patients, mais ça veut dire qu'on garde des données qui leur appartiennent. Tant que c'est fait dans l'industrie pharmaceutique régulée, où les autorités de santé sont solides et réelles, ça ne me fait pas peur parce que l'on ne peut pas faire n'importe quoi de ces données-là. On a accès à des données cliniques. Mais pour en faire autre chose que leur utilisation initiale, on est obligé de demander l'accord de l'autorité de santé.

Que pensez-vous du Machine Learning appliquée à la médecine préventive ?

Ce qui me pose un problème avec, c'est que c'est une technologie blackBox, c'est-à-dire qu'on y met tout un tas de trucs. On ne connaît pas l'algorithme qu'il y a derrière, alors qu'il prend des décisions pour toi. Quand ça prend une décision industrielle, au pire, on perd de l'argent,

ce n'est pas dramatique ; mais quand c'est une décision sur le traitement, ça l'est. Je crois que ce genre d'approches doit être évalué comme un médicament, avec des essais cliniques, des biostatistiques et une méthodologie correcte. On ne peut pas mettre sur le marché un outil comme ça. Je pense que c'est dangereux.

Est-ce que les politiques ou les institutions publiques s'intéressent à ces questions de bioéthique ?

Ça fait bien longtemps que nous ne faisons plus de bioéthique en France. Ces sujets ne sont plus du tout à la mode. Je pense qu'il faudrait à nouveau des états de la bioéthique pour traiter ces sujets-là, d'autant plus qu'on a de nouvelles technologies pour lesquelles on n'a jamais légiféré. Dans les conférences auxquelles je vais, on parle très peu de bioéthique, excepté par rapport aux données personnelles, notamment à cause du scandale du vol de données avec Facebook et Google. Mais encore une fois, ce sont de grandes firmes internationales qui possèdent ces données, ce sont de véritables lobbys, et on comprend pourquoi l'Etat ne veut pas entrer en conflit avec eux.

Pensez-vous que nous sommes à l'aube d'un hiver bio-informatique ?

Je pense que cet hiver a déjà commencé. Si on regarde les grandes compagnies pharmaceutiques, Roche, spécialisée notamment dans l'oncologie, commence petit à petit à abandonner les solutions biotechnologiques pour l'immunothérapie. Après, c'est peut-être une deuxième génération des biotech qui est en train d'arriver.

Interview Jérôme Hedin

Quelles études avez-vous faites ?

J'ai commencé par un Bac S option Maths, puis une prépa scientifique générale BCPST. Après, j'ai fait l'ENS de Lyon en Chimie-Physique. Je savais déjà que j'avais un intérêt pour les sciences en général, et je savais que BCPST était la seule prépa généraliste. Après ça, j'ai fait un master de Chimie-Physique ; et puis en thèse, j'ai fait vraiment de la simulation et j'ai utilisé l'outil informatique. Officiellement Chimie théorique, mais en fait c'était sur des thèmes biologiques. C'était des simulations de molécules biologiques. Finalement, c'est toujours ce que je fais et c'est là que je dois préciser que je ne suis pas un bio-informaticien au sens étroit du terme mais au sens large. J'utilise l'informatique pour faire de la biologie. Le bio-informaticien, c'est en général celui qui fait de l'analyse de séquence biologique et souvent dans un but de biologie évolutive, pour regarder des mutations pathogènes. Mais il y a l'idée qu'on traite l'information biologique qui a à voir avec le mot anglais « bioinformatics » ; et il y a en France cette dualité entre « est-ce que ce métier c'est traiter l'information biologique ? » ou alors « utiliser l'informatique dans la biologie ? ». Moi, je fais un petit peu de traitement des données ; mais c'est une part mineure de mon activité. Je fais plutôt des simulations, ce qu'on appelle des expériences in silico (simulation moléculaire).

Quelle est votre poste aujourd'hui ?

Je suis chargé de recherche au CNRS

Quelles sont vos prérogatives ?

D'une part, je gère complètement le travail de recherche que je fais (je suis autonome), je gère mon travail de recherche. Je décide des sujets sur lesquels je vais travailler. Si j'obtiens des financements, je monte moi-même des projets scientifiques. Je cherche à les faire financer, éventuellement en collaboration. Si j'ai des financements, j'embauche des jeunes chercheurs, des stagiaires de master, des doctorants... J'encadre ces jeunes chercheurs.

Quels types de données manipulez-vous ?

En simulation, les données sont des coordonnées moléculaires atomiques ; mais on les transforme. Si on a une longue simulation, on aura beaucoup d'instantanés du système à différents moments comme pour un film, qui correspondent à un gros volume de données qui ne seront pas vraiment exploitables. Une grosse partie de ce travail consiste à traiter ces données pour les projeter dans des espaces qui ont plus de sens intuitif et qui sont plus petits pour les projeter sur des espaces en une ou deux dimensions. On peut alors visualiser et intuitif.

Quels logiciels utilisez-vous ? Est-ce que vous les développez vous-même ?

Un peu des deux. Pour faire les simulations elles-mêmes, j'utilise des logiciels académiques qui sont développés par des groupes universitaires. Parfois, je contribue à ces logiciels au sein d'une communauté de développeurs qui appartiennent à des labos différents. La plupart des logiciels que l'on utilise sont des logiciels libres. L'important, c'est qu'on ait accès au code source. Souvent, on les compile nous-mêmes car, quand on fait du calcul haute performance, on a besoin de les recompiler sur la machine sur laquelle ça va tourner pour qu'ils soient réglés spécifiquement pour le matériel ; et puis éventuellement on modifie le logiciel. Mais le plus important, comme on fait de la recherche, on est redevable quand on publie un travail avec nos équipes. On est censé connaître l'outil et, si l'on n'a pas accès au code source et que ce sont des boîtes noires, on ne peut pas vraiment dire qu'on connaisse nos outils.

Quels langages utilisez-vous au quotidien ?

Pour les simulations, on fait une campagne qui peut durer des semaines sur des centaines de processeurs. Pour ce niveau de performance, l'outil est écrit en C ou C++ et, les plus vieux, en Fortran, mais de moins en moins. Pour la simulation, c'est toujours du langage compilé. Pour le traitement des données, ce qui est moins critique en termes de performance, on utilise de plus en plus Python. Les gens qui font des séquences biologiques font beaucoup de Java.

D'où vient le métier de bio-informaticien ?

La réponse la plus générale, c'est que l'informatique s'est immiscée dans absolument toutes les sciences et dans tous les domaines. En 2018, je dirais que l'informatique est dans la biologie par ce qu'elle est partout ; mais c'est vrai que ça a commencé assez tôt, et je pense par deux chemins différents. Quand on a commencé à séquencer certaines molécules dans les années 1960 et que ces séquences ont commencé à représenter une taille importante, il a été naturel d'utiliser l'informatique pour traiter ça. Séparément, il y a eu le côté physique. Dans les années 1960, on a commencé à utiliser les ordinateurs pour faire des simulations de petits systèmes physiques moléculaires ou physiques des fluides. Dans les années 1980, on a commencé à avoir des ordinateurs assez puissants pour faire des simulations de petites protéines (1980). Brusquement, l'informatique est rentrée dans la biologie par le biais de la physique.

Y a-t-il une singularité française dans le milieu ?

Il y a de très bonnes équipes côté traitement des données biologique. Il y a des gens qui développent de bons outils logiciels, qui font de belles études de biologie évolutive à partir de ça. De mon côté, dans la simulation moléculaire, il y a deux, trois labos qui se débrouillent ; mais globalement la France est un peu à la traîne. Les voisins européens, les Britanniques et les Allemands en particulier sont très forts.

Comment se porte le marché de l'emploi dans le secteur ?

Comme le marché global de l'emploi ne se porte pas super bien, il faut juger relativement. Globalement, la tendance côté publique est au léger déclin au niveau des recrutements d'enseignants-chercheurs. Je sais que le CNRS diminue ses recrutements de chercheurs permanents. Je dirais qu'on est en stagnation ou en léger déclin côté publique. L'industrie pharmaceutique est également un peu frileuse. Je pense que l'informatique est appelée à jouer un rôle croissant de ce domaine, par exemple en faisant du design de protéine, en créant de nouvelles protéines, en cherchant à leur donner des fonctions voulues.

Est-ce que l'intelligence artificielle va changer votre métier ?

C'est une question qu'on est tous en train de se poser. Je pense qu'on est actuellement dans une bulle de l'intelligence artificielle. On en met à toutes les sauces. On essaie de se convaincre que ça peut tout résoudre. Actuellement, on attend de voir. Tous les labos essaient d'implémenter de l'IA avec plus ou moins de compétences. Résultat, je pense que ça va plus ou moins décanter. Il y a des travaux qui ont l'air prometteur ; mais, en y regardant de plus près, ce n'est pas forcément plus intéressant que les outils qu'on avait avant. Il faudra encore quelques années pour faire le tri et pour qu'émergent les domaines où il y aura vraiment des avancées.

Sommes-nous aux prémices de l'hiver biotechnologique ?

Je pense que nous sommes dans un printemps frisquet qui n'en finit pas. Dans mon domaine, qui a connu une croissance exponentielle de 1980 à 2010, on promettait la simulation de molécules de plus en plus grosses et le remplacement des essais cliniques par des simulations. Et puis, c'est comme le tassement de la loi de Moore, on arrive à une certaine maturité et on retombe un peu. Et on se rend compte que beaucoup de choses qu'on a promises ne pourraient pas se faire à court terme.

Est-ce que votre métier vous pose des questions éthiques ou politiques ?

Indirectement, comme je suis chercheur indépendant au CNRS, les politiques de recherches que je mène sont une partie des politiques de recherches publiques. Donc, il y a une vraie citoyenneté à avoir quand on est chercheur. Sinon, au niveau éthique, moi qui ne fait pas d'application médicale, je ne suis pas confronté aux problématiques des données personnelles. Comme on fait ces énormes calculs dans des centres, ça représente une grande utilisation d'énergie et quelque part on est responsable de l'utilisation de ressources. On commence à se poser la question. Quand j'étais étudiant, ce n'était pas le cas. Quand je démarre ma campagne de calculs, j'estime son bilan carbone ou énergétique, et le bénéfice de l'avancée scientifique que j'en tire par rapport à l'énergie consommée. C'est une question assez nouvelle, celle de la pertinence écologique des ressources qu'on utilise.