WIKILEAKS ET BIOLOGIE, UTILISATION SIMILAIRE DES DONNÉES?

LE 5 DÉCEMBRE 2010 ***ROUD***

Le journalisme de données et la saga Wikileaks sont la transposition dans la société d'un phénomène récent en biologie : le déluge de données. Quels enseignements tirer de ce parallèle ?

TITRE ORIGINAL : OPINION : WIKILEAKS, BIOLOGIE DES DONNÉES, ÉMERGENCE

Le journalisme de données et la saga Wikileaks sont la transposition dans la société d'un phénomène récent en biologie : le déluge de données. Quels enseignements tirer de ce parallèle ?

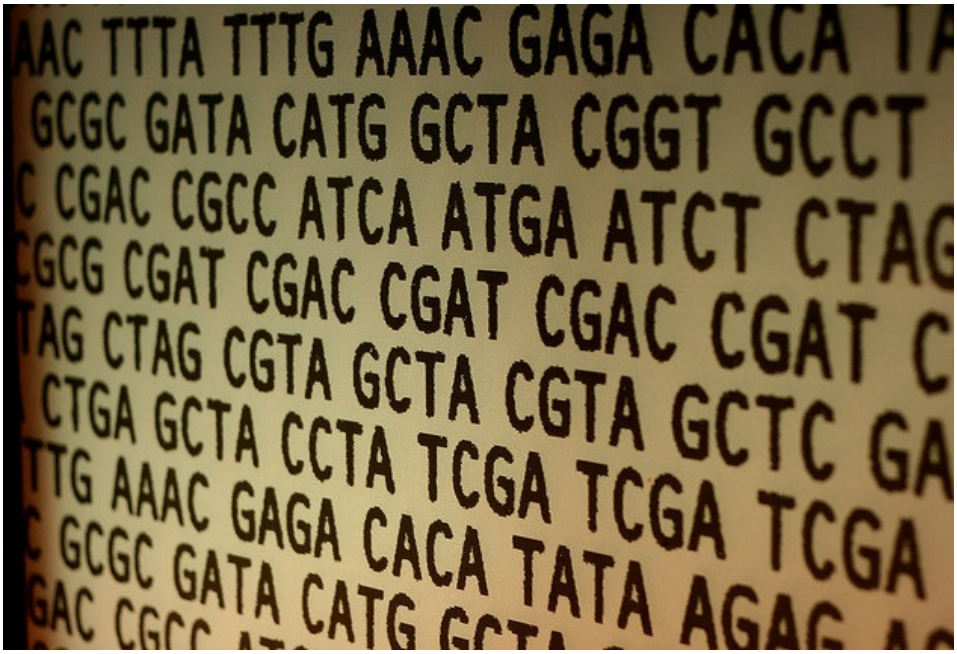
> ***Low-input, high-throughput, no-output biology***

Ainsi Sydney Brenner qualifiait-il dans **une conférence récente** le phénomène de "biologie des données" : en somme, la génération de données brutes ne serait qu'une démarche un peu paresseuse ("low-input""), coûteuse et n'apprenant au fond pas grand chose de neuf sur la biologie ("no output"). Brenner combat en réalité cette idée que la "science" peut émerger spontanément des données, par une analyse non biaisée et systématique, qu'au fond les données vont générer les théories scientifiques naturellement (et c'est aussi un peu le principe d'algorithmes d'analyse comme **Eureka**).

Emergence de la connaissance

La démarche de Wikileaks me semble relever de la même tendance : des données brutes et nombreuses, disponibles à tous, va surgir une vérité, obscure dans les détails mais éclatante vue de loin. **More is different** pour reprendre le titre du papier célèbre du Prix Nobel de physique Phil Anderson. Wikileaks, c'est l'émergence appliquée au journalisme, l'idée qu'un déluge quantitatif va changer la vision qualitative des faits.

Est-ce vrai ? La comparaison avec la biologie de données est éclairante à mon sens. Au-delà des critiques juridiques, sur ce que j'ai entendu, on entend que la majeure partie des mémos de Wikileaks sont sans aucun intérêt, que cette publication met l'accent sur des épiphénomènes ou que les télégrammes qui semblent un peu "croustillants" ne nous apprennent en fait **rien de vraiment nouveau ou rien dont on ne se serait douté**. Allez dans une conférence de biologie, et discutez avec des critiques de la biologie des données, vous entendrez exactement le même genre de critiques, à savoir que l'analyse est trop simple, biaisée, et qu'on ne trouve rien de vraiment étonnant ou neuf. Bref, dans les deux cas, ce saut qualitatif à la Anderson ne se produirait pas, les données sont jolies mais totalement inutiles au fond.



Le retour de l'expert

Il y a néanmoins une différence de taille : si je vous donne la séquence d'ADN d'un gène, vous n'êtes pas capable de dire ce que ce gène fait dans la cellule, c'est une information intéressante mais dont on ne saisit pas la portée exacte (aujourd'hui en tous cas), tandis que si je vous dis que "Sarkozy est autoritaire et colérique", d'une part, c'est une information considérée comme signifiante par l'analyste, donc son contenu informatif est maximisé dès la collecte de celle-ci[1] , d'autre part, vous êtes capable de replacer cette donnée immédiatement dans un contexte plus global, repensant au "Casse-toi pauvre con", à la brouille avec la commission européenne sur les Roms, et plus généralement à sa pratique politique globale.

En d'autres termes, dans le journalisme de données, nous pouvons bien comprendre les sens individuels des atomes de données, mais nous avons *déjà* une idée de l'image globale, du niveau supérieur *émergent*, et du coup, nous sommes tout à fait à même de comprendre comment des petits détails deviennent signifiants sur la vision et l'organisation du monde. Dans ce cadre, on a besoin de nouveaux experts, des personnes ayant une bonne maîtrise de ces petits détails, capables de mettre ensemble ce qui est signifiant a priori pour bien nous aider à visualiser cette réalité (cf. **cette tribune du Monde** signalée par **Enro** sur **twitter**).

Où sont ces experts dans la biologie des données ? Ils sont capables de comprendre les petits faits individuels apparemment anodins, de les mettre ensemble dans un cadre plus global, de les faire comprendre à tous par une représentation adéquate. Lisez ou relisez *L'origine des Espèces*, et vous verrez que c'est exactement la démarche suivie par Darwin. Pense-t-on vraiment que des robots soient capables de faire cela ? Ou n'est-ce pas plutôt le boulot des théoriciens, espèce qui demeure rare en biologie ?

>> Article initialement publié sur **Matières Vivantes**

>> Illustrations FlickR CC : **Garrettc**, Elliot Lepers pour OWNI

1. Un exemple de contenu informatif non maximisé serait une description pièces par pièces de la garde-robe du dit Sarkozy,et je ne suis pas loin de penser que certaines données biologiques abondantes ont à peu près le même intérêt. [↩]

WANDRILLE
le 5 décembre 2010 - 22:16 • SIGNALER UN ABUS - PERMALINK

Excellente convergence! Et nous y voilà, les journalistes sont confrontés aux mêmes problèmes que les biologistes! Qui l'aurait cru!
Aucune des deux professions refuseraient de dire, « non je ne souhaite pas de données de masses ! » Les journalistes des grands quotidiens qui ont eu accès aux données de wikileaks ont procédé de la même manière que lors qu'un chercheurs de biologie intégrative ce retrouve devant une nouvelle base de donnée dicté par une approche humaine, intuitive:
-Recherche dans la basse de données de l'information clé que l'on aurait souhaité obtenir pour confirmer la théorie sur une rumeur qui circulent mais dont personne à la preuve, les journaux de chaque pays on d'abord recherché des informations sur leurs pays et leurs politisent (exemple : Sarkozy, Clinton…)

-Recherche dans la basse de données des informations sur des domaines qui occupe la majorité de la communauté mais ou l'information filtre difficilement (exemple Irak et Afghanistan).
Ces deux approches génèrent les premiers articles, les pistes à approfondir et une première visions de ceux que va apporter la basse de donnée. L'inconvénient majeur c'est que ces deux approches limite les pistes de réflexions focalise la recherche et la limite. Les publications, qu'apportent ces deux approches, qui arrive très rapidement après la mise à disposition de la base de donnée font la une mais déçoit les communautés discourt du type « à quoi cela sert de mettre autant de personnes, de moyen sur le dossier ci les résultats sont peut nombreux et pas forcement convaincants. »
En biologie intégrative, une troisième approche suit les deux autres, celle ci consiste à traiter de manière informatique les données :
-confronter la nouvelle base de données à celles qui existe déjà
-regrouper les informations par cluster
-mettre en place des réseaux d'interaction entre les données obtenue
-Fournir un model dynamique à la communauté.
La matière qu'apportent les biologistes intégratifs n'est pas de « l'ordre de la garde robe de Sarkosy » mais nettement supérieur à celle –ci : organisé l'information pour la rendre accessible à tous.
Les données de Wikileaks doivent être maintenant traitées selon les mêmes approches méthodiques que celle proposé par les bio-informaticiens :
Confrontation semaine par semaine des données des média avec ceux de Wikileaks des dernières années.
Regroupement des télégraphes par groupe de mots répétés.
Cela pourrait peut être donné des informations intéressantes non issues du « pif » des journalistes ! De trier l'information et de la rendre plus accessible.

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

STEPHANE
le 6 décembre 2010 - 7:30 • SIGNALER UN ABUS - PERMALINK

Un article intéressant.
Il est évident qu'une nouvelle race de biologistes doit apparaitre, les biologistes théoriques. Des gens capables de trouver du sens à la masse des données actuellement générées par la biologie et ce n'est que le début. Il faut redonner de la place à la créativité et à l'imagination. Les données en masse ont tendance à anéantir toutes réflexions noyant l'esprit critique et l'analyse. La génération des données est capitale mais doit être suivi d'une réflexion pour lui donner du sens et une intelligibilité.
Sidney Brenner est dur, la biologie de données a fourni des données importantes, je ne citerai que deux exemples la visualisation en temps réel des 24 premières heures de la vie du zebrafish en spim (http://www.digital-embryo.org/) qui nous a permis de voir enfin les mouvements des différents feuillets et de comprendre les relations entre division cellulaire et déplacement.
Un autre exemple est fourni par l'effort de mitocheck (http://www.mitocheck.org/) pour identifier et organiser les gènes importants dans la mitose.
Ceux sont je croit des bons exemples de la bonne utilisation et de la pertinence du high through put en biologie. Le point important, il faut du temps pour la réflexion et la conceptualisation. Il nous faut sortir de l'immédiateté pour faire des projets sur le plus long terme. En biologie c'est encore possible, je ne sais pas pour combien de temps, dans le journalisme, j'ai l'impression surtout en france qu'il y a une vraie prime au vite fait, mal fait. Il faudra peut être passer par le journalisme version propublica (http://www.propublica.org/) subventionné pour donner à des journalistes le temps de l'analyse des données.
Nous verrons.

VOUS AIMEZ 0 VOUS N'AIMEZ PAS 0 LUI RÉPONDRE

1 ping

Tweets that mention Wikileaks et biologie, utilisation similaire des données? » Article » OwniSciences, Société, découvertes et culture scientifique -- Topsy.com
le 5 décembre 2010 - 15:15

[...] This post was mentioned on Twitter by jean marc, Martin Clavey, Pascal, Yoodoo, OWNIsciences and others. OWNIsciences said: Wikileaks et biologie, utilisation similaire des données? http://goo.gl/fb/OAfym [...]