

# Comment la bio-informatique et les droits de propriété intellectuelle contribuent au partage de l'information biologique

Peter Dawyndt, Tom Dedeurwaerdere et Jean Swings

## Objectifs

À mesure que les scientifiques et les cercles d'utilisateurs sont mieux reliés entre eux, notamment par l'Internet, et que la recherche se concentre sur des questions de portée mondiale comme le changement climatique, la santé humaine et la biodiversité, il devient de plus en plus nécessaire de traiter systématiquement les problèmes d'accès aux données et de partage des données dans un cadre dépassant les compétences nationales en bénéficiant ainsi de la valeur ajoutée créée par la coopération internationale. Le but devrait être de faire en sorte que les chercheurs et le grand public obtiennent un retour optimal sur les investissements publics, et de développer la chaîne de valeur des investissements en travaux et données de recherche (Stiglitz et coll., 2000). Le partage des données microbiologiques est essentiel si l'on veut accélérer la conversion des résultats de la recherche en savoirs, produits et procédures en vue de pouvoir progresser dans des domaines d'intérêt général tels que l'exploration, la conservation et l'exploitation de la biodiversité. De nos jours, le partage

national, international et transdisciplinaire des données de recherche n'est plus seulement une question de technologie, c'est aussi un processus social complexe dans lequel les chercheurs doivent trouver un équilibre entre des pressions et des intérêts divers. À défaut de tenir compte de ces facteurs, des approches purement réglementaires du partage des données ont peu de chance de succès, car la technologie ne réalisera pas d'elle-même les promesses de la cyberscience. Les technologies de l'information et de la communication fournissent l'infrastructure matérielle, mais il appartient aux gouvernements nationaux, aux organisations internationales, aux institutions de recherche et aux scientifiques eux-mêmes de s'assurer que les aspects institutionnels, financiers, économiques, juridiques, culturels et comportementaux du partage des données sont pris en considération (Arzberger et coll., 2004). Pour mettre en lumière carences technologiques et conflits d'intérêts – et découvrir les moyens de surmonter les oppositions au partage des données sur la biodiversité –, le moment est venu de rapprocher les points de vue des différentes parties prenantes à l'information.

Peter Dawyndt travaille au laboratoire de microbiologie de l'Université de Gand, avec un groupe de taxonomistes bactériens qui a reçu des récompenses internationales. Il a obtenu en 2004 son doctorat ès sciences. Il est l'auteur de plusieurs articles sur la conception d'un système de raisonnement autodidactique pour la description de la diversité bactérienne qui ont fait l'objet d'analyses critiques de confrères et ont été publiés dans des revues internationales. Email : Peter.Dawyndt@ugent.be

Tom Dedeurwaerdere est directeur de recherche au Centre de philosophie du droit de l'Université catholique de Louvain et professeur à la Faculté de philosophie de la même université. On trouvera des renseignements bibliographiques concernant ses publications sur le site Web : [www.cpd.ucl.ac.be/perso/dedeurwaerdere](http://www.cpd.ucl.ac.be/perso/dedeurwaerdere)

E-mail :

[Dedeurwaerdere@cpdr.ucl.ac.be](mailto:Dedeurwaerdere@cpdr.ucl.ac.be)

Jean Swings a été responsable des recherches en microbiologie chez Plant Genetic Systems, directeur de laboratoire de microbiologie, professeur titulaire et administrateur de la collection de bactéries BCCM<sup>TM</sup>/LMG, en Belgique, de 1992 à 2005. Il a publié dans des revues internationales plus de 280 articles qui ont été évalués par ses pairs. En 2002-2004, il a été élu président de la Fédération mondiale des collections de cultures (wfcc).

E-mail : [Jean.Swings@Ugent.be](mailto:Jean.Swings@Ugent.be)

Les articles rassemblés ici traitent des problèmes posés par le partage des données dans le cas particulier des microbes. Les microbes sont les plus petites formes de vie, mais pris ensemble ils représentent la plus grande masse vivante sur terre. En conséquence, ils sont souvent négligés dans les projets généraux intéressant la biodiversité. Cependant, à l'instar de la matière noire répartie de manière invisible dans tout l'univers, les microbes jouent un rôle crucial dans l'instauration, le maintien et la restauration de l'équilibre dans la quasi-totalité des écosystèmes. Toute vie terrestre est indissociable des micro-organismes, indispensables pour préserver la santé des organismes, qu'ils fournissent en éléments nutritifs et en minéraux et dont ils recyclent l'énergie ; à l'inverse, certains micro-organismes peuvent causer des maladies infectieuses s'ils trouvent un terrain vulnérable. Parmi toutes les espèces vivantes ce sont les microbes qui présentent la plus grande diversité ; ils recourent à des processus biologiques et chimiques qui n'ont pas d'équivalent dans la nature. Par conséquent, nous pouvons considérer le monde bactérien comme un vaste potentiel de ressources biotechnologiques, presque entièrement inexploité ; l'étude des microbes permettra de comprendre la majorité des processus vitaux et d'aller plus loin dans le décryptage des mécanismes fondamentaux de la vie sur terre.

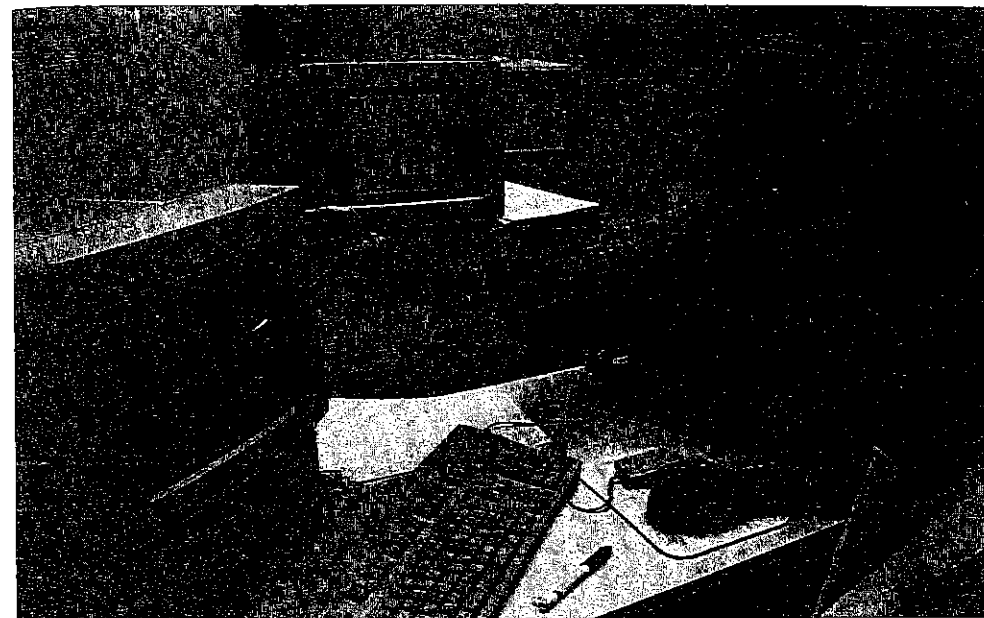
Cette interaction entre les micro-organismes et leur environnement (de la cellule à l'écosystème) demande à être transposée en une plateforme technologique qui intègre sans solution de continuité tout le savoir disponible et permette ainsi la construction de systèmes autodidactiques dynamiques se prêtant à l'automatisation de l'acquisition d'information et de la création de savoir. Les articles regroupés ici ont leur origine dans un projet pilote visant à rassembler des connaissances spécialisées et des idées neuves pour la création d'une plate-forme européenne de ressources biologiques. Pour plusieurs raisons, il semble que le monde bactérien soit le prototype idéal à avoir devant les yeux quand on évoque l'idée d'une plate-forme d'échange de données sur la biodiversité. Nonobstant leur grande diversité métabolique, il n'y a actuellement pas plus de 6 000 noms d'espèces valablement décrites, ce qui représente un nombre de références assez limité pour un système d'information. De plus, l'approche en plusieurs phases retenue pour la plupart des études de criblage microbien a permis de constituer de vastes ensembles de données

portant sur des observations normalisées qui reflètent la diversité phénotypique et génotypique rencontrée chez les bactéries. En outre, la taille limitée du génome constituant le canevas de la vie bactérienne a grandement facilité le séquençage du génome complet d'environ 200 organismes bactériens, et au moins 650 autres projets de séquençage génomique atteindront prochainement leur stade final ([www.genomesonline.org](http://www.genomesonline.org)).

L'accès intégré et combiné à ce champ informationnel aux multiples facettes ouvre des perspectives pour de nouvelles applications. Par ailleurs, cette nouvelle panoplie d'outils pour l'étude des composantes de la vie élémentaires et des cheminements biologiques permettra de jeter les bases de projets futurs encore plus complexes. Parmi ceux-ci pourraient figurer la cartographie complète des réseaux protéiques et métaboliques et la création de modèles biologiques susceptibles de préparer le terrain à des modèles théoriques sur la spéciation bactérienne et sa dynamique écologique complexe (Gevers et coll.). La mise au point d'outils d'identification automatisée d'espèces implique incontestablement l'accès à une batterie de compétences que l'on ne rencontre normalement pas chez les systématiciens ou dans les services et institutions où la plus grande partie des identifications taxonomiques officielles sont réalisées. La mise au point d'approches rigoureuses passe par des collaborations originales entre microbiologistes, ingénieurs, mathématiciens, informaticiens et personnels ayant de bonnes connaissances à la fois en biologie et en informatique, pour que ne soient pas perdus de vue les aspects juridiques du partage de ressources biologiques et d'outils logiciels se trouvant dans le domaine public. En faisant appel à des disciplines apparemment sans rapport, les lacunes traditionnelles de la terminologie, des approches et des méthodologies pourraient être graduellement comblées. Quand les obstacles aux possibilités de collaboration auront été levés, un véritable échange d'idées pourra s'instaurer : un échange qui élargira le champ des recherches sur les problèmes de biodiversité, ouvrira des perspectives nouvelles et peut-être inattendues, et donnera même peut-être naissance à de nouvelles disciplines hybrides capables d'analyses plus subtiles.

### État des lieux

La réalisation de ce programme de recherche interdisciplinaire n'a de sens que si on commence



Analyse du polymorphisme de l'ADN par la technique RFLP (restriction fragments length polymorphism) au Laboratoire LRGAFT à Montpellier, France. IRD/Alain Rival.

par établir un diagnostic commun des problèmes à résoudre et formuler une hypothèse fondamentale susceptible de guider la recherche.

En se basant sur de récents rapports de colloques de l'American Society of Microbiology ([www.asm.org](http://www.asm.org)), on peut déjà repérer certains éléments clés qui s'inscrivent dans un tel diagnostic. Le recours accru au ciblage génétique et à la bioinformatique en microbiologie modifie profondément l'organisation de la recherche et du développement dans les biotechnologies. Les ressources biologiques, en particulier, sont de plus en plus souvent explorées par des moyens informatiques et l'information est extraite en combinant une masse de données provenant de différentes sources et échelles d'interaction. Par ailleurs, la recherche contemporaine montre la nécessité de passer à une approche biologique systémique : la manière dont les gènes s'expriment dans un organisme dépend des propriétés écosystémiques de son environnement. Dans l'analyse des propriétés des micro-organismes, l'information génétique doit être combinée avec les données comportementales et environnemen-

tales. De plus, dans des domaines aussi divers que les antibiotiques, la recherche alimentaire et la biosécurité, de nouveaux aperçus d'une grande importance sont à attendre de l'accroissement de nos connaissances des principes présidant à la diversification microbiologique, permis par les nouvelles techniques informatiques et expérimentales.

Conséquence de cette réorganisation de la recherche et du développement dans les biotechnologies, utilisateurs et scientifiques sont désormais plus étroitement liés dans la chaîne de l'innovation. Cette interaction dynamique entre les utilisateurs et les scientifiques a été rendue indispensable par trois facteurs essentiels.

Premièrement, l'adoption d'une approche biologique systémique impose de combiner des informations venant d'un grand nombre d'acteurs. En particulier, pour connaître les propriétés comportementales des organismes dans le monde réel, il faut rassembler des données émanant d'une grande diversité de « parties prenantes à l'information », qui vont des communautés traditionnelles, pour les données concernant les

variables lentes des écosystèmes, aux médecins, pour les données relatives à la résistance aux antibiotiques, en passant par l'industrie, pour les données sur les processus de fermentation. Par exemple, une étude sur la diversité génétique des souches de *Vibrio cholerae* isolées dans diverses régions géographiques du Brésil a montré un lignage évolutif très semblable pour différentes souches responsables du choléra dans des régions géographiques complètement différentes (Thompson et coll., 2003). Cette étude reposait sur une combinaison de données cliniques sur le choléra, de données environnementales sur *Vibrio cholerae* et de données sur les empreintes génomiques des souches collectées. Comme le montre cet exemple, organiser et combiner des informations venant de différentes « parties prenantes » est devenu une question clé.

Deuxièmement, le fait que le processus de collecte des informations fasse appel à de multiples acteurs est la source de nouvelles interrogations intéressant la société. En effet, des inquiétudes se sont manifestées dans l'opinion publique touchant la protection convenable à assurer aux droits des parties prenantes à l'information, par exemple lors de l'utilisation de données cliniques ou pour les activités de bioprospection. Dans ce contexte, opter pour un certain mode d'organisation du processus de collecte des informations n'est pas seulement un choix technique mais aussi celui d'une combinaison particulière de valeurs sociales, comme le consentement préalable éclairé, la protection de la vie privée ou le partage des avantages. Par conséquent, la voie de l'innovation technique en biotechnologie est devenue plus « réflexive » (Beck, 1997, p. 11-19), à l'instar de ce qui s'est produit dans d'autres domaines comme l'Internet (Dedeurwaerdere, 2002). De plus, la responsabilité de la protection des droits des parties prenantes, dans ce processus à plusieurs acteurs, ne peut incomber aux seuls scientifiques : elle dépend aussi de la manière dont se répartit l'ensemble des droits accordés aux intermédiaires du processus de partage des données et aux utilisateurs finaux de celles-ci.

Enfin, le recours à la bio-informatique et à la gestion de bases de données fait surgir un nouveau type d'« acteur » technique dans le processus : les technologies de l'information et de la communication ou TIC. De fait, le rôle des TIC va très au-delà de leur utilisation en tant qu'outil passif de collecte et d'échange de données. Elles

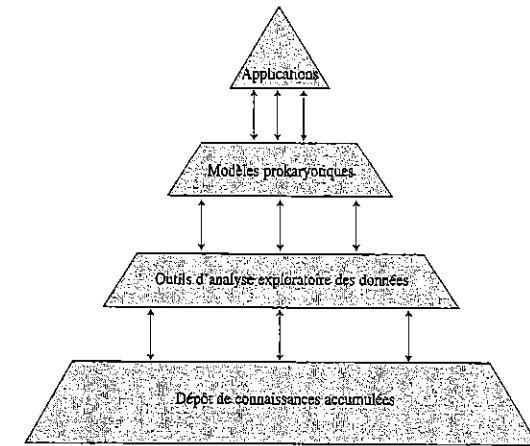
apportent, en réalité, une contribution active à la production même des connaissances. Dans le cas, par exemple, des bases de données intégrées de souches microbiennes, les systèmes autodidactes organisent les données sur différentes échelles et font apparaître de nouveaux types de liaison inattendus (Dawyndt et coll.). De même, les simulations sur ordinateur de réseaux auto-organiseurs produisent de nouvelles formes de biodiversité à partir de séries de données existantes, permettant ainsi d'étendre notre connaissance au-delà des micro-organismes cultivables connus (Kohonen, 1990 ; Abe et coll., 2003). Le caractère auto-organisateur de ces processus informatisés implique lui aussi une interaction plus étroite entre le fournisseur et l'utilisateur de l'information. En effet, ces processus génèrent une pluralité de voies de développement possibles et l'utilisateur joue un rôle clé en fournissant en retour les informations nécessaires au choix de la voie appropriée.

L'hypothèse de recherche fondamentale découlant de ces considérations est que cette réorganisation de la chaîne de l'innovation suppose une *interaction réversible* entre scientifiques et innovateurs, d'une part, et utilisateurs finaux des nouveaux produits (produits alimentaires, médicaments, technologies environnementales, etc.), d'autre part :

- d'un côté, la combinaison des informations venant du criblage génétique, de la bio-informatique, des savoir-faire et des savoirs traditionnels, etc. génère, souvent de façon inattendue, des connaissances nouvelles et différentes voies possibles d'innovation ;
- d'un autre côté, les utilisateurs finaux devraient : fournir les critères du *choix* entre les différentes voies de développement ; fournir des contributions au processus d'innovation en apportant des informations sur l'environnement comportemental des produits ; fournir des garanties adéquates pour la protection des droits des parties prenantes à l'information.

Certains projets ont déjà été mis sur pied en tirant parti de cette interaction réversible entre utilisateurs et scientifiques, comme l'Iceland Health Sector Database (IHD), qui combine des bases de données du secteur de la santé avec des données de génomique humaine et des données généalogiques pour générer de manière novatrice des connaissances sur les interactions entre gènes, environnement, maladie, traitements et effets<sup>1</sup>. Aucune étude plus systématique n'a été

Figure 1. La raison pour laquelle les applications novatrices des biotechnologies ont une si grande portée est qu'elles « reposent sur des épaules de géants », c'est-à-dire qu'elles bénéficient des travaux antérieurs de nombreux chercheurs.



entreprise sur le concept générique d'un système global d'information microbienne expressément destiné à la production de connaissances. Aussi étudierons-nous uniquement ici un cas particulier dans lequel existe un ensemble de données suffisamment complet, ce qui permet d'aborder ces questions dans un esprit plus systématique.

L'un des avantages qu'il y a à se concentrer sur le rôle de la bio-informatique et des droits de propriété intellectuelle s'agissant de production de connaissances, d'accès aux données et de partage des données est la disponibilité de grandes quantités de données sur des phénomènes d'observation normalisés et reproductibles, de nature à la fois génétique et phénotypique. En outre, du point de vue des droits de propriété intellectuelle, c'est aussi l'un des domaines clés où ont eu lieu les expériences les plus hardies d'institutions d'échange et de partage de données et de matériel biologiques (par exemple, les bases de données publiques sur le séquençage<sup>2</sup>, le projet Mosaics<sup>3</sup>, etc.). Ces nouvelles institutions représentent des efforts de collaboration pour mettre en place un mode approprié de partage des données permettant l'exploration de l'ensemble commun de ressources microbiologiques.

## La bio-informatique comme outil de production de connaissances

L'utilisation de la « bio-informatique » pour la construction de bases globales de données microbiologiques vise à repérer les technologies clés et les composantes élémentaires nécessaires pour permettre de : construire un fonds cumulatif de connaissances où seraient déposées les masses de données et de métadonnées expérimentales concernant les micro-organismes, développer des outils généraux d'exploration de données aptes à découvrir des connaissances nouvelles au milieu d'un environnement excessivement riche en données, afin de créer des portails flexibles et constamment mis à jour consacrés à la diversité bactérienne observée et aux innovations biotechnologiques concomitantes, dans le but ultime de valoriser les découvertes récentes en tant qu'applications nouvelles ou produits finaux. Ce cheminement est représenté schématiquement dans la figure 1. Dans la réalité, tous ceux qui sont impliqués dans les étapes initiales de la conception de modèles automatiques et dynamiques exploitant la matière brute qui est au cœur de l'investigation biologique sont dans une période d'expérimentation intense dont le résultat est dif-

ficile à prédire. Ils ont cependant la ferme conviction que – même si certaines conceptions peuvent changer dans le détail – le prototypage et l'absence de dogmatisme sont incontestablement la voie de l'avenir.

Avec le développement rapide de formats de données et d'applications bio-informatiques qui font travailler toute une industrie artisanale des bases de données et des services Internet, la conception et la mise en œuvre de formats et de langages d'interrogation communément acceptés devient primordiale dans un scénario holistique. La question de l'interrogation des bases de données dans des environnements où les sources de données réparties suivent différents schémas a été largement traitée dans les publications spécialisées et est connue sous le nom de problème de l'intégration des schémas. De multiples initiatives pour la conception de schémas communs permettant la normalisation des échanges de données entre fournisseurs de données microbiennes réparties ont vu le jour au cours des deux dernières décennies : le réseau MINE (Microbial Information Network Europe) et CABRI (Common Access to Biotechnological Resources and Information) sont des configurations standards spécialement conçues pour diffuser des informations sur les micro-organismes, tandis que le GBIF (système mondial d'information sur la biodiversité) supporte aussi bien ABCD (Access to Biological Collection Data) que Darwin Core en tant que formats standards d'échange et de communication de données couvrant la totalité de l'information sur la biodiversité dans le monde. Les normes de gestion des contenus concernant la biodiversité ont rarement été un sujet de fascination pour les chercheurs. Mais elles sont la clé d'une foule de problèmes confrontant les scientifiques et les groupes d'utilisateurs, comme la recherche, l'exploration de données, la fonctionnalité et la création d'archives stables et pérennes à partir des résultats de recherche.

Pour réussir l'intégration de bases de données, il ne faut pas seulement développer des schémas communs permettant de faire des recherches dans les différentes sources d'informations à partir d'un unique point d'accès logique ; il est aussi impératif de normaliser et de rectifier l'information collectée chaque fois que nécessaire. Les annotations de bases de données n'ont pas le prestige d'articles publiés, leur valeur est largement ignorée des indicateurs de citation, et leur mise à jour est souvent perçue comme une

tâche ingrate. L'administration des bases de données a, par conséquent, manqué du contrôle de qualité typiquement assuré dans les bonnes revues. L'intégration des données est la contrepartie et le complément de l'intégration des schémas, mais cette question ne paraît pas avoir été complètement traitée dans la problématique de la microbiologie ou des sciences de la vie en général. Au lieu de s'évertuer à mettre en place une base matérielle unique de connaissances contenant une grande quantité des informations accumulées sur la biodiversité bactérienne, il faudrait s'attendre à ce que le paysage futur de l'information microbienne se compose d'un grand nombre de fournisseurs d'informations à haute valeur ajoutée se superposant à des archives et à des bases de données largement automatisées. Cette observation incite à adopter, pour la gestion des fournisseurs d'information microbienne répartie, une stratégie résolue de « diviser pour régner ». Une telle stratégie holistique d'intégration des données est raisonnable car elle reconnaît le fait que l'information scientifique est hétérogène par sa valeur et sa nature.

La valeur ajoutée par les utilisateurs la plus notable à attendre du processus d'intégration du « bien commun » microbiologique est la création de passerelles d'information permettant de recoller parfaitement entre eux les morceaux apparentés du puzzle du savoir commun. Par elles-mêmes, ces passerelles peuvent améliorer : la navigation manuelle entre sources d'informations microbiennes réparties et hétérogènes, le recouplement et la fusion d'informations diffusées par divers fournisseurs de données, l'exécution automatique d'interrogations réparties dynamiques, et la réalisation d'activités d'exploration de données à grande échelle pour la découverte de nouveaux modèles et principes régissant les processus de diversification bactérienne. Cette quête nécessite la conception de stratégies exploratoires objectives d'analyse des données dont les applications en matière d'innovation biotechnologique sont évidentes. Les mathématiques et l'informatique pourraient en tant que disciplines tirer un parti croissant de leur association à la biologie, de même qu'elles ont déjà profité – et continueront à le faire – de leur engagement historique dans la solution des problèmes de physique (Cohen, 2004). En brisant les barrières terminologiques entre disciplines, cela devrait permettre aussi d'améliorer la compréhension interdisciplinaire et l'exploitation fortuite d'opportunités.

Bien qu'une foule de questions restent irrésolues, nous plaidons ici pour une action mondiale en faveur de l'intégration des sources de données microbiologiques, en espérant que cela ne sera pas un vœu pieu. Pour y parvenir, il faudra des formes de collaboration originales entre microbiologistes, mathématiciens, informaticiens et autres parties prenantes. Après tout, il ne serait pas prudent de mettre tous ses œufs dans le panier d'une « solution » unique, quelle qu'elle soit. La diversité est le meilleur pari que l'on puisse faire.

### Droits de propriété intellectuelle pour l'accès aux données et le partage des données

Notre hypothèse d'une interaction réversible entre groupes d'utilisateurs et scientifiques dans l'exploration et l'exploitation du bien commun microbiologique implique des réponses innovantes en matière de droits de propriété intellectuelle appliqués à l'accès aux données et à leur partage, ainsi que d'institutions dans ce domaine.

Dans la nature, la biodiversité microbienne a quelques caractéristiques en commun avec les biens privés – c'est un bien non renouvelable – et avec les biens publics : sa consommation est *de facto* publique et elle est fréquemment en accès public afin d'assurer son utilisation durable. En tant que telle, elle peut être décrite avec justesse comme un « pool de ressources mises en commun » (Polski, 2005). Cependant, l'importance croissante de l'infrastructure numérique pour l'exploration et l'exploitation du « commons » microbiologique et la possibilité concomitante de rendre plus exclusif l'accès aux données, appellent la création d'un second type de « bien commun » englobant le patrimoine commun de l'information microbiologique.

Il existe en matière de microbiologie des initiatives qui visent à partager des connaissances à travers des bases de données en rassemblant des savoirs provenant de différents domaines, telles que le réseau CABRI ou le projet GBIF en cours<sup>4</sup>. D'un point de vue de gouvernance, ces réseaux sont confrontés à une pression croissante liée à la mondialisation des marchés. En particulier, le développement des droits mondiaux de propriété intellectuelle a engendré une concurrence pour la propriété de ressources jusque-là partagées. Dans le même temps, le rôle de l'État dans la fourniture

de services d'intérêt général, tels que les collections et les bases de données publiques, consiste de moins en moins à intervenir directement et de plus en plus à réglementer les marchés ou les quasi-marchés. Dans le cadre de ce nouveau rôle de l'État, un accès peu coûteux aux bases de données peut être garanti, par exemple, en instaurant une exemption générale pour la recherche à but non commercial. De façon similaire, l'échange de matériel biologique peut être réglementé par l'insertion de clauses obligatoires dans les arrangements contractuels relatifs à l'échange de matériels biologiques, comme la mention de l'origine de la ressource et/ou de l'obtention préalable du consentement éclairé.

Les articles rassemblés ici analysent les conditions institutionnelles du développement du partage des bases de données dans le contexte de la mondialisation de la propriété intellectuelle. Nous nous appuyons, en particulier, sur les enseignements tirés depuis peu des théories de la gouvernance, qui montrent la nécessité de mettre au point de nouvelles formes d'action collective pour faire face à la fois aux insuffisances des solutions fondées sur le marché et aux limites des nouvelles formes de réglementation, dans la perspective de la constitution d'un ensemble commun de données scientifiques aux fins de la recherche (Reichman, 2003 ; Hess et Ostrom, 2003). Par exemple, dans le domaine de la communication numérique, le développement de dépôts de textes électroniques comme arXiv.org et BioMedCentral et de dépôts numériques fiables de connaissances d'intérêt général reposent sur une coordination entre des groupes d'universitaires et des spécialistes de l'information pour mettre en place un pool de savoirs. La nouveauté de ces initiatives tient au fait que les auteurs font partie d'une communauté épistémique internationale résolue à créer une bibliothèque spécialisée mondiale interopérable, le but étant de permettre à tous d'en tirer un plus grand profit et de réduire les inconvénients dus à l'isolement dans lequel chacun travaille. En ce qui concerne la fusion des bases de données dans le domaine des ressources microbiologiques, il semble également nécessaire de recourir à des formules de collaboration de ce genre pour faire face aux problèmes liés à l'incertitude et à la complexité du processus d'innovation. En particulier, des arrangements collectifs au sein des réseaux de connaissance semblent nécessaires pour surmonter les insuffisances du marché dues au caractère imprévisible du pro-

cessus automatisé de génération des connaissances, et établir de nouveaux partenariats entre les divers acteurs publics et privés au long de la chaîne de l'innovation.

Ces aperçus sur les théories actuelles de la gouvernance nous permettent de concevoir les droits de propriété intellectuelle dans une perspective entièrement différente. En effet, si nous considérons le processus d'innovation représenté dans la pyramide de la figure 1, nous constatons que la valeur d'une ressource biologique se crée graduellement au fil des différentes étapes du processus de création de valeur – depuis l'extraction et l'accumulation d'informations concernant les ressources biologiques jusqu'au développement de produits et aux nouvelles applications, en passant par la détection en laboratoire et le processus de modélisation. Pourtant, le régime actuel de la propriété intellectuelle crée une incitation au sommet de la pyramide – les applications – mais ne s'attaque pas à la totalité de la chaîne de l'innovation. Dans ces conditions, il semble plus pertinent d'adopter une approche dynamique de la valorisation économique (Driesden, 2003), qui intègre des conditions de rationalité limitée et tient compte de la dynamique du changement économique en abandonnant toute référence à une situation d'équilibre statique. Dans ce cadre, la préoccupation majeure n'est plus tant l'allocation optimale des ressources existantes que les problèmes d'efficacité adaptative, comme l'acquisition de connaissances tout au long du processus de création de valeur et l'incitation à préserver une valeur d'option future dans des conditions d'incertitude (Dedeurwaerdere, 2004).

La nécessité de tenir compte d'une conception dynamique de l'efficacité économique dans la définition des droits de propriété intellectuelle coïncide avec les analyses d'auteurs tels que Jerome Reichman ou Timothy Swanson, pour qui de nouveaux outils de réglementation sont requis non seulement pour adapter le régime actuel des droits de la propriété intellectuelle à une situation nouvelle, mais aussi pour refléter une modification des opinions sous-tendant le paradigme classique de ces droits (Reichman, 1994 ; Swanson, 1997). Ces auteurs prennent leurs distances par rapport à une conception des difficultés posées par les droits de propriété intellectuelle sur les ressources génétiques comme étant un problème juridique purement technique. Pour saisir l'originalité des nouveaux outils juridiques à mettre en place, une autre lecture des changements en cours

est indispensable – une lecture qui ne les réduisent pas à un simple ajustement technique dans chaque secteur d'activité. Il nous faut réfléchir sérieusement – puis procéder à leur élaboration détaillée – à de nouveaux instruments juridiques complétant le régime applicable à la biodiversité et permettant de répondre à l'impératif d'une approche plus dynamique de l'efficacité. Reichman, par exemple, propose de passer d'un paradigme fonctionnant par hybridation d'outils existants et reposant essentiellement sur le brevet et le droit d'auteur, à un paradigme fondé sur un régime de responsabilité et permettant la rémunération *ex post* du maillon précédent de la chaîne de l'innovation (Reichman, 2000, p. 1776-1796). D'autres ont proposé de créer des sociétés de collecte de connaissances et/ou de savoir-faire traditionnels, dans la double perspective d'une large diffusion de ces connaissances et de leur protection appropriée (Drahoš, 2000). Ces propositions alternatives ont encore bien du chemin à faire avant d'être pleinement opérationnelles pour le partage des données dans le domaine commun microbiologique, mais elles sont sans aucun doute la voie d'avenir pour la création d'incitations à l'innovation au long du processus de création de valeur.

Les articles publiés ici sont les communications originales discutées au cours du premier atelier qui a été organisé sur le projet pilote de bien commun microbiologique. Depuis lors, ce nouveau champ de recherche a été approfondi à l'occasion de réunions de l'International Association of Common Property (IASCP) et au sein de réseaux de recherche interuniversitaires européens et belges. Dans une série de sessions parallèles, le premier atelier a cherché à rassembler les compétences spécialisées nécessaires à la mise au point d'un prototype de fusion de l'information (sessions consacrées à la bio-informatique) et à concevoir des droits de propriété intellectuelle et des institutions de partage de bases de données (sessions consacrées à la gouvernance).

Élaborer des concepts algorithmiques complexes est une chose, mais les mettre en pratique en est une autre. Les sessions techniques de l'atelier ont donc discuté de prototypes pour la cartographie du monde microbien, c'est-à-dire de la mise au point de systèmes d'information automatisés, dynamiques et interactifs destinés à l'accumulation, à l'exploration et à l'exploitation des connaissances. Bien des questions pratiques restent ouvertes et ont été au cœur des débats de la

première série de sessions de l'atelier (les sessions de « bio-informatique ») :

- Quelles sont, parmi les TIC, les principales technologies servant à la construction de réseaux d'information répartie ?
- Quels services sont nécessaires à la mise en œuvre d'un cadre intégré d'information biologique conçu comme une entreprise engageant toute la communauté ?
- Comment des méthodes d'exploration de données dernier cri peuvent-elles permettre la découverte de connaissances dans les bases de données et quels sont les indices précurseurs pour leur application dans la chaîne de l'innovation biotechnologique ?

L'analyse du rôle des droits de propriété intellectuelle et des réseaux de connaissances collaboratifs dans l'organisation d'un accès adéquat aux données microbiologiques et d'un partage approprié de celles-ci a été au centre de la seconde série de sessions de l'atelier (les sessions consacrées à la gouvernance). Trois sous-sessions ont traité respectivement : d'études de cas portant sur les institutions existantes de ges-

tion de base de données en collaboration (bases de données publiques sur le séquençage, GBIF et CABRI) ; de nouvelles approches pour la création d'ensembles de droits concernant la fusion de bases de données et le partage d'informations (accords de licence collectifs, protection *sui generis* de base de données, etc.) ; et de la conception d'un cadre institutionnel pour la mise en commun des informations microbiologiques, en puisant dans une liste de bases de données indispensables qui devraient être combinées pour la réalisation du projet pilote (données taxonomiques, données sur les ressources biologiques, publications scientifiques, données d'observation – ma 16S, FAME, MLSA, stocks d'ADN, etc.). Le présent dossier de la Revue internationale des sciences sociales est basé sur les communications originales, substantiellement remaniées, présentées à la seconde session de l'atelier, consacrée aux sciences sociales et aux problèmes institutionnels posés par la création de bien communs microbiologiques.

Traduit de l'anglais

## Notes

1. La base de données IHD rassemblera les informations tirées de dossiers de patients rendus non identifiables par codage, qui lui seront fournies par le Service national de santé islandais, et stockera les données dans un système informatique aux fins d'analyse clinique et statistique, une protection légale étant assurée contre les violations ou les utilisations abusives. L'IHD peut

être reliée à une base de données généalogiques existante. Cette initiative permet également le croisement de données de l'IHD avec des données génomiques obtenues et analysées avec le consentement éclairé de donateurs islandais (OCDE, 2000, p. 37).

2. International Nucleotide Sequence Database (base de données internationale sur les séquences de nucléotides) ; accès

public via les portails de la DNA Data Bank du Japon – DDBJ ([www.ddbj.nig.ac.jp/Welcome.html](http://www.ddbj.nig.ac.jp/Welcome.html)), du Laboratoire européen de biologie moléculaire – EMBL ([www.ebi.ac.uk/embl/index.html](http://www.ebi.ac.uk/embl/index.html)) et de la GenBank des États-Unis ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

3. Cf. <http://www.belspo.be/bccm/>.  
4. Cf. [www.cabri.org](http://www.cabri.org) et [www.gbif.org](http://www.gbif.org).

## Références

ABE, T. ; KANAYA, S. ; KINOCHI, M. ; ICHIBA, Y. ; KOZUKI, T. ; IKEMURA, T. 2003. « Genomics for unveiling hidden genome signatures », *Genomics Research*, 13, p. 693-702.

ARZBERGER, P. ; SCHROEDER, P. ; BEAULIEU, A. ; BOWKER, G. ; CASEY, K. ; LAAKSONEN, L. ; MOORMAN, D. ; UHLIR, P. ; WOUTERS, P. 2004. « Promoting access to public research data for

scientific, economic, and social development », *Data Science Journal*, 3, p. 135-152.  
BECK, U. 1997. *The Reinvention of Politics. Rethinking Modernity in*

- the *Global Social Order*, Polity Press, Cambridge.
- COHEN, J.E. 2004. « Mathematics is biology's next microscope, only better; Biology is mathematics' next physics, only better », *PLOS Biology*, 2(12), p. 2017-2023.
- DAWYNDDT, P. 2004. *Knowledge Accumulation of Microbial Data aiming at a Dynamic Taxonomic Framework*, Thèse de doctorat, Université de Gand, Belgique.
- DAWYNDDT, P. ; VANCANNEYT, M. ; DE MEYER, H. ; SWINGS, J. 2008. « Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources », *IEEE Transactions on Knowledge and Data Engineering*.
- DEDEURWAERDERE, T. 2002. « Ethics and learning. From state regulation towards reflexive self-regulation of the information society », dans K. Brunstein, J. Berleur (sous la direction de), *Human Choice and Computers*, Kluwer Academic Publishers, p. 121-130.
- DEDEURWAERDERE, T. 2004. « Bioprospecting, intellectual property law and evolutionary economics: the stake of a theory of reflexive governance », dans M.M. Watanabe et coll. (sous la direction de), *Innovative Roles of Biological Resources Centres*, World Federation of Culture Collections, Tsukuba, p. 389-395.
- DEDEURWAERDERE, T. 2005. « From bioprospecting to reflexive governance », *Ecological Economics*, avril 2005.
- DRAHOS, P. 2000. « Indigenous knowledge, intellectual property and biopiracy : is a global biocollecting society the answer? », *European Intellectual Property Right Review*, 6, p. 245-250.
- DRIESDEN, D.M. 2003. *The Economic Dynamics of Environmental Law*, Cambridge, Massachusetts: MIT Press.
- GEVERS, D. ; COHAN, F.M. ; LAWRENCE, J.G. ; SPRATT, B.G. ; COENYE, T. ; FELL, E.J. ; STACKEBRANDT, E. ; VAN DE PEER, Y. ; VANDAMME, P. ; THOMPSON, F.L. ; SWINGS, J. « Reevaluating prokaryotic species », *Nature Reviews Microbiology* (communication présentée).
- HESS, C. ; OSTROM, E. 2003. « Ideas, artifacts and facilities. Information as a common-pool resource », *Law and Contemporary Problems*, p. 66.
- KOHONEN, T. 1990. « The self-organizing map », *Proceedings of the IEEE*, 78(9), p. 1464-1480.
- OCDE, 2001. *Les centres de ressources biologiques : fondement des sciences de la vie et des biotechnologies*, Paris, OCDE.
- POLSKI, M. 2005. « The institutional economics of biodiversity, biological materials, and bioprospecting », *Ecological Economics*, avril 2005.
- REICHMAN, J.H. 1994. « Legal hybrids between the patent and the copyright paradigms », *Columbia Law Review*, 94, p. 2432-2558.
- REICHMAN, J.H. 2000. « Of green tulips and legal kudzu: repackaging rights in subpatentable innovation », *Vanderbilt Law Review*, 53, p. 1743.
- REICHMAN, J.H. 2003. « A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment », *Law & Contemporary Problems*, 66, p. 315-440.
- STIGLITZ, J. ; ORSZAG, P.R. ; ORSZAG, J.M. 2000. *The Role of Government in a Digital Age.*, Report for the Computer and Communications Industry Association, États-Unis.
- SWANSON, T. 1997. *Global Action for Biodiversity*, Earthscan.
- THOMPSON, F.L. ; THOMPSON, C.C. ; VICENTE, A.C.P. ; THEOPHILO, G.N.D. ; HOFER, E. ; SWINGS, J. 2003. « Genomic diversity of clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 2001 as revealed by fluorescent amplified length polymorphism analysis », *Journal of Clinical Microbiology*, 41, p. 1946-1950.

# 生物信息共享中 生物信息学和知识产权的贡献

彼得·多因特 汤姆·德得威尔德 让·斯温斯

## 目 标

现在,科学家和使用者群体彼此之间的联系通过互联网而变得更加紧密。科研重点也集中在诸如气候变化、人类健康和生物多样性等全球性问题上,因此人们越来越需要更加系统地解决超越国界的信息获取和共享问题,从而更多地得益于国际合作。我们的目标应保证研究者和更多的公众从公共投资中得到最优的回报,并打造研究和研究数据的价值链(Stiglitz *et al.*, 2000)。为了迅速地将研究成果转化为知识、产品和流程以提高公众对生物多样性的探索、保护和利用的兴趣,共享微生物信息至关重要。当前,广泛共享国家、国际和跨学科研究数据的问题已不仅仅是一个技术问题,还是一个复杂的社会化过程,其中研究

者必须去平衡不同的压力和利益。不考虑这些因素,单纯的数据共享管理很难成功,因为技术本身不会实现科研信息化的目标。信息通讯技术(ICT)只能提供物质基础设施。只有各国政府、国际机构、科研机构 and 科学家们才能保证数据共享所需的体制、资金、经济、法律、文化和行为的各个方面(Arzberger *et al.*, 2004)。

为了描述不同技术的弱点和利益的冲突,找出妨碍生物多样性信息共享的阻力并加以克服,我们需要及时将信息拥有者联合在一起。工作坊(后文将解释——编者注)探讨了微生物这一特定案例中信息共享所遇到的问题本质。微生物是最小的生命形式,但它们共同代表了地球上最大的单一生命群体。正因如此,它们常常受到忽视,生物多样性计划中也很少顾及它们。但就像宇宙中隐藏的暗物质所起的作用一样,微生物

彼得·多因特(Peter Dawyndt)是根特大学的研究教授,在应用数学和计算机科学系任教。他的主要研究方向是为地表细菌多样性设计一个自学习分析系统,即生物分类模型。他被认为是 StrainInfo.net 这一门户网站的创始人。Email: Peter.Dawyndt@ugent.be

汤姆·德得威尔德(Tom Dedeurwaerdere)是鲁汶天主教大学法律哲学中心主任和哲学系教授。他发表的著作可见 [www.cpdr.ucl.ac.be/perso/dedeurwaerdere](http://www.cpdr.ucl.ac.be/perso/dedeurwaerdere)。Email: Dedeurwaerdere@cpdr.ucl.ac.be

让·斯温斯(Jean Swings)于1992—2005年担任“植物基因系统”微生物学研究项目经理,为生物实验室主任和全职教授及比利时 BCCM/LMG 细菌采集主任。他在国际同行刊物上发表了280多篇文章。他在2002—2004年间担任世界文化收集联盟主席。Email: Jean.Swings@ugent.be

实际上承担所有生态系统中创造、维持和恢复平衡的作用,所以是不能被忽视的。地球上的所有生命都与微生物密不可分,而后者在维持有机体健康上起着至关重要的作用,为其提供营养、矿物质和能量循环,同时,它们也会造成易感宿主的疾病感染。在所有生命体中,微生物具有最大的多样性,其使用的生物和化学过程在大自然中也是独一无二的。所以,我们可以把细菌世界看作是一个巨大的、基本上未被开拓的生物技术资源,可以通过研究微生物来了解生命的大部分过程,以便进一步解开地球生命的基本机制之谜。

正是由于微生物与其环境——从单个细胞到整个生态系统——的互动使我们想到这样一个非常吸引人的想法,即将微生物的研究传递到一个能够无缝融合所有知识的技术平台上。这将使我们能够建立一个能自动获取信息和创造知识的动态自主学习系统。为此而开展了一个试点工程,为的是在建立欧洲生物资源平台①的过程中整合各种技术专长和新思路,为探索和利用微生物共用资源而组织的工作坊正是这项工程的一部分。工作坊一个特定目标是组成一个战略研究计划咨询小组,服务于整个平台上兴趣交叉的课题。

细菌世界作为生物多样性数据交换平台的一个理想的原型有几个理由。尽管它们具有广泛的新陈代谢多样性,到现在为止有效划分的细菌种类不超过 6000 种,因此适合作为信息系统的锚点数量相当有限。此外,细菌鉴定研究中主要使用的多相方法能够集成庞大的和标准化的数据群,展现我们能观察到的细菌所具有的表型和属型特征。另外,由于作为细菌生命研究基础的染色体组其数量较为有限,所以现在已经完成了约 200 种细菌有机体的染色体组完整排序,还有至少 650 多种将

在不久之后完成 (Genomes online database, 日期不明)。

将这一多重信息领域整合在一起使我们能够实现新的应用。此外,这一研究生物基本构成和途径的新工具包将为以后更加复杂的研究打下基础。今后的研究可能包括完整描绘一个有机体的蛋白质,以及建立生物学模型,为以后建立细菌的物种形成和其复杂生态活动 (Gevers *et al.*, 2005) 的模型铺平道路。研发自动化的物种鉴定工具无疑需要利用系统专家或是从事正规分类工作的机构或院系通常不具备的一系列技能。这种有效工具的研制需要微生物学家、工程师、数学家、计算机专家和对应用生物学与计算机都具有相当知识的人员共同进行创新性的合作。同时还不要忘记在公众领域中共享生物资源与工具软件所涉及的法律问题。将看起来不相关的学科联合起来可能会逐步消除专业术语和方式方法上一直存在的隔阂。扫除了合作的障碍之后,就可以实现真正的思想交流,开阔生物多样性问题的调查领域,产生新鲜和可能意料之外的见解,甚至可能催生新的混合型和复杂分析型学科。

## 最新科技

想要完成这样一种交叉研究项目就必须对所解决的问题作出共同的诊断并明确指导研究的最基本假设。

浏览美国微生物学会的会议报告就可以看出这样的共同诊断所应包含的一些要点。微生物学研究越来越依赖于遗传学鉴定和生物信息学,这一点正在使生物技术的研发组织工作产生深刻的变革。具体而言,人们正在越来越多地通过计算手段来探究生物资源,并整合多种不同的渠道和互动方式来获取信息。当代研究还表明:

我们应该采取一种系统化的视角:如何表述有机体的基因取决于有机体在所在环境中所具有的生态属性。在分析微生物时,遗传信息需要与行为和环境数据结合在一起。此外,我们可以期待诸如抗生素、食品研究或是生物安全等不同领域出现重要的新发现,使我们有可能通过新的计算和实验技术增强对微生物多样性原则的了解。

重新组织生物科技的研发使使用者和科学家与创新链的联系更加紧密。在使用者—科学家动态互动中有三个重要的因素。

第一,采用系统生物的视角要求将非常广泛的参与者所提供的信息结合起来。具体而言,有关真实世界中有机体行为属性的知识需要从各种各样的“信息股东”那里收集数据,从传统研究机构采集的缓慢生态变量数据到物理学家得到的抗生素抗性数据和制造商发酵过程的数据。例如,研究从巴西不同地理环境中分离出来的霍乱弧菌菌株所具有的遗传多样性,表明来自完全不同地理环境中的引发霍乱的菌株具有很近的进化关系 (Thompson *et al.*, 2003)。这项研究结合了霍乱临床数据、霍乱弧菌的环境数据和采集到的菌株基因组指纹数据。这一例证说明,组织和结合不同信息持有人的信息已成为了一个关键性的问题。

第二,信息采集过程中的多方参与造成了新的社会问题。实际上,恰当保护信息持有者权益的问题已经引起了关注,如使用临床数据或在生物开发方面。在这种情况下,支持某种形式的信息采集方式就不仅只是一个技术抉择问题,还是一个涉及一系列社会价值观的抉择问题——如事先获得允许、保护隐私或分享利益等。所以,与诸如互联网等领域中已经发生的情

况相似 (Dedeurwaerdere, 2002), 生物技术的创新道路已经变得更加相互关联 (Beck, 1997, pp. 11-19)。此外,在这个多方参与的过程中,保护股东权利不只是科学家的责任,还有赖于如何分配在数据共享过程中的中介和终端使用者所拥有的一系列权利。

最后,对生物信息学和数据库管理的依赖造就了这一过程中的一个新的参与者:信息与通讯技术本身。实际上,信息和通讯技术的作用远远超出了其作为一个被动的数据收集和交换工具的作用。它在知识创造过程中作出了主动的贡献。例如,在集成的菌株数据库中,具有自学习能力的系统将不同层面的数据组织在一起,并显示出了未曾预料到的新联系 (Dawyndt *et al.*, 2005)。此外,以自我组织网络为基础的计算机仿真从现有的数据集集中生成了新的生物多样性模式,使我们的知识扩展到了已知微生物之外 (Abe *et al.*, 2003; Kohonen, 1990)。计算过程中的自我组织特性还需要信息提供者 and 使用者之间更加紧密的互动。实际上,这些过程会产生出多种可能的发展路径,而使用者在提供选择这些路径所需的反馈信息上发挥着关键性的作用。

依据这些见解得出的最基本假设是:这一创新链的重组意味着科学家/创新者与新产品(如食品、药品或环境技术)的最终使用者之间的双向互动:

- 一方面,遗传鉴定、生物信息学、技能和传统知识等相互结合能够产生新的知识和新的创新路径,且常常是以一种出乎意料的方式。
- 另一方面,终端使用者应当:(1)提供选择不同发展路径的标准,(2)通过提供产品的行为环境信息来帮助创新过程,



(3) 为保护信息持有人的权利提供恰当的保证。

依赖使用者和科学家双向互动的一些项目已经开发出来, 如冰岛卫生数据库。为了生成有关基因、环境、疾病、医疗和结果之间相互作用的知识, 该数据库以创新的方式将卫生数据库与基因和人类染色体组的数据结合在一起。②从知识生成的角度考察全球微生物信息系统的一般概念, 以往还没有过更系统的研究。这也是我们在这一试点项目中想要讨论这个具体个案的原因。这一个案中已经有了足够全面的数据库, 使我们能够更加系统地讨论这些问题。

利用微生物共用资源的试点项目将注意力集中在这一创新链上的一个重要组成部分。这就是生物信息学和知识产权在知识生成、数据检索和数据共享中所起的作用。如前所述, 关注这一个案的一个好处是因为在基因和系谱方面已经有了记录标准的、可复制和观察到特征的庞大数据库。此外, 从知识产权的角度看, 这也是一个已经研发出了机构之间交换与共享数据和生物资料最先进方法(如公共序列数据库③和比利时微生物协调采集嵌合体工程)的重要领域。这些新机构的出现是为了利用微生物共同资源而分享数据的结果。

### 服务于知识生成的生物信息学

在建设全球微生物信息库的过程中利用生物信息学的目的, 是要找出关键的技术和建设单元以(1)建立一个能够收录有关微生物庞大试验数据和研究数据的积累性知识库,(2)在这一丰富的数据环境中开发能够发现知识的普遍性数据发掘工

具, 从而(3)为所观察到的细菌多样性和相关的生物技术创新建立动态更新门户网站, 以便最终(4)将新发现转化为新的应用或终端产品(Dawyndt *et al.*, 2001)。图1系统地描述了这一宗旨。现实情况是, 作为生物开发研究的核心, 处于初始阶段的原始资料自动动态模型设计正处在深入的研究阶段, 其结果很难预测。不过, 尽管有些模型的细节可能会改变, 但我们坚信创造原型和消除教条主义无疑是前进的方向。生物信息研究的一个基本目标是改进这些开创性的计划, 将其得出的不同发现塑造为更具整合性的方法。

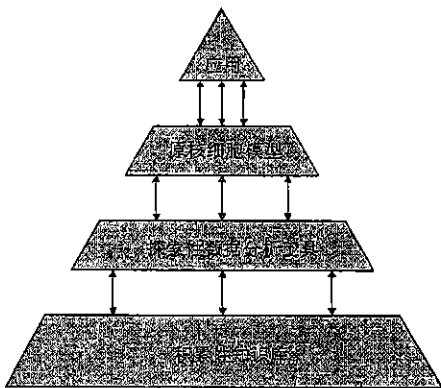


图1. 创新性生物技术的应用能够达到当前的水平是因为它们站在了巨人的肩膀上, 就是说, 站在了很多为其铺路的研究者获得的科学成果之上。

随着数据格式和支持可验证的数据库与网络服务产业生物信息学应用技术的出现, 要想使它们能够通用, 最重要的问题是推行能够得到普遍认可和接受的数据格式和咨询语言。人们已就数据资源体

系不同的情况下咨询数据库的问题, 即体系兼容问题进行了广泛的讨论。过去20年中出现了很多旨在将公开的微生物数据标准化的通用体系设计工程。欧洲微生物信息网(MINE)和生物技术资源与信息通用入口(CABRI)都是专门为传播微生物信息而设计标准体系。全球生物多样性信息服务(GBIF)支持生物采集数据入口(ABCD)和达尔文核心, 将它们作为查询地球生物多样性全部信息的标准体系。管理生物多样性信息的问题对研究者来说并不引人注目, 但对一系列问题都很重要。这些问题会影响到科学家和使用群体, 如检索、数据发掘、功能性和为研究结果建立长期稳定档案等问题。

不过, 整合数据库不光需要开发能够从单一门户起点搜索不同信息资源的共同格式, 还需要将所收集的信息规范化并在必要时进行纠错。数据库的价值由于点击率的计量而在很大程度上被忽视, 而数据库的更新则常常被视为是一件得不到感谢的工作, 所以它们无法与出版物相比。因此, 数据库缺乏高质量刊物那样的质量管理。数据整合问题与体系整合问题相辅相成, 但在微生物学和生命科学领域似乎没有得到足够的重视。我们在未来应该预见的不是一个收集了庞大细菌多样性信息的单一物理知识库, 而是大量增值信息提供者参与的庞大但在相当程度上自动化的知识档案和数据库。这促使我们制定一个能够有效地管理微生物信息提供者的分类和检索战略。这种整体性数据整合战略的合理性在于它承认这一事实, 即科学信息的价值和性质是多元的。

使用者附加值的一个最明显的例证见于整合微生物公共资源过程中所建立的信息门户网站。它们将零散的共用知识顺畅地拼接在一起。这样, 它们就能够增强公

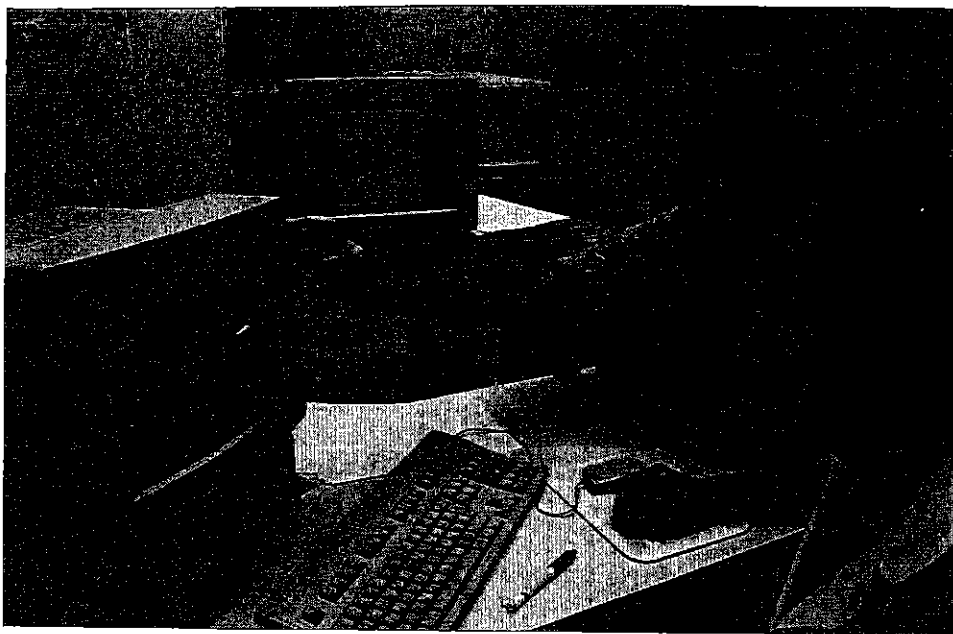
开发的和不同性质的微生物信息资源之间的手动浏览、交叉检索, 提高不同数据提供者所发布信息融汇程度, 改进动态分布咨询的自动执行和推动旨在发现细菌多样化过程中的新模式新原则的大规模数据发掘的利用。这样做要求明确地利用生物技术创新来制定客观的探索性数据分析战略。如此, 就像数学和计算机科学参与解决了并仍将参与解决物理学问题一样, 它们也可能越来越多地有助于生物学(Cohen, 2004)。消除学科术语之间的隔阂也能够增强跨学科的理解和提高发现新宝藏的几率。

尽管仍存在很多未解决的问题, 但可预见的是: 工作坊集合起来的人和思想会进一步推动微生物数据资源的全球整合, 而不是一厢情愿想想而已。做到这一点将需要生物学家、计算机科学家和其他参与者彼此之间进行创新性的合作。毕竟, 将你所有的鸡蛋放到一个篮子中是不明智的。多样化才是最好的选择。

### 使用和共享数据中的知识产权

在探索和利用微生物公共资源的过程中, 我们关于使用者群体和科学家之间双向互动的假设, 要求在知识产权领域和数据获取与共享制度中找到创新性的答案。

作为一种消耗性产品, 微生物多样性在本质上具有私人产品的某些属性, 而在其公共消费和经常被公众使用以保持其可持续性的意义上也具有公共产品的某些属性。所以, 它可以被当作“公共资源群”(Polski, 2005)。不过, 在开发和利用微生物共用资源中数字基础建设的日益重要性和相关的获取数据的可能性, 更加突出了建立第二种“共用资源”, 即微生物信息共用资源的需求。



法国蒙彼利埃的 LRGAPT 实验室中正在通过 RFLP (限制片段长度多态性) 技术分析 DNA 的多态性。IRD/Alain Rival.

在微生物学领域，我们已经有了通过数据库共享知识和从不同领域收集知识的先例，如 CABRI 网络（生物资源信息公共入口）或是正在进行之中的 GBIF 工程（全球生物多样性信息服务）。从管理的角度看，这些网络面临全球市场所带来的越来越多的压力。全球知识产权的发展特别导致了对此前共享资源所有权的竞争。同时，国家在提供公益服务——如公共资料 and 数据库——时正在逐渐从直接干预转为市场或类似市场的监管。国家的这一角色转变可以让非商业研究无偿使用数据库，从而保证了获取信息的成本效益等。同样，可以规定生物资料的交换必须附有强制性的协议条款，注明资源出处和/或所

有者的事先同意。

在这期刊中，我们要分析在全球知识产权的背景下发展数据共享的体制条件。我们尤其倚重现代管理理论中的一些深刻见解。它们强调必须采取新的共同行动以解决市场方案的不足和在建设科学数据的共用研究时新监管形式的局限性问题 (Hess and Ostrom, 2003; Reichman and Ulhir, 2003)。例如，在数字通讯领域，诸如康奈尔大学图书馆和“生物医学中心” (BioMedCentral) 的电子出版物存储库或是公益知识共同存储库的发展依靠的是学者和信息专家的合作。这些创新的不同之处在于作者共同参与了旨在建设一个在全球范围可以操作的学术图书馆的国

际性认知团体，其目的是获得更高的共同效益和减少封闭性操作带来的共同损害。在微生物资源领域，数据库整合看来也需要依靠这样的合作方式。这样才能应对创新过程中的不确定因素和复杂问题。我们尤其需要在知识网络中采取共同措施以克服自动知识生成过程的不可测性所造成的市场不足，并在参与了整个创新链的公共和个人参与者之间建立新的伙伴关系。

现代管理理论中的这些看法还使我们从一个完全不同的视角来看待知识产权问题。确实，如果我们审视图 1 所描绘的金字塔式创新过程，就可以看到生物资源的价值来自价值创造的不同渐进步骤——从资源信息的选取和积累，实验室的甄别和模型过程，到产品的研发和新的应用。不错，当前的知识产权制度只关注金字塔的塔尖部分——产品的应用——而不关注创新链中的全部参与者。在这种情况下，可能应该采取一种动态的经济价值框架 (Driesden, 2003)，这种动态方式能够考虑到有限供给的情况，也能考虑到静止平衡状态之外的经济变化。这样，在这一框架之内，重心将从现有资源的优化分配转移到适应效率问题上，如价值创造整个过程中的知识获取和在不稳定条件下保留未来选择价值的动机问题 (Dedeurwaerdere, 2004)。

在定义知识产权中是否有必要考虑一种对经济效率的动态概念，杰罗姆·赖克曼 (Jerome Reichman) 和提摩泰·史万森 (Timothy Swanson) 等人对此都进行过讨论。他们认为需要用新的监管工具来使现有知识产权体制适应新的形势，同时指出作为知识产权传统范式的基础的信念已发生变化 (Reichman, 1994; Swanson, 1997)。这些作者不同意某些人的看法，他们将基因资源中知识产权导致的困难视

为简单的技术法律问题。为了解所需的新法律工具的创新之处，有必要再次审视当前的变化而不是将其视为依据产业活动类型进行的简单技术调整。为此，在生物多样性领域中，我们需要使用一种更加动态性的效率方式来开发新的法律工具。例如，杰罗姆提出将主要以专利和版权为基础、混合现有工具运作的模式进化为一个义务制为基础的新模式，允许事先连接创新链并事后补偿 (Reichman, 2000, pp. 1776-1796)。也有人提出建立传统知识和/或技术协会来推动知识的广泛传播与恰当的保护 (Drahos, 2000)。要想在微生物共用资源的数据共享方面完全应用这些提议还有很长的路要走，但它们无疑是在价值创造的全过程中推动创新的正确方法。

这期刊收入了在第一次微生物共用资源研讨会上的一些新论文。作为一个新的研究领域，这一课题此后在国际共同财产协会的会议上和比利时校际研究网上得到了进一步的发展。研讨会组织了两个平行系列会议，目的是收集相关的专业技能以进一步发展一个整合知识（微生物信息学会议）和设计恰当知识产权与数据库共享机构（管理会议）的基本模式。

构思复杂的计算机概念是一回事，而将其付诸实践则是另一个问题。所以，研讨会的技术会议讨论了勾勒微生物世界的基本模式，即为知识积累、探索和利用而开发的自动动态互动信息系统。研讨会的信息学会议（第一系列会议）中讨论的很多实际问题仍没有答案。其中一些重要的议题包括：(1) 推动信息发布网络的主要 ICT 技术是什么，(2) 完成生物信息整合框架所需的服务是什么，(3) 最新科技数据发掘方式如何才能实现数据库中的知识发现，这些方式应用于生物科技创新链的

前兆是什么。

第二系列会议的主题。该系列会议分别讨论了三个议题：(1) 现有合作式数据库管理机构（公共排序数据库，GBIF 和 CABRI）个案研究；(2) 建立数据库整合和信息共享（如合作式许可协议和特殊数据库保护）的新方式；(3) 利用一系列在实现试点工程（分类数据，生物资源数据，科学论文和可观察数据，如 16S

## 注 释

1. 这一平台的策划源于一些欧洲研究项目（EBRCN, EUROAGENTEST, MOSAICC, TEDDY）参与者的商榷。他们在 2004 年 12 月讨论了建立未来技术平台的问题后提出了这一计划。2005 年举行了若干次会议，以讨论逐步扩大和构建这一平台的核心社群。
2. 冰岛卫生数据库（IHS）将从冰岛国家卫生部采集匿名病人的病历信息并将数据储存在计算机系统中以进行临床和统计分析。该数

rRNA, FAME, MLSA 和 DNA 群) 的过程中应予整合的必要的数据库来设计微生物信息共用资源的体制。这期刊重新整理了第二系列会议上提交的有关社会科学和微生物共用资源所面临的制度挑战的新论文。

[项 龙译]

据库受到法律保护，以防止侵权或是滥用。它可以被连接到一个已有的系谱数据库。这一工程还允许 HIS 数据与基因组研究数据相互检索。基因组数据的采集和分析来自经过本人同意的捐献者(OECD 2001, p. 37)。

3. 公众可以通过日本的 DNA Databank, EMBL Nucleotide Sequence Database 和 GenBank 等门户网站进入国际核酸序列数据库 (International Nucleotide Sequence database)。

## References (参考文献)

ABE, T., KANAYA, S., KINOCHI, M., ICHIBA, Y., KOZUKI, T., AND IKEMURA, T. 2003. "Genomics for unveiling hidden genome signatures", *Genomics Research*, 13 (4), 693-702.

AMERICAN SOCIETY OF MICROBIOLOGY n.d. *Colloquia Reports*. Available from: <http://www.asm.org> [Accessed 12 December 2006] Washington DC: American Society of Microbiology.

ARZBERGER, P., SCHROEDER, P., BEAULIEU, A., BOWKER, G., CASEY, K., LAAKSONEN, L., MOORMAN, D., UHLIR, P., AND WOUTERS, P. 2004. "Promoting access to public research data for scientific, economic, and social development", *Data Science Journal*, 3, 135-152.

BECK, U. 1997. *The reinvention of politics. Rethinking modernity in the global social order*. Cambridge: Polity Press.

BELGIUM CO-ORDINATED COLLECTIONS OF MICRO-ORGANISMS n.d. Available from: <http://www.belspo.be/bcom/> [Accessed 12 December 2006] Brussels: Belgium Federal Sciences Policy.

BIOMEDCENTRAL n.d. Available from: <http://www.biomedcentral.com/> [Accessed 12 December 2006].

COHEN, J. E. 2004. "Mathematics is biology's next microscope, only better; biology is mathematics' next

physics, only better", *PLOS Biology*, 2 (12), 2017-2023.

COMMON ACCESS TO BIOLOGICAL RESOURCES AND INFORMATION n.d. Available from: <http://www.cabri.org> [Accessed 12 December 2006].

CORNELL UNIVERSITY LIBRARY n.d. Available from: <http://arXiv.org> [Accessed 12 December 2006] Ithaca, NY: Cornell University.

DAWYNDT, P. 2004. *Knowledge accumulation of microbial data aiming at a dynamic taxonomic framework*. Ph.D. thesis, J. Swings, H. De Meyer (promoters), Faculty of Sciences, Department of Biochemistry, Physiology and Microbiology, Ghent University.

DAWYNDT, P., VANCANNEYT, M., DE MEYER, H., AND SWINGS, J. 2005. "Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources", *IEEE Transactions on Knowledge and Data Engineering*, 17 (8), 1111-1126.

DEDEURWAERDERE, T. 2002. Ethics and learning. From state regulation towards reflexive self-regulation of the information society. In: Brunstein, K. and Berleur, J., eds. *Human choice and computers: issues of choice and quality of life in the information society*. Proceedings of the IFIP-TC9 HCC Conference, Montreal August 2002. Dordrecht: Kluwer Academic Publishers, 121-130.

DEDEURWAERDERE, T. 2004. Bioprospecting, intellectual property law and evolutionary economics: the stake of a theory (of reflexive governance). In: Watanabe, M. M., Suzuki, K., and Seki, T., eds. *Innovative roles of biological resources centres*. Tsukuba: World Federation For Culture Collections, 389-395.

DNA DATABANK OF JAPAN (DDBJ) n.d. Available from: <http://www.ddbj.nig.ac.jp/Welcom.html>

[Accessed 12 December 2006] Shizuoka: Japan.

DRAHOS, P. 2000. "Indigenous knowledge, intellectual property and biopiracy: is a global biocollecting society the answer?", *European Intellectual Property Right Review*, 6, 245-250.

DRIESDEN, D. M. 2003. *The economic dynamics of environmental law*. Cambridge, MA: MIT Press.

EMBL NUCLEOTIDE SEQUENCE DATABASE n.d. Available from: <http://www.ebi.ac.uk/embl/index.html> [Accessed 12 December 2006] Produced in collaboration between GenBank (USA) and DNA Data Bank of Japan (DDBJ).

GENBANK n.d. Available from: <http://www.ncbi.nlm.nih.gov> [Accessed 12 December 2006] Bethesda, MD: National Centre for Biotechnological Information.

GENOMES ONLINE DATABASE n.d. Available from: <http://www.genomesonline.org> [Accessed 12 December 2006].

GEYERS, D., COHAN, F. M., LAWRENCE, J. G., SPRATT, B. G., COENYE, T., FEIL, E. J., STACKEBRANDT, E., VAN DE PEER, Y., VANDAMPE, P., THOMPSON, F. L., AND SWINGS, J. 2005. "Reevaluating prokaryotic species", *Nature Reviews Microbiology*, 3 (9), 733-739.

GLOBAL BIODIVERSITY INFORMATION FACILITY n.d. Available from: <http://www.gbif.org> [Accessed 12 December 2006] Copenhagen: Denmark.

HESS, C., AND OSTROM, E. 2003. "Ideas, artifacts and facilities. Information as a common-pool resource", *Law and Contemporary Problems*, 66 (1/2), 111-146.

KOHONEN, T. 1990. "The self-organizing map", *Proceedings of the IEEE*, 78 (9), 1464-1480.

OECD 2001. *Biological resource centres. Underpinning the future of life sciences and biotechnology*. Paris: OECD.

POLSKI, M. 2005. "The institutional economics of biodiversity, biological materials, and bioprospecting", *Ecological Economics*, 53 (4), 543-557.

REICHMAN, J. H. 1994. "Legal hybrids between the patent and the copyright paradigms", *Columbia Law Review*, 94 (8), 2432-2558.

REICHMAN, J. H. 2000. "Of green tulips and legal kudzu: repackaging rights in subpatentable innovation", *Vanderbilt Law Review*, 53 (6), 1743-1798.

REICHMAN, J. H., AND ULHIR, P. F. 2003. "A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment", *Law and Contemporary Problems*, 66 (1/2), 315-440.

STIGLITZ, J., ORSZAG, P. R., AND ORSZAG, J. M. 2000. *The role of government in a digital age*. Washington, DC: Report for the Computer and Communications Industry Association.

SWANSON, T. 1997. *Global action for biodiversity*. London: Earthscan.

THOMPSON, F. L., THOMPSON, C. C., VICENTE, A. C. P., THEOPHILO, G. N. D., HOFER, E., AND SWINGS, J. 2003. "Genomic diversity of clinical and environmental *Vibrio cholerae* strains isolated in Brazil between 1991 and 2001 as revealed by fluorescent amplified length polymorphism analysis", *Journal of Clinical Microbiology*, 41 (5), 1946-1950.