



**Institut national
de la santé et de la recherche médicale**

Esquisse d'un projet de « publications informatisées » pour les Comptes Rendus de l'Académie des Sciences

Philippe Kourilsky, Suzy Mouchet, Colette Brezin
Février 1992

I – LE CONSTAT

Nombreux sont les scientifiques qui éprouvent un malaise croissant devant les problèmes liés à la communication et à la publication, et leurs corollaires qui concernent l'évaluation des hommes, des équipes et des projets. L'accroissement considérable du volume des publications dans le domaine des sciences de la vie et de la santé constitue un phénomène caractéristique de ce domaine de la science : 10 à 100 fois plus de pages sont publiées chaque année en biologie qu'en chimie ou en physique. Ceci reflète non seulement la « productivité » intrinsèque très élevée du domaine, mais aussi l'augmentation du nombre des chercheurs engagés dans les disciplines biologiques –étant entendu que, du fait de l'unicité structurale profonde des êtres vivants, l'étude de la drosophile comme celle du bégonia peuvent contribuer à la compréhension de l'homme.

La multiplication des publications a pour conséquence que l'information scientifique est de plus en plus difficile à maîtriser par un individu, même dans son domaine de spécialisation, et devient de plus en plus l'affaire d'un groupe. (C'est l'une des considérations qui doit inspirer la politique de sites). Cette multiplication est telle que nul ne peut espérer lire plus d'une douzaine de revues – et ceci renforce l'impérialisme » des grandes revues internationales telles que *Nature*, *Cell*, etc... En fait, l'abondance de l'écrit est telle que beaucoup feuilletent les revues et s'instruisent de façon impressionniste, voire « pointilliste », par la lecture des résumés et des titres et, souvent, par la lecture des seuls titres. Un comportement assez

répandu chez les chercheurs consiste à empiler des photocopies d'articles jugés importants qui auront été parcourus mais rarement véritablement lus –leur accumulation fournissant une fausse sécurité ou sécurisation vis-à-vis d'un problème, en fait, mal contrôlé. Une conséquence de cet état de fait est le poids accru pris par la communication orale, source d'informations qualifiées par d'autres et donc d'économies considérables de temps de lecture. La communication orale constitue, en outre, une solution adaptée à la vie en groupe. Elle présente de nombreux avantages mais aussi les inconvénients d'une certaine subjectivité : les phénomènes de rumeur, la « cote » des chercheurs, etc... sont ainsi véhiculés par les systèmes de communication oraux dont on notera aussi le caractère fortement aléatoire (le hasard des rencontres, des voyages, des invitations aux séminaires, aux conférences, etc...).

On pourrait penser qu'une spécialisation accrue devrait suffire à régler un problème qui serait conçu, par essence, comme individuel, *i.e.* qui renverrait à la capacité des individus à le résoudre. Il n'en est rien : la baisse de la qualification de l'information a des conséquences beaucoup plus perverses, parce que **la qualité des évaluations se trouve altérée**. Ainsi, des erreurs ou des imprécisions de plus en plus fréquentes sont commises lors de l'évaluation anonyme des articles soumis pour publication. Les référents n'ont plus le temps d'analyser à fond le projet de publication et n'ont pas toujours la compétence requise. En gros, nul ne se préoccupe plus de vérifier la validité des citations des travaux antérieurs. Faute de place, auteurs et éditeurs sont implicitement d'accord pour diminuer le nombre des citations et renvoyer à des revues de synthèse. Ces dernières, malheureusement, oscillent entre le plaidoyer *pro domo* en faveur des auteurs eux-mêmes, et la compilation non critique de plusieurs centaines d'articles. Trop rares sont les synthèses honnêtes, critiques et équilibrées. Au demeurant, la lecture des articles par des référents anonymes n'induit qu'une diminution marginale du nombre des articles publiés. Un article rejeté par une revue finit presque toujours par être publié dans une autre. Les critiques des référents permettent, en revanche, d'améliorer la qualité des publications –à condition que les référents se livrent avec sérieux et générosité à cet apostolat de la critique scientifique.

Au-delà de l'évaluation des articles, c'est évidemment celle des hommes, des équipes et des projets qui est en jeu. Dans les commissions spécialisées et les divers comités ou conseils scientifiques, l'évaluation des personnels ou des projets prend en compte, à un certain degré, le niveau supposé des publications selon la « cote » des revues dans lesquelles elles ont été acceptées. Le prestige de la revue rejaille sur l'article et ses auteurs. Les grandes revues ne

publient, en général, que de « bons » articles, mais tous les « bons » articles ne sont pas publiés dans les grandes revues –pour lesquelles la sélection ultime dépend parfois de critères non scientifiques (modes, cote des auteurs, réseaux d’amitiés, politique éditoriale, etc.). Ainsi, une partie de l’évaluation des chercheurs, des équipes et des projets (mais une partie seulement) se trouve *de facto* déléguée à des systèmes connus pour être imparfaits (voire partiaux) et qui conservent, néanmoins, une influence considérable.

Sans vouloir nous étendre sur cet autre aspect des problèmes, nous ne pouvons passer sous silence les **problèmes culturels** liés à un système de publication très largement dominé par les anglo-saxons et les américains en particulier. Même si l’emploi de l’anglais comme langue véhiculaire pour les données primaires est, pour l’instant, un fait acquis, on peut s’interroger sur les contraintes ainsi introduites au niveau de la pensée, de l’écriture et de la communication pour les non-anglophones de naissance. Mais au-delà de la question de la langue, la prégnance des traits culturels (la « spéculation » est une activité intellectuelle vue très négativement aux Etats-Unis) et sociologiques (publier constitue, dans le système américain actuel, et dans d’autres, une véritable lutte pour la vie) retentit fortement sur nos propres échelles de valeur. Ceci comporte des aspects tout à fait positifs en ce que nous corrigeons ainsi certains travers nationaux mais, à l’inverse, nous subissons ou importons certains travers des autres –et rien ne doit nous interdire de tenter d’y résister.

Si nous essayons d’imaginer l’évolution du système actuel, abstraction faite des nouveaux procédés dont nous allons parler, nous ne voyons rien qui permette d’espérer une amélioration, bien au contraire. Il nous semble que, depuis une dizaine d’années environ le système a commencé à dériver dans un sens où la perte de l’efficacité dans la transmission de l’information et la dégradation de la qualification de l’information se conjuguent **avec un certain relâchement de la déontologie scientifique** et l’intrusion croissante de petites malhonnêtetés. Il importe donc, en premier chef, **que la communauté regagne une meilleure maîtrise de l’information scientifique**, d’où doivent découler des améliorations de la recherche et de son évaluation.

II – LA DEFINITION DU FAIT SCIENTIFIQUE

Les nouvelles techniques de communication fondées sur l'emploi de l'informatique et de l'électronique offrent des pistes de solutions que nous allons analyser. Cependant, nous souhaitons auparavant introduire une notion forte : il nous apparaît beaucoup trop réducteur de penser que ces nouvelles techniques doivent servir à gérer plus facilement un nombre toujours croissant de publications sans nous interroger sur le contenu de ces dernières. Plutôt que de gloser sur l'« utilité » ou l'« inutilité » de beaucoup d'entre elles, il nous semble essentiel de revenir à la **notion de fait scientifique. Une publication doit contenir un fait scientifique nouveau.** Et l'on doit bien évidemment s'interroger sur la question de savoir si la multiplication des publications (qui est source, comme nous l'avons vu de très nombreux maux) est due au seul accroissement de la connaissance scientifique ou, pour partie au moins, à la **dégradation du contenu de ce que l'on reconnaît pour un fait scientifique.** Et même si l'on n'accepte pas qu'il y ait eu dégradation du contenu, on doit s'interroger sur la vertu qu'il y aurait à **renforcer le contenu du fait scientifique** méritant publication. Cette démarche rejoint une opinion élitiste assez répandue que S. Tonegawa, par exemple, formule de la façon suivante : chaque individu ne fait, dans sa vie de chercheur, qu'un petit nombre de découvertes importantes. Donnons donc à chacun un carnet de 50 tickets l'autorisant à publier un maximum de 50 articles dans sa vie entière. Il est intéressant de noter que la tendance actuelle de l'édition est exactement inverse : on assiste à une floraison croissante de petites revues relativement spécialisées qui touchent un public de scientifiques limité mais bien cerné, mais qui rapportent rarement des faits « majeurs ». L'évolution des techniques, qui diminuent le coût de production des publications écrites, a ouvert une multitude de petits marchés pour des revues tirées à 700-1000 exemplaires et diffusées quasi exclusivement dans les bibliothèques. Ainsi, la réponse du système, face à la difficulté de publier dans les « grandes » revues, n'est pas de revoir le contenu du fait scientifique mais de multiplier les vecteurs de diffusion.

Une question princeps est donc de déterminer si l'emploi des **technologies nouvelles constituera simplement un mode de gestion de la prolifération actuelle ou si il sera accompagné par une redéfinition ou une requalification du fait scientifique.**

III – LES POSSIBILITES OUVERTES PAR LES TECHNOLOGIES NOUVELLES

Ces possibilités sont ouvertes par deux types de technologies : l'informatisation des banques de données et l'amélioration des communications, notamment des communications en réseau.

A) Les banques de données produites à partir de l'écrit

L'extension des banques de données scientifiques informatisées constitue un tournant relativement récent dont les retombées joueront probablement un rôle majeur pour les problèmes évoqués plus haut.

A l'heure actuelle, les titres des articles publiés dans la quasi-totalité des revues mondiales (indexées par les *Current contents*, par exemple) sont disponibles, sous forme de disquettes expédiées à un rythme hebdomadaire, avec un délai de saisie faible (une à deux semaines). Les résumés des articles sont, depuis peu, fournis avec un délai de quelques semaines. Une banque telle que Medline propose, sur CD, à un rythme mensuel, les titres et les résumés, mais le délai de saisie atteint 3 à 6 mois. D'autres banques de données, dont des banques françaises et européennes, sont disponibles sous diverses formes. Beaucoup sont accessibles « on-line », mais à des coûts souvent prohibitifs pour le simple particulier (les tarifs étant souvent ajustés à l'usage fait en bibliothèque par un nombre important d'utilisateurs divers).

Les algorithmes permettant d'exploiter de façon de plus en plus performante et « intelligente » les banques de données se sont perfectionnées de façon spectaculaire. Ces perfectionnements, joints à la puissance croissante des micro-ordinateurs, rendent la recherche bibliographique de plus en plus performante.

En principe, la généralisation -probable- des méthodes d'interrogation informatisées devrait, à terme, diminuer le poids des très grandes revues internationales, en ce sens qu'un travail publié dans une revue de moindre prestige ne devrait pas échapper à la recherche bibliographique. On notera que cette embellie hypothétique dépend très étroitement de la finesse des logiciels d'interrogation et/ou des sous-structures des banques elles-mêmes (on semble évoluer vers la constitution de banques spécialisées plus faciles à interroger que les banques générales). En effet, si, comme c'est souvent le cas, une recherche par mots clés livre plusieurs centaines de références et de résumés, l'interrogateur est confronté à la lecture de l'équivalent d'un livre, c'est-à-dire qu'il est quasiment ramené au problème précédent. Reste,

en outre, le fait que l'information est restituée (pour l'instant sous forme de résumé) de façon non qualifiée, si bien que le lecteur pourra toujours, sans réelle raison objective, considérer que l'article publié dans *Nature* est de meilleure qualité que celui publié dans une revue plus spécialisée. Ce qui est certain, c'est que ce dernier ne pourra pas être totalement ignoré.

En dépit de ces réserves, il est évident que la généralisation des banques de données informatisées constitue un progrès tout-à-fait majeur, qui permet de professionnaliser et de systématiser la gestion de l'information scientifique.

B) L'amélioration des communications en réseau

Le succès foudroyant de la télécopie indique à quel point l'amélioration des communications en réseau est susceptible de modifier rapidement les habitudes des utilisateurs. Mais la télécopie n'est, en un sens, qu'une version primitive de la communication électronique. Dans certains secteurs (physique et informatique notamment) l'usage du courrier électronique domine toutes les autres formes de communication. Le point essentiel est, bien évidemment, que la communication peut quitter le champ du dialogue entre deux individus pour devenir communication entre n individus constituant un réseau.

Nous fondant sur l'expérience acquise depuis une dizaine d'années par un groupe international de physiciens théoriciens, ainsi que sur les débats qui commencent à se faire nourrir autour des publications purement électroniques (dont nous discuterons plus bas), nous pouvons esquisser le scénario suivant : des groupes de scientifiques dont les activités sont centrées sur un thème relativement précis, peuvent communiquer par voie électronique, au sein d'un réseau, de façon beaucoup plus rapide et efficace que par tout autre moyen. Ce mode de communication peut être corrélié à un changement radical des règles de la communication. Ainsi, les informations neuves sont-elles injectées dans le réseau sans aucun contrôle préalable et bien avant leur publication sous forme d'article. L'injection dans le réseau, à une date donnée, est considérée comme donnant la priorité de la découverte. Le réseau agit, par ses réactions explicites, comme un ensemble de référents, si bien que l'évaluation du fait nouveau injecté dans le réseau est, en réalité, collective. En fonction de l'évolution du travail fait par l'auteur lui-même, et des réactions du réseau, les quelques pages de texte qui décrivaient initialement le travail peuvent être modifiées, au vu et au su de tous. Lorsque le travail est jugé mûr, il est envoyé à une revue et paraît sous forme d'article imprimé, souvent un an et plus après sa divulgation première dans le réseau. Ce système

évoque, dans une certaine mesure, celui des envois de « pre-prints » avant soumission pour publication à un club de collègues qui peuvent communiquer leurs réactions. Ce système, assez en vogue chez les biologistes dans les années 1970, a plus ou moins disparu, mais tend à resurgir.

Détaillons, ci-dessous, quelques caractéristiques des messageries électroniques :

- 1) L'information est disponible instantanément dans tous les centres internationaux disposant de messagerie électronique.
- 2) L'information est à la portée des directeurs de laboratoires, chercheurs, étudiants, sans rétention ni distorsion.
- 3) Les micro-ordinateurs usuels peuvent servir de terminaux reliés à un réseau local. Le serveur de communication qui pilote le système pour un ensemble de laboratoires d'un site donné, est une station de travail courante dont les prix ne cessent de baisser (quelques dizaines de milliers de francs aujourd'hui).
- 4) L'information est de coût faible, et met sur pied de quasi-égalité petits et grands instituts, pays pauvres et développés.
- 5) La plupart des centres universitaires mondiaux disposent déjà de tels réseaux de communication (en raison, notamment, de leur utilisation systématique par les physiciens).
- 6) Les capacités de stockage des articles sur les disques informatiques sont considérables, même en regard du gigantisme de la quantité d'informations à stocker (voir calculs du physicien P. Ginsparg).
- 7) Néanmoins, **la communication orale restera essentielle**, en particulier pour déborder des cadres de l'hyperspécialisation. Les revues scientifiques consacrées à des articles de mise au point joueront presque certainement un rôle croissant (et ces articles de revue devraient être pleinement pris en compte dans les procédures d'évaluation).

Plusieurs remarques peuvent être ici formulées.

- 1) En premier lieu, le système évoqué ci-dessus constitue un outil de travail relationnel avant d'être en quoi que ce soit un « journal » électronique. Mais on voit bien qu'il serait facile de franchir le pas.
- 2) En deuxième lieu, si nous imaginons que de tels réseaux se créent dans le champ des sciences de la vie et de la santé, ils doivent être de dimension relativement limitée, à la fois en raison des limitations techniques propres au réseau de communication (mais qui peuvent évoluer dans le temps) et du fait, que pour être efficaces, de tels réseaux ne

peuvent être abondés que par une quantité relativement limitée d'informations. Même dans ces conditions, il conviendra de doter la messagerie des logiciels d'interrogation performants qui permettent de faire efficacement des recherches bibliographiques.

- 3) En dernier lieu, et dès lors qu'un réseau, pour bien fonctionner, doit être de dimensions limitée, il importe de bien réfléchir aux règles constitutives afin d'éviter toutes sortes de déviations et de perversions. Par exemple, un réseau devrait être centré sur un thème précis, mais ouvert à tous. L'inverse (le réseau fermé) pourrait conduire à un club de caciques bénéficiant d'informations privilégiées, mettant plus ou moins hors compétition ceux qui n'appartiennent pas aux « happy few » (et notamment les plus jeunes). Enfin, si l'on imagine la naissance d'un certain nombre de tels réseaux, on peut penser qu'un même individu participe à plusieurs réseaux, que les réseaux doivent s'interfacier, voire se hiérarchiser en réseaux de réseaux, etc...

IV – UN ROLE POSSIBLE POUR L'ACADEMIE DES SCIENCES

Ainsi qu'il a été mentionné plus haut (Section II), l'un des points critiques pour l'avenir concerne une éventuelle redéfinition, ou requalification, du fait scientifique. Il est évident que ce problème dépasse le cadre de l'INSERM et ne peut être traité que par des instances nationales, européennes ou internationales. **Nous faisons, ici, l'hypothèse que l'Académie des Sciences serait en mesure d'aborder cette question, toujours, répétons-le, dans une approche de type expérimental.** Cela nous conduit à esquisser une proposition de collaboration entre l'INSERM et l'Académie des Sciences (qui pourrait apparaître comme une suite donnée au Prix commun) (et qui pourrait, bien évidemment, être étendue à d'autres organismes).

Nous partons de l'idée que, d'une part, l'Académie des Sciences possède le statut et la stature morale adaptés à l'ampleur du problème et que, d'autre part, les notes communiquées aux Comptes Rendus ont un format adapté à la communication d'un fait scientifique neuf –en conformité avec leur mission initiale. L'expérience dont les principes sont énoncés ci-dessous (sous réserve de réflexions et de critiques plus approfondies) pourrait aider à un renouvellement des CRAS sans que le risque soit pris, pour cette publication, d'un changement radical. Enfin, l'Académie pourrait jouer pleinement son rôle, dans un esprit novateur, à visée éducative, qui pourrait avoir une portée profonde dans la communauté

scientifique. Bien entendu, il pourrait apparaître, à l'analyse, que le projet est inadéquat, et il conviendrait peut-être d'en étudier la réalisation à l'échelle européenne.

V – UN PROJET DE MESSAGERIE ET DE PUBLICATION ELECTRONIQUES

En bref, il s'agirait de mettre en place, sous l'égide de l'Académie des Sciences, une messagerie électronique fonctionnant selon les principes suivants :

- 1) La messagerie peut être consultée par tous, de façon aussi simple que possible, et à coût réduit ou quasi nul.
- 2) La messagerie est alimentée par des notes scientifiques dont le format correspond, en gros, aux notes soumises à ce jour aux CRAS (soit environ 3 ou 4 pages dactylographiées et quelques figures ou tableaux).
- 3) En principe, ces notes doivent contenir un « fait scientifique » neuf. Il arrivera souvent que le format retenu ne permette pas d'exposer la totalité des données expérimentales qui établissent la validité du fait scientifique. L'auteur de la note devra alors, en vertu d'une règle déontologique clairement énoncée, tenir à la disposition de n'importe quel demandeur, les données scientifiques complémentaires.
- 4) Les notes sont transmises à l'Académie et peuvent appartenir successivement à trois catégories A, B et C. Toute note qui parvient à l'Académie est dite de catégorie C. Elle doit être validée quant à son format et son contenu général pour être admise en catégorie B.
- 5) L'admission en catégorie B signifie que la note entre dans la base de données sans aucun examen par des référents. En d'autres termes, les notes entrent dans la base de données sous la responsabilité principale de leurs auteurs. Elles deviennent alors accessibles à tous.
- 6) Les auteurs peuvent demander le classement de la note en catégorie A, ce qui impliquera l'évaluation de celle-ci par des référents anonymes. Des lecteurs de la note (accessible dans la base de données) peuvent également demander, s'ils jugent le résultat important, le classement en catégorie A.
- 7) Les notes classées en catégorie A constituent le sous-ensemble de la base de données qui est destiné à être publié dans les CRAS. La base de données elle-même contient une image du patrimoine scientifique national.
- 8) Les lecteurs des notes dans la base de données peuvent, par voie électronique, qualifier les notes (en A ou B). Ils peuvent donc contester un classement en A ou en B, et ce même

plusieurs années après le dépôt de la note dans la base de données. Ainsi, il peut apparaître qu'un résultat important a été méjugé et cette sous-estimation peut être corrigée.

- 9) Il serait opportun que les notes fassent l'objet d'un double dépôt systématique en français et en anglais. Le format modeste des notes fait de cette clause une contrainte minimale. Dans ce cas, la forme écrite des CRAS devrait exister sous forme de deux produits distincts, l'un en français, l'autre en anglais. La banque de données en anglais permet évidemment une bien meilleure ouverture internationale.

Il est clair que ce schéma, pour ambitieux qu'il puisse paraître, aborde de front les problèmes évoqués au début de ce document. En particulier, la requalification progressive du « fait scientifique » devient un problème traité par la collectivité des scientifiques, sous couvert de l'autorité morale de l'Académie. Le caractère pluri-disciplinaire de l'Académie (car ce schéma engloberait toutes les disciplines) faciliterait sans doute une évolution spécifique des Sciences de la Vie. L'effort entrepris au sein même de l'INSERM pour moderniser la communication scientifique permettrait à l'organisme d'appuyer efficacement un projet de ce type pour ce qui concerne la série III (Sciences de la Vie) des CRAS.