

Impact sociétal de la recherche : évaluation et perspectives

par **Laura BEDOURET**

*Docteur en chimie-physique
Diplômée de l'Université de Bordeaux*

et **Anna DUPLEIX**

*Post-doctorante dans l'équipe Bois du Laboratoire de Mécanique et Génie Civil
UMR CNRS, Université de Montpellier 5508*

1. Vers une évaluation de l'impact sociétal de la recherche.....	AG 117 - 2
1.1 Évolution de la place de la recherche scientifique dans la société	— 2
1.2 Évolution des enjeux de la recherche scientifique	— 3
2. Impact sociétal : une difficile définition	— 4
2.1 Un point sur l'étymologie	— 4
2.2 Une vaste gamme de domaines et de critères possibles.....	— 5
2.3 Questions épistémologiques et enjeux éthiques	— 5
3. Méthodologies d'évaluation de l'impact sociétal	— 8
3.1 Méthodes d'évaluation ex ante : planifier la recherche pour en optimiser l'impact.....	— 8
3.2 Méthodes d'évaluation ex post : vérifier l'atteinte d'impacts sociétaux	— 9
3.3 Systèmes d'évaluation nationaux des établissements de recherche	— 11
4. Mesure de l'impact sociétal : l'apport des métriques	— 14
4.1 Les différents types de métriques	— 14
4.2 Avantages et limites des métriques	— 15
4.3 Vers des métriques responsables.....	— 16
4.4 Des métriques évaluatives aux métriques descriptives	— 17
5. Bilan et perspectives de l'évaluation de l'impact sociétal	— 18
5.1 Orientation générale de l'évaluation des impacts sociétaux	— 18
5.2 Perspectives d'amélioration méthodologique.....	— 18
5.3 Limites et perspectives d'évolution des évaluations d'impact	— 19
6. Conclusion.....	— 21
Pour en savoir plus	Doc. AG 117

Les trois dernières décennies ont connu un processus de réorganisation lent mais continu de la recherche publique, mêlant compétition globale et coupes budgétaires, demande de transparence et de responsabilité, évaluation et financement. À l'aune de cette réorganisation, les missions de la recherche académique sont progressivement en train d'évoluer d'une science priorisant le développement de la connaissance à une science devant être « utile » et apporter la démonstration de son impact sur la société. Les critères d'évaluation qui reposaient jadis essentiellement sur la qualité et l'impact scientifique des travaux de recherche se sont donc élargis à la considération de l'ensemble des retombées sociétales que ces travaux permettent de réaliser.

Cependant, en raison, notamment, du très large spectre que peut englober le terme « impact sociétal », des interrogations continuent à peser sur sa définition. Par ailleurs, ces impacts devant être démontrés, la manière dont ils peuvent être mesurés soulève tout autant de questions et fait l'objet de nombreuses études.

Cet article propose de retracer les facteurs ayant amené l'évaluation de la recherche à intégrer l'impact sociétal parmi ses critères et de faire le point sur les connaissances et quelques-unes des méthodologies actuellement utilisées. Seront également évoquées l'évolution de la place allouée à la scientométrie dans la mesure des impacts sociétaux ainsi que les perspectives d'évolution de l'évaluation de la recherche qui émergent peu à peu.

1. Vers une évaluation de l'impact sociétal de la recherche

L'Histoire fut témoin de fréquentes variations dans la relation entre la recherche et la société, notamment via l'évolution du point de vue adopté par ses institutions politiques. Parfois indifférentes ou critiques, celles-ci furent également à l'origine de la structuration d'organismes scientifiques, leur étant de fait souvent inféodées. Ferdinand II de Médicis et l'un de ses frères fondèrent par exemple l'Accademia del Cimento, de même que Charles II concéda une chartre à la Royal Society of London et parraina l'Observatoire de Greenwich, que l'Académie des Sciences française fut créée par Colbert puis placée sous la protection de Louis XIV et que celle de Saint-Petersbourg fut instituée par Pierre le Grand. Revendiquant une relative autonomie au milieu du XX^e siècle, la recherche publique fut peu à peu contrainte de rendre des comptes quant à ses performances et sa qualité à la fin de celui-ci, ce qui se traduisit par une modification de la nature des évaluations.

À partir des années 90, de nouvelles pratiques scientifiques centrées sur l'application et l'interaction avec des parties prenantes non académiques émergent, remettant en question tant le paradigme d'une recherche désintéressée et relativement autonome que le lien systématique entre recherches jugées de haut niveau scientifique et utilité pour la société. Dans le contexte d'une restriction des budgets publics et de l'émergence d'une volonté de planification stratégique des recherches afin de répondre aux grands défis sociétaux, la justification de leur valeur pour la société devint alors de plus en plus prégnante, se traduisant finalement par l'introduction d'un impact sociétal en tant que critère d'évaluation de la recherche.

1.1 Évolution de la place de la recherche scientifique dans la société

1.1.1 D'un système social distinct, autonome et autorégulé...

Point prenant aujourd'hui une importance capitale dans l'organisation de nombreuses sociétés modernes, la question de la place de la recherche scientifique dans la société fit l'objet de nombreux travaux. Tour à tour conspuée, ignorée ou encouragée au fil des systèmes de valeurs en vigueur, elle atteint selon R. K. Merton au cours de la première moitié du XX^e siècle le statut de fin en soi, statut qui s'accompagna d'une tendance pour « le scientifique à se considérer comme indépendant de la société et voir la science comme une entreprise auto-validée » [1].

Néanmoins, des attaques portant sur son intégrité conduisirent les chercheurs à reconnaître leur dépendance à certains types de structures sociales, ce qui se traduisit par la formulation d'un système de valeurs et de normes visant à « stabiliser et réguler [la recherche scientifique] en la protégeant d'abus internes et en lui permettant de résister aux influences et intrusions des acteurs politiques et économiques ». Dans un paradigme où l'objectif de la science est l'exten-

sion d'une connaissance confirmée expérimentalement et résultant d'une logique cohérente, R. K. Merton se basa ainsi sur une observation des usages et écrits de l'époque pour définir quatre ensembles d'impératifs à la recherche scientifique : l'universalisme, le communisme, le désintéressement et le scepticisme organisé [1] :

- L'universalisme garantit une objectivité vis-à-vis de la source d'une découverte scientifique, un jugement indépendant des attributs personnels ou sociaux de son ou ses auteurs.
- Le communisme considère les résultats substantiels de la recherche comme un patrimoine commun qu'il est nécessaire de communiquer et est par exemple incompatible avec la définition de la technologie comme « propriété privée » dans une économie capitaliste.
- Le désintéressement attribue aux chercheurs une passion pour la connaissance, une curiosité gratuite ou encore une préoccupation altruiste au bénéfice de l'humanité.
- Enfin, le scepticisme consiste en une suspension temporaire du jugement et un examen minutieux des croyances en termes de critères empiriques et logiques, source fréquente de conflits avec d'autres institutions telles que l'Église, le monde économique ou l'État dont certaines découvertes scientifiques semblent invalider les dogmes.

Ces normes font donc de la science un système social distinct, autorégulé et relativement autonome, et leur respect apparaît fondamentalement incompatible avec la réponse à des demandes d'utilité sociale venues par exemple d'acteurs du monde économique ou politique. Elles ne nient pas le fait que la recherche a des implications pratiques mais n'en font pas ses objets principaux.

1.1.2 ... à un élément stratégique dont les performances et la qualité sont évalués

Les implications pratiques de la recherche commencèrent à être codifiées peu après la Seconde Guerre mondiale, lorsque le contexte de réalisation des activités de recherche mit en exergue son rôle de facteur de compétitivité économique. La résolution de problèmes pratiques était alors déjà centrale au fait d'investir dans la recherche, mais le processus liant la science et ses effets, considéré comme essentiellement linéaire, reposait sur le fait de laisser une totale liberté aux chercheurs. Dans son livre « *Science, the endless frontier* » publié en 1945, V. Bush affirmait ainsi que « tant que les [organismes de recherche] sont vigoureux et sains et que leurs scientifiques sont libres de poursuivre leur vérité où qu'elle puisse mener, il y aura un courant de nouveaux savoirs scientifiques pour ceux qui peuvent l'appliquer aux problèmes pratiques dans le gouvernement, l'industrie ou ailleurs » [2]. Jusque dans les années 70, il n'y eut aucun doute sur le fait que l'investissement public dans la recherche aurait un impact positif. Bien que l'on note l'existence d'indicateurs bibliométriques tels que le Journal Impact Factor (JIF) ou le Science Citation Index (SCI) respectivement développés par Eugene Garfield en 1955 et en 1961, ils étaient principalement utilisés par les experts en scientométrie pour décrire le développement et la direction de la recherche scientifique, et non pas pour l'évaluer.

La question de l'évaluation de l'impact de la recherche avait déjà été abordée, par exemple au début des années 60, afin de plaider pour un financement à long terme d'une organisation

internationale spécialisée dans la recherche agronomique pour le développement, mais elle ne devint réellement centrale qu'à partir des années 80. Cette décennie connut l'émergence d'une gestion entrepreneuriale des universités et institutions de recherche publiques qui furent contraintes de rendre des comptes sur leurs accomplissements de manière à « offrir une vision objective et transparente » de leurs performances. Ces comptes rendus prirent la forme d'évaluations internes qualitatives via des revues par un panel de pairs, méthode déjà décrite par R. K. Merton comme un « système de vigilance institutionnalisé » dans l'autorégulation de la communauté de recherche [1] [3]. S'y ajoutèrent également des méthodes quantitatives visant à mesurer le rendement et l'impact des recherches sur le savoir académique et scientifique avec notamment le développement d'indicateurs bibliométriques renseignant sur le nombre de publications et de citations.

L'hypothèse demeurait alors que la société bénéficierait automatiquement de la recherche scientifique menée au plus haut niveau, notamment via un continuum entre recherche fondamentale et recherche appliquée.

1.2 Évolution des enjeux de la recherche scientifique

1.2.1 Émergence d'une nouvelle pratique de la recherche

Les années 90 furent témoin de changements dans la manière de générer de la connaissance. En 1994, M. Gibbons rapporta ainsi l'émergence, parallèlement aux usages traditionnels qualifiés de « mode 1 », de nouvelles pratiques qu'il définit comme un « mode 2 » de production du savoir [4]. Le mode 1 décrit une science gouvernée par les intérêts académiques d'une communauté spécifique (notamment les théoriciens), où la production de connaissances est indépendante de son contexte d'application. Le mode 2 se caractérise quant à lui par une collaboration non seulement au sein de la communauté scientifique mais également entre celle-ci et d'autres parties prenantes non issues du monde académique, une transdisciplinarité dans l'étude de problèmes du monde réel, et une recherche fondamentale menée dans le contexte de son application pour les utilisateurs de la recherche.

La distinction entre ces deux modes fut ultérieurement précisée comme consistant en une focalisation et une responsabilité différentes. Tandis que le mode 1 est centré sur la communauté des chercheurs et fait seulement l'objet d'une évaluation par des pairs sur des critères d'excellence scientifique et d'originalité [5], le mode 2 se focalise sur les besoins de la société et rend des comptes à un large panel de parties prenantes en fonction de valeurs et de critères utilitaires [5] [6]. Les scientifiques y sont donc amenés à définir leurs recherches en fonction des demandes sociales [7], ce qui remet en question la norme mertonienne d'autonomie de la science.

L'émergence de ce nouveau paradigme d'une production des connaissances orientée vers l'application généra de nombreuses critiques de la part des défenseurs de l'hégémonie de la science théorique ou expérimentale et de l'autonomie des scientifiques et de leurs institutions. Cependant, pour certains, ces deux modes de recherche ne sont pas forcément incompatibles. Il serait en effet possible pour les chercheurs d'opérer d'une manière satisfaisant les exigences et les normes de leurs pairs académiques tout en cherchant en même temps à résoudre des problèmes socialement pertinents avec des partenaires d'autres secteurs. Par ailleurs, la scission entre recherche fondamentale et recherche appliquée (ou finalisée) fut elle-même qualifiée de « notion simpliste » [8]. Plusieurs scientifiques avancèrent que la nature d'une recherche ne dépend pas de son objectif (avancée du savoir ou résolution de problèmes concrets) et que, « si la recherche ayant pour but de résoudre des problèmes réels dépend souvent de l'avancée de la compréhension fondamentale, la résolution de problèmes peut elle-même apporter des éléments nouveaux ». Autrement dit, les impacts scientifiques n'auraient pas toujours pour origines des recherches « désintéressées » [8].

1.2.2 Vers une justification de l'utilité sociétale de la recherche

Face à l'accroissement de la pression sur les budgets publics, la question de la distribution de l'argent public entre les différents domaines dépendants de lui a soulevé de plus en plus explicitement celle des objectifs que l'investissement dans chacun d'entre eux permet d'atteindre. Dans le cas de la recherche, cet investissement se trouve justifié au moins en partie par les divers bénéfices qu'une base scientifique efficace apporte à la société [9] [10].

Malgré le glissement progressif d'une recherche en mode 1 vers une recherche en mode 2, son évaluation et celle des chercheurs est longtemps restée focalisée sur une exigence de qualité définie et implémentée par la communauté académique. Or, se concentrant sur des solutions aux problèmes réels, les décideurs politiques sont difficilement à même d'en apprécier la valeur. L'importance pour la société de la recherche fondamentale n'est en particulier pas totalement comprise [5]. Par ailleurs, le postulat liant recherches de très haute qualité académique et bénéfiques pour la société commença à faire l'objet de questionnements. Des travaux très souvent cités ou publiés dans des journaux renommés peuvent avoir un excellent impact scientifique mais peu ou pas d'impact sociétal, tandis que d'autres, moins ambitieux d'un point de vue académique, peuvent en revanche conduire à d'importantes retombées. Estimées de grande qualité scientifique, les recherches sur l'apoptose n'avaient par exemple toujours pas d'impact sur la santé au bout de 30 ans. De même, bien que susceptibles d'apporter beaucoup à l'avancée d'autres recherches, les ondes gravitationnelles prédites en 1915 par Albert Einstein dans la théorie de la relativité générale et confirmées un siècle plus tard n'ont pour l'instant pas d'applications technologiques potentielles. En revanche, l'étude du rapport coût/efficacité des serviettes d'incontinence, peu valorisée par la communauté scientifique, a eu un impact sociétal immédiat et important [11].

Exemples de premières estimations de l'impact de la recherche sur la société

Parmi les premières investigations concernant les retombées de la recherche sur la société, beaucoup se sont concentrées sur le transfert de connaissances académiques vers les sphères privées et ont consisté en des sondages auprès d'utilisateurs potentiels de résultats scientifiques. Les exemples les plus cités de ce type d'études sont les travaux publiés en 1991 et 1998 par E. Mansfield. Suite à des enquêtes menées auprès de 76 et 77 firmes industrielles des États-Unis, il parvint à l'estimation que 10 % des produits ou processus commercialisés par ces firmes entre 1975 et 1985 et 13 % de ceux commercialisés entre 1986 et 1994 n'auraient pas pu être développés sans l'apport de recherches académiques récentes [12].

La méthode utilisée dans ces travaux fut par la suite reprise dans plusieurs autres dont notamment un sondage réalisé auprès de 2 300 compagnies allemandes qui mit en avant des résultats relativement similaires. Un peu moins de 9 % des firmes ayant introduit des innovations en termes de produits ou de processus entre 1993 et 1996 avaient reçu l'aide de recherches publiques pour leurs innovations [13].

Parallèlement, en 1997, une étude s'intéressa à l'influence de la recherche sur l'industrie en évaluant la fréquence à laquelle les publications scientifiques issues de la recherche publique ont été citées dans 400 000 brevets déposés aux États-Unis. Les références à ce type de publications ont été chiffrées comme étant passées de 17 000 en 1987-1988 à 50 000 en 1993-1994. Une tendance qui fut également observée dans d'autres pays dont le Royaume-Uni, la France, l'Allemagne et le Japon entre 1985 et 1995, témoignant d'une augmentation générale du flux de connaissances de la recherche vers l'industrie [14].

Exemples de premières estimations de l'impact de la recherche sur la société (Suite)

Dans le domaine médical, les bénéfices sur la santé et le PIB du Royaume-Uni attribués à la recherche publique sur les maladies cardiovasculaires entre 1975 et 1992 furent évalués comme correspondant à un taux de retour interne de 39 %. Autrement dit, 1 £ d'investissement aurait produit un flux de bénéfices équivalent à 0,39 £ par année à perpétuité. Le taux de retour sur investissement de la recherche sur les maladies mentales fut quant à lui estimé à 37 % [15].

Se développa alors l'idée que, dans un système dont le temps et les ressources sont limités, une focalisation excessive sur les critères du mode 1 dans l'évaluation de la recherche peut nuire à l'atteinte des objectifs du mode 2 [10]. Des preuves de la valeur de la science pour la société commencèrent alors à être attendues et les institutions firent face à d'importantes questions concernant comment diriger et guider le système de recherche dans le futur. Comme le résuma S. Hill en 2016, « d'un côté, la focalisation sur l'investigation indépendante et l'auto-gouvernance de la recherche a une histoire longue et productive. De l'autre, la complexité croissante des problèmes que la société doit résoudre augmente la pression sur la recherche pour délivrer des solutions et étendre le savoir. » [10]

1.2.3 Naissance d'un critère d'impact sociétal dans l'évaluation de la recherche

Conformément aux modifications, notamment en Europe, des méthodes de gestion des entités publiques visant à les rapprocher de celles des entreprises privées, les politiques de recherche connaissent depuis les trois dernières décennies un processus de réorganisation se traduisant par l'émergence de ce qui a été qualifié de « capitalisme académique », « science post-normale », « université de l'entreprise », « science post-académique », « recherche stratégique » ou encore « science de la finalisation » [9]. Alliées au développement de l'idée qu'une part du contrat social liant la recherche et la société implique que celle-ci devrait répondre et contribuer à des exigences et défis sociétaux [4] [7] [10], ces transformations ont engendré un glissement de l'évaluation de la recherche vers des perspectives enrobant à la fois pertinence scientifique et sociétale.

Jadis essentiellement jugés par leurs pairs à l'aune de critères d'excellence scientifique, les institutions de recherche et les chercheurs sont désormais évalués par leurs divers financeurs publics (agences de financement, collectivités territoriales, Europe) et privés sur une vision de la performance élargie à l'utilité sociétale de la connaissance produite et sa pertinence pour répondre à des défis spécifiques. Souvent regroupés sous l'appellation « impact sociétal de la recherche », ces nouveaux critères ont été intégrés dans les systèmes d'évaluation nationaux de nombreux pays dont l'Australie, le Danemark, la France, l'Italie, le Royaume-Uni ou encore les Pays-Bas. Dans le cadre d'une évaluation ex post, ils visent à étayer quantitativement et qualitativement un retour sur l'investissement engagé, vérifier l'atteinte d'impacts (en général ou prédéfinis en amont) et effectuer des comparaisons à différentes échelles ou par rapport à la moyenne mondiale. Liées à l'allocation de financement ou à la planification des recherches pour en optimiser l'impact, certaines évaluations dites « ex ante » intègrent également un examen de l'impact sociétal prospectif des projets scientifiques afin d'accroître ce type de retombées et d'orienter l'effort de recherche vers la réponse à des problématiques sociétales spécifiques (réchauffement climatique, développement durable...). Une partie importante des programmes de financement de la recherche est par ailleurs spécifiquement mise en place pour répondre à ce type de défis. Le programme Horizon 2020 de l'Union Européenne s'articule par

exemple autour de trois axes : « un soutien sans interruption tout au long du processus de recherche, de l'idée au produit commercialisable », « mettre les financements au service d'une réponse aux grands défis sociétaux » et « soutenir l'innovation et les activités proches du marché de façon à créer de nouvelles opportunités commerciales ».

2. Impact sociétal : une difficile définition

Les missions de la recherche ont progressivement évolué d'une science ayant comme seul objet la recherche de la connaissance à une science « utile », devant démontrer son impact sur la société. Cette démonstration nécessite toutefois un référentiel et des critères sur lesquels s'appuyer. Or, la définition même du terme « impact sociétal » et de ce qu'il englobe peut être envisagée de différentes manières selon la réponse à un certain nombre de questions : Quels domaines peuvent être impactés par la recherche ? Cet impact sociétal englobe-t-il l'ensemble des conséquences d'une recherche ou se limite-t-il à celles attendues et définies en amont ? Est-il lié à la qualité des travaux scientifiques dont il résulte ? S'inscrit-il dans une échelle de valeur ou est-il « absolu » ? Inclut-il seulement les effets pouvant être quantifiés ? Autant d'interrogations dont certains éléments de réponse actuellement discutés sont résumés ci-après.

2.1 Un point sur l'étymologie

Le terme « impact » serait apparu vers 1824 et dériverait du latin *impactus* signifiant « heurté ». Emprunté à la langue anglaise, il est souvent utilisé dans le sens de « effet » ou « influence », emploi qui s'est généralisé au sein de la francophonie bien qu'il demeure critiqué. Le Comité consultatif d'éthique Inra-Cirad note par exemple une ambiguïté entre les termes « impact » et « effet » que certains auteurs différencient : l'effet engloberait tant la conséquence visée de l'action envisagée que les conséquences inattendues tandis que l'impact relèverait de la transposition de l'effet sur une échelle de valeurs et correspondrait au résultat de la comparaison d'un état incluant l'action envisagée à un état de référence. Dès 1988, P. Wathern définissait par exemple l'impact environnemental comme l'effet, pendant un temps et sur un espace donnés, d'une activité humaine sur une composante de l'environnement pris dans son sens large (i. e. physique, biologique et humain). L'impact désigne alors le changement de la mesure d'une variable de l'environnement, compte tenu du contexte général (spatial et temporel) dans lequel il s'insère [16].

Le terme « évaluer » viendrait quant à lui du vieux français *avaler* dont l'origine serait une racine indo-européenne signifiant « exprimer la force », ou « être puissant » [16]. Le fait d'évaluer quelque chose correspondrait donc à en extraire la valeur, en montrer la force et la puissance. Considérant ces remarques étymologiques, évaluer l'impact sociétal d'une recherche consisterait donc à apprécier la force ou la valeur de son apport à la société en comparant la situation suite à cette recherche à celle en son absence. Selon cette définition, l'évaluation peut alors inclure tant les finalités assignées à la recherche que ses conséquences inattendues, heureuses ou non. Elle dépend de la valeur, potentiellement variable, accordée aux effets de la recherche et son importance relève d'un jugement pouvant s'appuyer sur différents critères. Tout le problème réside alors en l'identification de ces critères et la circonscription des domaines, ou dimensions, auxquels ils seront appliqués.

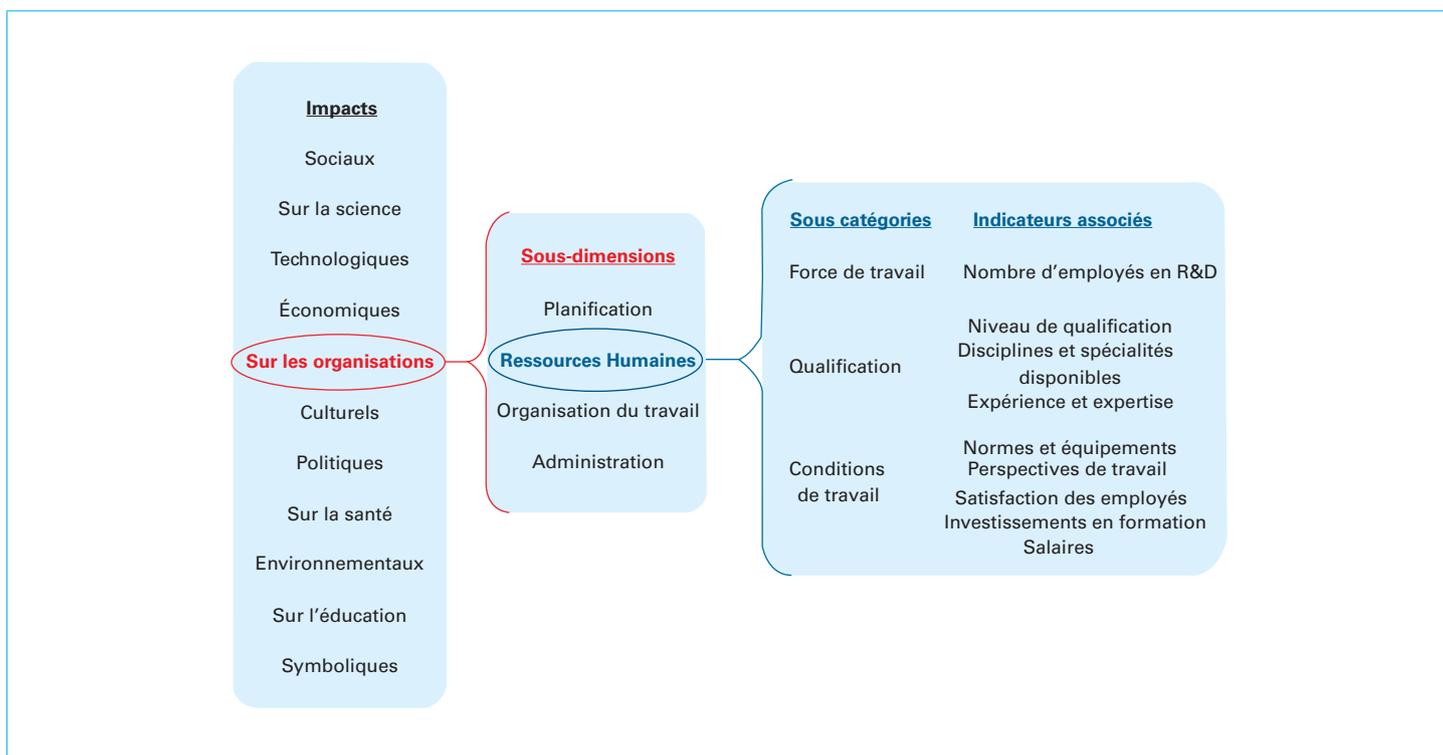


Figure 1 – Illustration d’une proposition d’organisation des impacts sociétaux potentiels [17].

2.2 Une vaste gamme de domaines et de critères possibles

Un très grand nombre de critères, se recouvrant parfois les uns les autres, peuvent entrer en jeu dans l’évaluation de l’impact de la recherche sur la société. Face à cette vaste gamme de paramètres, plusieurs tentatives de classifications ont été menées. En 2005, une série d’entretiens avec des scientifiques d’organisations publiques et des utilisateurs (effectifs ou potentiels) des résultats de la recherche mena à une liste de onze dimensions correspondant à autant de catégories d’impacts : les impacts sociaux mais aussi ceux sur la science elle-même, la technologie, l’économie, la culture, la politique, les organisations, la santé, l’environnement, l’éducation et la symbolique (comprise comme la notoriété, la crédibilité de l’organisme ayant mené les recherches) [17].

Comme l’illustre la figure 1, chacune de ces dimensions est subdivisée en plusieurs sous-dimensions. L’impact sur les organisations est par exemple scindé en 4 sous-catégories : la planification, l’organisation du travail, l’administration et les ressources humaines. Chaque sous-dimension peut elle-même comprendre plusieurs points caractérisés par un ou plusieurs indicateurs ou critères. Ainsi, les ressources humaines incluent par exemple un point « force de travail » dont l’indicateur est le nombre d’employés dans le pôle recherche et développement, un point « qualification des employés » caractérisé par trois indicateurs (niveau de qualification, disciplines et spécialités disponibles, expérience et expertise des employés) et, enfin, un point « conditions de travail » comptant cinq indicateurs (implémentation de nouvelles normes et de nouveaux équipements en lien avec la santé et la sécurité, perspectives de travail, taux de satisfaction des employés concernant les conditions de travail offertes, hauteur des investissements pour la formation, salaires).

Quelques années plus tard, la Commission européenne mit en place un groupe de 15 experts pour identifier les paramètres

avec lesquels la recherche menée au sein des universités européennes pourrait être évaluée. Ces experts, venus de 12 pays membres différents et d’Australie, remirent en 2010 un rapport dans lequel ils distinguèrent 4 catégories globales d’impacts sociétaux : bénéfiques économiques, sociaux, environnementaux et culturels, chacune caractérisée par des critères récapitulés dans le tableau 1 [18].

Si les critères cités dans ce rapport n’ont pas fait l’objet d’une généralisation, les grandes catégories d’impacts qu’il utilise sont souvent retrouvées dans les méthodes d’évaluation nationales ou les études portant sur l’impact sociétal de la recherche, avec toutefois quelques variations. L’impact environnemental est par exemple parfois inclus comme une part de l’impact social. Par ailleurs, si les impacts socio-économiques sont souvent les seuls considérés, il arrive également que les impacts technologiques, politiques et/ou concernant le domaine de la santé soient distingués en tant que catégories à part entière. Enfin, notons l’émergence de critères et de domaines d’impact jusque-là non abordés dont par exemple l’influence de la recherche sur le tourisme, les arts ou encore les droits de l’Homme, ainsi que l’importance accordée au transfert de connaissances aux différentes parties prenantes sociétales via par exemple une couverture médiatique, l’enseignement, la tenue de conférences ou la constitution de réseaux.

2.3 Questions épistémologiques et enjeux éthiques

Outre une difficile exhaustivité en termes de critères, la perception de ce qu’est l’impact sociétal et ce qui devrait être pris en compte dans sa définition est l’objet de nombreux débats résultant la plupart du temps d’une conception différente de la mission de la recherche et du rôle des chercheurs.

Tableau 1 – Catégories et critères d’impacts sociétaux distingués en 2010 par un groupe d’experts mandatés par la Commission européenne [18]

Catégories	Critères d’impacts
Bénéfices économiques	Amélioration de la productivité Ajouts à la croissance économique et la création de richesses Renforcement de la base de compétences Réduction des coûts Augmentation de l’emploi Hausse de la capacité d’innovation et de la compétitivité globale Amélioration de la livraison de services Retours économiques des ajustements des politiques sociales et publiques
Bénéfices sociaux	Amélioration de la santé et la qualité de vie Stimulation de nouvelles approches des questions sociales Changements dans les attitudes de la communauté Influence sur les développements ou les questions de la société Éclairage du débat public et de la prise de décisions Renforcement de la connaissance et la compréhension de la Nation Amélioration de l’égalité Amélioration dans les domaines de la santé, sûreté et sécurité Amélioration du lien social Amélioration du niveau et de la garantie des droits politiques
Bénéfices environnementaux	Amélioration concernant l’environnement et le style de vie Réduction des déchets et de la pollution Amélioration de la gestion des ressources naturelles Réduction de la consommation d’énergies fossiles Assimilation des techniques de recyclage Réduction du risque environnemental Conservation de la biodiversité Renforcement des services écologiques Amélioration des variétés végétales et animales Adaptation au changement climatique
Bénéfices culturels	Compréhension des origines et identités des nations et sociétés Compréhension des liens aux autres cultures et sociétés Stimulation de la créativité à l’intérieur de la communauté Contribution à la préservation culturelle et son enrichissement Apport d’idées et de nouveaux modes d’expériences à la Nation

2.3.1 L’impact sociétal : résultat mesurable ou processus qualitatif ?

Le lien entre la recherche et ses retombées sociétales a longtemps été considéré comme un tunnel de transfert simple et direct : une découverte scientifique apparaît, elle est aspirée vers des tests d’application et résulte finalement en divers produits sociétaux. Ce modèle repose sur l’idée que « puisqu’une connaissance existe, elle sera utilisée » [19]. L’impact sociétal est alors supposé avoir lieu linéairement via l’utilisation des résultats de recherche par diverses parties prenantes. Il s’agit donc ici d’un résultat final, mesurable, parfois défini comme « l’information qui provient de publications de recherche ayant un effet quantifiable en dehors du groupe de provenance de la publication » [19].

Néanmoins, face à cette idée se dresse celle soutenant que la réalisation d’impacts sociétaux résulte non d’un tunnel de transfert direct allant de la recherche vers la société mais d’un processus désordonné au sein duquel interviennent une multitude de facteurs externes incontrôlables, parfois considérés comme relevant du hasard, et parfois associés à une barrière créée par la société. Cette barrière sous-jacente peut être liée à la visibilité des recherches, à des effets de mode ou encore à la réceptivité des parties prenantes sociétales qui peuvent choisir de ne pas se saisir de certains résultats. Une société pharmaceutique peut

par exemple acheter un produit mais décider de ne pas le développer pour des raisons commerciales ou financières. De même, un gouvernement peut négliger certains rapports qui lui sont remis et les résultats qu’ils contiennent peuvent alors disparaître jusqu’à ce que les instances politiques décident de se faire une idée sur le sujet, finir par influencer la réglementation lorsque le poids des trouvailles s’accumulant devient trop important, voire ne jamais être utilisés [19]. L’évaluation des impacts sociétaux implique alors de considérer si et comment les facteurs freinant leur réalisation peuvent être dépassés, ou, plus généralement, si les conditions permettant cette réalisation sont réunies.

Plutôt que des impacts conçus comme des résultats quantifiables, il s’agit alors d’examiner l’ensemble du processus de recherche et les chemins d’impact incluant des paramètres dont les effets sur la société ne sont pas explicitement mesurables. Cet examen peut par exemple porter sur la contribution à des « biens communs » susceptibles de nourrir la recherche appliquée (bases de données, ressources génétiques ou biologiques...) ou sur des « échecs d’impact » envisagés comme une source de renseignements quant à la délimitation des domaines à explorer dans le futur ou la définition de schémas d’occurrence d’impact. Ces renseignements seront particulièrement utiles pour les évaluations ex ante demandées aux chercheurs comme préalable à l’attribution de financements contractuels.

2.3.2 Les interactions productives sont-elles un critère d'impact sociétal ?

L'existence d'interactions productives avec le monde extérieur à la recherche est quasi unanimement reconnue comme importante pour la concrétisation d'impacts sociétaux. Les résultats du programme européen SIAMPI ont par exemple conclu que des interactions « productives et hautement professionnelles entre les scientifiques et les autres parties prenantes de la société » engendrent généralement un impact sociétal. Selon les conclusions du projet néerlandais ERIC [20], ces interactions sont même nécessaires à la réalisation d'impacts. En effet, les scientifiques ne transfèrent pas le savoir qu'ils génèrent à la société eux-mêmes. L'impact sociétal aurait en réalité lieu selon un processus itératif entre les chercheurs et les autres parties prenantes, processus pouvant inclure, pendant ou après le travail de recherche, des contacts directs (réseaux, projets collaboratifs...), des interactions indirectes via des artefacts (expositions, logiciels, site internet...) ou encore des interactions financières (échanges économiques) [21]. Cependant, des divergences se manifestent lorsqu'il s'agit de déterminer si ces activités doivent être considérées comme un critère entrant dans la définition de l'impact sociétal.

Ces divergences résultent notamment de différences de conception quant au rôle du chercheur. En supposant un tunnel de transfert entre recherche et société, le rôle des chercheurs se limite à mener des travaux scientifiques d'excellente qualité et en disséminer les résultats convenablement. L'utilisation de ces résultats relève ensuite de la responsabilité d'autres parties prenantes. Si des interactions entre le monde de la recherche et ces différentes parties prenantes peuvent intervenir dans le développement des impacts sociétaux, elles n'en constituent pas pour autant un résultat et ne peuvent donc pas être considérées comme un critère dans leur évaluation. En revanche, lorsque l'impact sociétal est considéré comme découlant d'un processus gêné par diverses barrières devant être dépassées, il peut être envisagé qu'il incombe aux chercheurs de « pousser » leurs résultats afin que la société les absorbe. Cette valorisation de leurs travaux est notamment considérée comme passant par la construction de relations ou de partenariats avec les diverses parties prenantes sociétales mais aussi par des activités de recommandations de recherche, c'est-à-dire d'explication de la manière dont peuvent être utilisées les idées générées, voire une forme de lobbying [19]. Ces interactions devraient donc être reconnues comme des précurseurs d'impact sociétal et prises en compte comme un critère dans son évaluation. Il est par exemple envisagé d'estimer la communication avec les destinataires d'une recherche, le choix d'un sujet avec des partenaires appropriés, les conditions de collaboration, l'aide à l'appropriation des produits intermédiaires d'une recherche ou encore la force des interactions.

Notons toutefois que la « force » des interactions nouées par un chercheur avec différentes parties prenantes est souvent indépendante de sa réputation au sein du monde de la recherche et de l'impact scientifique de ses travaux selon des critères traditionnels. La question de la prise en compte de ces interactions dans l'évaluation de l'impact sociétal soulève donc le problème de la place de la qualité de la recherche.

2.3.3 L'impact sociétal d'une recherche est-il lié à sa qualité scientifique ?

La conception de la place qui doit être accordée à la qualité d'une recherche dans l'évaluation de ses impacts sociétaux joue un grand rôle dans la définition de ceux-ci.

Une interrogation récurrente consiste en effet à déterminer si la qualité de la recherche doit être considérée comme une composante sous-jacente nécessaire à l'impact sociétal. Certains affirment qu'elle constitue un critère essentiel, voire la condition sine qua non de la réalisation d'impacts sociétaux et qu'elle doit à ce titre en constituer la base même. Ce point de vue, qui s'accom-

pagne parfois de l'idée que toutes les recherches scientifiquement excellentes mènent à un impact sociétal à plus ou moins long terme, reflète une hypothèse morale sous-jacente : seule une excellente qualité de recherche contribuerait à un impact sociétal reproductible et auquel on peut se fier. Sans ce pré-requis, il n'y aurait rien à évaluer puisque la recherche n'offrirait pas de connaissances valables sur lesquelles se reposer et ne pourrait donc que se traduire par des impacts sociétaux « creux », « manquant de fiabilité », voire « négatifs » et « potentiellement dangereux » [19]. Ces effets potentiellement dangereux sont souvent illustrés par l'exemple de la polémique entourant depuis près de 20 ans le vaccin rougeole-oreillons-rubéole (ROR).

Polémique sur le vaccin rougeole-oreillons-rubéole

En 1998, un rapport publié dans un journal médical britannique avançait l'hypothèse d'un lien entre l'administration du vaccin ROR et l'apparition d'une maladie inflammatoire de l'intestin pouvant provoquer l'autisme. Entre autres failles méthodologiques, cette hypothèse n'était basée que sur l'étude de douze enfants et ne pouvait pas démontrer de réels liens de causalité. Relayée par la presse britannique, l'étude prit une ampleur très importante et créa une vaste controverse qui se traduisit par une baisse significative de la couverture vaccinale du ROR au Royaume-Uni. Dès 2000, un retour de la progression du nombre de cas de rougeole fut constaté alors que celui-ci était pourtant en régression constante depuis 1994-1995. La polémique s'est par ailleurs propagée partout dans le monde et persiste toujours dans de nombreux esprits, 20 ans après, malgré des dizaines d'études épidémiologiques et physiopathologiques n'ayant pas trouvé d'associations entre l'autisme et la vaccination ROR ainsi que la mise en avant de fraudes dans l'étude initiale. Non seulement les douze enfants ne présentaient pas de lésions digestives, mais les délais entre la vaccination et les troubles autistiques ne correspondaient pas à ceux rapportés par les parents. Une enquête d'investigation journalistique a par ailleurs révélé que le principal auteur de la publication avait été rétribué par un cabinet d'avocats avant la parution de l'article dans le but d'asseoir un lien scientifique entre la vaccination et une maladie et d'ainsi faciliter la mise en œuvre d'une plainte à grande échelle.

Cependant, parallèlement à ce courant de pensée, existe également l'hypothèse que la qualité scientifique et l'impact sociétal d'une recherche ne sont en réalité que faiblement corrélés. Une étude du Centre médical universitaire de Leiden [22] visant à développer des méthodes de mesures de l'impact sociétal au niveau d'un groupe de recherche arriva par exemple à la conclusion que, bien que la qualité scientifique prévale et soit un pré-requis dans le domaine médical, elle n'est pas nécessairement liée à la communication avec la société. Autrement dit, l'importance des retombées sociétales ne serait pas simplement la conséquence d'un travail scientifique de qualité. De même, en reprenant l'exemple de la polémique ayant entouré le vaccin ROR, il est évident que, bien que l'étude ayant déclenché la polémique ne remplisse pas les critères de qualité scientifique, elle a eu d'importantes retombées au sein de la société. Il apparaît donc qu'une recherche de « mauvaise qualité » peut avoir un important impact sociétal, point sur lequel s'appuie une autre approche de l'évaluation de ce dernier. Cette approche consiste à totalement distinguer d'une part un impact scientifique et de l'autre un impact sociétal. Considérés comme deux entités séparées et indépendantes, ils doivent alors être dotés de jeux de critères différents. La qualité scientifique, jugée selon des critères traditionnels, intervient seulement dans l'évaluation de l'impact sociétal pour s'assurer que la recherche est « assez bonne pour être prise au sérieux » [19]. En supposant que le minimum de

qualité requise est atteint, l'impact sociétal résultant est alors considéré comme absolu, c'est-à-dire indépendant de toute classification morale en « bons » ou « mauvais » impacts.

2.3.4 Est-il de la responsabilité de la recherche de fournir de « bons » impacts ?

La question de l'attribution ou non d'une « valeur » aux retombées sociétales de la recherche est loin de faire l'unanimité. L'idée de récompenser tous les impacts sans tenir compte de leurs effets sur la société s'oppose vivement à celle de ne récompenser que les impacts « positifs » permettant des « avancées ». Cette notion de « récompense », probablement par des financements, sous-entend par ailleurs que les chercheurs portent la responsabilité morale des retombées de leurs recherches.

Tout au long de l'histoire, des membres du monde de la recherche se sont interrogés : la mesure des effets sociétaux de leur travail, l'évaluation de ses risques voire l'alerte sur ceux-ci relèvent-elles de leur responsabilité ? Si certains considèrent que les différentes parties prenantes se saisissant des résultats de la recherche sont responsables de l'utilisation qu'ils en font, l'économiste et sociologue M. Weber affirma quant à lui en 1919 qu'il revenait à la science, alors vue comme « axiologiquement neutre », c'est-à-dire libre de jugement de valeur, d'évaluer son impact possible, y compris les « effets pervers » de ses interventions dans les différents domaines du réel [23]. Dans cet esprit, aux premières heures du génie génétique, ce sont des scientifiques qui, conscients des enjeux liés à cette nouvelle technologie, ont appelé à la mise en place d'un moratoire lors de la conférence d'Asilomar de 1975, le temps d'évaluer les risques et d'encadrer les pratiques de laboratoire. Cette responsabilité morale a par ailleurs été mise en avant dès 1998 par les autorités françaises via une décision du Comité interministériel de la recherche scientifique et technologique proposant que les organismes de recherche, et notamment ceux ayant une mission d'expertise dans les domaines de santé et d'environnement, puissent « mener une réflexion relative aux conséquences des recherches qu'ils conduisent, aussi bien en termes de risque potentiel que d'acceptabilité pour la société ». Cette proposition s'est traduite par la mise en place de comités d'éthique dans la plupart des établissements de recherche français et par le rappel que les établissements de recherche ont aussi une responsabilité en matière d'information du public et de vulgarisation scientifique et technique.

Cependant, l'évolution actuelle de la science, avec le rapide développement de ses applications technologiques, pose désormais des problèmes éthiques inédits qui dépassent le champ d'application des méthodologies savantes et incluent bien plus que la maîtrise technique des conséquences de la recherche. Dans sa contribution à la réflexion menée lors de la 21^e conférence sur les changements climatiques, le Comité consultatif national d'éthique soulignait ainsi que « questionner les usages des applications de la recherche et plus largement de l'ingénierie liée aux applications des recherches est devenu un enjeu éthique majeur. Ne pas considérer toute avancée technique comme un progrès est une nécessité absolue lorsque certaines d'entre elles contribuent à dégrader la qualité de vie et de la santé d'une partie de l'humanité et à creuser les inégalités ». Cette idée est également soutenue par le Comité consultatif d'éthique commun Inra-Cirad qui rappelait déjà en 2010 que « l'importance de l'efficacité ne peut faire oublier celle des valeurs ». En 2016, ce comité a réaffirmé son avis que « les aboutissements techniques des nouveaux savoirs doivent faire l'objet d'une évaluation sociale et éthique, et pas seulement technique et économique, afin qu'un jugement soit porté sur le bien-fondé de leur usage » [16]. Souscrivant donc à l'idée d'un impact sociétal jugé selon sa valeur et soulignant que « dans un paysage de la recherche où les réseaux se complexifient, les questions de responsabilité deviennent diffuses », il avance qu'une évaluation basée sur ce critère peut constituer une

« prise en compte collective de l'intérêt général au regard des intérêts particuliers », même s'il reconnaît que la définition de l'intérêt général reste sujette à beaucoup d'interprétations.

Pour le Comité consultatif d'éthique Inra-Cirad, l'estimation de ses impacts réels par la recherche est « l'une des expressions de sa responsabilité sociale et environnementale » et la capacité de mesurer toutes les retombées renforce une exigence croissante de « responsabilité morale » [16]. Face à cette idée, de nombreuses méthodologies telles que celles de SIAMPI ou des projets ASIRPA [24] de l'Inra et ImpresS du Cirad considèrent toutefois que l'évaluation de la recherche devrait principalement viser l'apprentissage plutôt que la responsabilisation et se focaliser sur l'identification des processus pouvant conduire à un impact sociétal.

3. Méthodologies d'évaluation de l'impact sociétal

En raison de son spectre extrêmement large et des interrogations continuant à peser sur sa définition, l'impact sociétal de la recherche est beaucoup plus délicat à évaluer que son impact scientifique. À l'échelle d'un projet, d'un institut ou d'un chercheur, cette entreprise est rendue encore plus ardue par des questions d'attribution, de causalité et de temporalité qui compliquent l'accès à l'information complète sur les retombées sociétales. Malgré ces difficultés, de très nombreuses méthodologies ont été développées au fil des années et certaines d'entre elles sont même actuellement mises en place au niveau national.

3.1 Méthodes d'évaluation ex ante : planifier la recherche pour en optimiser l'impact

3.1.1 Exemples d'outils développés en Australie

De nombreux programmes de financement demandent traditionnellement aux chercheurs d'identifier les impacts sociétaux qui résulteraient de leur recherche comme préalable à l'allocation de ressources. C'est par exemple le cas du Cooperative Research Centre (CRC) du gouvernement australien qui, depuis 1992, a rendu obligatoire pour tous les postulants à ses bourses de modéliser en termes économiques l'impact net des recherches qu'ils proposent. Cette modélisation s'appuie sur un outil couvrant 7 domaines principaux : les contributions d'entrées, les activités, les résultats, l'utilisation, les impacts, l'analyse de risques et le rapport net/bénéfices. Cet outil a cependant spécifiquement été conçu pour évaluer l'impact économique de la recherche de larges programmes ou thématiques.

En 2016, une équipe du Lowitja Institute entreprit donc de construire un outil de planification de la recherche permettant un alignement entre les critères via lesquels les agences de financement de la recherche (CRC ou autres) priorisent les recherches et d'autres aspects essentiels tels que la satisfaction des utilisateurs finaux des résultats scientifiques. Développé dans le contexte de la recherche sur les autochtones et les insulaires du détroit de Torres via la consultation de nombreux chercheurs et destinataires potentiels de la recherche, cet outil décrit un jeu de six étapes pouvant guider les chercheurs dans la gestion de leur recherche afin qu'elle génère des retombées sociétales [25] :

1/ Définir clairement en amont les utilisateurs de la recherche et les informations dont ils ont besoin. Cette définition doit être étayée par des preuves et une explicitation de la manière dont il est envisagé d'utiliser les informations produites.

2/ Vérifier l'état des preuves existantes. Il s'agit ici d'effectuer un état des lieux des informations déjà disponibles, notamment via une recherche bibliographique complète, et de peser aussi objectivement et honnêtement que possible d'une part les avantages et inconvénients liés à l'utilisation d'informations additionnelles au lieu des connaissances déjà existantes pour informer les décideurs, et d'autre part les autres moyens d'obtenir des résultats similaires.

3/ Sélectionner et implémenter le type et la conception appropriés pour la recherche en s'assurant qu'ils sont tous deux rigoureux et adaptés aux objectifs définis en amont.

4/ Évaluer les coûts et bénéfices de la recherche. Les chercheurs sont ici invités à détailler pas à pas l'ensemble des coûts et bénéfices prospectifs de la recherche, en fournissant des données empiriques qualitatives et quantitatives pertinentes.

5/ Transférer les connaissances. Il s'agit ici, en se basant sur des méthodes dont l'efficacité est vérifiée, de développer l'implémentation du projet et les plans de transfert de connaissances afin d'influencer des décisions et actions au-delà du projet.

6/ Évaluer le rapport bénéfices/coûts au-delà du projet. Dans cette dernière étape, les chercheurs utilisent des processus et indicateurs d'impact pour contrôler et évaluer systématiquement ce rapport à chaque pas du processus.

Selon ses auteurs, cet outil de planification de l'impact pourrait être généralisé à tous les domaines scientifiques, tant appliqués que fondamentaux.

3.1.2 L'approche ex ante du projet ImpresS (Cirad)

Le groupe méthodologique « Impact of Research in the South » (ImpresS) du Cirad est actuellement en train de développer une approche systématique pour construire des chemins de l'impact (cf. lexique) ex ante participatifs au moment de la conception et avant la mise en œuvre « d'interventions », terme générique pouvant désigner aussi bien un projet qu'un programme ou encore des « grappes de projets » [26]. La construction du chemin de l'impact se décompose en six étapes :

1/ Construire le récit. Il s'agit de construire un premier narratif de l'innovation incluant l'ensemble de ses acteurs majeurs et les grandes phases de son histoire. Cette étape implique donc de formuler l'hypothèse initiale des impacts, définir les modes d'intervention de la recherche et identifier les problèmes, les solutions, les opportunités, les acteurs majeurs ou encore les potentiels impactés. Constituant le tout premier pas de la démarche, cette étape est basée sur un diagnostic initial du contexte (dont l'étude des projets préexistants) et une analyse fine des acteurs qui doivent ensuite être réitérés après la réalisation de chacune des étapes suivantes.

2/ Cartographier les résultats. Cette étape comprend la définition du rôle des acteurs dans la production des résultats et l'identification des changements clés souhaitables pour générer des résultats, des obstacles pouvant s'y opposer et des stratégies pour les surmonter. À partir de ces stratégies, le type d'intervention (analyse, expertise, service...), la méthode choisie et le degré de participation des acteurs dans sa mise en œuvre peuvent ensuite être élaborés.

3/ Prendre en compte les politiques publiques. Déjà abordé dans les étapes précédentes, ce point fait l'objet d'une étape dédiée en raison de son importance. Il s'agit de se pencher spécifiquement sur l'identification des acteurs de l'innovation dans le domaine public, des résultats qui seront liés à ces acteurs ou encore du moment et des modalités de leur intervention dans le processus en construction.

4/ Viser le renforcement des capacités. Cette étape consiste à repérer les besoins de renforcement en capacités diverses (techniques, managériales, à expérimenter, à apprendre, à interagir...) des individus ou organisations acteurs dans le chemin d'impact.

5/ Finaliser le chemin de l'impact et imaginer des chemins alternatifs. Le chemin d'impact principal est dessiné et les liens de causa-

lité explicités via l'identification des facteurs contextuels influents et l'imagination de chemins alternatifs.

6/ Suivi-évaluation participatif et apprentissage. Encore en construction, cette étape envisage l'élaboration d'indicateurs de suivi et finaux dans le cadre d'un système global d'évaluation et d'apprentissage mené en collaboration avec les partenaires de l'intervention considérée.

Lexique des termes utilisés dans l'approche ImpresS

Chemin de l'impact : description d'un processus d'innovation mettant en évidence des relations de causalité entre les moyens de la recherche, ses produits, ses résultats se matérialisant directement au niveau des utilisateurs des produits et ses impacts dits de 1^{er} et 2^e niveaux.

Moyens de la recherche : ensemble des interventions et des ressources mobilisées avant ou pendant la recherche qui permettent de la mener : ressources humaines et matérielles, budget de recherche, informations, connaissances générées antérieurement ou tacites, autres activités de recherche...

Produit de la recherche : production issue de la recherche ou des interactions avec des organisations ou individus extérieurs pouvant contribuer à l'émergence d'innovations lorsqu'elle fait l'objet d'une appropriation par des acteurs de la société : connaissances (scientifiques ou non), formations, expertises, technologies...

Résultat : appropriation d'un produit par des acteurs interagissant directement ou indirectement avec la recherche qui conduit à des changements de pratiques, organisationnels, d'interactions ou de règles.

Impacts : effets à long terme, positifs ou négatifs, intentionnels ou non, directs ou indirects, mesurés par des indicateurs.

Ceux de 1^{er} niveau sont mesurés sur et avec les acteurs interagissant directement ou indirectement avec la recherche et/ou les acteurs majeurs de l'innovation selon deux critères : l'intensité du changement et l'ampleur du changement.

Ceux de 2^e niveau résultent des effets d'impacts indirects ou de changement d'échelle horizontaux et/ou verticaux. Leur évaluation se base sur les mêmes critères que les impacts de 1^{er} niveau mais de manière plus approximative ou moins détaillée.

3.2 Méthodes d'évaluation ex post : vérifier l'atteinte d'impacts sociétaux

3.2.1 Le Payback Framework : un des précurseurs

Présenté au milieu des années 90, le Payback Framework a initialement été développé pour la recherche et le développement dans le domaine de la santé [27]. Il s'agit d'un outil compilant des données collectées via des enquêtes réalisées auprès de décideurs, et l'analyse de documents sur l'impact de la recherche d'une unité. Facilitant les analyses comparatives de cas croisés en fournissant une structure commune à chaque étude de cas, cet outil peut être utilisé pour dépeindre l'ensemble du processus de recherche. Il rend pour cela disponible une catégorisation multidimensionnelle des bénéfices de la recherche sur cinq points : la connaissance, les futures recherches et l'utilisation de la recherche, les domaines politiques et liés au développement de produits, la santé et le secteur de la santé, et, enfin, l'économie au sens large. Bien qu'ayant connu plusieurs développements et révisions au cours des années, son modèle a conservé sa structure et ses éléments originels qui peuvent être décrites selon la figure 2 : sept étapes et deux interfaces entre le système de

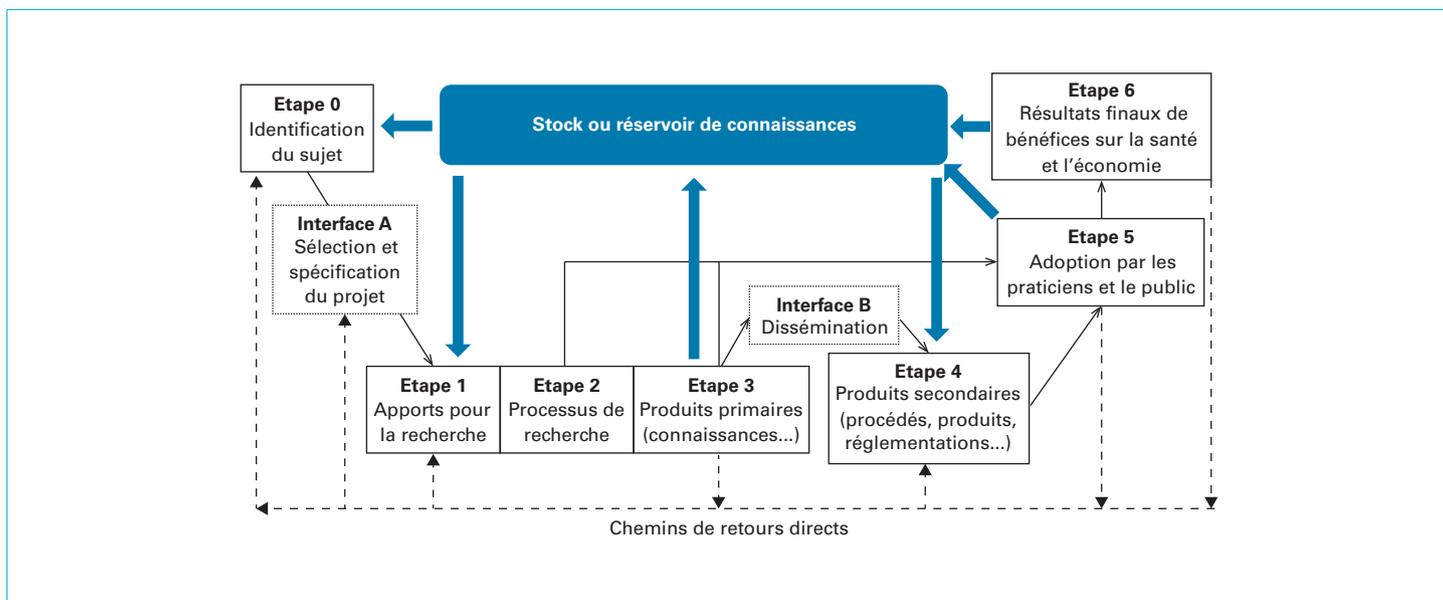


Figure 2 – Schématisation du Payback Framework (adapté de [28]).

recherche et son environnement politique et économique au sens large.

Malgré les grands efforts qu'il requiert, le Payback Framework a été appliqué dans de nombreux contextes à la fois dans des services de santé et au-delà. Il a par exemple été utilisé dans le cadre des sciences sociales [29] et a constitué la base de nombreuses recherches ultérieures. Une équipe australienne a notamment développé une forme modifiée de cet outil nommé « Health Service Research Impact Framework » (HSRIF). Conçue afin de mieux saisir l'impact de la recherche fondamentale dans le domaine de la santé des contextes ruraux, cette nouvelle approche suggère cette fois une répartition des preuves d'impact en six domaines clés : la production de savoir, le ciblage des recherches et la construction de capacités, l'information des décideurs et le développement de produits, les bénéfices sur la santé et le secteur de la santé, le transfert de la recherche et, enfin, les bénéfices plus larges sur l'économie.

3.2.2 ImpresS ex post : méthode construite et appliquée avec les parties prenantes

Avant d'être élargie à la planification ex ante des impacts de la recherche, la méthode ImpresS avait initialement été développée afin de fournir un cadre à leur évaluation ex post dans le contexte spécifique de l'étude des retombées de la recherche agronomique dans les pays du Sud [30]. Basée sur une recherche bibliographique, des réflexions conduites dans d'autres organisations (CGIAR, FAO, WUR, Inra), des travaux réalisés par des groupes de travail internes du Cirad, les résultats de deux ateliers méthodologiques et des retours d'évaluation, l'approche d'ImpresS repose sur l'analyse d'études de cas et consiste en une construction itérative et participative des évaluations. Elle utilise deux outils, le récit de l'innovation et le chemin d'impact, et se décompose en cinq phases : préparation, confrontation, construction, mesure et validation.

■ La phase de préparation consiste à délimiter le périmètre spatio-temporel de l'étude de cas et à identifier d'une part les différents acteurs et partenaires ayant participé, influencé ou été impactés par le processus d'innovation, et d'autre part les projets de recherche et de développement qui participent ou ont participé à l'innovation. Mobilisant notamment la bibliographie, les docu-

ments du projet et les connaissances des chercheurs, elle se traduit par la formulation d'hypothèses d'impact préliminaires et des premières versions du récit de l'innovation et du chemin de l'impact.

■ Durant la deuxième phase, les hypothèses formulées sont confrontées à la perception des acteurs associés au processus d'innovation via un premier atelier participatif et/ou des entretiens individuels. Elles sont affinées en précisant le récit de l'innovation et la nature des impacts puis une liste de descripteurs d'impact est établie de manière participative. En pratique, les acteurs évoquent les changements (positifs ou négatifs) qu'ils ont identifiés chez eux ou chez d'autres acteurs, et précisent la manière de collecter de l'information pertinente pour caractériser ces changements.

■ Intervient ensuite une phase de construction du récit de l'innovation et du chemin de l'impact. Les moyens, produits et résultats (voir lexique) sont documentés de manière systématique en accordant une attention particulière aux situations d'apprentissage et aux interactions avec les politiques publiques.

■ Une quatrième phase consiste à caractériser et quantifier les impacts de 1^{er} et 2^e niveau (voir lexique). La caractérisation porte à la fois sur leur intensité (le degré de changement) et leur amplitude (leur distribution), ce qui implique de rassembler les données via des entretiens, des enquêtes ou des focus group mais aussi en mobilisant des sources secondaires puis de trianguler les informations obtenues des différentes sources pour renseigner les éléments du chemin d'impact et analyser les liens de causalité entre la recherche et ces impacts.

■ Enfin, la cinquième et dernière phase permet de conclure l'étude via la validation des résultats de l'étape précédente par les acteurs du processus d'innovation. Ces résultats sont présentés aux acteurs et discutés durant un second atelier participatif à l'issue duquel sont entérinés le récit de l'innovation, le chemin de l'impact et la quantification de ceux-ci. Ils alimentent ensuite une base de données générique de manière à harmoniser et archiver les informations collectées dans chacune des études de cas, permettre une analyse transversale de celles-ci et faciliter la construction de démarches ex ante.

3.2.3 L'étude de cas semi-standardisée du projet ASIRPA (Inra)

Depuis 2011, l'Inra a mis en place un programme d'« Analyse des impacts de la recherche publique agronomique », plus connu sous le nom d'ASIRPA [24]. Comme la méthode ImpresS ex post du Cirad, ce programme vise un double objectif : évaluer et caractériser les impacts découlant de la recherche effectuée au sein de l'institut afin de justifier les investissements effectués, mais aussi améliorer la compréhension des mécanismes à l'origine de ces impacts pour optimiser ce processus. Néanmoins, si la position du Cirad est d'associer d'une part des partenaires à la co-construction de la méthode en amont et d'autre part les acteurs de l'innovation aux différentes phases de l'évaluation, celle de l'Inra est de s'appuyer sur le regard externe des pairs.

Dans le cadre de l'approche ASIRPA, l'impact sociétal peut être défini comme « une influence contrôlable et enregistrable sur une ou plusieurs organisations non académiques ou sur un ou plusieurs acteurs appartenant à un secteur extérieur au secteur universitaire lui-même ». Cela inclut par exemple des corporations d'affaires, des corps gouvernementaux, des organisations de la société civile, des organisations médiatiques ou encore des professionnels intervenant dans les médias ou le débat public [24]. Basée sur une méthode d'études de cas standardisées combinant analyse qualitative et quantitative, cette approche mobilise trois outils complémentaires :

- une chronologie qui identifie les principaux événements de l'étude de cas ;

- un chemin d'impact qui représente graphiquement les étapes et acteurs de la génération des impacts à partir des connaissances scientifiques produites. Il décrit les caractéristiques de la recherche, ses produits primaires, ses impacts de 1^{er} niveau, la généralisation de ces impacts au-delà du premier cercle de bénéficiaires (2^e niveau) mais aussi les intermédiaires impliqués à chaque étape du chemin d'impact ;

- des vecteurs d'impact synthétisés sous la forme d'un tableau décrivant la nature et l'intensité, sur une échelle de 5, des impacts générés sur cinq dimensions : les dimensions économique, environnementale, politique, sanitaire et socio-territoriale.

Si la notation sera effectuée par des pairs extérieurs, ce sont les équipes de chercheurs, d'ingénieurs et de techniciens directement concernées par l'étude de cas qui sont responsables de la reconstitution historique du chemin d'impact et de la collecte des données d'impact auprès des bénéficiaires finaux extérieurs à l'institut. Afin de rendre compte des retombées inattendues de la recherche, les impacts sont systématiquement explorés selon les cinq dimensions d'impact pour chaque étude de cas, quels qu'aient été les objectifs initiaux de la recherche. Cette exploration s'effectue à l'aide de barèmes communs construits via une combinaison de bibliographie, de descripteurs fournis par les utilisateurs et bénéficiaires finaux, et de discussions de panels d'experts. Chaque barème inclut un jeu de catégories subdivisées en sous-catégories pour lesquelles un certain nombre de critères sont définis. À titre d'exemple, les catégories, sous-catégories et critères proposés par le barème de la dimension politique des impacts sont présentés dans le tableau 2.

De manière à capitaliser sur les données d'impact disponibles et pertinentes pour les bénéficiaires eux-mêmes, les indicateurs étayant ces critères sont en revanche choisis au cas par cas pour chaque dimension, via les chemins d'impact et l'interrogation des parties prenantes. Un certain nombre d'arguments est fourni pour chaque sous-catégorie et la notation est ensuite réalisée par des tiers sur la base de ces informations. Le tableau 3 récapitule par exemple les arguments fournis pour la catégorie « Utilisation dans le débat public et la formulation de politiques » lors de l'évaluation de la dimension politique de l'impact des recherches de l'Inra ayant montré que le bisphénol A (BPA), perturbateur endocrinien utilisé dans la fabrication de plastiques et résines, peut s'extraire

de ces matériaux et traverser les barrières intestinales, cutanées et hépatiques. Cette catégorie a obtenu la note de 4 sur 5.

Notons que les barèmes construits dans le cadre du programme ASIRPA n'ont aucunement vocation à éclairer l'allocation de fonds dans la mesure, notamment, où le voyage entre l'initiation de la recherche et sa traduction en un impact socio-économique peut prendre plusieurs décennies.

3.3 Systèmes d'évaluation nationaux des établissements de recherche

3.3.1 Exemple du Research Excellence Framework du Royaume-Uni

Remplaçant le « Research Assessment Exercise » utilisé au Royaume-Uni depuis les années 80, le « Research of Excellence Framework » (REF) a constitué à partir de 2011 une première tentative d'établissement d'un système permettant à la fois l'allocation sélective de financements aux structures d'enseignement supérieur sur la base d'impacts non académiques, le comptage de l'investissement public dans la recherche via la production de preuves des bénéfices de cet investissement et le référencement d'informations (réputation...) concernant le secteur de l'enseignement supérieur. En plus d'une revue des résultats de recherche par les pairs (comptant pour 65 % de l'évaluation) et d'une appréciation de l'environnement de l'institut d'enseignement supérieur (15 % de l'évaluation), le REF inclut un critère d'impact sociétal défini comme « un effet, changement ou bénéfice sur l'économie, la société, la culture, la réglementation publique ou les services, la santé, l'environnement ou la qualité de vie au-delà de l'académie » [19].

Pour évaluer ce troisième critère, chaque unité (universités, groupes de recherche...) fut invitée à remettre un dossier incluant deux éléments. Le premier, un élément narratif (ou « template » d'impact), décrit l'approche et la stratégie générale de la structure pour obtenir un impact à partir de la recherche ou faciliter la transition de la recherche en impacts. Le second était constitué d'un certain nombre d'études de cas illustrant des exemples d'impact réalisés durant la période évaluée et atteignant au moins une note de 2 sur 5 en terme de qualité. Le nombre d'études de cas à fournir dépendait de la taille de l'unité (en moyenne une étude de cas pour 10 chercheurs mais le ratio pouvait être plus élevé pour les petites structures), et celles-ci étaient standardisées selon un schéma en 5 points : résumé de l'impact, description de la recherche étayant l'impact, références de la recherche, détails de l'impact et sources corroborant l'impact.

Les dossiers remis par les structures d'enseignement et de recherche publique furent évalués par des panels d'experts selon deux critères : l'importance, définie comme l'intensité de l'influence ou de l'effet, et la portée, décrite comme la propagation ou l'étendue de l'influence ou de l'effet sur les circonscriptions pertinentes. Il est important de souligner que les panels d'experts ont inclus aussi bien des experts internationaux ou nationaux et des évaluateurs académiques traditionnels que des évaluateurs représentant la communauté des utilisateurs de recherches venus de toute une gamme de secteurs publics ou privés (monde des affaires, journalisme, secteur tertiaire, société civile...). L'évaluation a couvert l'ensemble des activités scientifiques réparties en 4 catégories (sciences de la santé et biomédicales ; sciences de la vie et de la terre, ingénierie, sciences physiques et mathématiques ; sciences sociales ; sciences humaines et les arts), elles-mêmes subdivisées en 6 à 10 sous-panels. Ses résultats furent publiés en 2014 et une base de données d'études de cas rendue disponible en 2015.

Les retours sur cette expérience, considérée comme un succès, soulignèrent les avantages liés à l'élargissement des critères de l'impact sociétal et à l'ouverture du réseau d'évaluation à diverses

Tableau 2 – Catégories, sous-catégories et critères du barème de l’approche ASIRPA pour l’évaluation de la dimension politique des impacts de la recherche [31]

Sous-catégorie	Critères
Utilisation dans le débat public et la formulation de politiques	
Qualité et puissance des messages relayés	Excellence et nouveauté Adaptation aux audiences visées Originalité Traçabilité Robustesse Crédibilité Légitimité
Concomitance du débat et des agendas politiques	Production ou médiatisation des connaissances dans une fenêtre d’opportunité politique Modification des agendas politiques Portée de la fenêtre de politique
Intensité et qualité de la médiatisation	Volume, cible, diversité, pertinence de la couverture médiatique Intégrité et sélectivité des messages relayés
Intensité et qualité du débat	Volume du débat Diversité et pertinence des parties prenantes du débat
Utilisation dans le cycle politique	
Étapes du cycle affectées	Nombre d’étapes du cycle impactées Ampleur des impacts à chaque étape
Échelles territoriales des impacts	Exhaustivité et homogénéité des échelles pertinentes affectées Intensité des impacts sur les différentes échelles
Pertinence et nouveauté de la solution politique conçue	Nouveauté et originalité de la solution politique Rapidité et niveau d’adoption Niveau d’incorporation des connaissances scientifiques dans la solution
Percolation des idées sur le long terme	
Importance de la connaissance dans le débat	Nature et pluralité des effets de la connaissance : évolution des termes, des positions, des tensions Nature du débat (public/interne, ponctuel/croissant)
Percolation des idées dans les arènes distantes et les futurs débats et études	Vitesse et durabilité de la percolation Importance et exhaustivité des arènes influencées
Pertinence et intégrité des idées à long terme	Résilience des idées sur le long terme Facteurs protégeant l’intégrité des idées : renommée des chercheurs/institut, continuité de l’implication, traçabilité des idées
Importance sociétale du domaine politique considéré	
Enjeux : gravité potentielle et caractère systémique	Multiplicité et imbrication des enjeux Gravité
Ampleur de la population et politique affectée	Taille et nature de la population affectée Échelle du domaine politique
Émotion publique	Ampleur de l’émotion publique et politique Occurrence de crises Existence d’un consensus social sur le problème et/ou la solution Régularité et ampleur de la médiatisation

parties prenantes. Un rapport commissionné par le gouvernement du Royaume-Uni conclut par exemple qu’il avait permis de « fournir une image riche de la variété et de la qualité de la contribution de la recherche à l’économie et à la société du Royaume-Uni »

[32]. L’analyse de l’ensemble des études de cas montre de plus que chaque catégorie disciplinaire évaluée affiche des contributions à tous les types d’impacts : qu’ils soient technologiques, sociaux, politiques, légaux, sanitaires, environnementaux, écono-

Tableau 3 – Arguments fournis pour la catégorie « Utilisation dans le débat public et la formulation de politiques » de la dimension politique dans l'évaluation de l'impact des recherches de l'Inra sur le bisphénol A [31]

Qualité et puissance des messages relayés
<p>Les résultats de recherche mettent à l'agenda politique de nouvelles questions : alerte concernant le passage du BPA par la barrière cutanée.</p> <p>Les résultats de recherche de l'Inra sur les barrières cutanées, hépatiques et digestives sont explicitement mentionnés dans les médias et débats.</p> <p>La faiblesse de certains messages scientifiques affecte leur crédibilité (l'ANSES juge nécessaire la réplication des travaux) mais les résultats servent de caution scientifique : leur statut scientifique est utilisé par les députés pour interroger les méthodes d'évaluation des agences.</p> <p>L'INSERM est le principal référent d'expertise pour le ministère de la Santé, ce qui affecte la légitimité de l'Inra.</p>
Concomitance du débat et des agendas politiques
<p>Les résultats de recherche sont délivrés et médiatisés dans une fenêtre d'opportunité politique nationale et européenne : les résultats de l'Inra sont publiés en 2010, alors que la première évaluation des risques par l'EFSA a été réalisée en 2006 et qu'une série de saisines de l'ANSES est en cours (2009-2010).</p> <p>Ces résultats contribuent au déclenchement d'une commande d'une expertise collective commandée à l'INSERM sur la Reproduction et l'Environnement (en 2010).</p>
Intensité et qualité de la médiatisation
<p>Forte couverture médiatique (150 articles dans l'année de publication des résultats) à destination d'un public très large : professionnel, grand public et politique.</p> <p>Faible distorsion des messages, malgré une confusion entre les notions de danger et de risque.</p>
Intensité et qualité du débat
<p>Forte contribution des résultats de recherche aux débats au-delà de la sphère sectorielle, au niveau national et européen.</p> <p>Prise de positions des parties prenantes principales : françaises (parlement, ministères, ANSES, ONG dont Réseau Environnement Santé, secteur privé dont PlasticEurope) ou européennes.</p> <p>Renforcement des positions des acteurs souhaitant une révision des modalités d'évaluation des perturbateurs endocriniens.</p>

miques ou culturels. Plusieurs recommandations visant à améliorer le processus furent proposées dont notamment un nouvel élargissement de la définition d'impact et la possibilité de soumettre des études de cas au niveau institutionnel pour encourager et mettre en avant le travail multi et interdisciplinaire. L'importance de cette interdisciplinarité fut soulignée dans de nombreux rapports. S. Hill, responsable de l'évaluation du REF au Higher Education Funding Council for England de Bristol, nota par exemple des preuves de diversité disciplinaire dans près de 90 % des études de cas et souligna que, deux fois sur trois, les disciplines impliquées étaient raisonnablement différentes les unes des autres [10]. En concluant que la livraison d'impacts semble fortement liée à l'habileté des chercheurs à travailler au travers des frontières disciplinaires, il suggère donc par exemple de changer le niveau d'évaluation pour opter pour des domaines disciplinaires plus larges ou même un niveau institutionnel. Entre autres réflexions, S. Hill remarqua également une sous-représentation des recherches impliquant des collaborations dans les études de cas soumises à l'évaluation. Ces remarques ainsi que d'autres préconisations venues de nombreuses sources nourrirent une réflexion quant aux modalités du futur REF2021, modalités proposées et soumises à consultation jusqu'en octobre 2018.

3.3.2 L'évaluation des établissements publics de recherche en France

Jusqu'en 2011, aucune mention d'une catégorie appréciant « l'impact » des activités de recherche conduites par les établissements publics français ne pouvait être recensée. Valables pour tous les domaines scientifiques, les critères de l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES) concernaient quatre points centraux : « qualité scientifique et pro-

duction », « rayonnement et attractivité », « gouvernance et vie de l'unité » et « projet global ». Une réflexion quant à la redéfinition de ces critères fut cependant engagée suite à la demande de certaines parties prenantes de cette agence [33]. Portant notamment sur la prise en compte des activités conduites en partenariat avec des acteurs non académiques, cette demande traduisait « une insatisfaction quant aux critères existants [ne tenant pas suffisamment] compte de l'ensemble des spécificités des différents domaines disciplinaires ». Rappelant que les sciences humaines et sociales participent à la construction des politiques publiques, à la culture et à la société au sens large, les représentants de ces disciplines estimaient en particulier qu'une ouverture des missions de la recherche vers les partenaires économiques et sociaux leur conférerait « une place et une légitimité comparables à celles d'autres domaines scientifiques » déjà reconnus pour les avancées sociétales et économiques majeures auxquelles ils ont contribué.

Les réflexions quant à une redéfinition des critères d'évaluation pour y introduire les impacts sociétaux potentiels de la recherche mirent en avant un certain nombre d'obstacles. Il fut par exemple souligné que la nature imprévisible des résultats ou encore le fréquent décalage temporel entre les découvertes scientifiques et les attentes de la société s'opposent à une mesure formelle de ces impacts potentiels. Plutôt que de se focaliser sur les « effets », difficilement quantifiables, de la recherche sur la société, l'attention se reporta donc sur les relations et les partenariats entre science et société. Cela se traduisit par l'introduction en 2013 d'un nouveau critère d'évaluation portant sur « les interactions avec l'environnement social, économique et culturel », critère associé à une évolution des missions de l'Enseignement supérieur et de la Recherche.

Extrait de « Critères d'évaluation des entités de recherche : le référentiel de l'AERES » (version du 21 février 2013)

Critère 3 : Interactions avec l'environnement social, économique et culturel.

Champ d'évaluation : les différentes activités et réalisations par lesquelles la recherche contribue aux processus d'innovation et conduit à des impacts sur l'économie, la société ou la culture [...].

Faits observables : activités orientées vers des acteurs sociaux qui n'appartiennent pas au monde de la recherche. Ils dépendent de la nature et la finalité des activités développées par les entités de recherche et incluent :

- les produits destinés à différents acteurs non académiques (articles, rapports d'étude et d'expertise, logiciels, brevets...);
- l'engagement dans les relations partenariales et tout élément mettant en évidence l'intérêt et l'engagement des partenaires non académiques, ainsi que la visibilité de l'entité de recherche dans le domaine socio-économique ou culturel (exemple : la mise en place de structures d'aide au transfert technologique, la collaboration avec des institutions culturelles, les contrats obtenus avec des partenaires non académiques);
- des indications éventuelles sur l'impact des recherches et des collaborations partenariales, comme par exemple la création d'entreprises [...], des innovations (nouveaux produits, techniques et processus [...]), des effets sur la santé publique, sur l'environnement, sur l'aménagement du territoire, sur la législation, sur le débat public...

L'introduction de ce nouveau critère ouvrit la porte à de nombreuses modifications ultérieures de l'évaluation de la recherche publique en France, assurée depuis 2014 par le Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES). Depuis la campagne d'évaluation 2017-2018 (vague géographique D), elle s'est par exemple recentrée sur les activités proprement scientifiques (bilan et projet) et leurs interactions avec l'environnement économique, social, culturel et/ou sanitaire. Un nouveau référentiel a été établi, mettant prioritairement l'accent sur les activités et les résultats déjà obtenus ainsi que sur les activités projetées et leur stratégie de mise en œuvre.

Les critères d'évaluation pour les unités de recherche y ont été réduits à trois points principaux :

- organisation et vie de l'unité ;
- qualité des activités et des produits de la recherche décomposés en trois ensembles : ceux liés à la production de connaissances et concourant au rayonnement et à l'attractivité scientifique, ceux issus de la valorisation, du transfert et des interactions avec l'environnement, et ceux en relation avec la formation par la recherche ;
- qualité et faisabilité du projet scientifique à cinq ans incluant la stratégie au service de la politique scientifique et la stratégie scientifique.

Chaque critère doit s'appuyer sur des faits observables et en apprécier la valeur. Seront par exemple pris en compte les partenariats dans la recherche et l'enseignement supérieur ou avec le monde socio-économique et culturel, le développement des compétences (formation, mobilité, recrutement...), la recherche de moyens (financements, équipements...), la diffusion des résultats (stratégie de publication, politique en matière d'accès libre, processus et supports de transfert des connaissances et des savoir-faire) ou encore la politique en termes de propriété intellectuelle.

Les unités de recherche de chacun des trois grands domaines scientifiques (sciences et technologie, sciences du vivant et de l'environnement, et science humaines et sociales) eux-mêmes subdivisés en six sous-domaines seront évaluées à l'aune de ce

même référentiel et via un même processus. Les responsables des unités commencent par renseigner un dossier d'auto-évaluation standardisé qui est soumis à un comité d'experts. Ce comité produit un rapport provisoire et une liste de questions puis programme une visite de l'unité de recherche qui permettra de recueillir des informations complémentaires (réponse aux questions formulées, présentation des faits nouveaux), d'échanger sur le projet, de rencontrer les personnels de l'unité et de se faire présenter les locaux et les outils de la recherche. Un rapport final est ensuite remis au HCERES et relu par le conseiller scientifique et le conseiller scientifique coordonnateur référent. Enfin, ce rapport est transmis aux tutelles puis au directeur de l'unité évaluée qui disposent de trois semaines pour faire connaître leur relevé d'erreurs factuelles et leurs observations, lesquelles seront annexées au rapport définitif.

4. Mesure de l'impact sociétal : l'apport des métriques

Quelles que soient la définition et la méthodologie envisagées pour évaluer les impacts sociétaux, ceux-ci doivent être démontrés plutôt que supposés. Or, entres autres limitations, la collecte d'indicateurs représente un important investissement financier et temporel, notamment en raison des multiples formes que ceux-ci peuvent prendre. Cette multiplicité entrave de plus la comparaison des évaluations de différentes institutions ou disciplines nécessitant des indicateurs uniformes.

Les injonctions au développement de méthodes de collecte de données généralisables et susceptibles de réduire le poids économique et temporel des évaluations se sont donc multipliées. Il a par exemple été suggéré que la collecte de preuves d'impact dans le cadre du REF anglais s'effectuera probablement à l'avenir via l'utilisation d'outils de capture automatiques de données [10]. Ces outils reposeraient en grande partie sur les métriques dont certaines sont d'ores et déjà utilisées comme indicateurs.

4.1 Les différents types de métriques

4.1.1 Les analyses bibliométriques

Traditionnellement, la quantification de l'impact de la recherche s'est principalement appuyée sur la bibliométrie, c'est-à-dire l'analyse de la littérature scientifique via deux types de mesures : la productivité en termes de nombre de publications, et la qualité et l'impact scientifique de ces publications estimés par leur nombre de citations [34]. Ces deux types de métriques peuvent être agrégés à différents niveaux (chercheur, unité de recherche, institution, pays) et proviennent le plus souvent de bases de données bibliométriques telles que Web of Science ou Scopus. De nombreux indicateurs plus sophistiqués ont également été créés à partir de ces données brutes, dont par exemple le facteur d'impact du journal (Journal Impact factor, JIF) [35], le h-index [36], les indicateurs de citation normalisés par domaine [37], l'Eigenfacteur [38], le SCImago Journal Rank (SJR) [39], le SNIP (Source Normalized Impact per Paper) [40], l'impact par publication (IPP) ou encore le CiteScore.

Initialement centrée sur les retombées des travaux de recherche au sein du milieu académique, la bibliométrie s'est peu à peu intéressée à leur impact sur d'autres parties prenantes et au développement d'indicateurs permettant de mesurer des activités comme le travail avec l'industrie, les contributions aux rapports gouvernementaux ou encore la communication de résultats de recherche à d'autres audiences que les pairs. Le nombre de citations dans les

brevets, les documents réglementaires ou les lignes directrices de recherche clinique est par exemple utilisé pour mesurer l'impact des publications académiques dans les domaines technologique, politique et médical [9] [34] [41] [42]. La bibliométrie permet également d'étudier les collaborations en se basant sur les relations de co-auteur. Le Leiden Ranking fournit par exemple des informations sur le nombre de publications résultant de collaborations inter-institutions et le pourcentage de telles collaborations sur la totalité des publications d'une institution donnée. Un facteur d'impact sociétal (SIF) [43] basé sur l'auto-évaluation d'une publication par ses auteurs grâce à une série de questions fut par ailleurs développé en 2011. Élaborées durant des discussions de groupe, ces questions portent par exemple sur le but de la publication (hausse des connaissances, application du savoir...), les efforts effectués pour traduire les résultats de recherche en actions sociétales et l'ampleur de ce transfert.

Lexique lié aux métriques [46]

Scientométrie : analyse de mesures portant sur tous les aspects des sciences et technologies et incluant les informations sur n'importe quelle sorte de résultats de recherche (données, logiciels, interactions, financement, commercialisation)...

Bibliométrie : analyses scientométriques focalisées sur l'analyse quantitative du nombre et des citations par d'autres travaux de recherche de publications scientifiques et érudites.

Indicateur : quantité mesurable supposée associée à une donnée moins « lisible » sans directement la mesurer. Le nombre de citations peut par exemple être utilisé comme un indicateur de l'impact scientifique des articles d'un journal même si cet impact n'est pas automatiquement lié à des citations.

Métriques d'utilisation (ou métriques d'usage) : mesures du nombre de vues et de téléchargements de productions scientifiques disponibles sur internet.

Métriques alternatives ou altmétriques : métriques basées sur l'analyse de données liées aux productions scientifiques disponibles en ligne. Elles incluent les métriques d'utilisation et toutes les données issues des réseaux sociaux.

4.1.2 Les métriques d'utilisation (ou métriques d'usage)

Avec l'avancée des technologies du Web, de nouvelles possibilités sont apparues pour évaluer l'impact de publications scientifiques, non seulement dans des journaux scientifiques, des livres et des rapports, mais aussi dans d'autres types de publications non traditionnelles.

Accessibles bien avant l'introduction du concept des métriques alternatives, les métriques d'utilisation mesurent le nombre de vues ou de téléchargements d'une production scientifique présente sur internet. Contrairement aux métriques liées aux citations, elles permettent de couvrir l'utilisation du matériel scientifique par une très large audience. De nombreuses parties prenantes sociétales dont des étudiants, des décideurs réglementaires ou politiques voire le grand public en général lisent en effet des publications ou utilisent des données sans jamais publier à leur tour et donc sans générer de citations. De plus, les publications référencées par un chercheur ne sont pas représentatives de l'ensemble de ses lectures.

Les métriques basées sur l'utilisation, comme le facteur d'impact d'usage [44] ou libcitations [45] sont supposées mesurer l'attention et la compréhension. Elles sont pertinentes non seulement en terme d'utilisation des publications mais aussi pour tracer les publications non traditionnelles et la réutilisation d'open

data ou d'open software. Les éditeurs en accès libre fournissent généralement des informations liées à l'utilisation des articles (PLoS) et les éditeurs commerciaux commencent également à s'ouvrir à ces pratiques. Plusieurs d'entre eux dont Springer, Nature, IEEE ou ACM montrent désormais le nombre de téléchargements d'un article spécifique depuis leur plateforme.

4.1.3 Les métriques alternatives ou altmétriques

L'idée des métriques alternatives fut introduite en 2009 sous l'appellation de « métriques au niveau de l'article » (article-level metrics) [47]. Ce concept de métriques focalisées sur la mesure de l'impact de la recherche académique sur internet fut ensuite développé et étendu dans « Altmetrics : a manifesto » [48]. Pouvant être définies comme « un terme décrivant des métriques basées sur internet avec une emphase sur les médias sociaux comme source de données », les métriques alternatives, ou altmétriques, reposent sur le postulat que les chercheurs utilisent différentes plateformes incluant les réseaux sociaux pour disséminer et discuter leurs contributions. Faisant cela, ils ouvrent des conversations érudites à un plus large public et il est supposé que la collecte des discussions, mentions et autres données relatives aux contributions scientifiques dans les médias sociaux peut fournir une mesure de l'impact sociétal [49].

Très hétérogènes, ces métriques englobent une multitude de signaux mesurables indiquant chacun un niveau différent d'implication. Elles peuvent se baser sur un très large panel de sources primaires, allant de blogs à ResearchGate et Mendeley en passant par Twitter ou Facebook, et incluent aussi bien les métriques d'utilisation que le nombre de followers, les notes, les sauvegardes, les tweets, les partages, les likes, les recommandations, les tags, les posts, le suivi du retour, les discussions, les marque-pages, les commentaires ou encore les mentions dans des articles de blogs voire dans les médias grand public ou des documents réglementaires ou politiques. Leur dénominateur commun est qu'elles excluent les indicateurs bibliométriques traditionnels [34].

Il existe actuellement trois principaux agrégateurs d'altmétriques collectant chacun à partir des sources primaires un jeu d'indicateurs légèrement différents : Altmetrics.com, PLUMx et ImpactStory. ImpactStory et PLUMx proposent également des données altmétriques à l'échelle des auteurs ainsi que des unités et institutions de recherche.

4.2 Avantages et limites des métriques

Une partie de la promesse des métriques repose sur leur apparente neutralité et le fait qu'elles permettraient des comparaisons à différentes échelles (chercheurs, instituts, pays), quelles que soient les disciplines, sur la base de critères uniformes.

Le classement de Shanghai comparant les universités à l'échelle internationale repose par exemple à 60 % sur des indicateurs bibliométriques. Néanmoins, alors qu'ils étaient vus comme une réponse pour fournir des mesures de l'impact de la recherche à la fois dans et à l'extérieur de la sphère scientifique, ces indicateurs ont connu ces dernières années une révision critique de leurs limitations. En 2013, la « San Francisco Declaration on Research Assessment » (DORA) a par exemple appelé à évaluer la recherche sur ses propres mérites et cesser d'utiliser les facteurs d'impact des journaux dans les décisions de financement, embauches et promotions. Des chercheurs des Pays-Bas établirent par ailleurs le mouvement « Science in Transition » dans le but de s'attaquer aux problèmes systémiques dans la culture de la recherche et de l'université qui serait « devenu un système d'auto-référence, où la qualité est mesurée principalement via des paramètres bibliométriques et où la pertinence sociétale est sous-évaluée ».

Du point de vue des analyses bibliométriques, de fortes disparités disciplinaires ont tout d'abord été soulignées quant aux

pratiques de publication dans des revues scientifiques, rendant caduque l'idée d'évaluer la productivité de l'ensemble des recherches sur la seule base du nombre de publications. Quant au nombre de citations, l'hypothèse de la bibliométrie évaluative selon laquelle les scientifiques réfèrent seulement et automatiquement les publications qui les ont influencés cognitivement [1] [3] n'a pas été confirmée. Des études menées depuis les années 80 ont en effet observé que certaines publications sont utilisées mais peu ou pas citées, d'autres sont principalement ou uniquement citées à travers des sources secondaires et seulement une partie est systématiquement citée à chaque utilisation [50]. Ajoutant à cela que les citations positives ne sont pas distinguées des négatives et qu'un bon score de citations peut donc en réalité résulter d'une controverse ou d'une erreur, le nombre de citations sorti de son contexte ne paraît pas non plus permettre une mesure fiable de l'utilité d'une publication. La mauvaise utilisation des indicateurs construits à partir des analyses bibliométriques fut également vivement montrée du doigt. Le facteur d'impact d'un journal est par exemple souvent utilisé à tort comme un indicateur de l'impact d'un article donné alors qu'il s'agit en réalité d'une moyenne sur l'ensemble du journal. De même, le recours au h-index tient rarement compte de son biais envers les chercheurs seniors et du fait qu'il ne reflète pas convenablement l'impact des publications très citées. Enfin, des doutes furent émis quant à l'utilité de la bibliométrie pour évaluer l'impact sociétal de la recherche en raison de son échelle de temps longue et de sa restriction à certaines catégories d'utilisateurs (industrie pour les brevets, confrères pour les publications dans des journaux scientifiques, monde politique pour les textes réglementaires...).

De ce point de vue, les altmétriques, et en particulier les données issues des réseaux sociaux, sont présentées comme un domaine prometteur pour mesurer les retombées sociétales plus larges de la recherche scientifique [49]. Les données altmétriques s'accumulent tout d'abord beaucoup plus rapidement que les signaux bibliométriques et peuvent rendre compte d'un impact presque instantanément. Couvrant aussi bien les publications dans des journaux scientifiques que des jeux de données, du code informatique, des conceptions expérimentales ou encore des posts et des commentaires sur des blogs et divers réseaux sociaux, elles sont par ailleurs susceptibles de fournir une large variété de signaux pour le même objet et de mesurer les réactions et interactions d'un vaste panel de publics. Cette multiplicité d'audiences et d'indicateurs est vue comme particulièrement appropriée à l'évaluation des recherches interdisciplinaires et de l'impact de résultats scientifiques sur la société dans son ensemble. Elle permettrait également l'identification de scénarios liés à l'atteinte d'impact et l'analyse des bénéficiaires éventuels des produits de la recherche.

Néanmoins ce nouveau type de métriques a lui aussi soulevé de nombreuses questions. La plupart des analyses d'altmétriques extrapolent le modèle des citations bibliométriques aux réseaux sociaux dans un but de justification de la recherche et présentent des mesures linéaires (de la publication initiale au support en ligne étudié), universelles (comptées sans réel contexte) et basées sur un modèle d'attribution total [51]. Une première interrogation porte sur la validité de l'analogie entre la citation d'une publication dans un travail de recherche et la citation d'une publication sur les réseaux sociaux. D'une part, alors que la première implique des experts capables d'évaluer la qualité du travail qu'ils citent, la seconde intègre des interlocuteurs qui ne le sont pas forcément. D'autre part, quand il y a une motivation claire dans le premier cas, le second peut simplement résulter d'un acte mécanique irréflecti [51]. La signification des comportements et motivations liés à la génération de différents signaux altmétriques (tweets, partages...) a été étudiée mais aucune conclusion n'a pour l'instant été atteinte [52]. Elle peut en effet marquer aussi bien un impact sociétal ou les prémisses d'une attention scientifique, d'une utilisation éducative et/ou d'une visibilité qu'une grande résonance négative, et cela dans des proportions dépendant fortement de la plateforme et ses utilisateurs, du sujet de

recherche ou encore du contexte de la métrique. La signification des mesures altmétriques étant, dans le meilleur des cas, ambiguë, leur utilisation comme indicateurs universels d'impact est donc remise en question.

Le choix des indicateurs altmétriques est par ailleurs lui-même problématique. Si la couverture des chercheurs dans les médias sociaux s'améliore avec l'essor des plateformes académiques comme Mendeley, ResearchGate ou FigShare et des réseaux sociaux en général, elle varie grandement en fonction des disciplines, des âges ou encore des pays. Certaines plateformes sont de plus utilisées à la fois pour un usage professionnel et personnel, sans compter que le paysage des réseaux sociaux est en constante évolution (apparition de nouvelles plateformes, disparition d'autres, variation de popularité...). Un autre point fréquemment soulevé concerne le rôle des éditeurs dans le développement des altmétriques. Soulignant que « détenus par de grands éditeurs souhaitant faire des profits, Almetric, Plum Analytics et Mendeley opèrent sous une certaine pression de mettre en avant la valeur des altmétriques et les rendre profitables » [52], certains ont critiqué un manque d'accès aux données sous-jacentes, qui sont la propriété du secteur commercial, mais aussi de transparence quant aux méthodes de collecte de données et aux données auto-collectées. En résumé, le choix des indicateurs altmétriques est crucial car la durée de vie des indicateurs est variable (ce qui est pertinent à un moment donné peut être rapidement désuet) et la mesure en elle-même n'est jamais neutre.

4.3 Vers des métriques responsables

Au-delà des limites et des doutes liés à la légitimité des métriques, leur utilisation systématique durant le processus d'évaluation pourrait également engendrer des modifications dans les mœurs scientifiques, au détriment du progrès scientifique et sociétal [34]. L'attention des gestionnaires et des agences de financement pourrait en effet se restreindre aux seuls impacts sociétaux pouvant être mesurés, conduisant potentiellement les universités à adopter des priorités stratégiques similaires les unes aux autres et les chercheurs individuels à se focaliser sur des travaux à faible risque. Outre une restriction de la diversité des travaux, cela pourrait également exacerber les problèmes de qualité de recherche, d'intégrité et de reproductibilité. Dans le cas des altmétriques, les chercheurs devraient de plus suivre le développement des plateformes web, ce qui ajouterait un poids pouvant brider leur créativité.

Contrairement aux attentes initiales qui envisageaient les indicateurs métriques comme pouvant devenir une méthode générale de collecte de traces d'impact sociétal en vue de justifier les recherches, les résultats empiriques bibliographiés semblent plutôt suggérer que leur application universelle et généralisable est problématique. Face à ce constat, un groupe de scientométriciens publia en 2015 le « Leiden Manifesto » [53] posant 10 principes à l'utilisation d'indicateurs quantitatifs dans l'évaluation de la recherche. Parallèlement, le « Metrics Tide » [46], revue du rôle des métriques dans l'évaluation et la gestion de la recherche au Royaume-Uni, proposa également une liste de recommandations ciblées. Ni le Leiden Manifesto, ni le Metric Tide ne rejettent les métriques, ils en conseillent plutôt une utilisation responsable. Selon eux, les métriques responsables doivent être comprises en termes de « solidité, d'humilité, de transparence, de diversité et de réflexivité » [46]. La solidité implique de baser les métriques sur les meilleures données possibles en termes de pertinence et de portée. L'humilité correspond à la reconnaissance que l'évaluation quantitative devrait supporter et non supplanter l'évaluation qualitative experte. La transparence concerne les collectes de données et processus analytiques qui doivent permettre aux évalués d'être en mesure de tester et vérifier les résultats. La diversité correspond à une prise en compte des variations selon les domaines et les utilisations dans le panel d'indicateurs qui doit refléter et soutenir une pluralité de

recherche et de carrières. Et, enfin, la réflexivité préconise de reconnaître et anticiper les effets systémiques potentiels des indicateurs et les mettre à jour en conséquence.

La Commission européenne se pencha également sur le sujet via la production en 2017 d'un rapport sur les métriques responsables et l'open science [54]. Reconnaisant la légitimité des inquiétudes liées au détournement de certains indicateurs quantitatifs portant sur la qualité des recherches et la possibilité que les indicateurs émergents introduisent des biais et ignorent ce qui est difficile à quantifier, ce rapport réaffirme les conclusions du Leiden Manifesto et du Metric Tide. Conformément à leurs déclarations et à celle de DORA, il considère que les indicateurs actuels doivent être améliorés et coexister avec des indicateurs additionnels et la revue par les pairs dans la politique et le processus d'évaluation. « Les mesures et les récits, les métriques et les revues par les pairs devraient être traités comme des outils d'évaluation complémentaires ». Par ailleurs, dans le cadre de leur utilisation pour des décisions politiques ou le financement et l'évaluation de la recherche, les métriques dérivant de plateformes privées devraient toujours s'accompagner d'autres métriques transparentes pour permettre une validation. Le rapport de la Commission européenne préconise par exemple l'utilisation des identifiants uniques attribués aux chercheurs par l'ORCID, qui sont associés à une liste de publications régulièrement mise à jour, pour les demandes de subvention ainsi que les plateformes de gestion et de rapport. Les éditeurs scientifiques sont par ailleurs incités à encourager des pratiques d'auteur responsables en fournissant des informations plus détaillées sur la contribution spécifique de chaque co-auteur et à réduire l'emphasis sur le facteur d'impact du journal en tant qu'outil promotionnel. Celui-ci pourrait être seulement utilisé dans le contexte d'un jeu de métriques, incluant par exemple un facteur d'impact à 5 ans, l'EigenFactor, SCImago et les temps d'édition et de publication, qui offrirait une vue des performances plus riche.

4.4 Des métriques évaluatives aux métriques descriptives

Parallèlement à la remise en question des métriques conçues comme des indicateurs quantitatifs servant à la comparaison et la justification de la recherche, une distinction a émergé entre la scientométrie dite « évaluative » et une scientométrie dite « descriptive », centrée quant à elle sur les tendances, les cartes scientifiques et les réseaux collaboratifs. L'analyse de réseaux sous la forme de cartes est devenue un outil de plus en plus populaire pour transmettre des résultats de bibliométrie descriptive dans les réglementations scientifiques. La visualisation de réseaux permet en effet de synthétiser des informations complexes et dotées de plusieurs variables. S'inspirant de cette distinction, une différenciation entre les indicateurs altmétriques (liés à l'évaluation) et les analyses altmétriques décrivant les contenus et contextes a été proposée en 2016 [55] en prenant pour exemple la manière dont

les mentions Twitter des publications scientifiques peuvent informer sur les points d'intérêts sanitaires des différentes régions du monde (SIDA en Afrique, santé mentale et obésité en Europe). Cet exemple illustre que si l'utilisation des altmétriques en tant qu'indicateurs quantitatifs universels dans le cadre d'évaluations reste problématique, celles-ci n'en possèdent pas moins un fort potentiel.

Outre les possibilités qu'elles offrent pour identifier des problématiques scientifiques pertinentes, elles peuvent également être utilisées pour enrichir et trianguler les preuves d'impact fournies dans les études de cas, identifier en amont des impacts sociétaux potentiels ou encore renseigner sur les mécanismes menant à la réalisation de ces impacts. Les sites consacrés aux altmétriques offrent en effet beaucoup plus de possibilités de contextualisation que les bibliométriques. Les réseaux sociaux reflètent au moins partiellement les différents publics auprès desquels les chercheurs s'engagent, facilitant à la fois l'identification des différentes communautés de parties prenantes et celle des réseaux et activités des chercheurs sans oublier leurs interactions et le contexte dans lequel elles prennent place. Une méthode a ainsi été proposée en 2017 pour identifier et analyser systématiquement les chercheurs et leur champ de spécialisation en se basant sur les informations enregistrées par Twitter [56]. La pertinence de Twitter pour cartographier le réseau de chercheurs et développer leur profil a par ailleurs été testée via deux études de cas portant sur deux individus aux caractéristiques très dissemblables [51]. Les résultats mirent en avant une correspondance entre les attributs de profils obtenus via l'analyse des données altmétriques et celles, connues, des deux chercheurs analysés.

Un grand intérêt est actuellement accordé à ce type de méthodologies qualitatives basées sur le contexte et les interactions. Comme le met en évidence le tableau 4, leurs caractéristiques semblent plus en accord avec les principes des métriques responsables prescrites par les divers rapports récents répertoriés en 4.3. Cet intérêt repose pour l'instant principalement sur le fait que ces méthodologies permettraient de suivre en amont les interactions pouvant être productives [21] et de cartographier les contextes d'impacts sociétaux potentiels. Bien que l'ensemble des interactions réelles des chercheurs n'aient pas nécessairement lieu dans les médias sociaux, une partie d'entre elles peut être supposée ou déduite à partir de réseaux d'interactions denses avec certaines parties prenantes. En cas d'observation de traces de changement comportemental ou d'innovations chez ces parties prenantes, l'engagement des parties soumises à évaluation avec lesquelles elles interagissent peut alors être exploré qualitativement comme une instance potentielle de contribution à l'impact sociétal. Considérant le long laps de temps séparant souvent la recherche et ses impacts, ces indications pouvant intervenir tôt dans le processus de réalisation d'impact présentent une utilité non négligeable. L'analyse de ce processus peut par ailleurs aider à comprendre a posteriori les types d'activités, de contributions et d'interactions des chercheurs à la fois dans et hors de la sphère scientifique qui ont conduit à un impact sociétal.

Tableau 4 - Comparaison entre les indicateurs altmétriques extrapolés de la bibliométrie et les analyses altmétriques basées sur les interactions

	Indicateurs altmétriques	Analyses altmétriques basées sur les interactions
But de l'évaluation	Rendre des comptes	Apprentissage, amélioration
Porté d'application	Universel, généralisable	Contextuel, analyse de cas
Focalisation de l'analyse	Résultats	Processus, interactions
Interaction avec les parties prenantes	Influence indirecte	Activité (engagement)
Causalité	Attribution	Contribution
Opérations	Compte linéaire	Structure de réseau

5. Bilan et perspectives de l'évaluation de l'impact sociétal

Les recherches menées sur l'impact sociétal de la recherche nourrissent une réflexion permanente sur sa définition et ses enjeux ainsi que sur les méthodes et indicateurs pouvant permettre son évaluation. Si elles semblent avoir abouti à l'adoption d'un certain nombre de réponses aux considérations épistémologiques et éthiques évoquées plus haut (cf. § 2.3), de nombreuses perspectives d'amélioration méthodologiques restent ouvertes. Le processus d'évaluation demeure par ailleurs confronté à un certain nombre de limites difficilement contournables qui pourraient faire évoluer ses objectifs d'une justification des investissements liés à la recherche vers l'étude des processus permettant à celle-ci de générer des impacts sociétaux.

5.1 Orientation générale de l'évaluation des impacts sociétaux

Qu'il s'agisse d'évaluation ex ante ou ex post, la majorité des méthodologies actuellement appliquées semble s'accorder sur le fait que l'impact sociétal résulte d'un processus qu'il convient de considérer dans son ensemble plutôt que de s'intéresser aux seules retombées quantifiables de la recherche sur la société.

Étant donné l'infinité de chemins pouvant mener à la réalisation d'impacts et les retombées très diverses pouvant résulter des travaux, très différents, d'un ingénieur, d'un astrophysicien, d'un sociologue ou encore d'un historien, il ne paraît pas possible de dégager des indicateurs universels. Un consensus semble donc émerger sur le fait que les études de cas constituent l'approche la plus appropriée à l'évaluation de l'ensemble des retombées sociétales potentielles ou effectives de la recherche à l'échelle des institutions, des programmes, des projets ou des individus [9] [10] [51]. Elles présentent en effet l'avantage d'enregistrer la complexité des impacts touchant toute une gamme de disciplines allant des arts aux sciences naturelles et permettent le recours à une infinité d'indicateurs incluant à la fois métriques et données non mesurables.

Sans qu'elles en garantissent la réalisation, les interactions des chercheurs avec d'autres parties prenantes de la société sont considérées comme des indicateurs d'impact potentiel et ont une place importante dans son évaluation. La science et ses divers représentants (chercheurs, instituts) sont encouragés à les favoriser et le fait d'entretenir un réseau et de mettre en place des collaborations semble de plus en plus estimé comme une composante à part entière du travail de recherche.

Si la qualité de la recherche et ses retombées sociétales sont majoritairement admises comme n'étant pas corrélées, la première demeure considérée comme un facteur sous-jacent dans l'évaluation des secondes. Enfin, les impacts se voient la plupart du temps attribuer une valeur, définie en terme de « désirabilité » pour la société.

5.2 Perspectives d'amélioration méthodologique

5.2.1 Dépasser les biais personnels des évaluateurs

Si les études de cas sont considérées comme la méthode la plus appropriée à l'évaluation des impacts sociétaux de la recherche, elles souffrent néanmoins d'un certain nombre de limitations. Reposant sur l'expertise d'évaluateurs, elles impliquent un

temps et un coût relativement élevés, et présentent le risque que, face à une notion d'impact sociétal qu'ils ne maîtrisent pas, lesdits évaluateurs émettent des jugements partiels, grandement dépendants de leurs biais personnels.

Interrogés juste avant la mise en place du REF anglais, certains évaluateurs impliqués dans ce processus ont affirmé qu'ils envisageaient jusque-là l'impact sociétal comme « quelque chose dont ils ne tenaient pas compte ou considéraient comme peu important » [19]. Il est possible que cet aspect de l'évaluation continue donc tout simplement à être contourné, en particulier par les pairs. Une tendance à la persistance d'évaluations s'appuyant en pratique sur les seuls critères scientifiques traditionnels a été constatée à de multiples reprises et attribuée aussi bien à un manque de confiance des experts en leur jugement (concernant l'impact sociétal) qu'à un choix délibéré lié à leurs valeurs [57].

Par ailleurs, même lorsque le critère d'impact sociétal est pris au sérieux, les différences de perception quant à ce qui est cru ou non constituer un tel impact sont très prononcées. Ces différences peuvent de plus être amplifiées par divers facteurs, dont notamment la progression du développement de l'impact au moment de l'évaluation [19]. Cette idée peut être illustrée dans le cadre de la découverte d'un nouveau médicament, commençant par la mise en place d'une nouvelle formule chimique suivie par un travail préclinique puis une éventuelle validation réglementaire et une mise sur le marché [58]. Selon son expérience personnelle, et en particulier le moment habituellement impacté par son propre travail dans ce type de processus, chaque évaluateur attribuera une valeur différente à chaque étape menant à la commercialisation d'un nouveau remède. Pour simplifier, un évaluateur dont le travail consiste à pratiquer des tests précliniques considèrera par exemple qu'une étude de cas évaluée après la réussite de ce type de tests a un plus fort impact sociétal qu'une autre évaluée au moment de la découverte d'une nouvelle formule chimique. En revanche, un évaluateur œuvrant dans la chimie de synthèse attribuera la même note aux deux études de cas.

La réunion de panels d'évaluateurs discutant et débattant lors des évaluations pourrait éviter cet écueil. Cependant il est important de ne pas sous-estimer les problématiques liées aux effets de groupe telles que, par exemple, la possibilité que l'opinion d'évaluateurs occupant des positions importantes bénéficie automatiquement d'une meilleure considération. Des réflexions restent à mener pour limiter l'impact de ce type de problématiques sur les résultats d'évaluation.

5.2.2 Considérer une multitude de groupes cibles aux intérêts parfois contradictoires

Contrairement à l'impact scientifique s'adressant aux seuls pairs, la détermination des impacts sociétaux implique de considérer une multitude de groupes cibles dont les attentes et les intérêts sont souvent différents. Dans le cadre de la recherche médicale, les parties prenantes furent par exemple séparées en trois catégories : le public en général, les professionnels de santé et le secteur privé [22]. Une tentative de catégorisation plus générale [59] en distingua quant à elle trois autres :

- les politiciens dont le but est d'utiliser la recherche pour leur propre politique ou pour faciliter les transferts de savoir de la science à la société ;
- les professionnels, c'est-à-dire les industriels ou les organisations de la société civile (associations, ONG...), qui utilisent les connaissances scientifiques pour développer des produits et/ou des services ;
- les utilisateurs finaux, à savoir le grand public, considérés dans leur ensemble ou séparés en différents groupes (les agriculteurs, les victimes d'une maladie...).

Face à cette diversité de groupes cibles, S. Hill s'interrogea dans sa revue du REF [10] sur la pertinence de tenir compte de la nature de l'impact d'une recherche dans la notation de celle-ci. Parallèlement au cas de travaux mal menés, faux ou fraudu-

leux qui peuvent avoir un grand retentissement, se pose en effet le problème de recherches dont l'évaluation peut s'apparenter à un jugement de « valeur publique », c'est-à-dire dont la note correspond à une décision quant à la désirabilité de ses impacts sur la société. Certains panels d'évaluateurs du REF ont rapporté être tout à fait conscients de l'importance de leurs jugements de valeur dans leurs choix de notation. Or, l'inclusion de tels jugements basés sur la nature de l'impact soulève d'importantes questions au sujet, notamment, de leur évolution dans le temps mais surtout des parties prenantes dont le point de vue doit primer.

Chaque groupe cible ayant des intérêts propres et ces intérêts étant parfois contradictoires, il paraît délicat de parvenir à un consensus quant à la nature des impacts sociétaux d'une recherche. Trancher quant à la désirabilité ou non de ces impacts implique donc d'adopter le point de vue de l'un ou l'autre d'entre eux. Cela pose la question de l'étendue de l'implication des parties prenantes dans le processus d'évaluation, en particulier dans le cadre du REF dont les panels d'évaluateurs intègrent aussi bien des scientifiques que des représentants de divers secteurs publics ou privés (monde des affaires, secteur tertiaire, société civile...). Certains évaluateurs pourraient en effet être tentés de défendre les intérêts de leur propre communauté. Il apparaît alors important d'accroître la transparence quant aux statuts des évaluateurs et de développer des solutions visant à prémunir le processus d'évaluation contre les conflits d'intérêts.

5.2.3 Clarifier la signification des résultats

La plupart des évaluations sont basées sur un rapport d'auto-évaluation fourni par les évalués eux-mêmes. Or, ceux-ci font face aux mêmes interrogations quant à la définition et la nature de l'impact sociétal que leurs évaluateurs. Dans le cadre du développement d'ASIRPA, l'Inra a par exemple constaté une exploration plus systématique de certaines catégories que d'autres par les chercheurs, souvent en raison du fait que ceux-ci n'ont pas l'impression d'être légitimes à revendiquer un rôle dans des domaines très éloignés de celui de leurs recherches ou manquent de connaissances quant à la diversité des impacts pouvant être atteints. Entrent également en compte la difficulté et le coût (aussi bien financier que temporel) liés à la collecte des informations. Les rapports d'auto-évaluation peuvent alors n'offrir qu'une représentation partielle des impacts sociétaux découlant d'une recherche.

Par ailleurs, malgré la consultation de parties prenantes extérieures et la livraison de preuves d'impact, un manque d'objectivité est parfois reproché à ces auto-évaluations. En effet, les évalués ont le choix de ne rapporter que leurs résultats ayant eu le meilleur impact. Un des biais induits par une évaluation de l'impact de la recherche, notamment à l'échelle des instituts ou des pays, en ne s'appuyant que sur des « success stories » est le fait que des anomalies auront une grande influence sur le résultat global. De la même manière qu'il suffisait d'une seule publication extrêmement citée à une université pour parvenir à un bon classement basé sur des indicateurs de citations, quelques études de cas prolifiques pourront par exemple être entièrement responsables du résultat de l'évaluation de l'impact sociétal d'une institution de recherche. Autrement dit, une institution de recherche pourra obtenir un excellent score non pas en raison du travail de l'ensemble des scientifiques qui y œuvrent, mais seulement de quelques brillantes réussites de certains d'entre eux [42].

Un second biais, d'autant plus prégnant que les résultats d'évaluations sont souvent associés à l'allocation de financements nécessaires à la poursuite de recherches, est un glissement, dans les rapports d'auto-évaluation, vers ce qui peut être qualifié de « sensationnalisme d'impact ». Une étude basée sur le témoignage de 50 évaluateurs et candidats à des sources de financement externes de différentes disciplines (arts, humanités, sciences sociales, sciences naturelles et de la vie, sciences physiques incluant mathématiques et ingénierie) au Royaume-Uni et

en Australie a ainsi souligné une exagération des revendications d'impacts quand ceux-ci ne sont pas immédiatement évidents [57]. Parfois attribuée à la mise en place d'une « hyper-compétitivité » qui résulterait des nouveaux modes de gestion de la recherche, cette attitude est également associée à l'idée persistante que la description d'impact préalable à la matérialisation des résultats de recherches est « infaisable », « illogique » voire proche d'un « blasphème scientifique » [57]. Gênant ou décourageant l'évocation de projections d'impact plus modestes mais réalistes, la tendance à l'exagération s'autoalimente : surestimer les déclarations d'impact apparaît désormais pour beaucoup comme un moyen inévitable pour acquérir des fonds. Il apparaît alors crucial de tenir compte de cet aspect tant dans l'analyse des résultats d'impact que dans l'élaboration des méthodologies d'évaluation.

5.3 Limites et perspectives d'évolution des évaluations d'impact

5.3.1 L'insoluble problème de l'échelle de temps

Quelle que soit la méthodologie mise en place, la problématique de l'échelle temporelle demeure. Il existe en effet de très nombreux cas où l'importance d'une recherche n'est apparue que bien après leur publication. Accueillie de façon mitigée lors de sa publication en 1961, la limite de Shockley-Queisser est par exemple devenue l'un des papiers les plus cités dans son domaine [42]. De manière générale, de nombreux travaux ont observé un long temps de latence, estimé de l'ordre de 15-20 ans, entre le début d'une recherche et ses impacts sociétaux pouvant être qualifiés de « finaux ». Dans le cadre des recherches cardiovasculaires, un écart moyen d'environ 17 ans entre le financement et les impacts sur la santé a par exemple été mis en évidence [60]. De même, lors du développement de la méthode ASIRPA, l'analyse d'un portefeuille de 41 études de cas a conduit à l'estimation que le voyage entre l'initiation de la recherche et sa traduction en un impact socio-économique prend en moyenne 19 ans [31]. Comme cela fut également souligné dans le cadre du REF, ces temps sont trop longs pour pouvoir être pris en compte dans une évaluation [10].

Par ailleurs, l'impact sociétal d'une recherche peut évoluer avec le temps. Une recherche environnementale conduisant à la fermeture d'une entreprise peut par exemple avoir un impact immédiat négatif tandis qu'à plus long terme elle permettra de préserver une ressource qui redeviendra utilisable [9]. De même, des recherches affichant des conséquences a priori positives peuvent être suivies d'autres travaux scientifiques les remettant en question. La spécialisation des systèmes de production agricoles et l'utilisation accrue d'intrants de synthèse dans les cultures ont par exemple initialement eu un impact positif, notamment sur les rendements, mais des études ultérieures ont montré qu'à long terme ces nouvelles pratiques engendrent divers impacts négatifs (voir encadré). Dans le cadre d'évaluations des retombées de la recherche jugées en terme de « désirabilité », cette imprévisibilité de l'évolution des impacts dans le temps demeure une limite difficilement contournable.

Exemples d'impacts liés à la spécialisation des zones de production agricoles [16]

De manière générale, une spécialisation des systèmes de production agricoles et des territoires a conduit au recul de l'élevage dans les régions les plus propices aux grandes cultures et à la concentration de l'élevage dans les régions bénéficiant d'un appareil industriel performant.

Exemples d'impacts liés à la spécialisation des zones de production agricoles [16] (Suite)

Spécialisation des grandes cultures dans le bassin de la Seine

Durant les quarante dernières années, le territoire du bassin versant de la Seine qui représentait 15 % des exploitations agricoles et 23 % de la surface agricole utile en 2000 a par exemple connu un recul de près de 70 % des systèmes de polyculture-élevage au profit des systèmes de production de grandes cultures. En parallèle, les assolements se sont spécialisés, avec un recul des espèces végétales liées à la polyculture-élevage (prairies, luzerne...) au profit notamment du blé ou du colza et des successions culturales qui se simplifient et se raccourcissent.

Ces évolutions, positives en termes de rendements agricoles, ont cependant été permises par un usage accru intrants de synthèse (engrais minéraux, pesticides), avec des conséquences environnementales majeures concernant l'uniformisation des paysages, la réduction de la biodiversité, le gaspillage des ressources non renouvelables et la contamination des ressources en eau du bassin. L'analyse des impacts sur le long terme de ce type d'agriculture spécialisée et intensive souligne donc des conséquences non attendues, suggérant que le retour à une plus grande diversification des productions dans les territoires constitue un réel enjeu de recherches. Néanmoins, mettre en œuvre des systèmes de production compatibles avec des objectifs plus centrés sur l'aspect environnemental se traduirait par une baisse de la production physique d'environ 10 %. Réduire le recours aux pesticides dans le secteur des céréales, avec la probabilité d'une diminution de rendement, peut de plus faire craindre une baisse de revenus pour les agriculteurs.

Algues vertes et modèle agricole breton

À partir des années 60, les productions animales se sont développées de manière considérable en Bretagne. En 2009, cette région représentait 56 % des productions françaises de porcs ainsi que 33 % des volailles, 41 % des œufs et 22 % du lait.

L'élevage nécessite des apports alimentaires en grande partie importés mais aussi produits sur place : un tiers de la surface agricole utile est destinée à la production de céréales (maïs en particulier) pour l'alimentation animale. Ces cultures annuelles nécessitent également des apports d'engrais minéraux importés. Au total, on estime à environ 300 000 tonnes le flux annuel d'azote entrant en Bretagne : une partie est exportée à travers les productions animales et végétales, mais environ un tiers de ce flux est estimé se retrouver dans les milieux aquatiques et, de là, dans la zone littorale. La concentration en nitrates des rivières bretonnes a ainsi été multipliée au moins par 5 depuis les années 70.

Or, dans les milieux littoraux, l'azote est le principal facteur stimulant la production de macro-algues et en particulier d'algues vertes comme les ulves. Le rôle déterminant des apports d'azote d'origine agricole dans le phénomène de prolifération ayant conduit à des échouages massifs sur les plages (avec diverses conséquences sanitaires, environnementales et économiques) est aujourd'hui bien établi. La seule solution viable à long terme est donc la réduction à la source de ces flux d'azote. Des solutions techniques telles qu'une amélioration de l'alimentation des animaux, une meilleure utilisation des effluents d'élevage, leur traitement éventuel pour produire des engrais organiques exportables ou un recours accru aux légumineuses (en culture ou dans les prairies) ont commencé à être mises en place. Cependant, leurs effets sont lents à se manifester : au rythme actuel, il faudrait environ cinquante ans pour retrouver les teneurs en nitrates des rivières des années 70.

Bien que positif au niveau du rendement, le maintien à long terme d'une telle concentration de productions animales en Bretagne est donc remis en question aussi bien sur les plans économiques et sociaux qu'environnementaux, et réexaminé en prenant en compte non seulement l'agriculture, mais aussi l'ensemble des activités de cette région.

5.3.2 Une détermination incertaine de la causalité et l'attribution des impacts

La réalisation d'impacts sociétaux résulte rarement d'un chemin linéaire et organisé mais plutôt d'un mélange complexe de chance et d'interactions intervenant sur plusieurs niveaux de transfert de savoir. Cela soulève tout d'abord des problèmes de causalité, c'est-à-dire de liaison d'un impact à une cause spécifique. Les théories de l'innovation soulignent que la recherche contribue à l'innovation via des réseaux au sein desquels les interactions et leurs effets ne sont pas clairs. Cela signifie notamment que l'impact sociétal ne peut pas être attribué aux seuls efforts de recherche mais également à d'autres facteurs et qu'il est donc impossible d'en attribuer la causalité à des facteurs uniques.

Or, par sa définition même, un processus d'évaluation implique de déterminer quels travaux de recherche et/ou quelles interactions sont à l'origine de l'impact sociétal. Cette co-création d'impact soulève donc à son tour des difficultés d'attribution, c'est-à-dire de détermination de la proportion à laquelle l'impact doit être attribué à différentes recherches, organisations ou chercheurs. Des études ont suggéré que l'impact global peut être cartographié selon des contributions majeures et des contributions mineures [58]. Néanmoins, outre le problème de la collecte de preuves soutenant cette cartographie, se présente alors le danger potentiel que de nombreuses contributions mineures à l'impact sociétal durant le chemin d'impact, venant avant ou après l'impact le plus important, soient sous-évaluées. De la même manière que toutes les publications s'avérant décisives ne connaissent pas un regain de citations une fois leur valeur reconnue, ce sont souvent les chercheurs qui connectent les différents brins de connaissance nécessaires à une découverte qui en gardent tout le crédit et non pas les chercheurs à l'origine des brins [42]. Se focaliser sur les contributions majeures risquerait alors de favoriser les domaines et les organisations ayant la capacité de faire ces contributions, au détriment de contributions mineures sans lesquelles l'impact global n'aurait peut-être pas pu être atteint, ce qui serait contreproductif.

5.3.3 Une difficile prise en compte des spécificités de la recherche

La pratique de la recherche se caractérise par un certain nombre de spécificités dont il est délicat de tenir compte dans le cadre d'une évaluation. La résolution de questions, hypothèses et problèmes scientifiques par exemple ne se fait pas automatiquement via la confirmation empirique d'hypothèses préalablement formulées mais implique souvent un processus de tâtonnement à travers de multiples essais. Par définition, la recherche est donc sujette aux erreurs et associée au droit de tenter des tests faillibles ou erronés, de prendre des risques et d'essayer des méthodes non orthodoxes ou intuitives. Le hasard joue également un grand rôle : beaucoup de découvertes importantes, dont par exemple celles des rayons X ou de la pénicilline, sont le résultat de « coïncidences heureuses » (serendipity) grâce auxquelles les chercheurs sont parvenus à des résultats qu'ils n'avaient aucunement anticipés. De nombreuses avancées n'ont par conséquent pas été basées sur la production continue d'unités de connaissance mais se sont concrétisées brutalement à travers des « révolutions scientifiques », ce qui peut être illustré par l'omniprésence des ruptures de trajectoire dans les voyages d'innovation étudiés lors du développement de la méthode ASIRPA.

Or dans un processus d'évaluation, les mesures sont répétées à quelques années d'intervalle et les résultats obtenus doivent progresser d'année en année ou, au minimum, se maintenir. Il est donc attendu des scientifiques qu'ils produisent un flux continu et allant croissant d'importants résultats de recherche impactant la société, ce qui ne semble pas compatible avec les spécificités précédemment énoncées. La disparité existant entre les attentes de l'évaluation et la réalité du processus de recherche augmente alors le risque de mauvaise conduite des chercheurs et de leurs institutions pour ajuster la réalité aux aspirations. De nombreuses études ont déjà documenté le

fait que les scientifiques basent leurs actions sur les critères et indicateurs appliqués dans les évaluations [61] et qu'ils adoptent des « stratégies » afin d'optimiser les résultats d'évaluations associées à des décisions d'allocations (voir quelques exemples en encadré). Plusieurs cas de méconduites scientifiques graves ont par ailleurs été rapportés récemment. Suite à la publication au Canada des « Résumés des dossiers de conduite responsable de la recherche » en 2017, plusieurs étudiants ont perdu leur doctorat et une quinzaine de professeurs ont été congédiés, suspendus, rétrogradés ou ont démissionné face à 63 cas établis de malveillance scientifique allant du détournement de fonds à la fraude en passant par de fausses déclarations ou des plagiat. En France, bien que précisant que « les chiffres donnés doivent être considérés comme approximatifs », un rapport commandé par le Secrétariat d'État chargé de l'Enseignement supérieur et de la Recherche a recensé plus d'une centaine de cas de manquements entre 2010 et 2015 dont 2 fabrications, 32 falsifications, 50 plagiat et 8 conflits d'intérêt [63].

Quelques stratégies parmi les plus couramment adoptées pour optimiser les évaluations [42]

Les chercheurs ne résumant plus les résultats d'un projet dans une unique publication mais les distribuent dans plusieurs pour créer l'apparence d'une meilleure productivité.

Le choix du journal où soumettre une publication ne se porte plus sur le plus adapté en terme de contenu mais sur le plus prestigieux.

Certains chercheurs changent de sujet de recherche pour mieux s'aligner sur les tendances générales et les domaines les plus « porteurs » [62].

Les chercheurs établis rechignent à travailler avec des chercheurs en début de carrière.

Les chercheurs optent pour des activités pouvant être directement documentées plutôt que des activités plus dures à documenter mais parfois plus utiles à la société [61].

Les institutions financent préférentiellement des chercheurs ayant déjà une longue liste de résultats plutôt que des jeunes chercheurs, même talentueux, qui n'ont pas encore fait leurs preuves [62].

Certaines institutions ont reclassé des chercheurs dont le rendement était estimé insuffisant vers des champs sans rapport avec leurs compétences afin d'obtenir de meilleurs résultats dans les champs qu'ils occupaient initialement [62].

6. Conclusion

Reconnues comme des limites importantes lors du développement de toutes les méthodologies actuellement utilisées, ces problématiques de temporalité, de causalité et de prise en compte des spécificités de la recherche remettent en question le bien-fondé du conditionnement des investissements alloués à la science aux résultats d'évaluation d'impacts sociétaux. Les exemples précédemment évoqués illustrent un risque réel que ce conditionnement soumette la recherche aux effets de mode et aux caprices de la demande sociale, avec pour effet néfaste d'entacher la légitimité et la qualité des résultats scientifiques [33]. Les chercheurs pourraient également devenir plus soucieux du marketing de leur recherche que de son contenu [42], point qui faisait déjà l'objet de vives inquiétudes lors de la construction du critère évaluant les liens entre sciences et société dans le système d'évaluation national français [33]. Enfin, ce point soulève la question de la charge de travail associée à l'évaluation de l'impact sociétal de ces travaux pour les chercheurs. Comme l'a souligné le comité consultatif d'éthique Inra-Cirad, « l'importance du temps passé aux travaux et réflexions relatifs aux impacts de la recherche, et ceci au détriment de la production même de connaissances » pourrait poser problème [16]; sans compter que, pour s'assurer de la réalisation d'impacts, les chercheurs pourraient également se tourner vers des activités tout aussi chronophages d'expertise, de valorisation ou encore de participation au débat public.

Face à ces problématiques, l'idée se développe que l'évaluation de l'impact sociétal devrait se focaliser sur les processus conduisant à l'impact plutôt que sur les résultats ou les impacts eux-mêmes, et servir un but d'apprentissage quant au contexte de réalisation de ces processus plutôt que de justification de la recherche. Un nouveau paradigme qui pourrait développer des opportunités professionnelles dans le domaine de l'ingénierie de la gestion de la recherche.

Impact sociétal de la recherche : évaluation et perspectives

par **Laura BEDOURET**

Docteur en chimie-physique
Diplômée de l'Université de Bordeaux

et **Anna DUPLEIX**

Post-doctorante dans l'équipe Bois du Laboratoire de Mécanique et Génie Civil
UMR CNRS, Université de Montpellier 5508

Sources bibliographiques

- [1] MERTON (R.K.). – *The Normative Structure of Science*. Storer N.W. (1942).
- [2] BUSH (V.). – *Science, the endless frontier*. United States Government Printing Office, Washington (1945).
- [3] MERTON (R.K.). – *The sociology of science : Theoretical and empirical investigations*. Chicago, IL : University of Chicago Press (1973).
- [4] GIBBONS (M.T.), LIMOGES (C.), NOWOTNY (H.), SCHWARTZMAN (S.), SCOTT (P.) et TROW (M.). – *The new production of knowledge : The dynamics of science and research in contemporary societies*. Sage Publications Ltd, London (1994).
- [5] PETIT (J.C.). – *Why do we need fundamental research ?* European Review, 12(2), 191-207 (2004).
- [6] NOWOTNY (H.), SCOTT (P.) et GIBBONS (M.T.). – *Introduction : « Mode 2 » revisited : The new production of knowledge*. Minerva, 41(3), 179-194 (2003).
- [7] NOWOTNY (H.), SCOTT (P.) et GIBBONS (M.T.). – *Re-thinking science : knowledge and the public in an age of uncertainty*. Wiley Polity Press, Cambridge (2001).
- [8] STOKES (D.E.). – *Pasteur's Quadrant. Basic Science and Technological Innovation*. Brookings Institution Press, Washington DC (1997).
- [9] BORNEMANN (L.). – *What is societal impact of research and how can it be assessed ? A literature survey*. Journal of the American Society for information science and technology, 64(2), 217-233 (2013).
- [10] HILL (S.). – *Assessing (for) impact : future assessment of the societal impact of research*. Palgrave Communications, 2:16073 (2016).
- [11] SMITH (R.). – *Measuring the social impact of research*. British Medical Journal, 323(7312), 528 (2001).
- [12] MANSFIELD (E.). – *Academic research and industrial innovation. Research Policy, 20(1), 1-12 (1991). Academic research and industrial innovation : An update of empirical findings*. Research Policy, 26(7-8), 773-776 (1998).
- [13] BEISE (M.) et STAHL (H.). – *Public research and industrial innovations in Germany*. Research Policy, 28(4), 397-422 (1999).
- [14] NARIN (F.), HAMILTON (K.S.) et OLIVASTRO (D.). – *The increasing linkage between US technology and public science*. Research Policy, 26(3), 317-330 (1997).
- [15] Health Economics Research Group, Office of Health Economics et RAND Europe. – *Medical Research : What's it worth ? Estimating the economic benefits from medical research in the UK*. London : UK Evaluation Forum (2008).
- [16] Comité consultatif commun d'éthique pour la recherche agronomique Inra-Cirad. – *Avis 9 : Les enjeux éthiques de l'évaluation des impacts de la recherche publique agronomique* (2016).
- [17] GODIN (B.) et DORE (C.). – *Measuring the impacts of science ; beyond the economic dimension* (1997).
- [18] European Commission expert group on assessment of university-based research. – *Assessing Europe's university-based research*. Brussels, Belgium : Publications Office of the European Union (2010).
- [19] DERRICK (G.E.) et SAMUEL (G.N.). – *The evaluation scale : Exploring decisions about societal impact in peer review panels*. Minerva, 54 (1), 75-97 (2016).
- [20] Evaluating Research in Context (ERiC). – *Evaluating the societal relevance of academic research : A guide*. Delft, The Netherlands : Delft University of Technology (2010).
- [21] SPAAPEN (J.) et VAN DROOGE (L.). – *Introducing « productive interactions » in social impact assessment*. Research Evaluation, 20(3), 211-8 (2011).
- [22] MOSTERT (S.), ELLENBROEK (S.), MEIJER (I.), VAN ARK (G.) et KLASSEN (E.). – *Societal output and use of research performed by health research groups*. Health Research Policy and Systems, 8(1), 30 (2010).
- [23] WEBER (M.). – *Le Savant et le Politique*. Plan 1959, 10/18, colt. « Bibliothèques » (1963, édition originale 1919).
- [24] JOLY (P.B.), GAUNAND (A.), COLINET (L.), LAREDO (P.), LEMARIE (S.) et MATT (M.). – *ASIRPA : A comprehensive theory-based approach to assessing the societal impacts of a research organization*. Research Evaluation, 24(4), 440-453 (2015) – *Évaluer l'impact sociétal de la recherche pour apprendre à le gérer : l'approche ASIRPA et l'exemple de la recherche agronomique*. Gérer et Comprendre – Les Annales des Mines. L'épreuve des faits 122: 31-42 (2015).
- [25] TSEY (K.), LAWSON (K.), KINCHIN (I.), BAINBRIDGE (R.), McCALMAN (J.), WATKIN (F.), CADET-JAMES (Y.) et ROSSETTO (A.). – *Evaluating Research Impact : The Development of a Research for Impact Tool*. Front. Public Health 4:160 (2016).
- [26] BLUNDO CANTO (G.), BARRET (D.), FAURE (G.), HAINZELIN (E.), MONIER (C.), TRIOMPHE (B.) et VALL (E.). – *ImpresS ex ante. Une proposition de démarche pour construire ex ante les chemins de l'impact*. Montpellier, France : Cirad, 64 p. ISBN : 978-2-87614-734-8 (2018).
- [27] BUXTON (M.) et HANNEY (S.). – *Assessing payback from Department of Health Research and Development : Preliminary report*. Uxbridge, United Kingdom : HERG, Brunel University (1994). – *How can payback from health services research be assessed ?* Journal of Health Services Research & Policy, 1(1), 35-43 (1996).
- [28] DONOVAN (C.) et HANNEY (S.). – *The « Payback Framework » explained*. Research Evaluation, 20(3), 181-183 (2011).
- [29] KLAUTZER (L.), HANNEY (S.), NASON (E.), RUBIN (J.), GRANT (J.) et WOODING (S.). – *Assessing policy and practice impacts of social science research : The application of the Payback Framework to assess the Future of Work programme*. Research Evaluation, 20(3), 201-209 (2011).
- [30] BARRET (D.), BLUNDO CANTO (G.), DABAT (M.H.), DEVAUX-SPATARAKIS (A.), FAURE (G.), HAINZELIN (E.), MATHE (S.), TEMPLE (L.), TOILLIER (A.), TRIOMPHE (B.) et VALL (E.). – *Guide méthodologique ImpresS : Évaluation ex post des impacts de la recherche agronomique dans les pays du Sud*. Montpellier, France : CIRAD, 96 p. ISBN 978-2-87614-731-7 (2017).
- [31] COLINET (L.), GAUNAND (A.), JOLY (P.B.) et MATT (M.). – *Des barèmes génériques pour évaluer les impacts de la recherche sur la société : l'exemple des impacts politiques*. Cahiers Agricultures, EDP Sciences, 26(6), 1-10 (2017).

- [32] STERN (N.). – *Research Excellence Framework : Building on success and learning from experience, an independent review*. UK Department for Business, Energy & Industrial Strategy (2016).
- [33] GOZLAN (C.). – *L'autonomie de la recherche scientifique en débats : évaluer « l'impact » social de la science*. Sociologie du Travail, Elsevier Masson, 57(2), 151-174 (2015).
- [34] BORNMANN (L.) et HAUNSCHILD (R.). – *Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact ?* Scientometrics 110, 937-943 (2017).
- [35] GARFIELD (E.). – *Citation analysis as a tool in journal evaluation*. Science, 178, 471-479 (1972).
- [36] HIRSCH (J. E.). – *An index to quantify an individual's scientific research output*. Proceedings of the National Academy of Sciences of the United States of America, 16569-16572 (2005).
- [37] WALTMAN (L.) et VAN ECK (N.J.). – *Source normalized indicators of citation impact : An overview of different approaches and an empirical comparison*. Scientometrics, 96(3), 699-716 (2013).
- [38] BERGSTROM (C.T.), WEST (J.D.) et WISEMAN (M.A.). – *The Eigenfactor™ metrics*. The Journal of Neuroscience, 28(45), 11433-11434 (2008).
- [39] GONZALEZ-PEREIRA (B.), GUERRERO-BOTE (V.P.) et MOYA-ANEGON (F.). – *A new approach to the metric of journals' scientific prestige : The SJR indicator*. Journal of Informetrics, 4(3), 379-391 (2010).
- [40] MOED (H.F.). – *Measuring contextual citation impact of scientific journals*. Journal of Informetrics, 4(3), 265-277 (2010).
- [41] THELWALL (M.) et MAFLAHI (N.). – *Guideline references and academic citations as evidence of the clinical value of health research*. Journal of the Association for Information Science and Technology, 67(4), 960-966 (2015).
- [42] BORNMANN (L.). – *Measuring impact in research evaluations : a thorough discussion of methods for, effects of and problems with impact measurements*. High Educ 73, 775-787 (2017).
- [43] NIEDERKROTENTHALER (T.), DORNER (T.E.) et MAIER (M.). – *Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists*. BMC Public Health, 11, 588 (2011).
- [44] KURTZ (M.J.) et BOLLEN (J.). – *Usage bibliometrics*. Annual Review of Information Science and Technology, 44(1), 1-64 (2010).
- [45] WHITE (H.D.), BOELL (S.K.), YU (H.), DAVIS (M.), WILSON (C.S.) et COLE (F.T.H.). – *Libcitations : A measure for comparative assessment of book publications in the humanities and social sciences*. Journal of the American Society for Information Science and Technology, 60(6), 1083-1096 (2009).
- [46] WILSDON (J.), ALLEN (L.), BELFIORE (E.), CAMPBELL (P.), CURRY (S.) et HILL (S.). – *The Metric Tide : Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Bristol, UK : Higher Education Funding Council for England (HEFCE) (2015).
- [47] NEYLON (C.) et WU (S.). – *Article-level metrics and the evolution of scientific impact*. PLoS Biology, 7(11) (2009).
- [48] PRIEM (J.), TARABORELLI (P.), GROTH (C.) et NEYLON (C.). – *Altmetrics : a manifesto*.(2010).
- [49] BORNMANN (L.). – *Do altmetrics point to the broader impact of research ? An overview of benefits and disadvantages of altmetrics*. Journal of Informetrics, 8(4), 895-903 (2014).
- [50] MACROBERTS (M.H.) et MACROBERTS (B.R.). – *Quantitative measures of communication in science – a study of the formal level*. Social Studies of Science 16(1), 151-172 (1986) – *Another test of the normative theory of citing*. Journal of the American Society for Information Science, 38(4), 305-306 (1987) – *Author motivation for not citing influences : a methodological note*. Journal of the American Society for Information Science, 39(6), 432-433 (1988) – *Citation content analysis of a botany journal*. Journal of the American Society for Information Science, 48(3), 274-275 (1997) – *Problems of citation analysis : A study of uncited and seldom-cited influences*. Journal of the American Society for Information Science and Technology, 61(1), 1-12 (2010).
- [51] ROBINSON-GARCIA (N.), VAN LEEUWEN (T.N.) et RAFOLS (I.). – *Using altmetrics for contextualised mapping of societal impact : from hits to networks*. Science and Public Policy (2017).
- [52] HAUSTEIN (S.), BOWMAN (T.D.) et COSTAS (R.). – *Interpreting « altmetrics » : viewing acts on social media through the lens of citation and social theories*. Sugimoto, C. R. (ed.), Theories of Informetrics and Scholarly Communication : A Festschrift in honor of Blaise Cronin, 372-406 (2015).
- [53] HICKS (D.), WOUTERS (P.), WALTMAN (L.), DE RIJCKE (S.) et RAFOLS (I.). – *Bibliometrics : The Leiden Manifesto for research metrics*. Nature, 520 (7548), 429-431 (2015).
- [54] WILSDON (J.), BAR-ILAN (J.), FRODEMAN (R.), LEX (E.), PETERS (I.) et WOUTERS (P.). – *Report of the European Commission Expert Group on Altmetrics. Métriques de nouvelle générations : métriques responsables et évaluation pour l'open science*. European Union (2017).
- [55] COSTAS (R.), VAN HONK (J.), ZAHEDI (Z.) et CALERO-MEDINA (C.). – *Discussing practical applications for altmetrics : social media profiles for African, European and North American publications*. Publication issue d'une présentation à la conférence 3 : AM, Bucarest (2016).
- [56] KE (Q.), AHN (Y.Y.) et SUGIMOTO (C.R.). – *A systematic identification and analysis of scientists on Twitter*. Plos One, 12(4), e0175368 (2017).
- [57] CHUBB (J.) et WATERMEYER (R.). – *Artifice or integrity in the marketization of research impact ? Investigating the moral economy of (pathways to) impact statements within research funding proposals in the UK and Australia*. Studies in Higher Education (2016).
- [58] PENFIELD (T.), BAKER (M. J.), SCOBLE (R.) et WYKES (M.C.). – *Assessment, evaluations, and definitions of research impact : A review*. Research Evaluation 23(1), 21-32 (2014).
- [59] SPAAPEN (J.B.), DIJSTELBLOEM (H.) et WAMELINK (F.). – *Evaluating research in context : A method for comprehensive assessment*. The Hague, the Netherlands : Consultative Committee of Sector Councils for Research and Development (2007).
- [60] BUXTON (M.). – *The payback of « Payback » : Challenges in assessing research impact*. Research Evaluation, 20(3), 259-260 (2011).
- [61] ERNO-KJOLHEDE (E.) et HANSSON (F.). – *Measuring research performance during a changing relationship between science and society*. Research Evaluation, 20(2), 131-143 (2011).
- [62] MCGILVRAY A. – *Research assessment : The limits of excellence*. Nature, 511(7510), S64-S66 (2014).
- [63] CORVOL (P.). – *Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique* (juin 2016).

À lire également dans nos bases

MAHE (A.). – *Altmetrics – Nouvelles mesures de la visibilité des résultats pour la recherche*. [H 7 730], Usages et management stratégique des documents numériques (2015).

Annuaire

Sites internet

Avis 9 du Comité consultatif commun d'éthique pour la recherche agronomique Inra-Cirad
<http://institut.inra.fr/Missions/Promouvoir-ethique-et-deontologie/Avis-du-comite-d-ethique/Avis-9-Evaluation-des-impacts-de-la-recherche-publique-agronomique>

GODIN (B.) et DORE (C.). – *Measuring the impacts of science ; beyond the economic dimension*
http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf

SIAMPI – *Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions* (2011)
<http://www.siampi.eu/>

Contribution du Comité consultatif national d'éthique à la réflexion dans le contexte de la 21^e Conférence sur les changements climatiques (nov. 2015)

http://www.ccne-ethique.fr/sites/default/files/publications/texte_ccne-cop21-6_novembre_2015.pdf

ASIRPA

<https://www6.inra.fr/asirpa>

ImpresS – Impact of Research in the South (Cirad)

<https://impress-impact-recherche.cirad.fr/>

Research Excellence Framework (REF)

<http://www.ref.ac.uk/>

STERN (N.). – Research Excellence Framework : Building on success and learning from experience, an independent review

<https://www.gov.uk/government/publications/research-excellence-framework-review>

Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES)

<https://www.hceres.fr/>

Altmetrics : a manifesto

<http://altmetrics.org/manifesto/>

San Francisco Declaration on Research Assessment (DORA) (2012)

<https://sfdora.org/read/>

Leiden Ranking

<http://www.leidenranking.com/>

Classement de Shanghai

<http://www.shanghairanking.com/fr/>