

**Principes et lignes
directrices de l'OCDE
pour l'accès aux
données de la recherche
financée sur fonds
publics**

Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics



ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

L'OCDE est un forum unique en son genre où les gouvernements de 30 démocraties œuvrent ensemble pour relever les défis économiques, sociaux et environnementaux que pose la mondialisation. L'OCDE est aussi à l'avant-garde des efforts entrepris pour comprendre les évolutions du monde actuel et les préoccupations qu'elles font naître. Elle aide les gouvernements à faire face à des situations nouvelles en examinant des thèmes tels que le gouvernement d'entreprise, l'économie de l'information et les défis posés par le vieillissement de la population. L'Organisation offre aux gouvernements un cadre leur permettant de comparer leurs expériences en matière de politiques, de chercher des réponses à des problèmes communs, d'identifier les bonnes pratiques et de travailler à la coordination des politiques nationales et internationales.

Les pays membres de l'OCDE sont : l'Allemagne, l'Australie, l'Autriche, la Belgique, le Canada, la Corée, le Danemark, l'Espagne, les États-Unis, la Finlande, la France, la Grèce, la Hongrie, l'Irlande, l'Islande, l'Italie, le Japon, le Luxembourg, le Mexique, la Norvège, la Nouvelle-Zélande, les Pays-Bas, la Pologne, le Portugal, la République slovaque, la République tchèque, le Royaume-Uni, la Suède, la Suisse et la Turquie. La Commission des Communautés européennes participe aux travaux de l'OCDE.

Les Éditions OCDE assurent une large diffusion aux travaux de l'Organisation. Ces derniers comprennent les résultats de l'activité de collecte de statistiques, les travaux de recherche menés sur des questions économiques, sociales et environnementales, ainsi que les conventions, les principes directeurs et les modèles développés par les pays membres.

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les interprétations exprimées ne reflètent pas nécessairement les vues de l'OCDE ou des gouvernements de ses pays membres.

© OCDE 2007

Toute reproduction, copie, transmission ou traduction de cette publication doit faire l'objet d'une autorisation écrite. Les demandes doivent être adressées aux Éditions OCDE rights@oecd.org ou par fax 33 1 45 24 99 30. Les demandes d'autorisation de photocopie partielle doivent être adressées au Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, 75006 Paris, France, fax 33 1 46 34 67 19, contact@cfcopies.com ou (pour les États-Unis exclusivement) au Copyright Clearance Center (CCC), 222 Rosewood Drive Danvers, MA 01923, USA, fax 1 978 646 8600, info@copyright.com.

Avant-propos

La recherche scientifique au service de l'innovation est essentielle pour relever les défis qui se posent à l'échelle mondiale, qu'il s'agisse des soins de santé et du changement climatique, de l'énergie renouvelable ou de la gestion des ressources naturelles. Pour en accroître le rythme et l'intensité, il importe de promouvoir des échanges collaboratifs entre les différentes communautés et d'assurer la diffusion la plus large possible de ses résultats. L'échange des idées, des connaissances et des données qui se font jour est fondamental pour le progrès de l'humanité et fait partie des valeurs essentielles portées par l'OCDE. Je suis donc particulièrement heureux que l'OCDE ait ouvert la voie en élaborant des principes et des normes qui facilitent l'accès aux données de la recherche financée sur des fonds publics.

Le développement rapide de l'informatique et de l'Internet a permis l'émergence de nouvelles applications pour les *sources* mêmes de l'étude scientifique que constituent les *données de la recherche*, dopant ainsi puissamment l'activité scientifique de ces dernières années. Les bases de données deviennent rapidement un élément essentiel de l'infrastructure du système scientifique mondial. Le Projet international sur le génome humain est un bon exemple, parmi d'autres, d'une entreprise à grande échelle dans laquelle une information librement accessible est exploitée avec succès par un large éventail d'utilisateurs, partout dans le monde et à des fins les plus diverses.

Par ailleurs, l'accès aux données de la recherche accroît le retour sur l'investissement public dans ce domaine, il renforce la liberté de l'investigation scientifique, il encourage la diversité des études et opinions, il favorise de nouveaux domaines d'activité et permet l'exploration de thèmes qui n'avaient pas été envisagés par les chercheurs d'origine.

En 2004, les ministres de la science et de la technologie ont demandé à l'OCDE de définir un ensemble de lignes directrices à partir de principes établis d'un commun accord en vue de faciliter un accès au moindre coût aux données numériques de la recherche financée sur fonds publics. Les Principes et lignes directrices présentés dans ce document font suite à cette demande. Ils sont destinés à aider tous acteurs intéressés à améliorer le partage des données de la recherche et de l'accès à ces données à l'échelle internationale. Maintenant que nous disposons de ce précieux instrument, j'appelle les responsables politiques et scientifiques à l'adopter sans réserve. Son utilisation, j'en suis convaincu, facilitera l'activité scientifique et contribuera donc à l'amélioration de la société.



Angel Gurría
Secrétaire général de l'OCDE

Table des matières

Contexte	7
Préface	11
Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics	17
I. Objectifs	17
II. Champ d'application et définitions	17
<i>Données de la recherche</i>	18
<i>Données de la recherche financée sur fonds publics</i>	18
<i>Dispositifs d'accès</i>	18
III. Principes	19
<i>A. Ouverture</i>	19
<i>B. Flexibilité</i>	19
<i>C. Transparence</i>	19
<i>D. Conformité au droit</i>	20
<i>E. Protection de la propriété intellectuelle</i>	21
<i>F. Responsabilité formelle</i>	21
<i>G. Professionnalisme</i>	23
<i>H. Interopérabilité</i>	23
<i>I. Qualité</i>	24
<i>J. Sécurité</i>	25
<i>K. Efficience</i>	25
<i>L. Responsabilité de rendre compte</i>	26
<i>M. Pérennité</i>	27

Contexte

Les ministres de la science et de la technologie des pays membres de l'OCDE se sont réunis à Paris en janvier 2004 pour étudier la nécessité d'élaborer des lignes directrices internationales sur l'accès aux données de la recherche. Lors de cette réunion, les gouvernements des 30 pays membres de l'OCDE ainsi que ceux de l'Afrique du Sud, de la Chine, d'Israël et de la Russie ont adopté une *Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics*, dans laquelle ils reconnaissent l'importance de l'accès aux données de la recherche et invitent l'OCDE « à définir un ensemble de lignes directrices fondées sur un ensemble convenu de principes destiné à faciliter l'accès, dans des conditions optimales d'efficacité par rapport aux coûts, aux données numériques de la recherche, pour approbation par le Conseil de l'OCDE à un stade ultérieur ».

Le Comité de la politique scientifique et technologique de l'OCDE, qui s'est chargé de cette demande, a commencé par demander à un groupe d'experts d'élaborer un ensemble de principes et lignes directrices. Ces experts ont rédigé un premier ensemble de principes et lignes directrices et entrepris plusieurs cycles de consultation avec des institutions de recherche et des organes directeurs dans les pays membres de l'Organisation afin de parvenir à un consensus. Un atelier organisé à Paris en février 2006 avec des acteurs clés a également contribué à ce processus. Les travaux menés pour la version finale ont montré que les pays membres manquaient encore de cadres internationaux visant à faciliter l'accès aux données, ils ont également révélé qu'un meilleur accès était généralement perçu comme un atout pour faire avancer la recherche, pour renforcer sa qualité et faciliter la coopération scientifique pluridisciplinaire. Les parties intéressées ont estimé que des lignes directrices internationales seraient utiles pour donner des orientations aux institutions cherchant à définir une ligne de conduite ainsi que pour renforcer la coopération internationale dans le domaine de la recherche. Les principes et lignes directrices élaborés à la suite de ce vaste processus de consultation ont été approuvés par le Comité de la politique scientifique et technologique de l'OCDE en octobre 2006, puis joints à une Recommandation de l'OCDE et entérinés par le Conseil de l'OCDE le 14 décembre 2006.

Les Recommandations de l'OCDE définissent des normes ou des objectifs collectifs et précis que les pays membres doivent en principe mettre en œuvre. Une Recommandation de l'OCDE est un instrument de l'Organisation qui n'est pas juridiquement contraignant mais qui, selon un usage bien établi dans les pays membres, est considéré comme ayant un grand poids moral. Les Recommandations de l'OCDE sont adoptées lorsque les gouvernements des pays membres sont prêts à prendre l'engagement politique d'appliquer les principes (et/ou lignes directrices) ainsi définis. Ce type d'instrument relève de ce que l'on appelle souvent le « droit non contraignant ».

Bien qu'une certaine latitude dans la rédaction de la norme soit admise, elle ne doit pas être suffisamment grande pour laisser aux pays membres la possibilité de modifier la norme ou l'objectif, et compromettre ainsi l'idée d'engagement de leur part. Cela dit, les Recommandations de l'OCDE laissent souvent une grande marge de manœuvre aux pays membres pour ce qui est des moyens de mise en œuvre, afin de prendre en compte les différents contextes juridiques, culturels, économiques et sociaux. Certains pays peuvent ainsi les mettre en œuvre par voie réglementaire, alors que d'autres choisiront par exemple d'entreprendre une action concertée avec les parties intéressées au niveau national.

Les Recommandations sont considérées comme des vecteurs de changement, et les pays membres de l'OCDE ne sont pas obligés, le jour de l'adoption, d'en respecter déjà les termes. Le but est qu'ils s'emploient sérieusement à atteindre la norme ou l'objectif défini dans des délais raisonnables, compte tenu du degré de difficulté auquel chaque pays membre est confronté pour y parvenir.

Dans sa Recommandation, le Conseil de l'OCDE demande que les pays membres prennent en considération les Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics qui sont exposés à l'annexe de la Recommandation, selon le contexte de chaque pays membre, pour élaborer des politiques et bonnes pratiques en relation avec l'accessibilité, l'utilisation et la gestion des données de la recherche. Il charge également le Comité de la politique scientifique et technologique de l'OCDE d'examiner la mise en œuvre de ladite Recommandation en tant que de besoin, ainsi que de réexaminer les Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics et, le cas échéant, de prendre en compte les progrès de la technologie et des pratiques de la recherche, en vue de resserrer encore davantage la coopération internationale.

Les *Principes et Lignes directrices* présentés ci-après ont été élaborés pour s'appliquer aux données de la recherche recueillies à l'aide de fonds publics pour produire des connaissances publiquement accessibles. La nature du « financement public » de la recherche varie considérablement d'un pays à l'autre, tout comme les politiques et pratiques actuelles en matière d'accès aux données en fonction des pays, des disciplines et des institutions. Ces différences plaident en faveur d'une approche flexible de l'élaboration de dispositifs d'accès aux données. Il appartiendra à chaque gouvernement national et à sa communauté scientifique d'arbitrer entre les coûts d'un meilleur accès aux données de la recherche et les bénéfices pouvant en résulter.

Préface

Augmenter le rendement des investissements publics dans la recherche scientifique

Les systèmes scientifiques publics des pays membres de l'OCDE sont fondés sur le principe de l'ouverture et de la libre circulation des idées, de l'information et du savoir. Les nouvelles technologies de l'information et des communications (TIC), aujourd'hui omniprésentes dans toutes les disciplines de recherche, ont beaucoup favorisé ce système de libre circulation et ont ouvert de nouvelles voies pour la collaboration et le partage. Mais le progrès scientifique n'est pas tributaire des seules technologies. Les politiques et pratiques de la recherche, ainsi que les systèmes qui la soutiennent et les valeurs culturelles sont autant de facteurs qui influent sur la nature et la fréquence des découvertes et le degré avec lequel celles-ci sont rendues accessibles et utilisées.

La puissance des ordinateurs et l'internet ont donné naissance à de nouveaux champs d'application, non seulement pour les résultats de la recherche, mais pour ses sources mêmes, à savoir le matériau de base que constituent les données de la recherche. De plus, les données de la recherche sous forme numérique sont de plus en plus utilisées dans des travaux qui vont au delà du projet en vue duquel elles ont été recueillies à l'origine, ainsi que dans d'autres domaines de recherche et dans l'industrie. Les données administratives des institutions des pays membres de l'OCDE telles que les données sur l'emploi, sont aujourd'hui largement utilisées en sciences sociales ainsi que pour l'élaboration des politiques. Les données des organismes de santé publique jouent un rôle de plus en plus important dans l'avancement des sciences du vivant. De même, les données géospatiales recueillies par un grand nombre d'organismes gouvernementaux différents sont essentielles pour la recherche en environnement et autres domaines. Et la liste est longue.

Les bases de données scientifiques deviennent rapidement un élément crucial de l'infrastructure du système scientifique mondial. Le Projet international sur le génome humain n'est qu'un exemple parmi d'autres d'une entreprise de recherche à grande échelle dans laquelle une source de données librement accessible est largement utilisée par de nombreux chercheurs différents, partout dans le monde, à différentes fins et dans différents

contextes. On peut facilement trouver un grand nombre d'autres exemples impliquant des entreprises de recherche, à grande ou à petite échelle.

Un accès efficace aux données de la recherche, utilisé de façon responsable et efficiente, est nécessaire pour tirer pleinement parti des nouvelles possibilités et retombées offertes par les TIC. L'accessibilité aux données de la recherche est devenue une importante condition pour :

- Une gestion avisée de l'investissement public dans l'information factuelle ;
- Le développement de chaînes d'innovation à forte valeur ;
- L'accroissement de la valeur procurée par la coopération internationale.

Plus spécifiquement, l'amélioration de l'accès aux données et du partage de celles-ci :

- Renforce la liberté de l'investigation scientifique ;
- Encourage la diversité d'analyse et d'opinion ;
- Stimule de nouvelles recherches ;
- Rend possible l'expérimentation d'hypothèses et de méthodes d'analyse nouvelles ou différentes ;
- Contribue aux études sur les méthodes de collecte de données et leur mesure ;
- Facilite la formation des nouveaux chercheurs ;
- Permet l'exploration de thèmes qui n'avaient pas été envisagés par les chercheurs d'origine ;
- Conduit à la création de nouveaux ensembles de données par la combinaison de données provenant de multiples sources.

L'accès ouvert aux données de la recherche financée sur fonds publics et leur partage contribuent non seulement à maximiser l'impact des nouvelles technologies et des nouveaux réseaux numériques sur le potentiel de recherche, mais permettent aussi un retour plus important sur l'investissement public dans la recherche.

Dans tous les pays membres de l'OCDE, des volumes sans cesse croissants de données sont collectés par les chercheurs et établissements de recherche financés sur fonds publics. Cette somme de données de recherche en expansion rapide représente à la fois un investissement massif de fonds publics et l'une des sources potentielles des connaissances nécessaires pour relever les multiples défis auxquels est confrontée l'humanité.

Pour promouvoir un meilleur rendement scientifique et social des investissements publics dans les données de la recherche, les pays membres de l'OCDE ont établi diverses lois, politiques et pratiques sur l'accès aux données de la recherche au niveau national. Dans ce contexte, des lignes directrices internationales constitueraient un atout important pour encourager les échanges et l'utilisation à l'échelle mondiale des données de la recherche.

Les présents *Principes et Lignes directrices* visent les données de la recherche qui sont recueillies à l'aide de fonds publics pour produire des connaissances publiquement accessibles. La nature du « financement public » de la recherche varie considérablement d'un pays à l'autre, tout comme les politiques et pratiques actuelles en matière d'accès aux données en fonction des pays, des disciplines et des institutions. Ces différences appellent une approche flexible de l'accès aux données, et la reconnaissance du fait qu'il n'y a pas de solution unique. En outre, il appartiendra aux différents gouvernements nationaux et à leur milieu scientifique d'arbitrer entre les coûts d'un meilleur accès aux données de recherche et les bénéfices pouvant en résulter.

Indépendamment des différences qui peuvent exister entre les pratiques et politiques en matière de partage de données, ainsi que des restrictions légitimes auxquelles peut être soumis l'accès aux données, un partage plus systématique serait bénéfique pour pratiquement tous les types de recherche. Comme le relèvent les auteurs de l'étude de l'US National Research Council, intitulée *Bits of Power* :

La valeur des données réside dans leur exploitation. L'accès total et ouvert aux données scientifiques devrait devenir la norme internationale pour l'échange des données scientifiques issues de la recherche financée sur fonds publics.

Les présents *Principes et Lignes directrices* ont pour finalités et objectifs spécifiques :

- De promouvoir une culture d'ouverture et de partage des données de la recherche entre les divers milieux de la recherche publique au sein des pays membres et à l'égard d'autres pays ;
- D'encourager l'échange de bonnes pratiques en matière d'accès et de partage des données ;
- De contribuer à une meilleure prise de conscience des coûts et bénéfices potentiels des restrictions et limitations actuelles visant l'accès aux données de la recherche financée sur fonds publics et le partage de ces données ;

- De mettre en évidence la nécessité de prendre en compte les réglementations et les pratiques en matière d'accès aux données et de partage des données dans l'élaboration des politiques et programmes scientifiques des pays membres ;
- De proposer un cadre de principes opérationnels, défini d'un commun accord, pour la mise en place de dispositifs d'accès aux données de la recherche dans les pays membres ;
- De fournir aux pays membres des recommandations sur les moyens d'améliorer le contexte international pour le partage et la diffusion des données de la recherche.

Les *Principes et Lignes directrices* exposés dans le présent document devraient aider les gouvernements, les organismes d'aide à la recherche et de financement de la recherche, les établissements de recherche et les chercheurs eux-mêmes à surmonter les obstacles et relever les défis liés à l'amélioration du partage des données de la recherche et de l'accès aux données, au plan international. Ces *Principes et Lignes directrices* doivent être interprétés et appliqués à la lumière des grandes questions que pose la fourniture de l'accès aux données :

- Questions technologiques : l'accès aux données de recherche et l'exploitation optimale de ces données nécessitent une infrastructure technologique adaptée, une large convergence de vues au plan international en matière d'interopérabilité, ainsi que des mécanismes efficaces de contrôle de la qualité des données.
- Questions institutionnelles et de gestion : bien qu'une meilleure accessibilité soit importante pour toutes les communautés scientifiques, la diversité des activités scientifiques donne à penser que des approches sur mesure et des modèles institutionnels variés de la gestion des données seront sans doute plus efficaces pour répondre aux besoins des chercheurs.
- Questions financières et budgétaires : l'infrastructure des données scientifiques nécessite une planification budgétaire spécifique continue et un soutien financier adéquat. L'exploitation des données de la recherche ne pourra être optimisée si les coûts d'accès, de gestion et de conservation doivent être imputés après coup aux projets de recherche. Il faut néanmoins souligner que les coûts de stockage et de gestion des données ont fortement diminué ces dernières années, et le manque d'informations sur ces changements peut en soi représenter un obstacle au progrès.

- Questions juridiques et de politiques : les législations nationales et les accords internationaux, notamment dans des domaines comme les droits de propriété intellectuelle et la protection de la vie privée, influent directement sur les pratiques en matière d'accès aux données et de partage des données et doivent être pleinement pris en compte dans l'élaboration des dispositifs d'accès aux données.
- Questions de culture et de comportement : la mise en place de mécanismes de formation et de rétribution adaptés est une composante nécessaire de la promotion des pratiques en matière d'accès et de partage des données. Ces considérations valent pour ceux qui financent, produisent, gèrent et exploitent les données de la recherche.

En oeuvrant pour améliorer l'accès aux données de la recherche dans le cadre de ces *Principes et Lignes directrices*, les pays membres devront arbitrer entre les coûts d'un meilleur accès à ces données et les bénéfices pouvant en découler. Les efforts déployés pour améliorer l'accès devront évidemment tenir compte des limitations financières existantes.

PRINCIPES ET LIGNES DIRECTRICES POUR L'ACCÈS AUX DONNÉES DE LA RECHERCHE FINANCÉE SUR FONDS PUBLICS

I. Objectifs

Ces *Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics* (ci-après dénommés « *Principes et Lignes directrices* ») fournissent aux organismes publics d'élaboration de la politique scientifique et de financement des pays membres des recommandations générales sur l'accès aux données de la recherche financée sur fonds publics. Ils ont pour but de promouvoir l'accès aux données et leur mise en commun entre les chercheurs, établissements de recherche et organismes de recherche nationaux, tout en reconnaissant et en prenant en compte la diversité des lois, des politiques de recherche et structures administratives des pays membres.

L'objectif ultime des *Principes et Lignes directrices* est d'améliorer l'efficacité et l'efficacité du système scientifique mondial, et non d'entraver son développement par de pesantes obligations et réglementations, ou d'imposer de nouveaux coûts aux systèmes scientifiques nationaux.

II. Champ d'application et définitions

Ces *Principes et Lignes directrices* visent les données de la recherche, déjà existantes ou à venir, qui sont subventionnées par des fonds publics pour produire des recherches et des connaissances scientifiques publiquement accessibles. Les *Principes et Lignes directrices* ne sont donc pas destinés à s'appliquer aux données de la recherche recueillies en vue de commercialiser les résultats de la recherche, ni aux données de la recherche qui appartiennent à une entité du secteur privé. L'accès à ces données fait intervenir tout un ensemble de considérations qui sortent du cadre du présent document. De plus, dans certains cas, l'accès aux données ou leur usage peut être limité afin de protéger la vie privée, la confidentialité ou des résultats couverts par un droit de propriété ou par la sécurité nationale.

Données de la recherche

Dans le cadre de ces *Principes et Lignes directrices*, les « données de la recherche » sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche.

Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris). L'accès à tous ces produits ou résultats de la recherche est régi par d'autres considérations que celles abordées ici.

Ces Principes et Lignes directrices portent essentiellement sur les données de la recherche sur support numérique exploitable sur ordinateur. C'est en effet ce format qui offre le plus de possibilités d'améliorer la distribution efficiente des données et leur application pour la recherche, dans la mesure où les coûts marginaux de la transmission de données via l'internet sont pratiquement nuls. Les *Principes et Lignes directrices* pourraient également s'appliquer à des données de recherche sous forme analogique lorsque les coûts marginaux d'accès à ces données peuvent être maintenus à un niveau raisonnablement bas.

Données de la recherche financée sur fonds publics

Les données de la recherche financée sur fonds publics sont définies comme les données provenant de recherches menées par des organismes ou services publics ou à l'aide de fonds publics, quel que soit le niveau de gouvernement qui les fournit. La nature des « fonds publics » attribués pour la recherche étant très variable d'un pays à l'autre, ces *Principes et Lignes directrices* reconnaissent que ces différences plaident en faveur d'une approche souple de l'amélioration de l'accès aux données de la recherche.

Dispositifs d'accès

Les dispositifs d'accès sont définis comme étant le cadre réglementaire, stratégique et procédural mis en place par les établissements de recherche, les organismes de financement de la recherche et les autres acteurs concernés pour déterminer les conditions d'accès aux données de la recherche et d'utilisation de ces données.

III. Principes

A. Ouverture

Par ouverture, on entend l'accès dans des conditions d'égalité de la communauté scientifique internationale, à un coût le plus bas possible, de préférence ne dépassant pas le coût marginal de la diffusion. L'accès ouvert aux données de la recherche financée sur fonds publics devrait être aisé, rapide, convivial et passer de préférence par l'internet.

B. Flexibilité

La flexibilité suppose de prendre en compte les évolutions rapides et souvent imprévisibles des technologies de l'information, les caractéristiques de chaque domaine de recherche et la diversité des systèmes de recherche, des systèmes juridiques et des cultures de chaque pays membre. Les implications nationales, sociales, économiques et réglementaires spécifiques devraient être analysées lorsque des organisations conçoivent des dispositifs d'accès aux données de la recherche, et lorsque les gouvernements élaborent des politiques pour promouvoir l'accès aux données et examinent la mise en œuvre de ces *Principes et Lignes directrices*.

C. Transparence

L'information sur les données de la recherche et les organisations productrices de données, la documentation sur les données ainsi que les spécifications des conditions qui régissent leur utilisation devraient être accessibles au plan international, en toute transparence, dans l'idéal via l'internet. Le manque de visibilité des sources de données de recherche existantes et futures fait sérieusement obstacle à l'accès.

Les facteurs à considérer pour assurer la transparence sont notamment les suivants :

- L'information sur les organisations productrices de données ainsi que sur les données qu'elles détiennent, leur documentation concernant les ensembles de données disponibles et les conditions d'utilisation de ces données devraient pouvoir être aisément accessibles sur l'internet.
- Les organisations de recherche et les institutions publiques de recherche devraient activement diffuser l'information relative à leurs politiques en matière de données de la recherche auprès des chercheurs, des associations académiques, des universités et des autres acteurs de la recherche financée sur fonds publics.

- Chaque fois que cela peut être utile, tous les membres des divers milieux de la recherche devraient contribuer à établir des accords sur des normes de catalogage des données. L'application des normes existantes devrait être envisagée, dans la mesure du possible, pour éviter de solliciter davantage les ressources de la recherche et d'accroître la charge de travail des chercheurs et de leurs institutions.
- Des informations devraient être échangées sur la gestion des données et les conditions de leur accès entre les institutions chargées des archives de données et celles productrices de données, de manière à mettre en commun les pratiques exemplaires.

D. Conformité au droit

Les dispositifs d'accès aux données devraient respecter les droits et intérêts légitimes de tous les acteurs de l'activité de recherche publique.

L'accès à certaines données de la recherche et leur utilisation seront nécessairement limités par divers types de prescriptions légales, qui peuvent imposer des restrictions pour raisons de :

- Sécurité nationale : certaines données relatives au renseignement, aux activités militaires ou à la prise de décision politique peuvent être classifiées et partant, soumises à un accès limité.
- Protection de la vie privée et confidentialité : les données relatives aux sujets humains et d'autres données personnelles font l'objet d'un accès limité en vertu des législations et des politiques nationales de protection de la confidentialité et de la vie privée. Il convient toutefois que les détenteurs de ces données envisagent des procédures d'anonymisation ou de confidentialité permettant d'assurer un niveau de confidentialité satisfaisant afin de préserver autant que possible l'utilité des données pour les chercheurs.
- Secrets commerciaux et droits de propriété intellectuelle : les données concernant les entreprises ou autres parties, ou provenant de celles-ci, qui contiennent des informations confidentielles peuvent être inaccessibles pour la recherche.
- Protection d'espèces rares, menacées ou en danger : dans certains cas, il peut exister des raisons légitimes de limiter l'accès aux données sur la localisation de ressources biologiques en vue d'assurer leur préservation.
- Procédure légale : les données en cours d'examen dans le cadre de poursuites en justice (sub judice) peuvent être inaccessibles.

L'adhésion à des codes de conduite professionnels peut faciliter le respect des prescriptions légales.

E. Protection de la propriété intellectuelle

Les dispositifs d'accès aux données devraient tenir compte de l'applicabilité du droit d'auteur ou des autres législations sur la propriété intellectuelle pouvant concerner les bases de données de la recherche financée sur fonds publics. Les facteurs à prendre en compte sont les suivants :

- Compte tenu de la multiplication des partenariats public-privé pour le financement de la recherche et la collecte de données destinées à la recherche, des arrangements équilibrés public-privé devraient faciliter, le cas échéant, un large accès aux données de la recherche. La participation du secteur privé à la collecte des données ne devrait pas, en soi, être une raison pour restreindre l'accès aux données. Il conviendrait de prendre en considération des mesures qui favorisent l'accès et l'exploitation à des fins non commerciales tout en protégeant les intérêts commerciaux, comme la communication différée ou partielle des données ou l'adoption volontaire de mécanismes d'octroi de licences. De telles mesures peuvent permettre aux principaux participants d'exploiter pleinement les données de la recherche, sans en interdire inutilement l'accès.
- Dans les juridictions où les données et informations issues de la recherche publique sont protégées par des droits de propriété intellectuelle, les détenteurs de ces droits devraient néanmoins faciliter l'accès à ces données, en particulier pour la recherche publique ou pour d'autres fins répondant à l'intérêt général.

F. Responsabilité formelle

Les dispositifs d'accès devraient promouvoir des pratiques institutionnelles explicites et formalisées, telles que l'élaboration de règles et de réglementations, sur les responsabilités des diverses parties intervenant dans les activités liées aux données. Ces pratiques devraient concerner la paternité des données, la mention des producteurs, la propriété, la diffusion, les restrictions d'utilisation, les modalités financières, les règles éthiques, les conditions de licence, la responsabilité civile et l'archivage durable.

Les dispositifs d'accès, au niveau gouvernemental comme institutionnel, devraient être élaborés en consultation avec des représentants de tous les acteurs directement concernés. Dans les programmes ou projets de recherche en collaboration, et particulièrement dans le cadre de la coopération scientifique internationale ou de projets de recherche fondés sur des partenariats public-privé s'appuyant sur des cadres réglementaires différents, les parties

intéressées devraient négocier des accords de partage de données le plus tôt possible au cours du projet de recherche, et dans l'idéal pendant la phase de proposition initiale. On pourra ainsi veiller à ce qu'il soit tenu compte comme il convient et en temps opportun de questions telles que l'allocation de ressources pour le partage et la préservation durable des données de la recherche, les différences dans les législations nationales sur les droits de propriété intellectuelle, les limitations motivées par des considérations de sécurité nationale, ainsi que la protection de la vie privée et de la confidentialité.

Les dispositifs d'accès devraient également prendre en compte des facteurs tels que les caractéristiques des données, leur valeur potentielle pour la recherche, le niveau de traitement des données (données brutes, partiellement traitées ou finales), le fait qu'il s'agisse de données homogènes issues des instruments ou capteurs d'une installation ou de données hétérogènes collectées sur le terrain par des chercheurs isolés, de données sur des sujets humains ou de paramètres physiques, ou encore de données qui sont ou non générées directement par une entité gouvernementale ou grâce à un financement public. Ces variantes dans l'origine ou la nature des données devraient être prises en considération lors de l'établissement des dispositifs d'accès aux données.

De plus, il convient de prendre en compte les éléments suivants :

- Bon nombre de problèmes liés à l'accès, à la diffusion et au partage de données résultent d'un manque d'accords institutionnels explicites sur les modalités d'accès et d'utilisation. La gestion des données devenant toujours plus complexe dans certains domaines de recherche, les arrangements informels conclus traditionnellement entre les chercheurs risquent de ne plus être adaptés et de devoir être complétés par des pratiques et des procédures convenues en bonne et due forme.
- La responsabilité des divers aspects de l'accès aux données et de leur gestion devrait être établie dans les documents concernant les tâches officielles des instituts, les demandes de subventions, les contrats de recherche, les accords de publication et les licences, par exemple.
- La pérennité de l'infrastructure requise pour l'accès aux données revêt une importance particulière. Les établissements de recherche et les organismes publics devraient assumer officiellement la responsabilité de faire en sorte que les données de la recherche soient efficacement préservées, gérées et rendues accessibles de façon à pouvoir être exploitées de manière efficiente et adéquate sur le long terme.

G. Professionnalisme

Les dispositifs institutionnels pour la gestion des données de la recherche devraient être fondés sur les normes professionnelles applicables et sur les valeurs inscrites dans les codes de conduite des milieux scientifiques concernés.

Les facteurs à prendre en compte sont les suivants :

- L'utilisation de codes de conduite/de déontologie par les chercheurs et la communauté scientifique pourrait contribuer à simplifier et à réduire la réglementation régissant l'accès.
- La confiance mutuelle entre les chercheurs, ainsi qu'entre ces derniers et leurs institutions et autres organisations, joue un rôle important dans l'élaboration et le maintien de ces codes de conduite.
- Dans la pratique courante de la recherche, le chercheur ou l'institution qui produit initialement les données est parfois récompensé par l'utilisation exclusive temporaire des données. Les règles applicables à de tels arrangements incitatifs devraient être élaborées et explicitement précisées par les organismes sources de financement, en coopération avec les milieux de la recherche concernés.

Dans certains domaines scientifiques, le manque de préparation et de mise en œuvre de la documentation adéquate et de l'archivage des ensembles de données constitue l'un des obstacles majeurs à l'exploitation maximale de l'investissement dans les données de recherche. Lors de la planification des projets et programmes, à tous les niveaux, les problèmes de données devraient être abordés dès les tout premiers stades afin de prendre en compte les besoins en termes de fonds ou d'assistance technique pour les activités essentielles d'organisation et de conservation de ces ensembles de données. Il conviendrait de prêter attention aux incitations et au développement des compétences professionnelles dans tous les domaines de la gestion des données de recherche.

H. Interopérabilité

L'interopérabilité technologique et sémantique est essentielle pour faciliter et encourager l'accessibilité et l'utilisation des données de la recherche dans un contexte international et interdisciplinaire. Les dispositifs d'accès devraient tenir dûment compte des normes internationales pertinentes applicables en matière de documentation des données. Les pays membres et les établissements de recherche devraient coopérer avec des organisations internationales chargées de l'élaboration de nouvelles normes.

Bien que la science devienne une entreprise fortement mondialisée, l'incompatibilité des normes techniques et procédurales utilisées peut être un obstacle très sérieux aux multiples usages des ensembles de données.

Les facteurs qui devraient notamment être pris en compte sont les suivants :

- Il conviendrait de mentionner explicitement les normes employées dans la mesure où il s'agit de la première exigence de l'interopérabilité.
- Le fait de prendre pour modèle les disciplines les plus avancées à cet égard devrait être encouragé, en particulier par les organisations professionnelles internationales se consacrant à la science ainsi qu'à la collecte et à la conservation des données à des fins scientifiques et technologiques.
- Les travaux d'organisations chargées de définir des normes plus générales concernant les technologies de l'information et des communications devraient également être pris en considération.

I. Qualité

La valeur et l'utilité des données de recherche dépendent pour une large part de la qualité des données elles-mêmes. Les gestionnaires de données et les organisations de collecte de données devraient particulièrement veiller au respect de normes de qualité explicites. Lorsque ces normes n'existent pas encore, les institutions et les associations de recherche devraient, avec leurs chercheurs, se consacrer à leur élaboration. Si tous les domaines de la recherche peuvent tirer parti d'une meilleure qualité des données, certains exigent des normes beaucoup plus rigoureuses que d'autres. Pour cette seule raison, il n'est pas réaliste d'envisager des normes universelles de qualité des données. Les normes devraient être élaborées en consultation avec les chercheurs pour s'assurer que le niveau de qualité et de précision répond aux besoins des différentes disciplines.

Plus spécifiquement,

- Les dispositifs d'accès aux données devraient décrire les pratiques exemplaires quant aux méthodes, techniques et instruments employés pour le recueil, la diffusion et l'archivage accessible des données, afin de permettre un contrôle de qualité par un examen mutuel et d'autres moyens assurant la qualité et l'authenticité.
- L'origine des sources devrait être établie et spécifiée de façon vérifiable. Tous ceux qui souhaitent utiliser les données devraient pouvoir obtenir facilement ces informations, qu'il conviendrait

d'incorporer dans les métadonnées accompagnant les ensembles de données. Il est important de constituer ces métadonnées pour permettre aux scientifiques de comprendre les implications exactes des ensembles de données.

- Autant que possible, l'accès aux données devrait être lié à l'accès aux documents de recherche originaux, et les reproductions des ensembles de données devraient être liées aux originaux pour faciliter la validation des données et la détection d'erreurs dans ces ensembles.
- Les établissements de recherche et les associations professionnelles devraient concevoir des pratiques adéquates concernant les références de données et l'enregistrement de ces références sous forme d'index, dans la mesure où ce sont des indicateurs précieux de la qualité des données.

J. Sécurité

Il convient de s'attacher en particulier à encourager l'utilisation de techniques et d'instruments destinés à garantir l'intégrité et la sécurité des données de recherche. En ce qui concerne la garantie de l'intégrité d'un ensemble de données, tout devrait être mis en œuvre pour s'assurer du caractère complet des données et de l'absence d'erreurs. En ce qui concerne la sécurité, les données, de même que les métadonnées et descriptions correspondantes, devraient être protégées contre la perte, la destruction, la modification et l'accès non autorisé, intentionnels ou non, en conformité avec des protocoles de sécurité explicites. Les ensembles de données et les équipements servant à leur conservation devraient également être protégés contre les risques environnementaux tels que chaleur, poussières, surtensions, champs magnétiques et décharges électrostatiques.

K. Efficience

L'un des buts essentiels poursuivis en s'attachant à promouvoir l'accès et le partage des données est d'améliorer l'efficience globale de la recherche scientifique financée sur fonds publics afin d'éviter une duplication inutile et coûteuse des efforts de collecte de données.

Il convient de prendre en compte les aspects suivants :

- Les dispositifs d'accès aux données devraient favoriser l'amélioration du rapport coût-efficacité au sein du système scientifique mondial par la description des pratiques exemplaires dans les services de gestion de données et de soutien spécialisé.

- Si les données de la recherche financée sur fonds publics sont soumises à la règle par défaut d'ouverture énoncée en vertu du Principe A, cela ne signifie pas pour autant que toutes ces données doivent être conservées de manière permanente. Les acteurs de l'archivage de données devraient mener des évaluations coûts-avantages régulières et concevoir et affiner en permanence des protocoles visant à s'assurer que les ensembles de données ayant la plus grande utilité potentielle sont conservés et rendus accessibles. Le recours à des protocoles acceptés de conservation et à une information détaillée sur les données devrait contribuer à réduire les doubles emplois ainsi qu'à établir la sélectivité nécessaire en matière de conservation.
- Des services de soutien spécialisés, par exemple dans le cadre d'une collaboration sur des projets de recherche spécifiques avec des spécialistes hors de la sphère universitaire ou le recours à des organisations spécialisées dans la gestion des données, devraient être envisagés pour veiller à la rentabilité de la production, de l'utilisation, de la gestion et de l'archivage des données de la recherche.
- Si les chercheurs ou les producteurs de bases de données ne bénéficient pas d'incitations suffisantes, ils risquent de relâcher leurs efforts dans les activités liées aux données. Il conviendrait pour éviter ce problème d'envisager l'élaboration de nouvelles structures d'incitations et l'adaptation des structures existantes, notamment la prise en compte des activités de gestion de données dans les procédures de nomination et de promotion.

L. Responsabilité de rendre compte

Le fonctionnement des dispositifs d'accès aux données devrait faire l'objet d'une évaluation périodique par les groupes d'utilisateurs, les institutions responsables et les organismes de financement de la recherche. Même si chaque partie utilisera sans doute des critères d'évaluation quelque peu différents, la somme totale des résultats devrait donner une image détaillée de la valeur des données et des régimes d'accès aux données. Ces évaluations devraient contribuer à accroître le soutien en faveur du libre accès aux données par le milieu scientifique et la collectivité.

Il convient de prendre en compte les aspects suivants pour l'élaboration des critères d'évaluation :

- Les investissements publics globaux dans la production et la gestion des données de la recherche.
- Les performances de gestion des organismes de collecte et d'archivage de données.
- Le degré de réutilisation des ensembles de données existants.
- Les connaissances générées par le réemploi de données existantes.
- Le recours à des exercices de prospective ciblés afin de déterminer la nature et le champ des activités de préservation des données et les catégories de données les plus susceptibles d'être nécessaires à l'avenir.

Même s'il ne sera pas facile de se faire une idée claire et précise des coûts, des avantages et des résultats des dispositifs d'accès aux données, les responsables des dispositifs d'accès aux données devraient s'attacher à montrer les avantages d'un libre accès aux données afin de justifier et d'obtenir un soutien durable de la part de tous les niveaux de gouvernement.

M. Pérennité

Il conviendrait de tenir dûment compte de la pérennité de l'accès aux données de la recherche financée sur fonds publics comme l'un des éléments-clés des infrastructures de recherche. Cela suppose d'assumer la responsabilité administrative des mesures destinées à garantir un accès permanent aux données jugées comme devant être durablement conservées. La tâche peut être difficile étant donné que la plupart des projets de recherche et les financements publics accordés sont de durée limitée, alors que garantir l'accès aux données produites est une entreprise qui s'inscrit dans la durée. Les organismes de financement de la recherche et les établissements de recherche devraient donc étudier la préservation à long terme des données dès le début de chaque nouveau projet, et notamment rechercher les structures d'archivage les plus appropriées pour les données.

LES ÉDITIONS DE L'OCDE, 2, rue André-Pascal, 75775 PARIS CEDEX 16
IMPRIMÉ EN FRANCE
(00 2007 39 3 P) – n° 88180 2007