

Fiche-projet pour un exercice de prospective inter-organismes

Big&Open Data en recherche à l'horizon 2040

La vision prospective d'acteurs de la recherche publique

Le Réseau PROSPER propose de mettre en place et d'animer un groupe de travail inter-organismes de recherche, constitué d'acteurs de la recherche motivés pour se projeter dans ce que la combinaison de Big et Open Data pourrait induire comme transformations à l'horizon 2040, tant dans l'activité de recherche scientifique que dans la nature des connaissances produites, dans le statut des savoirs ou encore dans le rapport que la société entretient avec la science et les institutions de recherche, afin d'éclairer la communauté nationale pour mieux anticiper ces changements.

Le positionnement de l'exercice de prospective

Les bouleversements induits par la conjonction du *Big Data* et de l'*Open Data* impactent globalement la société et sont l'objet de réflexions dans de multiples cercles. Dans le monde de la recherche, chacune et chacun pressent, plus ou moins confusément, qu'au-delà de ce qui est déjà visible aujourd'hui, c'est à une transformation profonde des pratiques, des organisations, des sujets de recherche et des interactions avec la société que l'on va assister. Comment peut-on se représenter les futurs possibles de la recherche au-delà de l'horizon de visibilité actuel ? Comment notre communauté nationale de recherche, et singulièrement nos organismes publics de recherche, peuvent-ils s'y préparer dès maintenant pour se positionner au mieux dans ce que ce monde deviendra ? C'est l'objet de l'exercice de prospective proposée par le réseau PROSPER.

L'idée directrice de cet exercice est de construire ce panorama prospectif en impliquant une variété d'acteurs issus de nos organismes, constituant un échantillonnage le plus large possible des différents profils opérationnels et fonctionnels concernés, pour apporter une diversité de visions et de préoccupations vis-à-vis de la question posée.

Construire une vision prospective capable de s'échapper du prêt-à-penser ambiant et, d'une manière générale, de tout ce qui bride notre capacité d'imaginer des futurs qui ne ressemblent pas à ce que nous connaissons aujourd'hui, nécessite des méthodes appropriées. C'est la fonction d'accompagnement méthodologique qu'assurera un groupe de pilotage transversal, constitué de membres du Réseau PROSPER expérimentés.

Le mode opératoire

C'est à chacun des membres du Réseau PROSPER qu'il appartient d'être le promoteur de cet exercice au sein de son organisme : il le fait connaître, en explique l'intérêt, veille à permettre l'implication d'acteurs de son organisme au niveau souhaité, puis assure le suivi et la valorisation des travaux au bénéfice de son organisme.

L'Atelier-séminaire du 14 juin 2017

Un exercice de prospective est d'autant plus utile qu'il répond à des attentes clairement exprimées. C'est d'autant plus vrai lorsqu'il s'agit de problématiques très vastes (et pour lesquelles il existe déjà des réflexions), comme peut l'être la combinaison du Big et de l'Open Data en recherche. La première étape a donc pour objet de construire précisément le questionnement que devra éclairer l'exercice de prospective. Pour cela, chacun des organismes intéressés par le projet sera invité, au cours du séminaire du 14 juin, à présenter les questions qu'il souhaite voir éclairer par l'exercice. Du débat qui suivra émergeront les principales dimensions de l'exercice de prospective, permettant à chaque organisme de mesurer son intérêt pour permettre à une ou plusieurs personnes de s'impliquer dans le groupe de travail qui sera mis sur pied pour la rentrée 2017.

Le déroulement ultérieur de l'exercice

Le dimensionnement exact de l'exercice de prospective sera défini à la suite de l'Atelier-séminaire de juin 2017. Sa durée devrait être de l'ordre d'un an, avec cinq à six réunions d'une demi-journée (ou d'une journée si l'ampleur prise par les travaux le nécessite). Le groupe de pilotage de l'exercice assurera tout le travail de back office, de manière à ce que les réunions soient très productives et la consommation de temps des participants minimale. Par contre, pour assurer la dynamique nécessaire à l'élaboration de la vision prospective, il sera attendu des participants qu'ils s'engagent à participer à l'ensemble de l'exercice.

Le Réseau PROSPER est la structure de rencontre, de concertation et d'action de la communauté des responsables de prospective de la recherche publique française. Etant à la fois un réseau-métier et un réseau-fonction, il développe une expertise qui porte aussi bien sur les processus mis en œuvre dans les actions de prospective, que sur les apports de ces travaux aux plans de leur intérêt scientifique et de leur utilité sociale.

Annexe

Big&Open Data en recherche : En quoi est-ce une « révolution » ?

La conjonction des phénomènes Big et Open Data ouvre sur une science de quatrième génération, de nature à générer de profondes transformations tant dans l'activité de recherche scientifique que dans la nature des connaissances produites, dans le statut des savoirs ou encore dans le rapport que la société entretient avec la science et les institutions de recherche. Quelques éléments d'éclairage en sont donnés ici.

► **Le quatrième paradigme de la science**

La place croissante prise par le numérique dans le monde contemporain impacte globalement l'écosystème des connaissances, et plus particulièrement le monde de la recherche. L'explosion des quantités de données disponibles et des moyens de les traiter d'une part (*Big Data*), l'ouverture des sources de données permettant au plus grand nombre de s'en saisir d'autre part (*Open Data*), font des données numériques une nouvelle matière de travail qui prend de plus en plus d'importance. Le monde de la recherche, dans ce qu'il était organisé autour de capacités d'observation et d'expérimentation par domaines disciplinaires, se trouve interpellé par cette révolution qui ouvre de nouvelles perspectives à l'activité scientifique, en considérant que les masses de données numérisées constituent par elles-mêmes une matière nouvelle essentielle pour la recherche, accessibles par chacun bien au-delà de son champ d'expérimentation et d'expertise habituel.

Après la science empirique des temps anciens, puis le développement d'une branche théorique suivie d'une autre de modélisation numérique, on entre donc aujourd'hui dans une science de quatrième génération avec la *e-science* reposant sur l'exploration des données. Cette capacité d'explorer massivement les données révolutionne complètement la recherche scientifique, en ouvrant sur de nouveaux horizons, de nouvelles méthodes, de nouveaux regards scientifiques, de nouveaux instruments ou encore de nouvelles expériences scientifiques.

Ce sont ces nouveaux territoires qu'il s'agit d'explorer par l'exercice de prospective proposé, avec toutes les conséquences sur les communautés scientifiques en matière de communication, d'éthique, d'intégrité scientifique..., mais aussi plus largement sur les relations entre science et société, où on retrouve bien sûr les questions d'éthique, mais aussi les aspects légaux, les questions de responsabilités individuelles et collectives, ou encore l'éducation et l'apprentissage dans un monde nouveau, etc.

Cette révolution étant fondée sur l'exploration de masses de données numérisées, les premières questions portent évidemment sur la production, la gestion et l'exploitation de celles-ci : Comment et par qui toute la matière numérique produite dans le cadre des activités de recherche va-t-elle être gérée ? Ayant facilement accès à toute cette matière, des chercheurs amateurs vont-ils déborder les professionnels ? Des programmes "intelligents" vont-ils eux-mêmes produire des *connaissances de synthèse* de nature nouvelle, difficilement vérifiables par l'homme ? Qui va qualifier et valider les résultats produits ? Comment seront-ils diffusés ? Quelle sera leur valeur scientifique, mais aussi économique ? Etc.

► **Différents volets de la révolution *Big&Open Data***

La transformation de l'activité de recherche

Depuis l'antiquité, l'essentiel de l'activité scientifique repose sur l'observation directe du monde qui nous entoure, afin de le décrire précisément, d'en comprendre les mécanismes d'évolution et d'en formuler les lois. Organisée en tant que métier, cette activité va de la paillasse ou du terrain d'observation jusqu'à la publication des résultats dans des médias appropriés, après validation par les pairs de la discipline concernée. Il s'agit donc d'une activité plutôt verticalisée, réservée aux spécialistes du domaine, produisant des résultats validés par le rebouclage au réel et la reproductibilité avérée.

La conjonction des phénomènes Big et Open Data apporte une rupture en ce qu'elle permet au plus grand nombre de travailler transversalement sur de multiples sources de données numériques existantes, sans nécessiter l'expertise de la réalité qu'elles codifient. Tous les métiers liés à la gestion et au traitement de

données prennent de l'ampleur, au détriment des sciences de l'observation pouvant parfois être considérées comme simples pourvoyeuses de données. Ces métiers débordent largement le monde traditionnel de la recherche, englobant des professionnels du monde économique mais aussi de simples citoyens : chacun peut aujourd'hui devenir acteur de la recherche.

La transformation dans la nature des connaissances produites

Le développement des outils statistiques permet de traiter de plus en plus finement des masses de données considérables, pour faire apparaître des résultats invisibles à l'échelle des traitements manuels. Ceci constitue une opportunité importante de repousser les limites de la connaissance dans un certain nombre de domaines (comme en épidémiologie, en sciences humaines et sociales ou en astronomie et astrophysique). Mais la question des limites de validité de ces traitements se pose, avec le risque de confusion entre causalité et corrélation, c'est-à-dire d'ériger en loi ce qui n'est qu'une concomitance liée à des conditions particulières. Des questions d'éthique sont aussi soulevées lors du traitement de données concernant les personnes.

Les technologies de *Text and Data Mining*, en constante évolution, permettent aussi d'ouvrir de nouveaux champs d'analyse pour la recherche : après le microscope et le télescope – qui n'étaient que des prolongements de l'œil –, les outils de traitement du *Big Data* offrent de nouvelles modalités d'appréhension de la réalité de l'état de la connaissance ou de la société, tout en permettant la production de connaissances nouvelles. On arrive ainsi à produire des méta-connaissances prenant comme données d'entrée des connaissances déjà élaborées (mais pas forcément elles-mêmes validées avant mise en ligne), avec tous les risques associés aux travaux en cascade.

La transformation du statut des savoirs

Dans une économie fondée sur la connaissance, la nature opérationnelle des savoirs prend le pas sur leur importance cognitive, sur leur fonction explicative des mécanismes qui régissent le monde. De ce fait, une affirmation à la portée pratique évidente pour tous, fondée sur une pseudo argumentation scientifique, prend la même valeur qu'une connaissance résultant d'un travail scientifique validé par les pairs. Faute d'une éducation suffisante à ce qu'est réellement une connaissance scientifique validée, celle-ci est trop souvent considérée, dans les débats, comme un opinion au même titre que d'autres.

Les questions de régulation et de validation des savoirs, dont les sources vont se multiplier, sont donc essentielles. Les données, accessibles sur internet, utilisées pour générer ces nouveaux savoirs soulèvent elles-mêmes question, dans la mesure où c'est de leur exactitude et de leurs limites d'utilisation que dépendent les résultats que l'on va pouvoir en tirer.

La détention tant des données que des résultats de recherche devient d'importance stratégique : tous les acteurs de l'économie tentent de se positionner sur ces créneaux, qui pour partie échappent au monde scientifique. Mais ce qui échappe pourrait être dans un second temps récupéré aussi par la science pour ouvrir de nouveaux champs...

La transformation des rapports que la société entretient avec la science et les institutions de recherche

Dans de multiples domaines où la collecte des données se fait de manière très distribuée (environnement, santé...), la contribution du public est essentielle pour progresser, elle est rendue possible par l'ouverture apportée par internet. Les travaux de recherche conduits en interaction avec les différents acteurs de la société pourraient, par l'implication de ces acteurs, apporter une dimension nouvelle aux connaissances produites, et faciliter le passage aux applications et innovations, faisant évoluer l'image de la science dans la société. La société pourrait de ce fait se réinvestir dans la science, par une participation active à la création de nouvelles connaissances, mais aussi en redonnant à la science une place significative dans le projet de société.

D'un autre côté, le risque existe que la société se désintéresse de la compréhension des lois d'évolution du monde, qui fonde la science actuelle, au profit du développement d'algorithmes prédictifs : pourquoi financer une recherche de la compréhension des mécanismes à l'œuvre dans un domaine, si un système expert sait directement prédire ce qui va se passer ?

