

Quelle(s) stratégie(s) de recherche face à la nouvelle massification des données ?

Mardi 2 décembre 2014 – Auditorium de la BULAC

65 Rue des Grands Moulins – Paris 13^{ème}

INTRODUCTION

Au moins depuis l'après-guerre, la science produit un très grand nombre de données au point que l'on a pu parler de Data Deluge pour nommer l'accroissement significatif du volume de données disponibles à la fois pour la recherche et le grand public.

Avec la révolution numérique et la généralisation de l'Internet au début des années quatre-vingt-dix, un nouvel accroissement sans précédent de l'information disponible et sa circulation facilitée sont venus bousculer notre perception même de ces données. Favoriser encore cette circulation et l'interconnexion des données, au-delà des formats et des fonctionnements en silos est aujourd'hui un enjeu primordial pour la recherche et les stratégies d'innovation.

Open data, open access, que faire des immenses masses de données disponibles et qui sont à 98 % numériques ? Comment gérer, indexer, conserver, actionner, recouper toutes ces données dans un monde où la « datafication » n'en est qu'à ces débuts ? Quelles conséquences sur la façon de faire de la science, sur les stratégies scientifiques, sur l'économie numérique et l'innovation ?

PROGRAMME

8h30 - 9h00 : Café d'accueil

9h00 - 09h30 : Ouverture

Introduction de Christophe Pérales, président de l'ADBU et **Jean-Michel Catin**, directeur de la rédaction d'AEF

Renaud Fabre, directeur, DIST-CNRS

Jean-François Balaudé, CPU, président de l'université Paris-Ouest Nanterre La Défense

9h30 - 11h00 : Session 1

"Les données numériques, un phénomène massif difficile à maîtriser"

La recherche fait aujourd'hui face à une problématique critique : l'inflation exponentielle de l'IST qui questionne les moyens à mettre en œuvre pour gérer les stocks et les flux de données, et le rôle des professionnels de la documentation auprès des chercheurs. Comment se caractérise ce mouvement massif de production de données ? Quels sont les enjeux en termes de pérennité des données et de pratiques d'indexation associées ? Que faut-il archiver ? Comment et combien de temps ? Avec quelles garanties ? Il est devenu fondamental que les données, pour qu'elles soient exploitables par le chercheur, soient documentées. Les jeux de données, produits dans des contextes et sous des formats différents, doivent aussi pouvoir être mis en relation et donc être techniquement interopérables (web sémantique).

Gildas Illien, directeur du département de l'information bibliographique et numérique à la BNF

Christophe Pérales, président de l'ADBU

Les technologies de big data et de web sémantique peuvent-elles changer la donne et ouvrir de nouveaux horizons ? Les industriels sont par exemple en avance sur le traitement des données en agroalimentaire et climat (pour anticiper les récoltes et donc les flux de marché) avec des approches de type data mining. Comment créer des systèmes interopérables et parfaitement documentés pour que les données puissent être correctement utilisées ?

Françoise Genova, directrice du centre de données astronomiques de Strasbourg

Jérôme Weinbach, (Inserm U933, Hôpital Trousseau), directeur scientifique et opérationnel du programme national cohortes maladies rares RaDiCo

"Exploitation des données : un environnement législatif en évolution"

Derrière la gestion des données se posent les questions relatives aux possibilités de les stocker, de consulter, de réutiliser, de se les approprier. Les débats autour des droits d'auteur investissent maintenant le champ des bases de données de la recherche, à l'instar du TDM (text data mining) discuté à l'échelle européenne notamment.

Les chercheurs ont besoin de croiser et d'analyser des données issues de sources différentes. Le droit des bases de données doit-il aménager pour faire face à ces nouvelles pratiques ? L'activité scientifique dépend aussi de la reproductibilité des résultats obtenus : quelles conséquences alors pour les labos ? Les éditeurs proposent aujourd'hui un arrangement par voie contractuelle : quelles conséquences pour les chercheurs ?

Négocier avec les éditeurs l'ouverture de droits se ferait en soumettant les projets de text and data mining pour susciter leur accord. Avoir un droit de regard sur les recherches en cours ne dépendrait donc plus uniquement des chercheurs, et les éditeurs se trouveraient alors dans la situation de connaître à l'avance des orientations de recherche. L'indépendance des acteurs de recherche est-elle sur la sellette ?

Lionel Maurel, juriste et bibliothécaire, BDIC (Bibliothèque de Documentation Internationale Contemporaine), animateur du blog S.I.Lex

Sarah Jones, Senior Institutional Support Officer, DCC (Digital Curation Center) du JISC, Glasgow University

11h00 - 11h30 : Pause

11h30 - 12h45 : Session 2

"Open science: entre géostratégie et intérêts économiques, l'enjeu est-il raisonnable?"

Le 7ème PCRD prévoit que les publications financées par l'Union européenne doivent être placées en open access. Dans le cadre de "Horizon 2020" le principe sera étendu aux données, à titre expérimental. Cette idée pose problème pour certains industriels notamment, et interroge sur l'équilibre géopolitique: l'open data est-il possible si certains seulement jouent le jeu ? Doit-on ouvrir les données de la recherche dans tous les domaines ?

L'open data concerne également la transparence des budgets alloués, les données relatives aux laboratoires eux-mêmes. Comment intéresser les labos et les faire adhérer à l'approche ? Comment structurer les stratégies d'open data pour augmenter leur efficacité et faciliter le mouvement ?

En accédant librement aux données, la communauté des chercheurs peut vérifier que les résultats de la recherche n'ont pas été falsifiés et s'assurer de la qualité d'une publication. En outre, la très grande majorité des données issues de la recherche n'apparaît pas dans des publications et reste donc ignorée. Certains chercheurs peuvent-ils a contrario avoir intérêt à ne pas ouvrir les données ? Comment encourager une logique de "data driven science" ?

Renaud Fabre, directeur de l'IST du CNRS

Bruno Latour, professeur des universités, directeur scientifique de SciencesPo et membre de médialab (*sous réserve*)

Jean Chambaz, président de l'UPMC

12H45 - 14H : Cocktail déjeunatoire

14h15 - 15h15 : Session 3

"Les éditeurs face aux stratégies de recherche, quelles sont les politiques à l'œuvre ?"

Les éditeurs comme Elsevier ont bien compris que l'avenir se trouve dans les outils de gestion de la recherche, et proposent désormais des systèmes d'information clé en main. Un logiciel comme Scival pour la gestion de la veille et des stratégies scientifiques permettent de tracker les comportements des chercheurs dans leurs utilisations des données et des publications. Quels sont les besoins des communautés en outils de gestion ? Comment se positionnent-ils face à l'offre des éditeurs ? En quoi cela interroge-t-il les pratiques actuelles de veille et d'orientation stratégiques de la recherche ?

Mondane Marchand, Sales Manager Europe TR Scientific & Scholarly Research, Thomson Reuters

Valerie Thiel Mba, Regional Solution Sales Director Europe, research intelligence, Elsevier

Marie Farge, physicienne, directrice de recherche, CNRS

15h15 - 15h30 : Pause

15h30 - 17h15 : Session 4

"Les stratégies de recherche à l'heure des mutations..."

Le nouveau data déluge et l'open data ouvrent des perspectives inédites en termes de pilotage scientifique. Comment identifier les axes de travail, les partenariats potentiels, les concurrences à veiller, à l'échelle d'une région, d'un pays, d'un continent, et différencier les activités des chercheurs au sein des grandes thématiques classiques (santé, environnement, etc.) ? L'exploitation des données massives peut-elle guider une stratégie de différenciation à l'intérieur d'un domaine de recherche ? Avec quels outils, quels processus, quels acteurs, à quel niveau ?

Comment se construisent les collaborations entre labos, dans le cadre des préconisations des programmeurs de la recherche ? Quelles mobilisations des réseaux d'acteurs de la recherche ? Quelles coordinations des actions ? Quelles conséquences sur la nature même du travail du chercheur ?

Alain Beretz, président, UNISTRA et LERU

Jean-Pierre Finance, délégué permanent pour la CPU à Bruxelles et président de Couperin

Francis Jutand, directeur scientifique de l'Institut Mines Télécom, membre du Conseil National du Numérique et coordinateur de l'ouvrage collectif "La métamorphose numérique".

17h15 - 17h45 : Allocution du MENESR