DAUPHINE RECHERCHES

Le magazine de la recherche à Dauphine

DAUPHINE UNIVERSITÉ PARIS

PSL RESEARCH UNIVERSITY

N°18
NOVEMBRE 2014

Big Data: traiter les données efficacement et à moindre coût

→ Page 3/4

La régulation bancaire et financière à l'épreuve de la crise

→ Page 5/6

Quels sont les freins à l'indépendance des experts?

→ Page 7/8

EFMD EQUIS ACCREDITED

EDITO

Loin d'être cantonnée à l'enceinte de ses laboratoires, la recherche universitaire se développe en interaction avec son environnement économique et social. Que ce soit par la création de modèles, l'analyse de pratiques ou de textes, elle est en prise avec les problématiques les plus brûlantes.

Les travaux Dauphinois, mis en avant dans ce numéro, en sont une parfaite illustration. Le big data, autrement dit la collecte et le tri de données massives, devient un sujet majeur tant pour les politiques que pour les entreprises. Les nombreuses informations disponibles permettent d'acquérir une meilleure connaissance des citoyens et consommateurs. Encore faut-il pour cela disposer de méthodes de traitement efficaces ; sans parler des dispositifs éthiques à penser et à mettre en place.

La qualité et l'indépendance de l'expertise scientifique, de plus en plus sollicitée par les pouvoirs publics, représentent un enjeu tout aussi prégnant. Sans remettre en cause l'utilité de ces rapports, il est essentiel de les replacer dans leur contexte et de prendre conscience de l'impact de la structuration des connaissances sur leurs conclusions.

L'actualité de la question de la régulation financière n'est, elle, plus à prouver. Les secteurs bancaire et financier ont été profondément bouleversés par la crise de 2008. Ils doivent aujourd'hui opérer avec de nouvelles règles, répondre aux exigences de régulateurs aux compétences étendues, et faire face à des conflits de type inédit. Nul doute qu'une prise de recul sur ces évolutions est indispensable pour mieux appréhender leur fonctionnement et leur impact potentiel.

Enfin, les thématiques, abordées dans cette édition, se situent souvent à la croisée de plusieurs disciplines : informatique et statistique, droit et économie, sociologie et sciences politiques… Plusieurs recherches ont ainsi été menées dans le cadre de projets interdisciplinaires ; une façon de mêler les regards et compétences et d'enrichir l'éclairage sur un monde toujours plus complexe.

Elyès Jouini, Vice-président de la Recherche de l'Université Paris-Dauphine

DAUPHINE RECHERCHES est une publication tri-annuelle de l'Université Paris-Dauphine
• Directeur de la publication : Elyès Jouini (Vice-président Recherche) • Réalisation : Université Paris-Dauphine (Service Commun Recherche et Valorisation)
• Ont participé à ce numéro : Anne-Laure Chagnon, Valérie Fleurette et Elyès Jouini • Avec le concours de Coralie Bach pour la rédaction des articles publiés en pages 3 à 8 • Conception graphique : Anne-Laure Chagnon • Impression : Imprimerie Champagnac
N°ISSN : 2102-1422 - Dépôt légal : à parution
Contact : Service Commun Recherche et Valorisation – Université Paris-Dauphine tél. : 01 44 05 44 89

Big Data : traiter les données efficacement et à moindre coût

D'après un entretien avec Dario Colazzo et les articles cités en bibliographie.

La multiplication des données sur le net génère de nouveaux défis pour les spécialistes de l'informatique. Les méthodes de traitement traditionnelles doivent être adaptées afin d'assurer une analyse performante, et ce, même dans un contexte de données massives. C'est tout l'objectif des travaux de Dario Colazzo présentés dans cet article.

Depuis septembre 2013, Dario Colazzo est Professeur à l'Université Paris-Dauphine et chercheur au LAMSADE. Ses activités de recherche portent sur le traitement sûr et efficace de grandes masses données semi-structurées. Précédemment, il a été chercheur visiteur à l'École Normale Supérieure de Paris, et a effectué un post-doctorat, d'abord à l'Université de Venise puis à l'Université Paris-Sud, où il a été Maître de conférences de 2005 à 2013.
Dario Colazzo est diplômé de l'Université de Pise, où il a obtenu son Doctorat en Informatique en 2004.

Site web :
http://www.lamsade.dauphine.fr/

Réseaux sociaux, sites e-commerce, applications bancaires, le web génère une quantité d'information colossale et en constante augmentation. En parallèle, les progrès technologiques, en termes de puissance de calcul ou de qualité du réseau Internet, ont accompagné cette explosion des données. Les échanges se sont ainsi multipliés et accélérés. Mais au-delà de l'aspect quantitatif, le « big data » témoigne également d'une évolution qualitative. En effet, la nature même des données a changé puisque ces dernières proviennent d'une multitude de sources hétérogènes et mêlent images, textes, vidéos, liens… Les données structurées ont ainsi fait place à des données dites « semi-structurées » ou « dynamiques » qui ne comportent pas toutes les mêmes attributs, et dont la structure évolue parfois au cours du temps. Certaines peuvent comprendre des propriétés de couleur et de poids ; tandis que d'autres auront des propriétés de taille et de titre par exemple. En outre, les données semi-structurées sont parfois « complexes ». Autrement dit, une propriété peut consister en une donnée décrivant une transaction financière, ayant elle même plusieurs propriétés, éventuellement complexes…

Cette hétérogénéité et cette irrégularité dans la nature des informations représentent un véritable défi pour l'informatique qui, traditionnellement, se base sur la régularité de la structure pour assurer l'efficacité du stockage et de l'analyse des données. De nouvelles problématiques sont donc apparues. Comment gérer une masse de données disparates ? Comment les traiter rapidement dans des conditions financières raisonnables ? Les travaux de Dario Colazzo et ses coauteurs s'inscrivent dans ce contexte et ont pour but de définir une méthode adaptée à l'environnement du big data.

La puissance du *cloud*…

Le chercheur a centré son étude sur la gestion d'une grande quantité de données XML (*Extensible Markup Language*) : un format de données couramment utilisé sur le web qui s'appuie sur une présentation par balisage. Les systèmes traditionnels ne parviennent pas à interroger une grande collection de documents XML : soit la requête est trop longue à aboutir, soit elle échoue totalement… Dario Colazzo s'est donc appuyé sur la technologie du *cloud* , qui permet d'utiliser à distance plusieurs ordinateurs simultanément. L'avantage de cette solution est d'offrir une grande souplesse à l'entreprise, qui, selon ses besoins peut augmenter ou réduire les ressources informatiques utilisées. À la puissance matérielle du *cloud*, le chercheur a ajouté la pertinence des stratégies d'indexation. Semblables aux index des livres, qui recensent pour chaque mot clé l'ensemble des entrées possibles, les index informatiques catégorisent et classent les données afin de faciliter leur traitement. L'auteur a ainsi élaboré plusieurs stratégies d'indexation plus ou moins sophistiquées. La plus simple indique, pour chaque type d'information recherchée, l'emplacement des documents concernés. La plus sophistiquée donne non seulement le nom des documents pertinents,

mais fournit également une information contextualisée : l'utilisateur visualise, par exemple, le mot recherché dans son paragraphe.

… mêlée à une indexation pertinente

Les résultats des différentes stratégies sont ensuite comparés en termes d'efficacité. L'analyse englobe, bien sûr, le délai de réponse mais aussi le temps préalable nécessaire à l'élaboration de l'index ; sachant que les plus simples sont rapides à construire mais plus lents dans l'exécution. Et étonnamment, les index simples remportent le test : ils assurent une très bonne performance sur le plan de la rapidité de la requête, pour un temps de construction plus court que les index complexes. « Ce résultat est rendu possible grâce au *cloud* et au principe du parallélisme, explique Dario Colazzo. Les données à analyser sont découpées en petits paquets, puis adressées à plusieurs ordinateurs. Le traitement est ainsi accéléré et demeure efficace, même face à une quantité importante de documents.»

Évaluer le coût du traitement

Dans un deuxième temps, le chercheur a souhaité comparer les stratégies sur le plan

financier. Il a ainsi développé un modèle de coût monétaire afin d'estimer les frais engendrés par le stockage des données et l'exécution d'une requête. Un tel modèle permet d'évaluer le coût de traitement de l'information par le *cloud* et d'opter pour la stratégie d'indexation la plus adaptée. Résultat, l'indexation simple assure une nouvelle fois la meilleure performance, en offrant un bon compromis entre rapidité de réponse et coût de développement.

Quel que soit le type de données à analyser, le modèle présenté donne la possibilité de comparer les stratégies d'indexation en termes de temps de réponse et de coût. Il peut ainsi être adapté à d'autres langages informatiques que le XML, et a notamment été utilisé avec des données RDF (*Resource Description Framework*). Un outil pour trouver le meilleur « rapport qualité prix » en quelque sorte…

Bibliographie

• « *Web Data Indexing in the Cloud: Efficiency and Cost Reductions* », Jesús Camacho-Rodríguez, Dario Colazzo et Ioana Manolescu. *International Conference on Extending Database Technology,* 2013.

• « *Building Large XML Stores in the Amazon Cloud* ». Jesús Camacho-Rodríguez, Dario Colazzo et Ioana Manolescu. *International Workshop on Data Management in the Cloud*, 2012.

• « *AMADA: web data repositories in the Amazon Cloud.*» Andrés Aranda-Andújar, Francesca Bugiotti, Jesús Camacho-Rodríguez, Dario Colazzo, François Goasdoué, Zoi Kaoudi, Ioana Manolescu. *International Conference on Information and Knowledge Management*, 2012.

APPLICATIONS PRATIQUES

Face à une masse de données disparates, les techniques informatiques traditionnelles n'assurent plus une analyse rapide et efficace. Dario Colazzo propose une nouvelle technique de traitement des données adaptée au contexte du big data. Il a également défini un modèle financier capable d'évaluer le coût du traitement des données par le *cloud*. Grâce à ces informations, les professionnels peuvent comparer les différentes stratégies et opter pour la plus adaptée.

La régulation bancaire et financière à l'épreuve de la crise

D'après un entretien avec Sophie Lemaire et Juliette Morel-Maroger.

L'ampleur de la crise de 2008 a contraint les acteurs politiques et économiques à revoir la régulation bancaire et financière. Modification des règles, évolution des organismes de contrôle, apparition de nouveaux types de conflits, la régulation a été fortement modifiée. Sous l'égide de PSL, un projet de recherche interdisciplinaire, regroupant économistes, juristes et sociologues, a analysé ces changements.

Sophie Lemaire est Professeur de droit à l'Université Paris-Dauphine, où elle enseigne le droit international privé et le droit de l'arbitrage international. Ses travaux de recherche menés au sein du Cr2D, portent principalement sur les questions posées par l'essor mondial de l'arbitrage d'investissement.

Juliette Morel-Maroger est Maître de conférences à l'Université Paris-Dauphine. Elle est co-directrice du Master de droit international et européen des affaires. Ses travaux de recherche menés au sein du Cr2D, portent principalement sur le droit bancaire, plus spécialement dans sa dimension européenne et internationale.

Site web :
http://institut-droit.dauphine.fr/

La crise des *subprimes* et la faillite de la banque Lehman Brothers ont généré une véritable remise à plat des systèmes bancaires et financiers. Les mécanismes de régulation ont ainsi été profondément modifiés afin d'éviter la reproduction d'une telle crise mais aussi de s'adapter au nouvel environnement économique.

Des conflits internationaux…

Les économies des différents États sont en effet beaucoup plus liées que par le passé. En conséquence, les effets systémiques des crises sont bien plus forts et accentuent, de fait, la nature internationale des conflits. L'architecture des organismes de régulation a donc été revue pour mieux répondre à cet enjeu. « La surveillance des principaux établissements bancaires, jusqu'à présent réalisée dans un cadre national, relève désormais des autorités européennes » explique Juliette Morel-Maroger, Maître de conférences en droit privé à l'Université Paris-Dauphine. « À partir du 4 novembre 2014, elle relèvera de la Banque centrale européenne dans le cadre du Mécanisme de surveillance unique, première étape de l'Union bancaire ». Le basculement de la régulation au niveau européen a été conçu comme une réponse à la crise de ces dernières années, et constitue également une suite logique à la construction de la zone euro : « Une monnaie unique implique des règles uniques, poursuit la chercheuse. Or, jusqu'à présent, bien que des règles aient été édictées au niveau européen, leur mise en œuvre dépendait des autorités nationales.»

…et impliquant les États

Autre évolution engendrée par la crise : la naissance de nouveaux types de conflit. C'est ainsi que l'actualité a mis en lumière les problématiques liées aux dettes souveraines, avec les contentieux opposant l'État débiteur à ses créanciers. En la matière, une brèche a été ouverte avec le cas argentin. Ne pouvant plus assumer sa dette, l'Argentine a, en 2005, proposé à ses créanciers d'échanger leurs titres contre des titres de moindre valeur. Une proposition rejetée par une partie des investisseurs, dont une association de banques italiennes baptisée la Task Force Argentina. Regroupant près de 180 000 personnes, ce collectif a engagé un contentieux contre l'État argentin, via une procédure d'arbitrage engagée sur le fondement du Traité de protection des investissements conclu entre l'Argentine et l'Italie en 1990, devant le Centre international de règlement des différends relatifs aux investissements (CIRDI), un organisme adossé à la Banque mondiale. « C'est la première fois que l'arbitrage est utilisé à l'encontre d'un État qui restructure sa dette, souligne Sophie Lemaire, professeur de droit privé à l'Université Paris-Dauphine. Auparavant, les outils juridiques ouverts à ses créanciers étaient plus classiques et souvent inefficaces : il s'agissait ou bien d'intenter sans grand espoir une action devant les juges de l'État débiteur, ou bien d'agir à l'étranger pour obtenir une décision difficilement exécutoire dans l'État débiteur. » Aussi, les traités de protection des investissements internationaux et les offres d'arbitrage qui y figurent se présentent comme une solution alternative et alléchante pour les investisseurs.

La régulation bancaire et financière à l'épreuve de la crise

Dettes souveraines : les créanciers attaquent

Si l'issue de l'affaire demeure encore inconnue, puisque l'action est toujours en cours, elle pourrait avoir d'importantes retombées. Certes, offrir un accès à la justice aux créanciers est en soi louable, et peut inciter les États à adopter un comportement plus vertueux en favorisant par exemple les négociations. Mais les risques ne doivent pas être écartés. Si l'action de la Task Force Argentina aboutissait, des procédures similaires pourraient se multiplier, notamment en Europe. La Grèce fait ainsi déjà l'objet d'une plainte déposée par une banque slovaque devant le CIRDI.

En outre, le pays condamné pourrait se voir encore plus affaibli au niveau international, en particulier dans ses relations avec la Banque mondiale, à laquelle le CIRDI est adossé. Un levier politique viendrait se greffer à la pression économique. Face à de telles hypothèses, il est utile de s'interroger sur les bénéfices de l'arbitrage en matière de dette souveraine : sont-ils supérieurs aux risques ? Les arbitrages doivent-ils concerner des contentieux ayant de tels impacts sur la souveraineté étatique ?

L'enjeu des traités internationaux

Bien qu'aucune réponse ne puisse être apportée avec certitude aujourd'hui, les travaux de Sophie Lemaire invitent à la prudence. La possibilité de recourir à l'arbitrage est prévue dans le cadre des traités de protection des investissements internationaux. Il importe donc d'être attentif à la rédaction de ces textes. « Il est possible de prévoir des exemptions à l'arbitrage pour les opérations de restructuration de dette souveraine, précise Sophie Lemaire. Le traité établi entre le Canada et l'Union Européenne, par exemple, rend impossible toute procédure à la condition que 75% des créanciers valident l'accord de restructuration. »

Si les détenteurs de dette souveraine doivent avoir un accès à la justice, l'instauration de certaines conditions à l'exercice de l'arbitrage est certainement nécessaire pour éviter de trop lourdes conséquences internationales. Un juste équilibre doit être trouvé entre la protection des créanciers et le respect de la souveraineté des États.

Bibliographie

- « L'arbitrage d'investissement et la restructuration de dettes souveraines: De l'expérience argentine au cas grec », Lemaire Sophie, Revue de l'arbitrage, 2014, p. 53.

- « Droit bancaire » Decocq Georges, Gérard Yves, Morel-Maroger Juliette, Revue Banque, 2014.

APPLICATIONS PRATIQUES

Sous l'impulsion de PSL, économistes, sociologues et juristes, se sont engagés dans un projet de recherche interdisciplinaire sur le règlement des conflits et la régulation bancaire et financière à l'épreuve de la crise. Comment les autorités et les États ont-ils réagi à la crise ? Quels outils ont été mis en place ? Comment la régulation a-t-elle évolué en termes de règles et d'organisation ? L'objectif des chercheurs est de proposer une synthèse et une analyse des nouveaux textes, en s'appuyant sur la diversité des points de vue des différentes disciplines. Ce travail s'est concrétisé par la présentation de travaux et l'organisation d'un colloque en présence des professionnels du secteur.

Quels sont les freins à l'indépendance des experts?

D'après un entretien avec Emmanuel Henry et les articles cités en bibliographie.

Depuis 2011, un collectif de chercheurs en sciences sociales apporte une nouvelle réflexion sur la problématique de l'indépendance des experts dans le cadre du programme de recherche INDEX financé par l'ANR. Emmanuel Henry, professeur de sociologie à Dauphine, est responsable de ce projet. Il démontre que les processus d'expertise sont contraints, très en amont, par les politiques de recherche et la structuration des connaissances qui en découle.

Emmanuel Henry est professeur de sociologie à l'Université Paris-Dauphine et chercheur à l'IRISSO. Ses travaux portent sur les dynamiques de construction des problèmes publics, que ce soit les conditions qui permettent à un problème d'acquérir une certaine publicité, ou encore les logiques qui permettent une gestion discrète de certains problèmes. Il a dirigé plusieurs programmes de recherche sur les politiques de santé publique et de santé au travail qui interrogent notamment la manière dont les connaissances scientifiques et l'expertise sont associées à la décision publique. Le programme de recherche INDEX donnera prochainement lieu à la publication d'un Dictionnaire critique de l'expertise associant les principaux spécialistes de ces questions.

Site web :
https://www.irisso.dauphine.fr/

Les experts scientifiques ont acquis une place majeure dans le débat public. Régulièrement consultés par les pouvoirs politiques ou les régulateurs, leurs avis pèsent sur les décisions économiques et sociales. La question de leur indépendance est devenue cruciale, d'autant plus que plusieurs affaires ont mis à jour des relations de proximité entre scientifiques et industriels. Les experts sont-ils toujours objectifs ? Leurs avis sont-ils orientés par tel ou tel lobby ? « La problématique de l'indépendance des experts est généralement abordée sous l'angle des conflits d'intérêts, explique Emmanuel Henry. Or, cette vision est assez réductrice. Nous avons donc créé un collectif de chercheurs en sciences sociales, réunis au sein du programme ANR INDEX, sur l'indépendance des experts. L'objectif est de renouveler les réflexions sur l'expertise en ne se limitant pas à la question des procédures. »

Une expertise conditionnée par l'état des connaissances

Les recherches en sciences sociales montrent en effet qu'au delà des éventuelles pressions, les questionnements et la vision des experts sont contraints bien en amont par la structuration même des connaissances scientifiques. « Contrairement à une idée reçue, les connaissances ne se déploient pas de façon linéaire, depuis l'ignorance vers une connaissance croissante, souligne Emmanuel Henry. Leur développement dépend de l'intérêt, plus ou moins grand, des acteurs politiques ou économiques pour les divers pans de la recherche.» Or, les groupes sociaux ne disposent pas tous des mêmes ressources, et n'ont, par conséquent, pas la même capacité d'influence sur la recherche scientifique… Les sociologues constatent ainsi une inégalité entre les différents domaines couverts par les sciences. « Les connaissances scientifiques les plus développées sont souvent celles qui répondent à des intérêts industriels, économiques ou politiques, poursuit l'auteur. Au contraire, celles qui répondent uniquement aux besoins du mouvement social ne trouvent généralement pas d'acteurs suffisamment puissants pour les encourager et les transformer en connaissances scientifiques solides. »

Une production d'ignorance

Ainsi, dans le secteur de la santé au travail, les connaissances scientifiques se développent avec énormément de retard par rapport aux enjeux sanitaires en question. Les syndicats ne sont pas en mesure de peser sur les agendas de la recherche ; leurs ressources d'expertises étant limitées. Les industriels, au contraire, disposent de capacités financières nettement supérieures, mais leurs intérêts ne vont pas toujours dans le sens de la transparence : « C'est ce que nous appelons la production de l'ignorance, précise le chercheur. Ne souhaitant pas rendre visibles les impacts éventuels de tel ou tel produit chimique sur la santé, les entreprises ne financent pas les travaux. » Or, sans donnée fiable, les experts sont dans l'incapacité de formuler des recommandations pertinentes, comme dans le cas des valeurs limites

Quels sont les freins à l'indépendance des experts?

d'exposition professionnelle. Ces valeurs, établies par le ministère du Travail, réglementent le niveau d'exposition des salariés aux polluants. Mais comment déterminer la dangerosité d'un produit sans données scientifiques ? C'est une des problématiques soulevées par Emmanuel Henry.

Les agences sanitaires : une réponse insuffisante

La création des agences publiques d'expertise a eu pour but de donner à l'État les moyens de produire des connaissances et des expertises de façon autonome. Si l'initiative va dans le bon sens, ses résultats semblent encore modestes… « Les agences arrivent trop tard dans le processus, indique le chercheur. Se doter des meilleurs moyens d'expertise ne résout pas l'absence de connaissance scientifique en amont. Ce sont les politiques scientifiques qui sont à revoir.» Selon lui, l'expertise publique reste, en partie, prisonnière de la structuration historique des savoirs. « Les questions de santé au travail, par exemple, font l'objet de trop peu d'études scientifiques. Elles ne sont pas définies comme des enjeux de santé publique et sont perçues comme relevant exclusivement de la négociation entre partenaires sociaux. »

Toutefois, des pistes d'amélioration existent et nécessitent de prendre en compte le caractère partiel des savoirs scientifiques. L'expertise ne doit pas se restreindre aux seuls savoirs académiques et à la seule expertise institutionnelle. Elle doit au contraire considérer les autres sources d'information. Associations de consommateurs, syndicats, ONG, sont autant de producteurs de connaissances pouvant éclairer des points aveugles. La capacité à expertiser une question doit ensuite toujours être relativisée en prenant en compte ce qui est connu et ce qui ne l'est pas. En allant au bout de cette logique, il serait peut être possible de dire que l'expertise n'est pas là pour « clore les controverses ». Elle devrait, au contraire, permettre de doter l'ensemble des acteurs concernés des ressources nécessaires à la construction de leur point de vue. L'expertise assurerait ainsi un débat à armes égales afin que la voix la plus forte ne soit pas la seule entendue.

Bibliographie

• « Nouvelles dynamiques de savoirs et permanence des rapports de pouvoir. L'impact – limité – des transformations – importantes – de l'expertise en santé au travail », E. Henry, Revue française de science politique, vol. 61, 4, août 2011, p. 707-726.

• « *"License to Expose"? Occupational Exposure Limits, Scientific Expertise and State in Contemporary France* », E. Henry, in Soraya Boudia, Nathalie Jas, direction, *Toxicants, Health and Regulation since 1945*, Pickering and Chatto, 2013, p. 89-102.

• « La définition des problèmes publics: entre publicité et discrétion », E. Henry, C. Gilbert, Revue française de sociologie, volume 53, 1, 2012, p. 35-59.

• « Amiante : un scandale improbable. Sociologie d'un problème public », E. Henry, Rennes, Presses Universitaires de Rennes, coll. Res Publica, 2007.

APPLICATIONS PRATIQUES

Entouré par un collectif de chercheurs en sciences sociales, Emmanuel Henry étudie l'indépendance de l'expertise scientifique. Il montre, qu'au delà, des conflits d'intérêts, l'expertise est contrainte par les contextes économiques et politiques qui influencent, en amont de l'expertise, la structuration des connaissances.

Pour le chercheur, les rapports de force entre les groupes sociaux doivent être rééquilibrés en dotant chacun des ressources nécessaires à la construction d'une connaissance scientifique solide. L'expertise ne devrait pas se restreindre aux seuls savoirs académiques et à la seule expertise institutionnelle mais au contraire pouvoir s'approprier des savoirs produits par des associations de consommateurs ou des syndicats.