



Au cœur des données

Entretien avec Serge Abiteboul (réalisé par Valérie Schafer)

Des bases de données au « data déluge » actuel, Serge Abiteboul se passionne depuis plus de trente ans pour un domaine qui ne cesse d'évoluer et de nous confronter à de nouveaux défis, techniques, mais aussi sociétaux. Il a reçu en 2013 le prix Milner de la Royal Society et vient d'achever le projet Webdam¹ sur la gestion des données, pour lequel il avait été lauréat en 2008 d'une bourse ERC. Membre de l'Académie des sciences et du Conseil national du numérique, président du Conseil scientifique de la SIF, co-fondateur du blog BINAIRE², Serge Abiteboul, chercheur à Inria, revient sur sa carrière, ses recherches et éclaire pour nous quelques-uns des enjeux que pose aujourd'hui le numérique.

Valérie Schafer³

Valérie Schafer : Vous avez démarré votre carrière scientifique au milieu des années 1970, en faisant des études à Telecom Paris et au Département d'informatique du Technion de Haïfa (Israël). Vous obtenez un diplôme de Telecom Paris en 1978, faites votre service militaire à l'ESEAT⁴ de Rennes, puis vous partez pour la Californie

1. <http://webdam.inria.fr>

2. <http://binaire.blog.lemonde.fr>

3. Valérie Schafer, CNRS, ISCC. Entretien co-publié avec la revue Technique et Science Informatique (TSI).

4. Maintenant, École des transmissions (de l'armée de terre).

et vous réalisez un PhD sous la direction de Seymour Ginsburg⁵. Revenons pour commencer sur ces années si vous le voulez bien. Qu'est-ce qui vous pousse à vous tourner vers l'informatique et à partir aux États-Unis ?

Serge Abiteboul : Les gens parlent souvent de vocations précoces... Ce n'est pas le cas pour moi. Je ne savais rien de l'informatique quand j'ai commencé mes études. J'ai fait des classes préparatoires scientifiques alors que j'aurais préféré des études littéraires. Puis je suis rentré à Télécom, tout bêtement parce que c'était la meilleure école que j'avais obtenue. Lors de ma troisième année je suis parti en Israël, dans le cadre d'un échange d'étudiants. J'ai choisi l'informatique parce que cela me permettait de passer un an en Israël. Et c'est là-bas que j'ai vraiment commencé à comprendre ce que c'était. À l'époque, il faut reconnaître que ce qu'on nous enseignait à Télécom n'était pas ragoutant. C'était l'« informatique bidouille » ; on apprenait à écrire des programmes en bricolant ; on nous laissait dans une ignorance crasse des fondements théoriques du domaine. La théorie des langages, l'algorithmique, la complexité, c'est en Israël que j'ai découvert tout ça. Je suis ensuite revenu en France à l'ESEAT de Rennes pour mon service militaire. Après, je n'avais pas envie de devenir ingénieur ; j'avais plutôt envie de voyager. Alors, j'ai obtenu un poste d'assistant d'enseignement à Los Angeles et je suis parti faire un PhD. C'était la vie d'étudiant en Californie... et j'y ai découvert la recherche avec Seymour Ginsburg.

V.S. : Parlez-nous un peu de Seymour Ginsburg si vous le voulez bien.

S.A. : Seymour Ginsburg avait une vraie densité scientifique. C'était une personnalité pas ordinaire, un personnage étonnant par des côtés parfois ridicules, pas très sociable, mais quand il expliquait une preuve... Tout à coup il devenait passionnant ! Ses travaux traitaient de théorie des langages. Je ne me suis pas engagé dans cette voie, et ce pour une raison simple : la Mecque mondiale de la théorie des langages, c'était l'école de Maurice Nivat en France ! Pour Ginsburg, Nivat était l'un des meilleurs informaticiens au monde. Je ne connaissais pas Nivat personnellement à l'époque, nous nous sommes rencontrés depuis. Toujours est-il qu'il aurait été ridicule d'aller aux États-Unis pour travailler sur un sujet où l'excellence était en France. Seymour Ginsburg voulait regarder du côté des bases de données, un domaine qui démarrait à peine. Alors pour quelqu'un de fainéant comme moi, une quinzaine de papiers à lire, face à des centaines de publications dans la théorie des langages ! Je n'ai pas hésité. Et puis, il y avait l'attrait de la nouveauté. Certains choisissent la difficulté, aiment s'attaquer à des problèmes auxquels d'autres se sont confrontés pendant des années sans succès, moi je préfère la facilité et la nouveauté, définir

5. Pionnier de la théorie des langages formels et de la théorie des bases de données, voir notamment l'hommage qui lui a été rendu en 2004 et qui retrace son parcours dans *SIGMOD Record*, vol. 34, n° 1, March 2005. <http://www.sigmod.org/publications/sigmod-record/0503/p5.special.vianu.pdf>.

mes propres problèmes. De plus, le domaine des bases de données était à la frontière entre informatique théorique et appliquée, ce qui m'attirait aussi. J'aime pousser les limites de la formalisation, investiguer de nouvelles frontières.

« For me, Seymour was at first the professor with a strange hat and immutable clothes going down the corridor of the department, a bit abrupt but always so helpful to the young ones. Seymour Ginsburg, the famous professor, a pioneer of formal languages, a knight of computer science theory, always found seemingly unlimited time for his PhD students. I remember my visits to his place on Saturdays to go over my thesis together, after the invariable spaghetti lunch. He would check every single definition, result, every single proof, sentence, every single comma. [...]

After many years, when I look back, I see that no one influenced my work more than Seymour. I realize when I talk to my PhD students, that I am repeating his lessons. When I write a paper, I try to follow his very precise guidelines. When I get started on some new problem, I wonder : what questions would have Seymour asked ? »

Serge Abiteboul, in
« In memory of Seymour Ginsburg 1928-2004 »,
SIGMOD Record, vol. 34, n° 1, March 2005. <http://goo.gl/F7FMoJ>

V.S. : *Sur quoi travaillait-on concrètement à l'époque quand on faisait de la recherche sur les bases de données ?*

S.A. : Les bases de données à l'époque, c'étaient les bases de données relationnelles, les tableaux à deux dimensions, des logiciels comme System R d'IBM ou Oracle. Ma thèse a porté sur les propriétés algébriques des bases de données relationnelles. Très rasoir. Il n'y avait aucune motivation pratique. J'ai appris des tas de choses en rédigeant ma thèse. Mais j'ai surtout appris que je ne me satisfaisais pas de tels travaux purement mathématiques. J'avais besoin d'accroches sur des aspects plus pratiques, des applications. Je n'affirme ici aucun dogme, aucune hiérarchie de valeur. Il n'y a pas une seule façon de faire la recherche en informatique. Il y a des recherches très théoriques et d'autres plus expérimentales. Pour moi, le premier moteur de recherche de Google et son PageRank par Sergei Brin et Larry Page, c'est ni plus ni moins de la recherche que, par exemple, la preuve de l'indécidabilité de l'implication (finie) pour les dépendances fonctionnelles et d'inclusion par Ashok Chandra et Moshe Vardi. Mon goût personnel, c'est d'essayer de concilier des aspects à la fois applicatifs et de recherche fondamentale ; c'est tout à fait possible en bases de

données, pas dans tous les domaines ; c'est aussi pour ça que je me suis plu en bases de données. Et puis, c'est un domaine qui bouge, qui vit. C'est important de bien choisir son champ, de ne pas se laisser enfermer dans un domaine en train de mourir. Toute la problématique des données, de l'information, des connaissances, ne cesse de poser de nouveaux enjeux, d'évoluer ; il y a eu un cheminement du domaine et ma recherche évolue avec.

V.S. : *Vous rentrez ensuite en France ?*

S.A. : Je suis entré à Inria Rocquencourt en 1982 dans l'équipe de François Bancillon, qui est passé assez vite après dans l'industrie, et qui est resté un ami. C'est l'actuel PDG de Datapublica. J'avais l'idée de passer un an ou deux en France et de retourner aux États-Unis...

V.S. : *...et vous êtes toujours chez Inria 30 ans après...*

S.A. : C'est un super cadre de travail pour faire de la recherche en informatique. Ce n'est peut-être pas le confort de Stanford ou Oxford, mais en France c'est ce qui se fait de mieux. Et puis, j'ai bougé de Rocquencourt, à Saclay, puis à Cachan. (rire) Plus sérieusement, Inria offre un très bon support d'aide à la recherche. Par exemple, quand vous montez une entreprise, Inria sait vraiment aider.

V.S. : *Ce que vous avez fait avec Xyleme SA, devenue Xyleme Inc.*

S.A. : Oui, avec des amis, notamment Sophie Cluet, nous avons lancé Xyleme. Je m'y suis investi à partir de 2000, pendant cinq ans environ, avant même la création de la startup. Déjà un an et demi avant, nous étions sur la recherche qui a conduit à l'entreprise et sur le montage de celle-ci, son business plan, etc. Mais il faut souligner que pendant ces années où nos activités étaient axées sur le développement d'un logiciel à transférer, je n'ai pas vu de diminution de la qualité scientifique de l'équipe. Si l'on prend des critères comme le nombre de publications, de doctorants, etc., critères qui valent certes ce qu'ils valent, mais constituent tout de même des indications, il n'y a pas eu de baisse de productivité de recherche. Le projet a attiré d'excellents étudiants, motivés par cet aspect appliqué et industriel. Le logiciel portait sur la gestion de documents semi-structurés. Parmi les clients importants que nous avons eus chez Xyleme, il y avait par exemple *Le Monde*, qui a mis toutes ses archives sur notre système. Mais dès que j'ai pu, je suis retourné à la recherche. Je reste fondamentalement un chercheur.

V.S. : *Vous aimez aussi transmettre ? Vous avez des activités d'enseignement ?*

S.A. : Oui, en effet. Actuellement, j'enseigne la gestion de données sur le web⁶ au master MPRI⁷ et les bases de données relationnelles en licence à Cachan.

V.S. : *Et vous avez été un an au Collège de France en 2011-2012*⁸.

S.A. : Oui. Ça a été une expérience extraordinaire avec un public très différent. Cette année au Collège de France m'a amené à réfléchir sur des aspects plus sociétaux, qui maintenant me passionnent ; une partie de mon temps aujourd'hui est dédiée à des questions moins techniques. Je suis notamment au Conseil national du numérique. À l'intérieur du conseil, le boulot est vraiment différent. J'ai fait par exemple partie d'un groupe sur l'inclusion numérique⁹. Nous avons rencontré des tas de gens. Je me doutais bien des problèmes d'exclusion générés par le numérique, mais ils étaient restés abstraits. De rencontrer des gens qui sont confrontés au quotidien à ces problèmes-là, cela change votre perspective. Et puis après, il faut regarder ce qui se fait, imaginer des solutions. C'est passionnant. Les chercheurs devraient regarder davantage autour d'eux ce qui se passe dans la vraie vie.

V.S. : *Sur ce point les pays anglo-saxons n'ont-ils pas pris un peu d'avance sur nous ? S'occupe-t-on suffisamment en France d'éthique des TIC ?*

S.A. : La France ne me semble pas en retard. Elle est même très en avance sur certains aspects comme la protection des données avec la Cnil par exemple. L'Europe en général me paraît plus que les États-Unis sensible à ces enjeux. En France, nous avons un système d'associations très dynamiques avec, par exemple, la FING¹⁰ ou encore la Quadrature du Net¹¹. Les associations font énormément sur le terrain. Plus récemment, en informatique, en France, un comité piloté par Max Dauchet, la Cerna¹², a été mis en place pour travailler sur les problèmes d'éthique. On aimerait que les chercheurs en informatique s'impliquent plus dans la société. Évidemment, c'est une généralité. Certains le font déjà énormément.

6. Voir notamment Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset et Pierre Senellart, *Web Data Management*, Cambridge University Press, 2011.

7. Master parisien de recherche en informatique.

8. La leçon inaugurale du 8 mars 2012 de Serge Abiteboul, intitulée « Sciences des données : de la logique du premier ordre à la toile » est disponible à l'adresse : <http://www.college-de-france.fr/site/serge-abiteboul/inaugural-lecture-2012-03-08-18h00.htm> Voir également sa *Royal Society Milner Award Lecture* « From data and information to knowledge : the Web of tomorrow », intervention à la *Royal Society* de Londres du 12 novembre 2013 : <http://royalsociety.org/events/2013/web-of-tomorrow/>.

9. Voir le rapport « Citoyens d'une société numérique, Accès, littératie, médiations, pouvoir d'agir : pour une nouvelle politique d'inclusion ». <http://www.cnumerique.fr/inclusion/>

10. Fondation Internet Nouvelle Génération. <http://fing.org/>

11. <https://www.laquadrature.net/fr>

12. Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique (CERNA). <http://www.inria.fr/actualite/mediacenter/creation-de-la-cerna>

V.S. : *En parlant des États-Unis, j'ai été frappée de voir que vous aviez été consultant pour la NASA. Les liens entre le monde de la recherche en informatique et le monde militaire semblent tellement plus forts outre-Atlantique...*

S.A. : Oui, dans ma carrière aux États-Unis, j'ai été payé par la NASA mais j'ai aussi travaillé sur des contrats avec ARPA et plusieurs autres organismes publics américains. Pendant les six ans que j'ai passés aux États-Unis, j'ai été régulièrement payé par les militaires, mais toujours pour faire la recherche que je voulais, par exemple sur les modèles de données semi-structurés (les ancêtres de XML) ou des modèles logiques du e-commerce. C'est une vraie différence entre la France et les États-Unis. J'en ajouterai une deuxième : aux États-Unis, les industriels venaient souvent voir ce que je faisais dans mon labo, y compris d'ailleurs des industriels français, alors qu'en France, les mêmes, s'ils viennent me voir, c'est qu'ils cherchent une subvention de l'ANR ou de la CE. Évidemment, mon point de vue est biaisé. Mes copains industriels vous diront peut-être que les enseignants-chercheurs américains sont plus faciles à approcher que les français. Il faudrait travailler pour changer les mentalités, fluidifier les liens entre les labos académiques et les équipes de R&D des entreprises.

V.S. : *C'est aux États-Unis que vous découvrez Internet ?*

S.A. : Effectivement quand j'ai commencé ma thèse à Los Angeles, il y avait Internet, on s'envoyait des emails. Quand je suis revenu en France, on m'a dit qu'on ne pouvait envoyer des emails qu'à l'intérieur d'Inria. . . La préhistoire quoi ! Enfin, grâce à des copains au Bâtiment 8, j'ai pu garder le contact par mël avec les copains aux États-Unis. Quant au web, je l'ai découvert à sa naissance. J'ai tout de suite pensé que cela pouvait tout changer pour les bases de données, les rendre disponibles de partout. On pouvait sortir une base de données du deuxième sous-sol où elle était cantonnée pour la mettre à disposition dans le monde entier ! François Bancilhon avait créé une start-up qui avait développé un système de gestion de bases de données objet, O₂ ; avec un étudiant, on a fait O₂Web pour accéder par le web à une base de données O₂. Cette notion d'information mondialisée explosait le paradigme de la gestion de données. À l'époque, le web restait confidentiel. Il apparaissait comme le deuxième alinéa d'une sous-rubrique dans le rapport prospectif d'Inria. Puis, c'est allé très vite. Quelques années après la naissance du web, tout le monde s'est mis à publier sur le web et on a vu arriver les premiers moteurs de recherche, puis Google. . . J'étais dans la même équipe de recherche de Stanford que les « Google Guys », Lawrence Page et Sergeï Brin. Ils nous ont présenté leurs idées. Elles n'étaient pas si révolutionnaires au premier abord. Ils reprenaient un algorithme d'IBM. Mais ils montraient qu'on pouvait faire ça sur le web. Ils nous ont fait part de leur idée de la première architecture et des algorithmes. Nous étions sceptiques face à la taille de l'index nécessaire. Quelques mois après, ils me faisaient leur première démo. J'ai testé leur système en posant la requête « Rodin » sur Infoseek et Google. Il n'y avait

pas photo, leur moteur de recherche était bien meilleur¹³. Qui se rappelle encore d'Infoseek, un des moteurs de recherche du web les plus utilisés à l'époque ?

V.S. : *On parle aujourd'hui de tsunami de données, de déluge informationnel (Data Deluge). Ce ne sont pas des termes très positifs. Faut-il craindre la révolution annoncée ?*

S.A. : Évidemment, il y a la masse, la déferlante. Nous générons de plus en plus d'information. Plus de gens génèrent des données. De plus en plus d'objets connectés génèrent des données. Ce qu'on produit comme informations et données double tous les 18 mois. Ce doublement est-il soutenable ? On n'a pas le choix ! Cela pose des questions : Que faire de toutes ces données ? Comment les gérer ? Prenez les données personnelles, mon sujet de recherche actuel. Cela devient impossible aujourd'hui de gérer ses données perso. Par exemple, avant on pouvait retrouver une information sur son PC en faisant une recherche avec *Google desktop*. Maintenant, Google a arrêté *Google desktop*, parce que nos données ne sont plus dans le desktop ; elles sont sur dix systèmes différents, dans les nuages. Vous ne pouvez plus faire de recherche globale sur vos données personnelles. Le problème n'est pas technique. La question, c'est que vous ne contrôlez plus vos données, qu'elles sont ailleurs. Comment faire coopérer ces données que vous avez sur dix ou cent systèmes différents, comment gérer par exemple les contrôles d'accès ? On est en plein dans la gestion de données distribuées. Mais on est aussi dans des problèmes de société. Comment faire pour que Prism ne sache pas tout de moi ? Comment faire disparaître une photographie que je souhaite ne plus voir apparaître sur le web ? Il faut des lois, il faut de meilleures pratiques, il faut des utilisateurs avec une vraie littératie numérique, il faut aussi de nouvelles solutions techniques. C'est drôle, parfois j'entends presque la même question dans un séminaire de recherche et dans une réunion du Conseil national du numérique. La recherche en gestion de données découvre sans cesse de nouveaux challenges.

V.S. : *Puisque nous nous rencontrons en ce tout début d'année 2014, je ne peux résister à la tentation de vous demander si l'homme de science que vous êtes a pris des « bonnes résolutions » professionnelles pour la nouvelle année... Et si l'on devait souhaiter une bonne année à la science informatique, que pourrait-on lui souhaiter ? Quels vœux lui adresser ?*

S.A. : Mes bonnes résolutions ? Travailler moins (rire). Pour ce qui est de la recherche, c'est un peu tôt pour répondre. J'ai fini le mois dernier le projet Webdam¹⁴

13. Voir notamment Sergueï Brin et Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Proceedings of the 7th International Conference on World Wide Web*, Amsterdam, Elsevier, Computer Networks and ISDN Systems, vol. 30, n° 1-7, 1998, p. 107-117, doi : 10.1016/S0169-7552(98)00110-X.

14. <http://webdam.inria.fr>

de l'*European Research Council*. Quant à la science informatique, je souhaite le décollement véritable de l'enseignement de l'informatique en France^{15, 16}, qu'on sorte de l'ornière ; cela ne peut venir que d'une décision politique. Il faut la volonté politique de former des écoliers, des collégiens, des lycéens, des citoyens qui sachent tous programmer, des littéraires qui sachent ce qu'est un moteur de recherche. Il faut former toute la population à l'informatique. C'est indispensable. Pas juste pour former une génération d'informaticiens, mais pour avoir aussi des scientifiques, des ingénieurs plus performants, plus innovants. Avions, voitures, appareils photo, tous ces objets incluent maintenant des tas d'informatique. Mais un artiste aussi a besoin aujourd'hui de maîtriser l'informatique. Un médecin, un plombier, un cultivateur. Tout le monde. L'informatique doit entrer dans l'école au même titre que les mathématiques, la physique, la chimie ; l'informatique doit faire partie du socle fondamental. Chacun doit apprendre à maîtriser l'informatique, pour ne pas être juste un utilisateur de logiciels, ne pas devenir l'esclave de logiciels développés par d'autres. On a l'impression de radoter quand on dit ça tellement ça a été répété. Mais je crois que c'est vraiment essentiel.

15. Rapport de l'académie des sciences, *L'enseignement de l'informatique en France, Il est urgent de ne plus attendre*, mai 2013, www.academie-sciences.fr/activite/rapport/rads_0513.pdf

16. Voir l'article de Serge Abiteboul, Gilles Dowek et Colin de la Higuera du 24 janvier 2013 « Recherche profs d'informatique désespérément », dans *01 Business*, disponible sur : <http://pro.01net.com/editorial/585095/recherche-profs-dinformatique-desesperement/>