



Le logiciel libre, de l'usage à la recherche

Entretien avec Roberto Di Cosmo (réalisé par Valérie Schafer)

Actuellement détaché auprès d'Inria de l'université Paris-Diderot où il est professeur d'informatique, Roberto Di Cosmo se consacre pleinement à Software Heritage dont il est le directeur. Ce projet patrimonial dédié aux codes et logiciels libres concrétise des années de pratique, d'investissement et de recherches dans le domaine des logiciels libres.

Retour avec Roberto Di Cosmo sur le parcours qui l'a mené d'Italie en France, de la découverte des ordinateurs à la recherche en informatique, enfin de l'agacement face aux systèmes fermés et propriétaires à l'action dans le domaine des logiciels libres et à leur préservation.

Valérie Schafer¹

Valérie Schafer : *Avant que nous abordions le projet Software Heritage pour lequel vous êtes en détachement depuis un an chez Inria, j'aimerais que l'on revienne sur votre parcours dans l'informatique, débuté en Italie.*

Roberto Di Cosmo : Le contexte italien est alors différent du contexte français. Cela va sans doute vous surprendre, mais, à mon époque, en Italie on mettait beaucoup plus l'accent sur les filières littéraires que scientifiques, contrairement à l'usage en France.

Au lycée je poursuis des études de philosophie, d'histoire, de lettres, de latin et de grec. Et c'est par ce biais que je vais me pencher sur l'informatique. En effet je remporte un prix lors d'un concours de traduction de textes du grec au latin. Et

1. Valérie Schafer, ISCC, CNRS/Paris-Sorbonne/UPMC. Entretien co-publié avec la revue *Technique et science informatiques* (TSI).

c'est avec l'argent de ce prix que je m'achète mon premier ordinateur ! On est en 1979 et c'est un ZX 80, avec un processeur 8080 cadencé à quelques mégahertz qui génère aussi le signal vidéo, et un seul kilo-octet de RAM, partagé entre mémoire de programme et mémoire vidéo. C'est un ordinateur extrêmement rudimentaire : si le programme que l'on écrit est trop long, la partie d'écran affichée se réduit ; quand on appuie sur une touche, on désynchronise l'écran, etc. Et pourtant, c'était absolument magique !

V. S. : *Vous apprenez sur le tas ?*

R. D. C. : Il n'y avait pas de cours d'informatique au lycée, et il y avait une seule université d'Informatique dans toute l'Italie, mais pour un lycéen motivé, quelques magazines informatiques, nés des milieux hobbyistes notamment, suffisaient pour commencer à écrire quelques petits programmes, et faire surgir plein de questions qui attendaient réponse.

Très naturellement, quand j'intègre la *Scuola Normale Superiore di Pisa* (École normale supérieure de Pise) je choisis l'option informatique. Nous devions être trois ou quatre en informatique et les promotions étaient très petites, autour de 25 personnes. Contrairement au système français où l'entrée à l'ENS se fait après une classe préparatoire, en Italie le concours était (et est encore) à la sortie du lycée : il n'y a pas de classe préparatoire, parce qu'il y a très peu de grandes écoles.

La sélection doit donc se faire parmi tous les lycéens, peu importe qu'ils soient issus de filières littéraires ou scientifiques. Et les épreuves ne peuvent pas s'appuyer sur un programme touffu et balisé comme en France pour déceler les meilleurs candidats. Si en France les sujets en mathématiques font plusieurs pages et demandent maîtrise et rapidité, à Pise tout tient sur un feuillet et il faut faire preuve de créativité et d'imagination. Si en France l'admission à une grande école est la preuve qu'on est parmi les meilleurs, à Pise après le passage du concours on doit constamment prouver qu'on est bien le bon candidat, et pas quelqu'un qui a eu la bonne idée au bon moment par hasard, avec des conditions draconiennes sur les résultats attendus à la fois dans les cours à la Faculté et les cours supplémentaires dispensés en interne. Dans ma promotion à Pise, nous étions une vingtaine à entrer, et on était un peu plus d'une dizaine à la fin.

Je fais mes études d'informatique entre 1982 et 1986. On pouvait avoir encore l'illusion de tout comprendre de l'informatique. Je dis bien l'*illusion* de tout savoir. Les matières enseignées l'étaient chacune alors par un luminaire, un maître dans sa discipline. Bien sûr aucun d'entre eux ne venait directement de l'informatique, ce champ était en construction. Ils n'avaient pas étudié l'informatique, mais ils l'avaient créée ! La plupart d'entre eux étaient partis à l'étranger puis revenus ensuite pour créer cette formation à l'université de Pise. Cette période est véritablement révolutionnaire ! Deux ans avant que je commence ces études d'informatique à la Faculté, la promotion était seulement d'une cinquantaine d'étudiants. En 1982, nous étions

environ 1200. Le lancement du PC IBM en 1981 avait notamment contribué à l'explosion du nombre d'étudiants.

V. S. : *Pas évident à gérer pour les enseignants...*

R. D. C. : Les moyens informatiques à disposition des étudiants d'informatique de l'université étant effectivement totalement sous-dimensionnés pour cette nouvelle foule, les cours étaient essentiellement théoriques dans ces années, ce qui n'est pas forcément une chose si mauvaise, vu le grand nombre d'étudiants qui sont devenus des informaticiens théoriques de niveau mondial.

À l'université, nous avions accès à un ordinateur, mais nous étions encore dans un système de *batch processing*². On pouvait entrer des calculs, mais nous n'avions les résultats que le lendemain.

Mais heureusement, à l'École Normale de Pise notre tout petit groupe de normaliens informaticiens avait sympathisé avec les responsables du centre de calcul, qui se révéla une véritable caverne d'Ali Baba. On a passé des heures à analyser le fonctionnement d'un vieil ordinateur HP décommissionné : il y avait une magnifique plaque en bronze avec gravées les instructions à insérer péniblement à la main pour mettre en marche le lecteur de rubans perforés, et heureusement une mémoire à noyaux magnétiques qui maintenait l'information des mois après avoir éteint la machine. De la même façon lorsque le PDP-11 de la comptabilité a été remplacé, nous l'avons récupéré, en découvrant les subtilités d'un système multitâche et multi-utilisateur.

Et très rapidement on rentre dans la période de développement ultrarapide de l'informatique personnelle de l'Apple 2 au Macintosh, des premiers PC IBM aux Atari et Amiga. C'est aussi le moment de la découverte des premiers jeux d'aventures, des premiers programmes d'intelligence artificielle et si nous sommes bien sûr tous intéressés à y jouer, on finit par s'intéresser beaucoup plus à comprendre comment tout cela fonctionne, en faisant du *reverse engineering* et en s'amusant comme des fous.

Il y avait un autre aspect intéressant qui compliquait la tâche des enseignants : l'apparition des premiers ordinateurs « de poche », comme le Sharp PC-1500, qui pouvaient servir à calculer automatiquement les réponses aux exercices de mathématiques.

Je me rappelle que je trouvais insupportable l'idée de devoir faire à la main des calculs qui sont automatisables, comme les dérivées. J'avais donc programmé un dérivateur formel pour faire le calcul des examens plutôt que de les effectuer à la main : comme il n'y avait pas assez de mémoire, je ne pouvais pas programmer toutes les fonctions prévues au programme, et j'ai dû finir par me résigner à me

2. Dans le *batch processing* ou traitement par lots, les programmes ou travaux devant être exécutés sont placés les uns derrière les autres. Ils constituent ainsi des lots qui sont traités par l'unité centrale de manière séquentielle.

convertir en calculette humaine pour passer l'examen d'analyse. Mais cette envie d'automatiser n'en était que plus vivante en moi.

V. S. : *Vous finissez ces études en 1986 ?*

R. D. C. : Absolument, et je pars faire mon service militaire ensuite, pendant lequel, après des longs mois d'entraînement, je me retrouve aussi à... enseigner l'informatique aux sous-officiers. En 1988 je commence une thèse sur la théorie des types sous la direction de Giuseppe Longo, professeur de logique à Pise et aujourd'hui directeur de recherche CNRS à Ulm. En 1989 j'ai passé une période à Cornell, une excellente université américaine, mais à mon retour mon directeur de thèse était en visite à l'ENS Ulm, donc je l'ai suivi, comme bien d'autres de ses doctorants, et je me suis retrouvé à Paris. J'ai découvert l'Inria Rocquencourt, où l'on allait religieusement suivre le séminaire du vendredi organisé par Gérard Huet, et j'ai pu croiser des chercheurs de grande renommée comme Jean-Jacques Lévy, Maurice Nivat, Gérard Berry, Pierre-Louis Curien, Jean Berstel, Jean Vuillemin, et bien d'autres qui j'espère ne se vexeront pas du fait que j'arrête ici la liste à la Prévert. C'était inespéré, et j'ai naturellement prolongé mon séjour, devenant d'abord ATER et ensuite Maître de Conférences à l'ENS Ulm, vivant quelques années magiques, dans une atmosphère de réelle pluridisciplinarité.

Je me souviens par exemple d'avoir un jour entendu des étudiantes littéraires qui parlaient des œuvres de Shakespeare et cherchaient à étudier la fréquence des mots utilisés : certaines avaient passé des jours entiers à pointer le nombre d'occurrences d'un mot donné dans toutes les œuvres. On était alors en 1990 ou 1991, et j'utilisais un ordinateur NeXT : nouvelle aventure de Steve Jobs après avoir été éjecté d'Apple, cette machine magnifique intégrait non seulement des logiciels en avance sur leur temps (certains écrits par Jean-Marie Hulot, que Steve Jobs était venu chercher exprès chez Inria), mais aussi une copie numérique de toutes les œuvres de Shakespeare. Bref en une heure j'avais construit la table des fréquences non pas d'un, mais de tous les mots qui apparaissaient dans les textes de Shakespeare, et convaincu un certain nombre de littéraires que l'Informatique n'était pas qu'un petit jeu de *geeks*.

Je suis resté à Ulm, jusqu'en 1999, et pendant ces années j'ai eu le privilège de rencontrer des étudiants exceptionnels, auxquels j'ai donné tout ce que j'avais : je garde encore le contact avec beaucoup d'entre eux. Ma recherche est très théorique : sémantique des langages de programmation, réécriture, théorie des nombres et théorie des catégories. Mais en parallèle je conserve un indomptable esprit *hacker* : j'adore les machines, j'adore jouer avec et je ne supporte pas qu'elles soient verrouillées, m'empêchant de corriger les erreurs qu'on rencontre, et me forçant à suivre un chemin pré-tracé.

Je commence donc à m'intéresser aux logiciels libres : ce n'était pas du pain béni, la variété et la qualité des logiciels libres disponibles à l'époque étant plutôt limitées,

mais c'était bien un moyen pour récupérer un peu de sa liberté, et il y avait une vraie dynamique qui émergeait. C'est à ce moment-là que je découvre que, même si je ne veux pas du système d'exploitation propriétaire qui est vendu avec les machines, je dois le payer quand même. C'était un comble.

V. S. : *Ce sont ces constats qui vous conduiront à écrire en 1998 Piège dans le cyberspace puis, avec la journaliste Dominique Nora, Le Hold-up planétaire : la face cachée de Microsoft*³ ?

R. D. C. : Je me rendais compte que l'informatique était en train de changer le monde. Mais on ne nous proposait pas les meilleurs outils. Au contraire on essayait de rentabiliser ce qui était déjà sur le marché et de freiner les concurrents. *Piège dans le cyberspace* était d'abord une thérapie psychologique personnelle qui se voulait un manifeste de révolte, plutôt humoristique, destiné à circuler entre amis.

Mais les amis ont commencé à le faire circuler : il se retrouve en ligne sur un site web au Canada et le nombre de lecteurs – dont témoigne alors le compteur de visites – explose, avec des milliers de téléchargements par jour. Il est ensuite traduit en plusieurs langues, et je me retrouve propulsé contre toute attente porte-parole d'une majorité silencieuse qui s'ignorait. Ce texte comportait deux volets : une analyse critique fine des problèmes de l'industrie informatique de l'époque, et des propositions alternatives, comme l'accès au code source et l'adoption des logiciels libres.

Piège dans le cyberspace

Ce texte est depuis le 20 mars 1998 sur le Web
<http://www.mmedium.com/dossiers/piege>

« Pendant les vacances de Noël, j'ai été frappé une fois de plus par l'engouement croissant des médias pour cet obscur objet du désir qui se cache derrière les mots « ordinateur », « multimédia », « web », « Internet » et leurs dérivés. À croire ces médias et bon nombre d'experts improvisés, on ne peut se prétendre citoyen à part entière sans posséder le matériel informatique flambant neuf (et très cher) donnant accès au paradis féérique du « cyberspace ».

Difficile aussi de ne pas remarquer l'étrange et omniprésent amalgame qui nous incite à penser que le seul type d'ordinateur existant est le PC, à condition bien entendu qu'il soit équipé d'une puce Intel, et que sur ce PC, il y a un seul logiciel indispensable, Windows produit par Microsoft [...].

Par contre, notre public à nous est bien loin de se réveiller : bercé par la voix douce du conformisme ambiant, il s'endort de plus en plus dans les bras de Microsoft. Il rêve

3. R. Di Cosmo et D. Nora, *Le Hold-up planétaire : la face cachée de Microsoft*, Paris, Calmann-Levy, 1998. Le livre est disponible sous licence Creative Commons Attribution-NoDerivs-NonCommercial. Voir <http://www.dicosmo.org/HoldUp/HoldUpPlanetaire.pdf>

d'un monde joyeux où un grand philanthrope distribue à tous les écoliers de France des copies gratuites de Windows 95, dans le seul but de les aider à rattraper leur retard technologique ; il sourit en pensant aux écrans bleus pleins de messages rassurants qui expliquent comment le programme machin a provoqué l'exception bidule dans le module truc non pas par la faute de Windows, bien entendu, mais par celle du programme ; il dort heureux sans se demander pourquoi un ordinateur bien plus puissant que celui qui a servi à envoyer des hommes sur la lune, et à les ramener vivants, n'est pas en mesure de manipuler correctement un document d'une centaine de pages quand il est équipé par ce Microsoft Office qui fait la joie de tous nos commentateurs [...] ».

Retrouvez l'intégralité du texte sur

<http://www.dicosmo.org/Piege/cybersnare/piege.html>

La critique à l'époque portait surtout sur Microsoft, qui était assise dans une position dominante exploitée sans réserve, et malheureusement beaucoup de lecteurs ont juste retenu un message négatif en général vis-à-vis de cette entreprise. En réalité, l'analyse était plus fine : ce que je critiquais n'était pas tant l'entreprise Microsoft en tant que telle, mais la stratégie de verrouillage et de contrôle absolu de toute la chaîne de l'information qu'elle poursuivait à l'époque, et les moyens de la mettre en œuvre. Aujourd'hui, bien d'autres acteurs sont plus inquiétants, et l'analyse de 1999 s'appliquerait à eux bien plus qu'à ce que Microsoft est devenue.

Le succès de ce livre m'a valu d'être invité un peu partout pendant plus d'un an, et même si je n'étais pas formé à toute cette exposition médiatique, je considérais que j'avais une responsabilité, une mission d'explication, et aussi celle de proposer une autre vision de la société de l'information : pour une fois qu'on écoutait le message, je n'avais pas le droit de me dérober. C'était pourtant épuisant pour moi et éprouvant pour ma famille : j'avais une fille de deux ans, une épouse brillante aussi impliquée que moi dans la recherche, que je remercie infiniment de m'avoir soutenu et supporté pendant cette période. J'ai donc décidé de m'arrêter en juin 1999 et de revenir à une vie et un rythme plus normaux. Mais c'était sans compter une dernière interview avec Dominique Nora : ce rendez-vous devait durer 30 minutes. Il a duré la journée ! Elle est sortie en me disant qu'il fallait faire quelque chose ! Et c'est comme cela que nous avons écrit *Le Hold-up planétaire : la face cachée de Microsoft*.

En relisant ce livre, je peux dire que même si nous avons une compréhension imparfaite du futur, on y retrouve une analyse lucide et fine des dangers, et notamment du risque d'une maîtrise de la chaîne complète de l'information qui est aujourd'hui enfin devenu un sujet central du débat public.

Le plan de retour à une vie normale avait définitivement échoué : ces ouvrages m'avaient transformé en une sorte d'apôtre du logiciel libre et cela a été dur de ne pas tomber dans le travers des « experts permanents ».

Piège dans le cyberspace

« Une alternative possible : les logiciels libres »

Quand il s'agit de choisir le logiciel à fournir à nos collègues et à nos lycées pour initier nos enfants à l'informatique, on n'est pas obligés de s'en tenir aux cadeaux louches des cybermonopolistes : plutôt qu'un système propriétaire qui se plante très souvent, change tout le temps de version sans raison et dont le code source n'est pas disponible, on peut choisir un système libre ouvert et stable (il faut savoir que contrairement au préjugé populaire, le logiciel libre a eu amplement l'occasion de faire ses preuves [...]) qui permettrait à tout jeune de travailler et apprendre en toute sécurité et à tout esprit curieux d'acquérir une formation informatique avancée et intelligente, car la disponibilité du code source lui permet d'ouvrir le capot et même, s'il le veut, de démonter le moteur.

Et quand il s'agit d'équiper des grands comptes, comme on les appelle, de systèmes informatiques, mieux vaut se fier à du logiciel dont on a le code source et la documentation, qui est constamment vérifié et mis à jour par une communauté techniquement compétente et que l'on peut adapter à ses besoins au moindre coût [...] ».

Retrouvez l'intégralité du texte sur

<http://www.dicosmo.org/Piege/cybersnare/piege.html>

V. S. : *Et en parallèle vous poursuivez votre carrière. Vous passez notamment à l'université Paris 7 sur un poste de professeur...*

R. D. C. : En effet il y a une loi non écrite à Ulm chez les mathématiciens, que l'on essayait d'appliquer aussi en informatique : l'idée qu'il ne faut pas que tous les meilleurs étudiants, qui convergent vers Ulm, soient exposés à la même école et donc que les enseignants tournent. Ainsi les enseignants ne restent pas plus de dix ans à Ulm, au moins en mathématiques, et souvent en informatique. Pierre-Louis Curien a donc décidé de partir à l'université Paris 7 créer en 1999 le laboratoire PPS (Preuves, Programmes, Systèmes), qui est devenu IRIF (Institut de recherche en informatique fondamentale⁴) par une fusion récente avec le LIAFA (Laboratoire d'informatique algorithmique : fondements et applications).

Nous avons été plusieurs à le suivre, et ce changement n'a fait qu'augmenter la charge. J'ai dû construire beaucoup de cours nouveaux et en parallèle faire des centaines d'exposés sur le logiciel libre, les standards ouverts, et l'importance de reprendre la maîtrise d'une technologie informatique qui change le monde. J'ai dépensé une énergie considérable avec mes deux premiers doctorants, Vincent Balat et Jean-Vincent Loddio, pour construire DémoLinux : la première distribution de Linux qui fournissait un environnement graphique complet en démarrant à partir d'un

4. <https://www.pps.jussieu.fr>

lecteur de CD, sans installation. Cela avait le but de donner la possibilité à tous de tester les logiciels libres sans changer leurs machines, mais cela s'est révélé utile dans d'autres occasions, y compris pour les cours à l'université.

Enseignant, chercheur, responsable de formation, vulgarisateur, hacktiviste, tout à la fois, cela ne pouvait pas durer. Il fallait rapidement réconcilier ces différentes casquettes et c'est comme ça que j'ai commencé à orienter ma recherche sur les problèmes scientifiques posés par les logiciels libres.

V. S. : *Quand vous dites que vous vous concentrez sur des recherches dédiées au logiciel libre, cela veut-il dire que les logiciels libres posent des questions particulières différentes des autres logiciels ?*

R. D. C. : Au tout début, beaucoup de collègues étaient surpris quand je tenais ce discours, mais maintenant on sait bien qu'il y a des problèmes scientifiques liés aux modalités de développement logiciel et à la complexité croissante de systèmes construits de façon décentralisée à partir de grandes masses de codes sources : beaucoup de ces problèmes se manifestent seulement, ou plus clairement, quand le code source est disponible et modifiable. Ce que nous entreprenons dans le cadre du projet *Software Heritage* est dans la continuité⁵. Il y a une croissance très rapide de la taille des codes dans le logiciel libre et d'énormes réseaux d'interdépendance. Bien sûr il faut se garder de cette vision illusoire d'une communauté magique qui déciderait démocratiquement. Je ne résume pas le monde du logiciel libre naïvement à l'idée que c'est un système parfait, mais c'est celui qui permet le plus de liberté. De même que la démocratie, c'est une base de construction, même si cela ne garantit pas forcément qu'il en sortira le meilleur.

V. S. : *Vous parlez de la communauté du logiciel libre. Participez-vous aux différentes associations qui se créent pour soutenir le logiciel libre ?*

R. D. C. : J'ai eu l'occasion de rencontrer beaucoup d'acteurs associatifs avec lesquels je garde des contacts étroits et parfois même une véritable amitié. Mais je n'ai pas été directement très impliqué faute de temps et d'énergie, et aussi parce que j'essaie d'être le plus objectif possible dans mes prises de parole : je n'ai aucune peur d'assumer une position atypique ou anticonformiste, et je l'ai montré pendant toute ma carrière, mais j'ai horreur qu'on me range dans une case. Cela a été le cas aussi lors du débat sur l'*open access* dans la loi sur la république numérique, où j'ai

5. « *Our ambition is to collect, preserve, and share all software that is publicly available in source code form. On this foundation, a wealth of applications can be built, ranging from cultural heritage to industry and research* ». Pour un descriptif complet des missions du projet *Software Heritage* voir le site : <https://www.softwareheritage.org/mission/>.

Voir également R. Di Cosmo, *Software Heritage : pourquoi et comment construire l'archive universelle du code source*, 1024 n° 10, avril 2017, pp. 67–72, <http://www.societe-informatique-de-france.fr/bulletin/1024-numero-10/>.

maintenu ma position⁶ jusqu'à la fin : pourtant plébiscitée par les collègues, elle n'a finalement pas été retenue.

En 2004-2005 j'ai introduit les premiers cours sur les logiciels libres dans la maquette de l'université et c'étaient des cours non techniques pour des étudiants en informatique dont je m'étais rendu compte qu'ils ne connaissaient rien au droit des logiciels, à son histoire, etc. Un des premiers projets a été de réaliser une matrice de compatibilité des licences dans le monde du logiciel libre, qui allait de la BSD⁷, qui offre un maximum de liberté, à la GPL⁸, qui dit que la liberté que l'on acquiert doit être compatible avec le fait de la donner aux autres. Cela crée bien sûr des débats par exemple sur quelle licence est au fond la plus libre.

V. S. : *Vous avez aussi une expérience dans l'industrie ?*

R. D. C. : Dans mon rôle de « sage » du logiciel libre, je me suis retrouvé mêlé au montage d'un pôle de compétitivité en 2005-2006, qui est finalement devenu le Groupe Thématique Logiciel Libre dans le pôle Systematic⁹. Pendant deux ans nous avons été plusieurs à porter ce projet, dont je me suis retrouvé finalement responsable afin d'assurer la confiance entre les différents acteurs de la galaxie un peu agitée des entreprises du logiciel libre de l'époque. Là aussi, cela ne correspondait pas à ma mission (le porteur et président d'un pôle doit être normalement un industriel, pas un académique), mais j'ai investi quatre années au service de la communauté du Logiciel libre française, avant de pouvoir enfin passer la main à un président industriel, Stefane Fermigier.

Malgré ce que cet investissement m'a coûté, presque dix ans après, je ne le regrette pas : cela a donné un groupe thématique qui compte sur un comité de pilotage de bénévoles varié et engagé, avec de nombreuses femmes, et on a plus de cent membres (entreprises et laboratoires) qui ont pu faire financer à travers le pôle une cinquantaine de projets de recherche et développement totalement orientés vers le logiciel libre. Nous avons joué un rôle bénévole, mais précieux de *technological gatekeeper* en mettant en relation les différents acteurs de la recherche et de l'innovation dans la région Île-de-France.

V. S. : *J'ai vu aussi que vous aviez reçu un Microsoft Research Distinguished Artefact Award, cela peut sembler ironique !*

6. Voir <http://www.dicosmo.org/FSP>

7. *Berkeley Software Distribution License*.

8. *General Public License*.

9. Voir encart page 48.

R. D. C. : En effet en 2011 avec Jérôme Vouillon nous avons envoyé à la conférence ESEC/FSE¹⁰ un article qui décrit un algorithme sophistiqué lié à certains problèmes dans les logiciels libres, même si nous n'étions pas dans notre domaine traditionnel. Les organisateurs de cette conférence ont été les premiers à noter un problème : trop souvent, les articles décrivaient des résultats obtenus avec des outils logiciels mais ne fournissaient pas le moyen de les vérifier. En 2011, ils ont décidé de créer un prix pour les artefacts logiciels associés aux articles scientifiques. L'idée était de soumettre, si on le souhaitait, l'artefact logiciel, qui serait évalué par d'autres que ceux ayant évalué les papiers. Le jury devait dire si l'artefact *meets expectations*, *exceeds expectations*, etc. Jérôme a fait un travail spectaculaire, comme d'habitude, et quelques mois après nous avons appris que nous avions remporté le prix du concours et découvert que ce prix était financé par... Microsoft. C'était bien un signe que les temps étaient en train de changer !

Le Groupe thématique logiciel libre (GTLL)

« [...] Le logiciel libre ou *open source* est utilisé de plus en plus largement par les entreprises. Il représente en France un marché de plus de 3 milliards d'euros et en croissance annuelle de l'ordre de 30 %. En fournissant des briques technologiques sur lesquelles il est possible de construire des offres à valeur ajoutée, en accélérant l'émergence et l'adoption de nouveaux standards, en catalysant la collaboration entre des acteurs complémentaires, voire dans certains cas concurrents, le logiciel libre s'est imposé aussi bien dans des secteurs industriels comme le militaire et l'aérospatial, que dans l'informatique d'entreprise, où il joue un rôle primordial dans l'émergence de l'offre autour du *cloud*, du *Big Data*.

La France, et plus spécifiquement l'Île-de-France, possède un leadership reconnu internationalement dans le domaine. Des centaines de PME spécialisées, éditrices ou prestataires de services en logiciel libre, plusieurs grands intégrateurs, et un tissu d'universités et d'écoles d'ingénieurs constituent un écosystème actif qui s'est rassemblé pour constituer le Groupe Thématique Logiciel Libre (GTLL) au sein du Pôle Systematic.

Dans ce contexte, la mission du GTLL est de développer durablement l'écosystème du logiciel libre en Île-de-France, à travers 4 axes d'actions :

- Animer l'écosystème du logiciel libre ;
- Susciter et faire croître des projets innovants sur le logiciel libre (et doter les entreprises de technologies innovantes les rendant compétitives) ;
- Aider à pérenniser et développer les industriels du logiciel libre ;

10. R. Di Cosmo & J. Vouillon, « *On software component co-installability* », in Tibor Gyimóthy and Andreas Zeller, editors, SIGSOFT/FSE'11 19th ACM SIGSOFT Symposium on the Foundations of Software Engineering (FSE-19) and ESEC'11 : 13rd European Software Engineering Conference (ESEC-13), Szeged, Hungary, September 5–9, 2011, pp. 256–266. ACM, 2011. <http://www.dicosmo.org/Articles/coinstallability.pdf>

— Faire progresser la réflexion et communiquer sur les fondamentaux du logiciel libre.

Pour atteindre ces objectifs, le Groupe thématique s'appuie sur la communauté du Logiciel libre francilien, caractérisée par la diversité de ses acteurs : PME, grands groupes industriels et universités et établissements de recherche, associations et individus ».

<http://www.systematic-paris-region.org/fr/logiciel-libre/propos/presentation>

V. S. : *Je vous propose d'en venir au projet qui vous occupe actuellement et que j'ai découvert avec beaucoup d'intérêt, à savoir le projet Software Heritage.*

R. D. C. : On commence au printemps 2014 une série de discussions, d'abord avec Guillaume Rousseau, et un peu plus tard avec Stefano Zacchiroli, autour de la machine à café et de la thèse de l'Irill¹¹. C'est juste une idée vague au départ : on se rend compte qu'on aurait bien besoin d'un endroit unique qui serve de référence à tout le code source de la planète, comme base pour une série d'applications.

Mais c'était juste une idée : dans ce monde il y a plein d'idées qui n'ont rien donné, la difficulté est de les transformer en réalité. D'autant que nous nous sommes dit que cette archive devait forcément exister, que c'était déjà fait.

V. S. : *Par exemple chez Internet Archive ?*

R. D. C. : Je me souviens très bien de la façon dont j'ai découvert Internet Archive. Ted Nelson¹² faisait une tournée en France et il nous a dit : « J'ai un ami qui a lancé un truc... ». C'était le tout début d'Internet Archive, un projet magnifique qui nous permet de revisiter vingt ans d'histoire du Web à travers leur *WayBack machine*. On a donc tout naturellement regardé ce qui était fait pour les logiciels, tant par l'Internet Archive que par d'autres initiatives de préservation numérique.

Mais on a vite compris qu'archiver du code source pose des problèmes spécifiques qu'on ne rencontre pas dans d'autres domaines. Le code source est une connaissance à la fois exécutable par une machine et faite pour être lue et comprise par un être humain (le programmeur). Et il a vocation à changer très rapidement : le code source est vivant, et il est important de savoir ce qu'on y a changé et quand, si on veut comprendre son évolution.

On s'est finalement rendu compte que la préservation du code source avec ses spécificités n'était vraiment au cœur de la mission de personne : on préservait des logiciels exécutables, jouables, des jeux vidéos, c'était notamment fait par Internet

11. Initiative de recherche et innovation sur le logiciel libre.

12. Ted Nelson a forgé le terme *hypertext*. Ce sociologue américain a lancé le projet Xanadu dès les années 1960, rêve de bibliothèque universelle, dans laquelle tous les documents seraient reliés les uns aux autres, préfigurant le Web. Voir notamment : <http://xanadu.com/xUniverse-D6>

Archive qui a une grosse sélection de jeux vidéo. On préservait des pages web qui parlaient de logiciels et de codes sources. Mais les codes sources, comme objet noble, non.

On s'est aussi tourné vers le *Computer History Museum* : Len Shustek¹³, qui préside son conseil d'administration, avait déjà tout compris de ce problème, et écrit sa vision, extrêmement proche de la nôtre, dans un article de 2006 dans *IEEE Annals of the History of Computing*. Mais 2006 était trop tôt pour pouvoir imaginer archiver tout le code source de la planète : le *Computer History Museum* s'est donc spécialisé dans la conservation d'une petite sélection de codes sources de logiciels, analysés en profondeur, avec des interviews des architectes et développeurs.

Nous espérons tisser bientôt des liens avec tous ces acteurs, qui partagent en partie notre mission et notre vision, tout en adoptant des approches différentes, tant sur le plan technique qu'organisationnel.

« When we started Software Heritage we did not expect that publicly available source code was going to be endangered so quickly. Among other major source code hosters, Gitorious¹⁴ was already being taken down in mid 2015, before our initial infrastructure was ready!

Luckily fellow hackers of the Archive Team¹⁵ were already working full speed, and they managed to salvage Gitorious content in time ».

*Gitorious content is safe thanks to the Archive Team,
site Software Heritage, 21/07/2016*

<https://www.softwareheritage.org/2016/07/21/gitorious-retrieved/>

V. S. : *Ce projet est soutenu par Inria. Cela a-t-il été difficile de les convaincre ?*

R. D. C. : Inria est un institut de recherche unique. D'abord parce que c'est un institut entièrement dédié à l'Informatique, mais qui reste à taille humaine et où les décisions peuvent être prises rapidement. Il y a plus de vingt ans, Inria a accepté de devenir un des trois piliers du *World Wide Web Consortium* (W3C), en donnant notamment mission à Jean-François Abramatic.

Il y a un peu plus de deux ans, l'actuel PDG d'Inria, Antoine Petit, a immédiatement compris la portée du projet, et l'adéquation parfaite de l'institut pour l'amorcer, avec la même volonté de donner son envol à une structure qui a vocation à devenir internationale et multipartenaire.

13. Il préside le *Board of trustees* du *Computer History Museum*. Voir <http://www.computerhistory.org/atchm/author/lshustek/>, Len Shustek, « What Should We Collect to Preserve the History of Software ? », *IEEE Annals of the History of Computing*, 2006, vol. 28, n° 4.

14. <https://gitorious.org>

15. http://www.archive-team.org/index.php?title=Main_Page

Et les planètes se sont rapidement alignées : j'ai eu la chance d'attirer dans l'équipe Stefano Zacchirolli, enseignant-chercheur passionné de logiciels libres et ancien leader du projet Debian¹⁶, et puis d'autres passionnés. Jean-François Abramatic, de retour de l'industrie, est devenu président du comité de pilotage, où siègent Serge Abiteboul, Gérard Berry et Julia Lawall. Nous avons à la fois réalisé un prototype, qui montre qu'on prend la chose au sérieux, mais aussi expliqué comment on comptait faire et suscité la confiance. L'équipe actuelle est petite, mais le support précieux d'Inria permet de faire des miracles.

V. S. : *Quels sont vos objectifs en patrimonialisant le code source ?*

R. D. C. : Il y a essentiellement trois objectifs : l'archivage, le référencement et l'étude du code source.

La *mission d'archivage* est claire : il s'agit de ne plus perdre le code source des logiciels qui changent notre société.

La *mission de référencement* est concernée par le problème d'associer à chaque code source un identifiant unique pérenne, et de construire un catalogue de tous les logiciels existants, avec les informations sur leur développement et l'origine depuis laquelle on les a obtenus.

Enfin la *mission d'étude* est probablement celle qui a le plus de potentiel d'impact sur notre futur : nous voulons construire un grand instrument de recherche qui permette d'observer l'évolution du développement du logiciel à l'échelle planétaire ; cela permettra de développer un grand nombre d'applications pour améliorer la qualité et la sécurité de nos logiciels et comprendre comment mieux développer notre code.

V. S. : *Être exhaustif ?*

R. D. C. : Nous pensons que c'est nécessaire. Contrairement au regard que portent par exemple les bibliothèques sur le patrimoine, qui se méfient de l'auto-édition et où l'on préfère sélectionner plutôt que tout archiver, nous avons une optique différente, et cela pour deux raisons. D'abord, techniquement, nous ne sommes pas obligés de faire une sélection : la taille du code source est grande, mais petite par rapport aux images ou aux vidéos. Ensuite, d'un point de vue plus philosophique, c'est vraiment ce que nous voulons faire : pouvoir observer la naissance du code source et le voir grandir. Avec les logiciels libres, c'est possible, on peut voir le code à sa naissance, comme pour un embryon, et il est trop tôt pour savoir si ce sera un paysan ou un roi.

Donc nous archivons tout, et nous pensons que le filtrage se fera *a posteriori* : les logiciels intéressants vont être mentionnés et référencés, et donc émerger d'eux-mêmes, alors que les scories ne seront pas mentionnées et vont lentement décanter.

16. <http://upsilon.cc/~zack/>

Enfin, n'oublions pas l'aspect *Big Data* : on ne sait pas *a priori* ce qu'on voudra chercher dans cette masse de logiciels, peut-être même les raisons des échecs de certains d'entre eux, et pour cela il faut tout garder.

V. S. : *Quels sont les publics envisagés ?*

R. D. C. : Ce projet intéresse plusieurs publics. Ceux qui sont intéressés au patrimoine de l'Humanité bien évidemment découvrent petit à petit qu'une partie grandissante de notre connaissance scientifique, technique, organisationnelle et sociétale est inextricablement embarquée dans les logiciels, et commencent à s'y intéresser. Il faut non seulement le préserver, mais aussi le rendre facilement accessible afin d'améliorer l'éducation.

Ensuite, il y a le monde de la recherche scientifique. Rendre disponible son code source est en train de devenir indispensable pour permettre de comprendre les procédés décrits dans les articles et contribuer à la reproductibilité scientifique. Il faut une archive de référence où retrouver ces codes source : *Software Heritage* est tout désigné pour cela et répond en même temps aux besoins des scientifiques pour lesquels le code est un objet d'étude.

Il y a enfin un très fort intérêt industriel : *Software Heritage* peut remplir le rôle de base de connaissance universelle ouverte sur tous les logiciels libres qui sont massivement utilisés aujourd'hui, avec un grand nombre d'applications à la clé.

V. S. : *Avez-vous déjà une garantie de pérennité du projet ?*

R. D. C. : Pour pérenniser *Software Heritage*, nous souhaitons parler à tous ces publics et les réunir derrière une fondation indépendante, qui va leur permettre de contribuer à l'évolution du projet et de s'y appuyer pour leurs différentes applications. Nous espérons que le nombre de soutiens grandira très rapidement, avec la prise de conscience du fait qu'il s'agit d'une infrastructure essentielle.

Grâce à Inria, nous disposons de quelques années pour y arriver, et on compte déjà des soutiens variés, comme la Société Générale, Huawei, Microsoft, Nokia et DANS.

Mais il n'est pas nécessaire d'être une grande entreprise ou une institution pour aider : notre propre code source est ouvert, et a besoin de pas mal de contributions ; des défis scientifiques nouveaux se posent avec l'arrivée de cette archive unique et ont besoin de chercheurs qui s'y intéressent ; retrouver tout le code source, même ancien, qui est disponible est une tâche complexe et a besoin de personnes qui signalent les sources pertinentes ; expliquer l'intérêt du projet à tous les publics est important, et là il y a de la place pour tous. Enfin, nous savons que nous ne sommes pas à l'abri d'erreurs, et l'on en commettra sûrement, mais en adoptant une approche ouverte, en collaborant avec une multiplicité de publics, nous espérons pouvoir les réduire et les corriger rapidement.