Pour tout le monde : Tim Berners-Lee, lauréat du prix Turing 2016 pour avoir inventé... le Web

Fabien Gandon [1]

Introduction

« *Je suis moi et mes circonstances.* »

– Jose Ortega

Quel pourrait être un point commun entre se renseigner à propos d'un concert, effectuer un virement depuis son compte bancaire, publier une base de données génomiques, échanger avec ses enfants à l'autre bout du monde et accéder aux données de sa voiture [1] ? Le fait de pouvoir le faire à travers le Web. Il est en effet difficile de trouver une activité humaine qui n'ait pas été impactée par le Web et, alors que j'écris cet article en avril 2017, on estime que le Web compte plus de trois milliards d'utilisateurs directs de par le monde. Ce même mois, le britannique Sir Timothy John Berners-Lee est lauréat du prix Turing 2016 pour avoir inventé ce World Wide Web, le premier navigateur Web et les protocoles et algorithmes permettant le passage à l'échelle du Web [2]. Sir Tim, comme il est coutume de l'appeler, est actuellement professeur au MIT et à l'université d'Oxford. Ce prix est le dernier en date d'une longue liste de distinctions qu'il a reçues. Mais le prix Turing est considéré comme le prix Nobel de l'informatique et nous étions nombreux à attendre la nomination de Tim pour son invention du Web, cette invention qui a transformé notre société depuis sa création en 1989. C'est donc pour nous l'occasion dans cet article de revenir sur

1. Directeur de Recherche, Inria Sophia Antipolis. Responsable de l'équipe Wimmics (Université Côte d'Azur, Inria, CNRS, I3S). Représentant Inria au W3C (World Wide Web Consortium). Fabien.Gandon@inria.fr – http://fabien.info

son histoire en essayant de montrer un grand nombre des influences et courants qui se sont mêlés pour permettre une telle invention. Ce sera aussi l'occasion, pour nous, de démêler certaines notions et d'en réintégrer d'autres dans un effort de mise en cohérence des nombreuses influences qui ont tissé la toile.

De Turing à Berners-Lee : une brève préhistoire du Web

« [...] juchés sur des épaules de géants,
de telle sorte que nous puissions voir plus de choses
et de plus éloignées que n'en voyaient ces derniers. »

– Bernard de Chartres, XII^{e} siècle

Les travaux de Turing sont évidemment omniprésents en informatique mais lorsqu'il s'agit du Web, il y a un cas singulier qui est celui du CAPTCHA : ces insupportables tests que le Web nous fait passer à outrance au prétexte de vérifier que nous sommes humains. L'acronyme (« *Completely Automated Public Turing test to tell Computers and Humans Apart* ») signifie littéralement qu'il s'agit d'un cas particulier du jeu de l'imitation de Turing, complètement automatique et ayant pour but de différencier les humains des machines [26]. Accessoirement, c'est aussi, notamment pour certains géants du numérique, un moyen de se procurer du temps de cerveau disponible gratuitement [27] et de « *Web-sourcer* » (externaliser sur le Web) l'étiquetage massif de bases d'entraînement d'algorithmes d'apprentissage (*machine learning*), par exemple. Au-delà de cette anecdote établissant un lien direct entre Turing et le Web, on peut identifier quelques grandes influences plus fondamentales dans l'invention du Web et ceci dès les contemporains de Turing.

La recherche de techniques d'organisation et de moyens d'accès efficients aux masses d'informations que nous collectons a été une motivation omniprésente dans la préhistoire du Web. En juillet 1945, Vannevar Bush (MIT) écrit l'article « *As we may think* » (« Tel que nous pourrions penser ») [28] où il s'inquiète de ce que la capacité des scientifiques à assimiler les publications pertinentes pour leurs travaux soit dépassée par le volume de celles-ci. Vannevar pose alors comme un défi scientifique en soi le problème d'améliorer les moyens d'accès aux connaissances. Il propose comme premier élément de réponse un système (MEMEX [28]) utilisant des codes d'indexation mnémotechniques pour pointer et accéder rapidement à n'importe quelle partie de l'un des documents qui nous importent. Ces points d'accroche doivent aussi permettre de créer des liens entre deux éléments, externalisant ainsi le lien d'association. On trouve même littéralement la notion d'une toile d'idées (« web ») dans cet article historique lorsque Vannevar parle des liens d'association en ces termes : « *the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain* » [28].

Techniquement, pour réaliser un tel système, il faudra attendre vingt ans, l'apparition des ordinateurs et les débuts de leur utilisation pour l'édition de texte. Alors, dans un article à la conférence de l'ACM en 1965, Ted Nelson propose « une structure de fichier pour l'information complexe, changeante et indéterminée » [29] et il forge un néologisme pour nommer cet ensemble d'écrits interconnectés : l'hypertexte, et même l'hypermédia. Les années suivantes verront la réalisation des premiers éditeurs d'hypertexte utilisant, en particulier, une autre invention de cette décennie : la souris avec les interfaces et interactions qu'elle permet. On les doit notamment à Douglas Engelbart du Stanford Research Institute (SRI) qui recevra lui aussi le Prix Turing en 1997 pour sa vision pionnière et inspirante des interactions homme-machine et les technologies clefs qui ont permis son développement. En particulier, son système NLS (« *oN-Line System* ») combinera dans les années 1960, hypertexte, interface d'édition et souris. Le système sera plus tard renommé *Augment* en référence au programme de recherche d'Engelbart qui, en regard des travaux pionniers de cette époque en intelligence artificielle, propose de travailler sur l'intelligence augmentée [30].

Si le concept d'hypertexte est né, il reste, pendant des années, essentiellement limité à des applications s'exécutant localement à un ordinateur. La communication par commutation de paquets et inter-réseaux (*inter-networking* en anglais) se développe entre 1972 et 1975 avec les travaux de Louis Pouzin (IRIA puis Inria, [31]), de Vinton Cerf (SRI) et de Robert Kahn (DARPA, [32]) qui culminent en 1978 avec les protocoles standards (TCP/IP) et les débuts de l'Internet. Les applications de communication sur les réseaux se multiplient alors : le courrier électronique (SMTP), les listes de diffusion, le transfert de fichier (FTP), la connexion distante (Telnet), les forums de discussion, etc. Là encore nous trouvons deux récipiendaires du Prix Turing en 2004 : Vinton Cerf et Robert E. Kahn.

Cette trop courte généalogie intellectuelle place déjà Timothy Berners-Lee sur des épaules de géants lorsqu'il arrive en 1980 comme consultant pour le CERN (Centre européen de recherche nucléaire), alors jeune diplômé en physique (Queen's College, Oxford) et autodidacte de l'informatique. Devant la quantité d'informations et de documentations à gérer dans son travail et son équipe au CERN, Tim écrit un programme (*Enquire*) pour stocker des informations et les lier à volonté et notamment pour documenter les logiciels, ressources et personnes avec lesquels il travaille. Une différence importante d'*Enquire* avec d'autres systèmes hypertextes de l'époque est qu'il s'exécute sur un système d'exploitation multi-utilisateur et permet ainsi à plusieurs personnes d'accéder et de contribuer aux mêmes données [6]. Comparé au système de documentation CERNDOC basé sur une structure hiérarchique contraignante, Tim retient que les liens arbitraires et bidirectionnels de *Enquire* sont plus flexibles et évolutifs. De 1981 à 1983, Tim retourne en entreprise où il travaille sur l'appel de procédures à distance en temps réel (« *real-time remote procedure call* »),

donc dans le domaine des réseaux et de la programmation en réseau, avant de revenir au CERN en 1984.

Vague mais passionnant : les ruptures du filet et la naissance d'une toile

« The World Wide Web was precisely what we were trying to PREVENT – ever-breaking links, links going outward only, quotes you can't follow to their origins, no version management, no rights management. »

– Ted Nelson

En réalité, la première motivation de Tim pour créer le Web sera qu'il en avait lui-même besoin pour son travail au CERN, un campus ou plusieurs milliers de personnes se croisent avec de multiples spécialités et instruments [13]. Tim est convaincu qu'il y a un besoin pour un système d'hypertexte global au centre de recherche du CERN. En mars 1989, pour convaincre la direction du CERN, Tim écrit une proposition de projet intitulée « *Information Management : A Proposal* » [6]. Tim s'y fixe comme objectif de construire un espace où mettre toute information ou référence que l'on juge importante et où fournir le moyen de les retrouver ensuite [6]. À défaut d'un meilleur nom, il y décrit un système qu'il appelle « Mesh » (Filet) où il suggère l'utilisation d'un hypertexte distribué pour la gestion d'information au CERN avec notamment la notion d'ancrage (*hotspot*, équivalent aux entrées de Ted Nelson) permettant de déclarer un morceau de texte ou une icône comme le départ d'un lien activable à la souris. Cette proposition est un formidable exercice d'équilibre entre intégration et rupture avec les paradigmes existants et émergents à l'époque.

Dans son livre [4], Robert Caillau revient sur certains points importants du contexte général au CERN qui expliquent plusieurs choix faits dans la proposition de Tim. Commençons par quelques résultats existants à cette époque et que Tim intègre au cœur de l'architecture du Web.

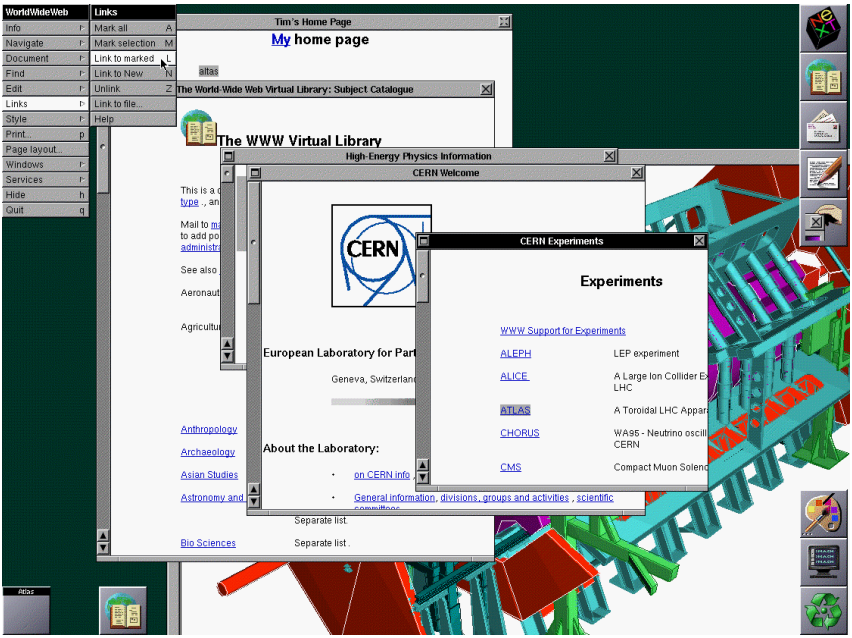
En 1989, le CERN est le plus grand nœud Internet d'Europe. Dans ce contexte, Tim a l'idée d'étendre les références des liens aux adresses réseau des documents afin de tisser ce « filet » (« *mesh* ») entre des documents mis à disposition par différentes machines. Conceptuellement, il reprend des principes des hypertextes et leur intègre TCP et le DNS. Plus précisément, Tim et l'équipe où il travaille s'investissent beaucoup pour l'acceptation et le déploiement au CERN de RPC (*Remote Procedure Call*, littéralement « appel de procédures à distance ») permettant à un programme s'exécutant sur une machine d'invoquer des procédures appartenant à des programmes s'exécutant sur d'autres machines. On peut voir RPC comme le chaînon manquant dans la distribution d'un hypertexte où le click sur un lien devient conceptuellement un appel de procédures à distance et le Web se conçoit alors moins comme un réseau

de documents mais comme un réseau de procédures. Des années plus tard, la thèse de Roy Fielding [33] introduira d'ailleurs le style d'architecture REST (*Representational State Transfer*) pour caractériser le système d'hypertexte distribué qu'est le Web. L'époque où Tim formule sa proposition est aussi celle où le système d'exploitation UNIX est très populaire et intègre nativement des fonctionnalités comme le support des protocoles Internet et les environnements multi-utilisateur.

Autre élément de contexte, en regard de la problématique de la documentation partagée notamment entre différentes disciplines au CERN, il règne déjà dans ce centre une culture de l'approche programmatique du document (p. ex. LaTeX, SGML) et ceci influence directement l'idée de s'orienter vers un langage de balisage simplifié. Une autre caractéristique assez peu connue du Web à sa conception, elle aussi héritée des systèmes d'édition d'hypertextes, est le fait qu'il était conceptuellement ouvert en écriture : tout dans son architecture permet non seulement de consulter (HTTP GET [34]) mais aussi de modifier le Web (HTTP PUT, POST, DELETE [34]). Le premier prototype de client imaginé par Tim est un navigateur-éditeur WYSIWYG [20] nommé W3. Cette fonction d'édition sera aussi longtemps portée et maintenue à l'Inria avec les travaux de Vincent Quint, Irène Vatton et leur équipe sur des bibliothèques et éditeurs Thot, Grif et Amaya [17]. Vincent avait par ailleurs travaillé avec Louis Pouzin sur le réseau Cyclades avant de s'intéresser aux documents structurés et en réseaux et, plus tard, Vincent deviendra le représentant d'Inria au consortium de standardisation de l'architecture du Web (W3C) avant de me passer le flambeau en 2012. Mais pour diverses raisons (sécurité, facilité, etc.) et par manque d'éditeurs [20], les navigateurs et serveurs les plus populaires ne vont pas mettre en œuvre cette possibilité de modifier les pages Web. Elle sera d'une certaine façon ré-introduite avec l'invention des Wikis (WikiWikiWeb) en 1994 par Ward Cunningham [44].

Enfin, la proposition de Tim relevant de la R&D informatique, elle ne peut être un projet du CERN en lui-même car elle est jugée hors du cœur de recherche de l'institut. La proposition est donc présentée comme un projet pour tester un nouvel ordinateur reçu au CERN [13], conçu et livré par la société NeXT dirigée alors par Steve Jobs, cofondateur d'Apple Computer, après sa démission forcée. Ce NeXT Cube était une station de travail haut de gamme, et elle va être utilisée par Tim comme premier serveur Web, ainsi que pour la conception du premier navigateur. Son système d'exploitation NeXTSTEP deviendra OPENSTEP, puis, à la suite du rachat de NeXT par Apple en 1996, Apple Rhapsody et enfin Mac OS et iOS. C'est aussi un environnement de programmation objet et graphique performant qui verra le développement de jeux mythiques comme Wolfenstein 3D, Doom et Quake. Pour ses premiers développements, trône donc sur le bureau de Tim un Cube NeXT qui matérialise la convergence des différentes influences que nous venons de mentionner et lui permet de s'appuyer sur l'expérience des hypertextes, l'avènement d'Internet et la mise

en réseau massive des ordinateurs (TCP/IP et Internet), les langages de programmation de haut niveau, les progrès des interactions homme-machine et notamment les interfaces graphiques et le multi-fenêtrage. Ce contexte de développement sera un accélérateur de la réalisation du premier serveur et du premier navigateur Web au-dessus d'une plateforme fournissant un grand nombre de briques de constructions de haut niveau. Ainsi fin 1990, le premier serveur et le premier navigateur sont testés à travers une connexion Internet. Le navigateur s'appelle World Wide Web ou « la toile d'envergure mondiale », qui deviendra le nom de l'hypertexte qu'il va engendrer et enverra aux oubliettes le « Filet » (« Mesh »).



Le premier site Web au monde est mis en ligne au CERN
le 20 décembre 1990 (Photothèque CERN).

Mais, nous l'avons dit, Tim Berners-Lee choisit aussi de rompre avec un certain nombre de caractéristiques des solutions existant à l'époque de sa proposition. Une raison est qu'il soutient l'idée qu'il faut travailler à un système d'information universel, dans lequel la généralité et la portabilité sont plus importantes que d'autres extensions.

La proposition est conçue pour couvrir des ressources publiques comme privées et leurs liens. La solution doit permettre des associations aléatoires entre des objets arbitraires et une contribution incrémentale et triviale pour les objets comme pour les liens par différents contributeurs, justifiant à nouveau une approche décentralisée

et non hiérarchique [5]. Dès le départ, il s'agit aussi de sortir des silos des applications existantes de prise de notes, publication, édition, documentation, d'aide, etc. et de rechercher l'indépendance au domaine, à la tâche, au matériel et aux systèmes d'exploitation [13]. On peut encore voir cette spécificité du Web dans le choix d'un couplage faible et notamment l'indépendance entre le serveur et le client. En effet, quel que soit le navigateur utilisé et quel que soit le serveur interrogé la communication fonctionne du moment que les standards sont respectés. Le navigateur masquera même l'utilisation de différents protocoles (HTTP, FTP, etc.) [20]. Notons que l'on retrouve cette préoccupation de briser les silos jusque dans les combats actuels de Tim Berners-Lee sur le maintien de la décentralisation du Web. La diversité des plateformes utilisées au CERN l'incitera non seulement à rechercher l'indépendance par rapport aux systèmes d'exploitation mais aussi, par rebond, à s'abstraire des chemins et systèmes de fichiers [5].

Tim Berners-Lee faisant une démonstration du World Wide Web
à la conférence Hypertext 1991, San Antonio, Texas (Photothèque CERN).

Enfin dans la proposition de Tim, l'hypertexte s'affranchit d'un serveur central : les données et les liens sont décentralisés sur Internet. De plus, les liens sont unidirectionnels et ne sont plus forcément maintenus : l'erreur 404 vient de naître. Cette rupture avec les fondamentaux des hypertextes est ce à quoi Ted Nelson fait référence dans la citation en début de cette section et explique aussi que la soumission de Tim à la troisième conférence ACM Hypertext en 1991 ne sera acceptée que comme une démo. Cependant ces ruptures avec l'existant sont autant de conditions

nécessaires au passage à l'échelle et à la viralité du Web. De plus, un certain nombre des ruptures du Web avec les hypertextes seront à l'origine de la création de services et d'applications Web majeurs. Par exemple, l'absence d'un index central et de liens bidirectionnels motivera la création d'annuaires, de répertoires (p. ex. Yahoo ! en 1994) et de moteurs de recherche (p. ex. Altavista, Google) pour (re)trouver des pages et des liens. Et, à bien y regarder, les pages comme celles générées par les résultats d'un moteur de recherche tissent maintenant à la demande une partie importante de la toile du Web.

In fine, on peut retenir trois notions fondamentales qui sont dès l'origine au cœur de l'architecture du Web :

— La première et la plus importante est celle des identifiants. Le prérequis majeur pour tisser une toile est d'avoir des points d'attache (*anchors*). La notion d'identification passera par les UDI, les URL, les URI et les IRI [13]. Les URL (*Uniform Resource Locator*) ou URI (*Universal Resource Identifier*) sont des formats d'adresses et d'identifiants permettant de localiser ou simplement nommer et faire référence sur le Web à n'importe quelle ressource. Si un identifiant (URI) donne un chemin d'accès pour obtenir une représentation de cette ressource sur le Web, alors on parle d'URL. Tout le monde connaît maintenant ces adresses, même si à l'origine elles n'étaient pas destinées à être manipulées directement par les usagers du Web. Par exemple, « `https://www.inria.fr/` » est l'URL de la page d'accueil d'Inria. Les URIs sont qualifiés d'universels au sens où ils doivent permettre d'identifier tout membre de l'ensemble universel des adresses réseau, dès l'instant où le protocole a une notion d'objet (la ressource). Il s'agit donc ici d'une vision orientée objet du réseau [20].

— La deuxième notion fondamentale de l'architecture du Web est celle de la communication et du transfert des données. Pour cela, le protocole HTTP permet notamment, à partir d'une adresse URL, de demander une représentation de la ressource identifiée et localisée par cet URL et d'obtenir en retour soit les données de cette représentation, soit des codes d'erreur indiquant un problème rencontré : par exemple, la célèbre erreur 404, qui indique que la ressource demandée n'a pas été trouvée.

— La troisième notion fondamentale est celle de la représentation des données échangées. Le Web étant initialement motivé par la représentation et l'échange de documents, le langage HTML sera le premier proposé pour représenter, stocker et communiquer les pages Web. Dès 1990, la documentation du Web est elle-même en HTML et le Web commence à s'auto-documenter ouvrant la possibilité à chacun d'augmenter cette documentation et d'apprendre à tisser en tissant. Cette forme de réflexivité confère au Web le statut d'un espace conçu pour qu'humains et machines y co-évoluent et y collaborent. Elle supporte la co-compréhension, co-documentation et la co-conception de tous les

sujets qui viennent s'y inscrire, à commencer par le Web lui-même.

Notons enfin que si URI, HTTP et HTML sont des inventions, la façon dont elles se combinent en une architecture simple et élégante l'est encore plus [13] et cela contribuera grandement à la viralité du Web.

Vague mais passionnant : début d'une viralité

« *Build small, but viral.* »

– Tim Berners-Lee

C'est avec la formule « vague mais passionnant » (« *vague but exciting* ») que fut acceptée la proposition d'hypertexte distribué de Tim Berners-Lee par son responsable Mike Sendall [35]. À l'heure où beaucoup de scientifiques s'interrogent sur l'efficacité de la gestion de la recherche par appels à projets, ce retour succinct mais assumant la prise de risque peut laisser songeur.

Nous l'avons dit, fin 1990, le premier serveur et le premier navigateur sont testés à travers une connexion Internet. Le 6 août 1991, Tim poste un résumé du projet World Wide Web sur plusieurs groupes de discussions en ligne (*newsgroups*), dont le forum dédié au sujet des hypertextes `alt.hypertext`. C'est le moment où le projet Web devient une application publiquement disponible sur Internet [18]. Le premier site Web est mis en ligne le même jour à l'adresse `http://info.cern.ch`. C'est aussi la source de documentation du Web lui-même, la graine de départ, le noyau auto-générateur en quelque sorte dont on trouve encore une archive en ligne [16]. Toujours en 1991, le premier serveur Web est installé hors d'Europe au Stanford Linear Accelerator Center. De ce point de départ, le Web va multiplier par dix le nombre de ses serveurs chaque année les deux premières années. Début 1992, on recense une dizaine de serveurs Web et de nouveaux navigateurs apparaissent au cours de l'année (Erwise, ViolaWWW, MidasWWW, Samba pour Macintosh, etc.). En 1993, les dirigeants du CERN annoncent officiellement que la technologie du Web sera gratuite et libre de droits [18], et en début d'année, on dénombre une cinquantaine de serveurs. De nouveaux navigateurs apparaissent (Lynx, Cello, Arena) mais le plus important est Mosaic, alors disponible sous Unix, Windows et Mac OS. Il permet notamment de visualiser les images directement dans le texte d'une page. Avec le navigateur Mosaic, le Web va réellement se répandre mondialement, laissant derrière lui ses ancêtres Gopher, WAIS et FTP. À Mosaic succéderont, dans l'arbre généalogique des navigateurs, Netscape puis Mozilla et enfin FireFox. En 1994, plus de 600 serveurs sont en ligne. L'année suivante, plus de 10 000 serveurs Web sont disponibles et Microsoft lance Internet Explorer, qui s'imposera comme le navigateur sous Windows, avec lequel il est diffusé. L'année 1995 voit aussi la naissance de JavaScript poussé par Netscape. En 1996 on passe la barre des 100 000 serveurs

et, en 1998 celle du million. Au début des années 2000 on en dénombre 26 millions et en 2004 les 46 millions ont été largement dépassés.

Un autre indicateur de cette viralité est donné par Tim dans un article de 1994 où il explique que la charge des accès à la documentation du Web sur le serveur `info.cern.ch` double tous les quatre mois entre avril 1991 et avril 1994 [20]. Cette viralité s'explique par plusieurs choix faits lors de la conception et la réalisation du Web. Outre les choix architecturaux déjà décrits, plusieurs éléments vont être décisifs pour permettre et maintenir la viralité du Web. Il y a des éléments techniques à commencer par la proposition de Tim de mettre l'accent sur l'importance de la généralité, de la portabilité et de l'extensibilité, plus importantes selon lui que la satisfaction d'utiliser les dernières capacités des ordinateurs (p. ex., le graphisme). Tim aurait pu intégrer des technologies plus complexes ou viser de multiples fonctionnalités supplémentaires mais si le Web est aujourd'hui une contribution d'importance technique durable et majeure à la communauté informatique c'est notamment en raison de sa simplicité, de son élégance et de son extensibilité [2].

Dans sa proposition, Tim prévient aussi que le résultat doit être suffisamment attrayant à l'usage pour que les informations contenues dépassent un seuil critique et qu'à son tour cette masse d'information encourage toujours plus l'utilisation et la contribution. Pour cela, il propose dès le début de prévoir une liaison entre les bases de données existantes et les nouvelles [6], et de rendre le Web systématiquement compatible avec l'existant. Dès sa création, le Web intègre des serveurs « *gateway* » pour importer des ressources légataires et donner accès à d'autres applications à travers des techniques comme CGI (*Common Gateway Interface*). Le Web va ainsi intégrer, interopérer et finalement absorber des systèmes existants, notamment WAIS et Gopher. Cette approche facilite le basculement complet de communautés d'utilisateurs d'anciens systèmes vers le Web. Tim conçoit aussi la rétrocompatibilité ou compatibilité descendante avec les protocoles précédents (FTP, Gopher, WAIS, etc.) comme une contrainte d'interopérabilité, une preuve de flexibilité et surtout une assurance d'évolutivité pour le futur [5, 20]. De plus, avec des approches comme les CGIs, la génération automatique de pages dynamiques et la possibilité de les référencer et de les lier joue immédiatement un rôle vital dans le tissage d'un graphe du Web qui soit un minimum fourni (nombreux nœuds), connecté et dense (nombreux liens) [20].

Tim a donc reconnu que la simplicité était nécessaire pour une adoption généralisée, en particulier dans la communauté scientifique qu'il ciblait. Ses simplifications de protocole, y compris son insistance sur l'absence d'états dans le protocole HTTP, ont rendu la conception facile à mettre en œuvre. De même, l'utilisation de scripts lisibles par l'humain rendait le système compréhensible, facile à déboguer, et viralement réutilisable par copier-coller-adapter [2]. Tim s'inspire systématiquement de l'existant aussi pour encourager l'adoption. Ainsi, HTTP s'inspire de SMTP et NNTP et les en-têtes utilisées dans les échanges HTTP (*headers*) sont une extension

des MIME (*Multipurpose Internet Mail Extensions*) afin d'ouvrir la porte à l'intégration des hypermédia, des mails, des news, etc. [20]. Du côté des URI, on peut noter l'intégration du DNS et des slash (/) pour leur structure [13]. De même, la volontaire proximité de HTML à SGML va encourager l'adoption de HTML par la communauté de la documentation [5]. Enfin, en termes d'interaction, le couplage de l'hypermédia et des formulaires textuels se révèle suffisamment simple et puissant pour couvrir de nombreux besoins tout en restant facile à mettre en œuvre et utiliser [20].

Mais au-delà des choix techniques d'autres choix économiques, légaux et sociaux vont aussi faire la différence et peser dans l'acceptation du Web. En premier lieu, l'architecture et les fondations technologiques du Web sont *open source* (« code source ouvert »), libres de droits et gratuites. Là encore, le contexte historique est important : la « Fondation pour le logiciel libre » (FSF) a été fondée par Richard Stallman en 1985 pour promouvoir le logiciel libre et elle sera à l'origine du projet GNU et des licences GPL qui, dès 1989, fixent les conditions légales de distribution d'un logiciel libre. Le début des année 90 verra aussi les négociations et le passage d'Unix (BSD) à l'*open source*. En rendant le Web gratuit et son code ouvert, le CERN fait littéralement don du Web au monde. En particulier, en 1992, Tim Berners-Lee et Jean-François Groff travaillent à une version en C qui aboutit à la création en code libre de la célèbre bibliothèque logicielle `libwww` qui sera utilisée par la suite dans de nombreuses implémentations [20]. Tim lui-même n'a jamais cherché à monétiser son travail et défend dès le départ un Web en logiciel libre et un code ouvert. Ce tournant important va permettre la pénétration virale des technologies Web dans toutes les organisations et dans leurs applications.

Une autre initiative décisive est l'établissement en 1994 du World Wide Web Consortium (W3C) qui va jouer un rôle primordial dans la normalisation des évolutions de l'architecture du Web, lui permettant de grandir sans perdre l'interopérabilité standard qui lui donne son universalité. Parce que le CERN est conduit à mettre toutes ses ressources dans la construction du LHC (*Large Hadron Collider*), il annonce fin 1994 qu'il ne peut plus continuer à s'investir dans le projet du Web (appelé projet WebCore). Avec le soutien de la Commission européenne, l'activité Web du CERN est transférée au W3C avec comme membres fondateurs : le MIT aux États-Unis, l'Inria en France et l'Université de Keio au Japon. Jean-François Abramatic, alors directeur des relations industrielles à l'Inria, va jouer un rôle crucial dans ce transfert et il deviendra quelques temps après président du W3C. Par la suite, Inria transférera le rôle d'hôte européen du W3C à l'ERCIM (European Research Consortium for Informatics and Mathematics) que l'institut a aussi contribué à créer avec des partenaires européens en 1989. Avant le W3C, les standards du Web étaient publiés sous forme de RFC (*Request For Comments*). Ils seront dès lors publiés comme des recommandations du W3C. Le consortium n'a cessé de chercher à améliorer son action depuis sa création en affinant et affichant sa mission [21], son

organisation [23], son fonctionnement [24] et son éthique [25]. Les travaux du W3C ont été, et le seront encore, décisifs pour permettre au Web de traverser des crises majeures, qu'elles fussent techniques (p. ex. la guerre des navigateurs et les incompatibilités), commerciales (p. ex. les extensions propriétaires), politiques (p. ex. le standard PICS pour répondre aux inquiétudes quant aux contenus inappropriés), etc. Le W3C ouvrira ainsi très vite des groupes de travail sur de multiples évolutions et facettes du Web comme l'accessibilité ou l'internationalisation. Maintenant, de nombreux groupes de standardisation ou simplement de discussion existent sur une grande variété de sujets.

Évolutions du Web : tisser une toile mondiale de ressources

« *When I took office, only high energy physicists*
had ever heard of what is called the World Wide Web...
Now even my cat has its own page. »

– Bill Clinton, 1996

La relecture que fait Tim de l'architecture du Web en 1996 [5] change un peu ses trois piliers en mettant l'accent sur l'adressage, le protocole et la négociation de contenu. Cette négociation de contenu est un mécanisme natif du protocol HTTP qui offre la possibilité de proposer, pour un même URI, différentes versions d'une même ressource et qui est directement inspiré du mécanisme de négociation de format du « System 33 » de Steve Putz au Xerox PARC. Le langage HTML n'est alors plus considéré que comme l'un des formats disponibles. Ainsi, dès 1994, Tim souligne lui-même que HTTP est peut-être mal nommé puisqu'il n'est pas tant limité au transfert de HTML que destiné à échanger des données arbitraires efficacement dans le contexte de liens et sauts dans un hypermédia distribué [20]. D'une certaine façon le mécanisme de négociation de contenu déclasse HTML en donnant la possibilité de négocier tout type de format dès le départ. Les URI eux permettent d'identifier tous les types de ressources. HTML redevient donc juste un prérequis pour un navigateur [20]. Il aura permis dans un premier temps de fournir un format uniforme de documents hypertextuels et de documentariser le réseau de ressources que devient le Web, avant d'évoluer vers un langage de programmation des applications Web. Inversement, le Web affirme ainsi son indépendance à un modèle ou une structure de données.

En 1994, Tim notait aussi déjà un besoin de faire évoluer le Web vers plus de fonctionnalités temps-réel (p. ex. téléconférence, réalité virtuelle) et d'envisager un support au commerce sur le Web, notamment pour le paiement en ligne [20]. Tim identifie de plus dans cet article de 1996 [5] des directions d'évolution qu'il considère importantes pour le Web et qui ouvrent encore actuellement de nombreuses

directions de recherche et développement. Il cite comme premier axe de travail l'infrastructure et les performances du Web ainsi que le Web comme un espace non seulement social mais aussi comme un espace pour les machines. Il suggère même que cet aspect, que l'on pourrait appeler maintenant des communautés hybrides [36], est vital et il explique que les machines sur le Web sont une nécessité du passage à l'échelle des interactions humaines. Il rappelle aussi un objectif initial toujours non atteint : la possibilité de lier facilement et sûrement les différentes échelles et sphères des documents privés et publics, allant des systèmes d'informations personnelles (PIM) aux systèmes de discussion globale, et d'offrir des outils de groupe à tous les niveaux (personnel, organisationnel, public) en préservant la capacité à lier à travers ces niveaux. Tim déplore aussi qu'en 1996 l'infrastructure de sécurité soit toujours absente du Web, car il est vrai qu'à part l'arrivée de SSL et HTTPS avec Netscape en 1994, peu de progrès ont été réalisés. Il faudra même attendre la RFC2818 en 2000 pour une première spécification formelle officielle de HTTPS. Enfin, Tim insiste encore et toujours sur l'importance de systématiquement rechercher plus de décentralisation : chercher l'ouverture à l'écriture et la contribution de contenus, l'externalisation des liens, l'externalisation des annotations, et l'externalisation des vérifications.

La standardisation de PICS comme l'une des premières recommandations du W3C en 1996 permet de filtrer les contenus inappropriés notamment pour les enfants. C'est aussi un exemple de cette décentralisation voulue : les filtres capturant les préférences des utilisateurs sont créés et stockés dans les clients (navigateur) ; les descripteurs sont stockés sur les sites consultés mais générés par des sites d'autorités tiers. Incidemment, PICS ouvre donc l'idée d'aller plus systématiquement dans l'architecture du Web vers une approche générique pour un problème spécifique, réutilisable pour d'autres scénarios d'usage. Ainsi, l'idée d'étiqueter par et en référence à un site tiers, propose plus génériquement de casser le lien binaire (navigateur/page)-(serveur/site) pour aller vers des liens plus complexes et aussi de ne pas limiter l'étiquetage aux problèmes d'acceptabilité du contenu. On s'ouvre alors à la notion de métadonnées en général sur le Web et on rejoint une autre évolution vers un Web de documents et données structurés.

Dans cette évolution, CSS est une étape importante qui marque le début de la séparation du fond et de la forme sur le Web (CSS vs. HTML). La notion de feuille de style permet de sortir et séparer la mise en forme de la structure du document et aussi de d'utiliser une même mise en forme pour plusieurs documents ou inversement de faire varier la mise en forme d'un même document. Cette étape va permettre aux contenus et aux présentations de se multiplier et de se diversifier de façon indépendante et créative. CSS va aussi marquer le passage du Web à des capacités de présentation professionnelles et très avancées rivalisant à terme avec les productions permises par les meilleurs outils d'édition de documents électroniques. Là encore la notion de feuille de style était présente dans le premier prototype de Tim, mais il

ne publie pas cette fonctionnalité ni la syntaxe qu'il utilise. Cette capacité est donc essentiellement perdue dans les premiers navigateurs qui suivront comme Mosaic. Håkon Wium Lie et Bert Bos seront les premiers à travailler sur une standardisation des feuilles de style qui débouchera en décembre 1996 sur la recommandation CSS level 1. Le succès est aussi dû à Thomas Reardon et Chris Wilson de Microsoft qui dès 1995 assurent qu'ils supporteront CSS. Et en effet, Microsoft Internet Explorer 3 sera la première implémentation de CSS alors que sa spécification n'est encore qu'un brouillon. L'histoire de CSS [45] est aussi l'un des premiers exemples d'activité qui demandera le développement de suites de tests dédiées et de démonstrateurs (p. ex. CSS Zen Garden) pour accélérer l'implémentation et l'adoption d'une recommandation et d'une évolution de l'architecture du Web.

Après la séparation du fond et de la forme, le contenu du Web va pouvoir évoluer en permettant de créer et gérer ses propres structures de documents et données. C'est la standardisation de XML en 1998 et, dans son sillage, de plusieurs langages permettant sa validation (DTD, XML Schemas), son interrogation (XPATH, XQuery), sa transformation (XSLT), sa gestion (XProc), etc. À titre personnel, je retiens en particulier à cette époque la feuille de route pour le Web sémantique que Tim publie en 1998 [37]. Cette feuille est dans la continuité de sa présentation de 1994 [19] et aussi de son article de 1994 où l'on peut lire qu'il souhaite une évolution des objets du Web, qui sont à l'époque essentiellement des documents destinés aux humains, vers des ressources avec une sémantique plus orientée vers les machines pour permettre des traitements plus automatisés [20]. La feuille de route de 1998 [37] ouvrira tous les travaux sur le Web de Données et le Web Sémantique (RDF, RDFS, SPARQL, OWL, etc.). En France, Rose Dieng-Kuntz, Olivier Corby et leur équipe à l'Inria identifient immédiatement cette évolution et lancent des recherches sur le sujet [10, 11]. Cette feuille de route sera décisive pour moi qui commence, au sein de cette équipe, en 1999, une thèse sur le couplage des architectures logicielles de l'intelligence artificielle distribuée (modèles multi-agents) et des modèles formels de connaissances distribuées (modèles du Web Sémantique) [12]. À ce jour, Tim est toujours un défenseur passionné des évolutions du Web comme le Web Sémantique [2].

En parallèle de cette évolution des contenus, a aussi lieu la naissance de la programmation Web et des applications Web qui vont devenir un pilier important de la toile et ses usages. Au départ, à la naissance du Web, le principe était très simple : une page générée à la demande suite à un clic sur un simple lien ou en réponse à la soumission d'un formulaire. À chaque interaction, la page était entièrement rechargée. Sur le serveur, des codes écrits en C, C++, Shell ou autres langages utilisent la méthode CGI (*Common Gateway Interface*) pour traiter ces requêtes et générer des pages en retour. Avec les *frames* vers 1996, on commence à pouvoir découper la page en cadres et par la suite ne recharger qu'une partie de ce qui est affiché. La même année, JavaScript fait son apparition et on peut commencer alors à utiliser dans les applications Web, même conjointement, de la programmation côté serveur et de

la programmation côté navigateur. Commence alors une généalogie de langages et techniques de programmation pour le Web, certains plutôt côté serveur (ASP, PHP, C#, Python, Ruby, Perl, JSP, etc.) d'autres plutôt client (p. ex., Plugins, ActiveX, CSS+HTML) et aussi d'autres utilisables des deux côtés (Java Servlet et Applet, JavaScript). Enfin, avec le composant *XMLHttpRequest* proposé par Microsoft en 1998 puis ajouté à JavaScript et rapidement adopté par la plupart des navigateurs, on peut échanger des données entre une page et son serveur, et, grâce au DOM de celle-ci, modifier la page affichée sans nécessairement la recharger. Cette technique sera nommée en 2005 AJAX et massivement adoptée dans les applications Web pour conjuguer du code s'exécutant sur le client et du code s'exécutant sur le serveur et permettre des interactions fluides avec l'utilisateur. En parallèle, l'architecture du Web est de plus en plus étudiée et formalisée comme avec la thèse de Roy Fielding qui introduit l'architecture REST pour la caractériser [34]. Avec le début des années 2000, c'est la proposition d'évoluer vers un Web de services (standards SOAP et WSDL) qui ouvre une nouvelle direction de travail pour l'utilisation du Web comme une plateforme programmatique et qui se réalise plus actuellement à travers les API et les langages liés comme JSON. Certains vont jusqu'à parler du Web comme un système d'exploitation au-dessus de la collection mondiale de services qui s'offrent sur Internet, et indépendant des ordinateurs et objets individuels qui s'y connectent. Ils positionnent le Web comme l'environnement de programmation et d'exécution par excellence des applications de l'Internet. Dans cette mouvance, l'une des prochaines évolutions actuellement étudiée est de faire du Web la plateforme applicative universelle de l'Internet des objets que l'on nomme le Web des Objets (*Web of Things* [38]). De nos jours, on envisage littéralement de faire une toile de tout. Mais si l'on revient cependant à cette influence initiale que RPC a eu sur la conception d'un hypertexte distribué, on peut en fait voir cette évolution comme un retour aux sources. De plus, en se représentant le Web comme une toile d'appels potentiels de procédures que l'on invoque à chaque lien que l'on suit, on comprend mieux pourquoi des ambitions comme celles de moissonner (*crawler*), indexer ou archiver le Web sont compliquées voire paradoxales.

Toujours en parallèle, dès 1996, des compagnies comme Nokia en Finlande vont s'intéresser à proposer un accès au Web sur les téléphones mobiles. En 1999, se sera NTT DoCoMo au Japon avec le i-mode. Cette même année, le code QR, créé par l'entreprise japonaise Denso-Wave en 1994, passe sous licence libre ce qui va contribuer à sa diffusion et sa mise à disposition dans des applications de reconnaissance sur les téléphones mobiles notamment pour glaner des URL affichés autour de nous. À cette époque aussi, WAP (*Wireless Application Protocol*) et WML (*Wireless Markup Language*) seront proposés pour adapter l'accès et les contenus Web aux contraintes des téléphones portables et de leur connectivité. Ils seront par la suite abandonnés lorsque les téléphones et réseaux atteindront des performances leur permettant de directement accéder au Web et à l'Internet classiques. Ces premiers essais,

ainsi que ceux faits avec des PDA ayant une carte réseau sans fil, devront attendre le milieu des années 2000 et l'avènement des smartphones pour trouver des plateformes au-dessus desquelles réaliser leur plein potentiel. C'est en 2004 que le W3C en lancera l'initiative Web mobile (MWI). En 2005, le nom de domaine de premier niveau « .mobi » est proposé mais sera critiqué, notamment par Tim, comme une solution qui casse l'indépendance du Web vis-à-vis du terminal. En 2007, dans une conférence invitée au congrès GSM, Tim se positionne aussi contre les prés carrés créés par les plateformes propriétaires et défend la plateforme ouverte qu'offre le Web [46]. Dans un premier temps, les problématiques auront donc été de pallier les limitations d'une connexion mobile (écran, bande passante, interactions limitées, puissance de calcul limitée, coût de connexion, etc.). Puis, soit avec la résolution de ces problèmes (p. ex. compression), soit avec leur disparition (p. ex. montée en puissance des terminaux et réseaux), les problématiques de cette évolution du Web vont progressivement passer de l'accès mobile simple à l'adaptation plus profonde à une utilisation en situation de mobilité (p. ex. géolocalisation, adaptation des interactions et interfaces, accès aux données personnelles, contextualisation, interaction audio et vocale, réalité augmentée, couplage de plusieurs terminaux, etc.). Nous avons maintenant de façon courante des applications Web mobiles, et beaucoup d'applications mobiles natives sont dans les faits grandement développées avec des langages et standards du Web. De plus, les prix d'entrée de gamme et la démocratisation des smartphones font que dans certaines régions du monde, le mobile est maintenant la première façon d'accéder au Web et le moyen le plus répandu d'avoir un premier contact avec le Web [39].

Entre Web de données, Web d'applications, Web de services, Web mobile, mais aussi Web multimédia, accessible, internationalisé, etc., le Web commence donc très tôt sa transformation vers une architecture de programmation et d'interaction hypermédia générique et surtout sa généralisation à une toile liant potentiellement tous types de ressources computationnelles ou non. Le Web peut toucher à tout puisque tout peut être identifié par un URI. Les principes du Web étant extensibles et génériques, ils nous ont permis de passer d'une vision documentaire de bibliothèque mondiale à un réseau de ressources protéiformes. L'une des plus grandes forces du Web est dans son universalité mais nous verrons qu'elle demande une constante attention pour être préservée.

Catégories et déclinaisons : une toile à facettes

« *As soon as you externalize an idea*
you see facets of it that weren't clear
when it was just floating around in your head. »

– Brian Eno

Pour parler des temps forts des évolutions du Web on trouve maintenant les appellations Web 1.0, Web 2.0, Web 3.0, etc., par lesquelles je ne suis pas convaincu car elles laissent penser qu'elles correspondent à des évolutions logicielles majeures du Web alors qu'elles sont plus souvent des évolutions des pratiques ou même de la compréhension que nous avons du Web et qu'elles ne rendent pas compte des multiples directions dans lesquelles le Web évolue en parallèle. Le Web 1.0 correspond essentiellement à la vision initialement documentaire et d'hypermédia distribué du Web. Le Web 2.0 est aussi appelé Web social et rend compte à la fois de la réouverture en écriture du Web, de l'approche AJAX pour l'interaction, et de la contribution et des échanges sociaux massifs qu'ils permettent avec les wiki, les blogs, les forums, les médias sociaux, etc., avec l'impact que l'on connaît. Le Web 3.0 recouvre en général l'intégration des pratiques du Web sémantique et du Web de données au Web 2.0, par exemple avec RDFa dans le protocole OGP de Facebook ou dans l'utilisation de `Schema.org` dans de nombreux sites intégrant des fonctionnalités sociales (p. ex. votes, critiques, etc.). Cependant, comme nous l'avons vu dans la section précédente, plus que des évolutions par sauts de versions, le Web vit en permanence un bouquet d'évolutions concurrentes qui demandent un travail constant et considérable au W3C pour rester compatibles, mais qui en même temps, dans une approche évolutionniste, lui permettent de lancer de multiples sondes à la recherche de ses prochaines mutations et de leurs croisements.

Un autre jeu de facettes que l'on reconnaît au Web sont : le Web de surface, le Web profond et le Web obscur. Le Web de surface est le Web indexable et parcouru par les moteurs de recherche (services appelés *crawlers*). Il est public et forme la partie la plus émergée du Web. Il compte aussi de nombreuses pages d'entrée de portails et applications qui eux ouvrent sur le Web profond (*deep Web*) avec des modalités d'accès, de parcours et de recherche dédiées. Le Web profond est aussi injustement qualifié de caché ou invisible alors que nous le voyons tous les jours. Il s'agit du Web essentiellement généré dynamiquement comme les pages de résultats de recherche qui nous sont bien visibles et accessibles mais qui ne sont pas indexées par les moteurs pour différentes raisons : contenus ou liens générés dynamiquement, contenus accessibles après une authentification ou tout autre formulaire ou interaction complexe allant au-delà du suivi d'un lien, contenus non liées au reste du Web, contenus avec une politique d'exclusion des robots d'indexation (fichier robots.txt), etc. Cette partie dite profonde du Web représente en fait la majeure partie des contenus de la toile.

Le terme de *deep Web* est parfois confondu à tort avec les termes *darknet* et *dark-web* qui ont pourtant un sens bien différent. Le Web obscur (*dark Web*), Web sombre ou Web noir par équivalence au terme de matière noire, est un Web de l'ombre utilisé dans des activités recherchant l'anonymat que ce soit, par exemple, un opposant politique cherchant à échapper à l'oppression de son pays ou une organisation criminelle cherchant à utiliser la puissance des outils du Web pour son activité. Les

toiles de ce Web sont volontairement déconnectées du Web classique pour ne pas être trouvées et indexées. Ces toiles sont tissées sur l'Internet obscur (*darknet*) qui utilise des techniques d'anonymisation, de cryptographie, de réseaux (p. ex. pair-à-pair) et de sécurité en général pour masquer les identités et échanges des usagers. En combinant ces techniques avec l'architecture Web, on peut tisser et naviguer sur des toiles cachées (p. ex. le navigateur TOR). Les termes *dark Web* et *darknet* sont souvent utilisés de façon interchangeable alors que la notion de *darknets* désignant des réseaux volontairement isolés existait déjà à l'époque d'Arpanet dans les années 1970, donc bien avant le Web, et leur utilisation applicative s'étend à bien d'autres services (mail, IRC, forum, etc.).

Enfin, on peut aussi mentionner les IntraWeb (l'utilisation du Web en intranet) qui, derrière des VPN et pare-feux (*firewall*) d'entreprise, utilisent au sein de nos organisations les solutions du Web pour créer des toiles réservées à leurs membres. Là encore, des techniques de sécurité sont utilisées pour contrôler l'accès à ces composantes du graphe du Web. On ne cherche pas forcément à éviter les liens entre ces toiles et le Web public, mais on en sécurise l'accès.

Les évolutions de la section précédente ou les facettes que nous venons de mentionner relèvent cependant d'une seule et même architecture : l'architecture standardisée du Web. Comme insiste la description de la mission du W3C (*One Web*), il ne s'agit pas de différents Webs mais de différentes facettes d'un et un seul Web, d'une et une seule architecture [21]. Au sein même du W3C, il existe d'ailleurs un groupe spécial appelé le TAG (Technical Architecture Group) [22] où siège notamment Tim au titre de Directeur du W3C et qui veille, notamment, à ce que toutes les évolutions du Web restent compatibles et continuent à respecter et se combiner en une architecture cohérente et universelle. Ces quelques exemples des évolutions et facettes du Web nous amènent tout de même à un nouveau besoin : celui de se doter de moyens d'étudier le Web et ses évolutions.

L'inconnu du Web

« *Web science – what makes the Web what it is, how it evolves and will evolve, what are the scenarios that could kill it or change it in ways that would be detrimental to its use.* »

– Dame Wendy Hall

Aujourd'hui encore, il est frappant de voir à quel point le Web est à la fois très connu et à la fois mal connu comme en témoigne la confusion tenace entre les termes Web et Internet que l'on rencontre encore bien trop souvent. Malgré le fait que leurs inventeurs respectifs aient reçu deux prix Turing bien distincts, respectivement en 2004 et en 2016 pour deux inventions bien différentes, Internet et Web sont encore

trop souvent utilisés de façon interchangeable. Redisons le : Internet permet l'interconnexion des réseaux d'ordinateurs et objets connectés en général. Il fournit une infrastructure de communication qui supporte au-dessus d'elle de nombreuses applications comme : la messagerie électronique (*mail*), la téléphonie et la vidéophonie... et le Web, cet hypermédia distribué qui devient l'architecture logicielle majoritaire des applications sur Internet.

Outre cette confusion, le terme Web est aussi souvent utilisé de façon indifférenciée pour se référer à la fois aux principes fondateurs de cette architecture logicielle et à l'objet qui en émerge, i.e. la toile tissée par des milliards d'utilisateurs. Dès 1994, Tim note que le terme de World Wide Web a très rapidement recouvert plusieurs choses et notamment : d'un côté une architecture et de l'autre un jeu de données mises à disposition sur Internet selon cette architecture [20]. L'architecture et l'objet qui en émerge ont deux histoires liées mais portent sur des aspects différents. Chacune des deux facettes exhibe cependant de façon différente une complexité qui nécessite à la fois recherche et développement.

En effet, l'architecture du Web repose sur des protocoles, des modèles, des langages et des algorithmes qui nécessitent d'être spécifiés, conçus, caractérisés et validés avec de plus, systématiquement, des contraintes comme celles du passage à l'échelle, de l'efficacité en temps et en mémoire, d'interopérabilité et d'internationalisation. Pour cela, l'architecture du Web et ses extensions sont des objets de recherche notamment dans les sciences du numérique. Au sein de l'informatique et des sciences du numérique, et nous verrons dans la section suivante que c'est valable pour de nombreux autres domaines aussi, le Web s'est répandu à la fois comme un nouvel outil de travail mais aussi comme un nouveau sujet apportant à la fois des solutions et des nouveaux problèmes et besoins.

Quant à l'objet qui émerge de l'usage de cette architecture, la complexité de ses usages, l'hétérogénéité et les volumes de ses contenus, services et données, la dynamicité de certains de ses flots ou les cycles de vie de ses ressources et communautés sont autant de sources de complexité qui à nouveau requièrent une approche scientifique et des recherches théoriques, appliquées, expérimentales et multidisciplinaires. Une proposition, parmi d'autres, illustrant cette idée d'étudier le Web comme un objet complexe est, par exemple, d'avoir des observatoires du Web [40] avec des instruments d'observation et des méthodes des sciences expérimentales pour étudier le Web ; se doter en quelque sorte de microscopes ou télescopes du Web et de la méthode scientifique pour « tourner ce télescope vers les toiles ».

Et plus le Web grandit en complexité architecturale et en ressources liées dans sa toile, plus il appelle des recherches et des développements transdisciplinaires [41]. Tim ira même jusqu'à dire qu'*in fine* le Web est plus une création sociale qu'une création technique [3].

Historiquement, le Web est rapidement devenu un objet de recherche et, là encore sous l'impulsion de Robert Cailliau et Tim Berners-Lee, commence au CERN dès

1994 le cycle des conférences WWW, « *The Web Conference* », devenu le rendez-vous annuel de la recherche, du développement et des industries et acteurs majeurs du Web. Pour avoir accepté d'être deux fois co-président de cette conférence, en 2012 et en 2018, je peux attester de l'importance et du niveau des recherches scientifiques que le Web suscite. Et cette communauté de recherche grandit et se diversifie avec des initiatives résolument multidisciplinaires comme « Web Science » ou des conférences plus spécialisées comme ISWC, WI, WebIST, etc. Pour conclure ce volet, notons que Tim est devenu Directeur en 2006 de la Web Science Research Initiative (WSRI), un programme de recherche scientifique portant sur les aspects techniques et sociaux qui sous-tendent les évolutions du Web [14] et qu'il est aussi fondateur du Web Science Trust [15], une organisation britannique à but non lucratif consacrée à l'étude interdisciplinaire du Web et de ses effets sur la société [2].

Extrait central du Panel final à la conférence WWW 1994. De gauche à droite :
Dr. Joseph Hardin, Robert Cailliau, Tim Berners-Lee et Dan Connolly (Photothèque CERN).

Où l'on bascule vers un Web-Wide World

« *We can only see a short distance ahead,*
but we can see plenty there that needs to be done. »

– Alan Turing

« *This is for everyone* », c'est avec ce message que Sir Tim Berners-Lee présente le Web lors de la cérémonie d'ouverture des Jeux olympiques d'été de 2012 à Londres. Ce tweet devenu célèbre a inspiré le titre volontairement ambivalent de cet article (« pour tout le monde ») car désormais, le Web s'adresse et touche tout le monde et ce faisant le Web s'intègre à tout autour de nous et se déploie dans les moindre recoins de notre monde. Du Web, pour tous, partout et pour tout. Cette toile d'envergure mondiale, appellation qui au départ a pu être perçue comme immodeste, se révèle en quelques années être une prophétie auto-réalisatrice où le fait de concevoir pour l'universalité aura effectivement permis de tendre à l'universel.

Le Web a modifié notre rapport avec le temps et l'espace en nous donnant la possibilité d'interagir avec des objets ou des personnes éloignées, en nous donnant accès à des informations non disponibles localement, en enregistrant les traces de nos actions, en documentarisant une variété d'activités et en nous permettant ainsi de nous y replonger et d'y naviguer *a posteriori* et à distance. On ne note plus l'adresse du dentiste, on la retrouve sur le Web. On ne programme plus son magnétoscope ou sa box, on cherche un enregistrement en ligne. On ne passe plus à la gare acheter ses billets mais on télécharge son billet électronique. L'omniprésence et l'hypermnésie du Web sont acquises au point où l'on ne supporte plus quand il n'est pas là pour nous répondre et où la page « *no results found* » des moteurs de recherche n'est presque plus jamais vue.

Le Web est aussi un formidable outil d'intégration et d'interopérabilité, un espace d'échange entre applications. Les lignes que vous lisez actuellement ont été partiellement écrites à travers l'interface Web d'une application d'édition collaborative de documents et partiellement dictées à travers une application mobile connectée au travers du Web à cette même application. L'orthographe et la grammaire ont été grandement corrigées à la volée par des outils ayant statistiquement appris du Web. Se connecter à de telles ressources, les intégrer, les synchroniser et en assurer l'accessibilité quels que soient l'application et le terminal utilisés, tout ceci est assuré par la standardisation du Web.

À l'échelle de l'histoire de l'informatique, le Web n'a pas uniquement fait ses preuves d'une architecture qui passe le test du temps, il a défini un nouveau temps : il existe une époque avant et une époque après le Web. Une époque où le problème est d'avoir des informations ou l'accès à un service, puis une époque où il faut pouvoir se retrouver dans leur masse. Une époque de fragmentation puis une époque d'hyper-intégration voire de sur-intégration avec ses risques.

Le Web a instauré un certain nombres de nouveaux objets et concepts maintenant courants comme : le serveur Web, le navigateur Web, la page Web, la page perso, le site Web, l'adresse Web ou l'application Web. Certains comme le « site Web » restent encore informellement définis. Le Web a aussi hybridé beaucoup d'objets et d'activités existantes : web documentaire, web series, webisode, web tv, web radio, webinar, web sourcing, web publishing, web ID, web mail, web commerce, web publicité,

webcam, webcast, weblog, web conf, web journal, webOS, etc. Le plus effrayant à écrire ces termes c'est que mon correcteur orthographique les reconnaît déjà. Dès qu'une ressource Web s'instancie dans un type d'objet, dès lors que ces objets s'hybrident et entrent en contact avec les principes et pratiques du Web, ils deviennent de nouveaux objets appelant de nouvelles pratiques. Et le phénomène est fractal : si une technologie Web se développe en restant au plus proche de son architecture et de ses principes universels (p. ex. wiki), elle acquiert sa viralité et le potentiel de se décliner de multiples façons (p. ex. wikipedia, wiktionnary, wiki travel, wikileaks, etc.). Sorti de la chrysalide initiale de la métaphore d'une bibliothèque universelle, le Web comme outil de partage est maintenant tellement vulgarisé qu'il fournit même des métaphores pour d'autres domaines, comme par exemple les réseaux mycorhiziens appelés par certains biologistes des *Wood*-Wide Web [9].

Comme on peut le voir au travers des références et allusions de cet article, Tim continue à défendre le Web et à s'investir sans compter pour amener le Web à son plein potentiel [2]. Ce prix Turing récompense non seulement le fait qu'il ait inventé le Web mais aussi le fait qu'il ait travaillé toute sa vie à le défendre. Car la défense du Web reste un enjeu. Il est universellement utile et utilisé mais reste fragile et son idéal de départ pourrait n'être qu'une parenthèse historique si l'on ne veille pas en permanence à sa préservation.

Tim Berners-Lee se bat encore actuellement contre toute forme de recentralisation, par exemple la centralisation applicative induite par des monopoles de certaines firmes sur certains services du Web, et pour la neutralité des réseaux en général et du Web en particulier. Les enjeux (neutralité, décentralisation, démocratisation, etc.), les dangers (recentraliser, classes d'accès, Web à plusieurs vitesses, etc.), les limitations (besoins en infrastructure, énergies, coûts, etc.) font du Web un interminable projet plus qu'une réalisation acquise. La plus grande crainte de Tim reste qu'une entité politique ou commerciale prenne le contrôle du Web ce qui, pour lui, signerait l'arrêt de mort du Web. Cela motive son engagement en faveur, entre autres, de la neutralité du net, de la redécentralisation du Web, de l'éclatement des bulles de filtrage [42] et de la réappropriation des données par les usagers [13]. L'architecture du Web est et doit rester robuste, même en milieu hostile, neutre, même si les couches basses venaient à être compromises, résiliente, même si les infrastructures venaient à manquer ou être limitées, etc. Pour le protéger, nous devons concevoir les araignées de l'architecture du Web comme des animaux extrêmophiles.

À l'inverse, le Web a pu lui-même être perçu comme un danger. Dès 1996, Tim écrit que la force de diversification qu'est la géographie est affaiblie par le Web [5]. Il le fait à une époque où, alors que je donnais mes premiers cours de Web à mes co-promotionnaires du Génie Mathématique de l'INSA de Rouen, on entendait en France certains s'inquiéter de voir le Web devenir un outil d'hégémonie de la langue

2. Le slogan du W3C : « *leading the Web to its full potential.* »

anglaise et de la culture anglophone. Cette année-là, Tim attire déjà l'attention de façon générale sur l'éthique et les enjeux sociétaux du Web : il rappelle l'impact des choix architecturaux du Web sur les formes de sociétés dans lesquelles nous vivons ; la nécessité de revisiter la notion de copyright dans un espace où la copie peut prendre bien des formes (p. ex. la mise en cache d'une copie d'un contenu et le statut juridique de cette copie) ; les problèmes que posent au respect de la vie privée les multiples opportunités de capter des données ; l'impact en termes d'information du Web sur un public de votants ; la nécessité de travailler main dans la main avec les systèmes de législation ; etc. [5].

À mon avis, ces sujets préfigurent en 1996 des besoins souvent encore plus impérieux deux décennies après, et la nécessité de chercher activement l'interdisciplinarité dans l'étude du Web et de ses évolutions. Nous reconnaissons tous que le World Wide Web a eu un impact social énorme [2]. Le Web est un objet de recherche, de développement, d'activité économique et commerciale (géants du numérique), un vecteur d'action et de structuration sociale, un sujet et un outil de politique, un nouvel objet et espace juridique, et même de questionnement philosophique [8]. Il est donc important pour comprendre le Web de façon holistique, dans toute sa complexité, d'encourager le mouvement « Web Science » vers la transdisciplinarité. Pour moi, les trois W du World Wide Wide appellent les trois M d'une Méthode Massivement Multidisciplinaire [41].

De plus, le Web doit aussi résolument devenir un sujet d'éducation et de formation en lui-même. Son utilisation (notions de base de navigation, de recherche, etc.), les bonnes pratiques (lecture critique, validation croisée, contribution active, etc.), la prévention (protection de la vie privée, protection des enfants, etc.), sont autant de sujets auxquels toute génération devrait être formée à l'école comme un élément important dans l'égalité des chances que l'on doit. En 2008, Tim est ainsi un des fondateurs de la World Wide Web Foundation, organisation à but non lucratif promouvant l'accès au Web pour tous [13] et un Web ouvert comme un bien public et un droit élémentaire [2].

L'universalité de l'approche et le pouvoir de lier tout ce qui est concevable et partout dans le monde étaient difficiles à comprendre au début, et peu de personnes voyait la différence avec les systèmes hypertextes existants. Si à l'origine, le problème était d'imaginer un monde avec le Web avant que celui-ci n'advienne, nous sommes maintenant dans le cas inverse où les gens oublient ou n'imaginent plus ce que serait un monde sans le Web [13]. Nous sommes à un point de bascule où l'on parle de réalité augmentée par le Web [43] et où la perception s'inverse tant et si bien que c'est le monde qui paraît s'inscrire dans le Web, plus ce dernier s'y déploie. Cette idée est pour moi résumée dans l'inversion des termes *Web-Wide World* où le Web dépasse le monde physique en 3D et ce dernier y rentre pour s'étendre et s'augmenter dans un nombre ouvert de dimensions. Et dans cette perspective, les échos de

la phrase de Tim sur le fait que nos choix dans l'architecture du Web ont un impact sur nos sociétés ne cesse de s'amplifier avec les évolutions du Web.

Le prix Turing récompense donc Sir Tim Berners-Lee non seulement pour l'invention du Web en 1989 mais aussi pour le fait que, depuis, Tim n'a jamais cessé de remettre la toile du Web sur de nombreux métiers.

« *The Web as I envisaged it, we have not seen it yet.*
The future is still so much bigger than the past. »

– Sir Tim Berners-Lee, WWW Conference 2009

De gauche à droite : Fabien Gandon, Sir Tim Berners-Lee, Louis Pouzin, Jean-François Abramatic, pour les 25 ans du Web à Futur en Seine, Paris, 2014.

Remerciements. Je tiens à remercier mes collègues et amis du W3C, de la SIF, d'Inria, d'UCA, du CNRS et le l'I3S pour leur relecture de cet article.

Références

[1] W3C Automotive Working Group Charter, accédé le 10 mai 2017, https://www.w3.org/2014/automotive/charter.html

[2] ACM Turing Award Sir Tim Berners-Lee, http://amturing.acm.org/award_winners/berners-lee_8087960.cfm

[3] Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor, Tim Berners-Lee, HarperCollins, 1999, ISBN 0062515861, 9780062515865

[4] How the Web was Born : The Story of the World Wide Web, James Gillies, Robert Cailliau, January 15, 2000, Oxford University Press, ISBN-10 : 0192862073 ISBN-13 : 978-0192862075

[5] The World Wide Web : Past, Present and Future, Tim Berners-Lee, August 1996, https://www.w3.org/People/Berners-Lee/1996/ppf.html

[6] "Information Management : A Proposal" (March 1989), the original proposal for the software project at CERN that became the World Wide Web. https://www.w3.org/History/1989/proposal.html

[7] Sergey Brin, Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks 30(1–7): 107–117 (1998)

[8] Philosophical Engineering and Ownerhip of URIs, Tim Berners-Lee, an interview in "Philosophical Engineering", collected set of writings from the PhiloWeb workshops, https://www.w3.org/DesignIssues/PhilosophicalEngineering.html

[9] La Vie secrète des arbres, Peter Wohlleben, Editions Les Arènes, 2017, 2352045932

[10] Cédric Hébert, Modèle de traitement de RDF basé sur les graphes conceptuels, Rapport de stage de DEA, I3S, université de Nice Sophia-Antipolis, 1999.

[11] Olivier Corby, Rose Dieng, Cédric Hébert. A Conceptual Graph Model for W3C Resource Description Framework. International Conference on Conceptual Structures, Aug 2000, Darmstadt, Germany. pp. 468–482, 2000

[12] Fabien Gandon, Distributed Artificial Intelligence And Knowledge Management : Ontologies And Multi-Agent Systems For A Corporate Semantic Web, Thèse, Université Nice Sophia Antipolis, 2002.

[13] Neil Savage, Weaving the Web, Communications of the ACM, Vol. 60 No. 6, Pages 20–22, 10.1145/3077334, https://cacm.acm.org/magazines/2017/6/217732-weaving-the-web/fulltext

[14] MIT and University of Southampton launch World Wide Web research collaboration Initiative will analyze and shape web's evolution, MIT News, November 2, 2006, http://news.mit.edu/2006/wsri

[15] Web Science Trust (WST), accédé le 7 juin 2017, http://www.webscience.org/

[16] "World Wide Web—Archive of world's first website". World Wide Web Consortium. https://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html

[17] Welcome to Amaya, https://www.w3.org/Amaya/

[18] https://timeline.web.cern.ch/timelines/The-birth-of-the-World-Wide-Web

[19] Tim Berners-Lee, Plenary Talk extracted slides, First WWW Conference, Geneva 94, https://www.w3.org/Talks/WWW94Tim/

[20] Berners-Lee, Tim and Cailliau, Robert and Luotonen, Ari and Nielsen, Henrik Frystyk and Secret, Arthur, The World-Wide Web, Commun. ACM, Aug. 1994, Vol. 37, nº 8, 0001-0782, pp. 76–82, 10.1145/179606.179671, ACM, New York, NY, USA

[21] W3C mission, accédé le 16 juin 2017, https://www.w3.org/Consortium/mission

[22] The W3C Technical Architecture Group (TAG), accédé le 16 juin 2017, https://www.w3.org/2001/tag/

[23] W3C organisation, accédé le 16 juin 2017, http://www.w3.org/Consortium/facts

[24] W3C Process, accédé le 16 juin 2017, http://www.w3.org/Consortium/Process/

[25] W3C Ethics, accédé le 16 juin 2017, http://www.w3.org/Consortium/cepc/

[26] Luis Von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford. 2003. CAPTCHA : using hard AI problems for security. In Proceedings of the 22nd international conference on Theory and applications of cryptographic techniques (EUROCRYPT'03), Eli Biham (Ed.). Springer-Verlag, Berlin, Heidelberg, pp. 294–311.

[27] Von Ahn, Luis, et al. "recaptcha : Human-based character recognition via web security measures", Science 321.5895 :1465–1468 (2008).

[28] Bush, Vannevar. "As we may think". The atlantic monthly 176.1:101–108 (1945).

[29] T. H. Nelson. 1965. Complex information processing : a file structure for the complex, the changing and the indeterminate. In Proceedings of the 1965 20th national conference (ACM '65), Lewis Winner (Ed.). ACM, New York, NY, USA, pp. 84–100. http://dx.doi.org/10.1145/800197.806036

[30] C. Engelbart, and William K. English, AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference, San Francisco, CA, December 1968, Vol. 33, pp. 395–410 (AUGMENT,3954,).

[31] Pouzin, L., Presentation and major design aspects of the Cyclades Computer Network. IN : Proc. 3rd Data Communications Symposium, Tampa, Fla., Nov. 1973, pp. 80–85.

[32] Cerf, V. G. and Kahn, R. E., A protocol for packet network intercommunication. IEEE Trans. Commun., Vol. COM-22, No. 5, pp. 637–648 (May 1974).

[33] Roy T. Fielding and Richard N. Taylor. 2000. Principled design of the modern Web architecture. In Proceedings of the 22nd international conference on Software engineering (ICSE '00). ACM, New York, NY, USA, pp. 407–416, http://dx.doi.org/10.1145/337180.337228

[34] R. Fielding, J. Reschke, Hypertext Transfer Protocol (HTTP/1.1) : Semantics and Content, RFC7231, Internet Engineering Task Force (IETF) , 2014

[35] Tim Berners-Lee's proposal, CERN, March 1989, http://info.cern.ch/Proposal.html

[36] Fabien Gandon, Michel Buffa, Elena Cabrio, Olivier Corby, Catherine Faron-Zucker, et al.. Challenges in Bridging Social Semantics and Formal Semantics on the Web. Hammoudi, S. and Cordeiro, J. and Maciaszek, L.A. and Filipe, J. 5h International Conference, ICEIS 2013, Jul 2013, Angers, France. Springer, 190, pp. 3–15, 2014, Lecture Notes in Business Information Processing. https://hal.inria.fr/hal-01059273

[37] Berners-Lee, Tim, "Semantic web road map", https://www.w3.org/DesignIssues/Semantic.html (1998).

[38] W3C Begins Standards Work on Web of Things to Reduce IoT Fragmentation, https://www.w3.org/WoT/

[39] The Mobile Web, Worl Wide Web Foundation, http://webfoundation.org/about/vision/the-mobile-web/

[40] Tiropanis, Thanassis, et al. "The web science observatory". IEEE Intelligent Systems 28.2 (2013), pp. 100–104.

[41] Fabien Gandon. The three 'W' of the World Wide Web callfor the three 'M' of a Massively Multidisciplinary Methodology. Valérie Monfort ; Karl-Heinz Krempels. 10th International Conference, WEBIST 2014, Apr 2014, Barcelona, Spain. Springer International Publishing, 226, Web Information Systems and Technologies. http://www.springer.com/fr/book/9783319270296, https://hal.inria.fr/hal-01223236

[42] Pariser, Eli. The filter bubble : What the Internet is hiding from you. Penguin UK, 2011.

[43] Fabien Gandon, Alain Giboin. Paving the WAI : Defining Web-Augmented Interactions. Web Science 2017 (WebSci17), Jun 2017, Troy, NY, United States. pp. 381–382, 2017, https://hal.inria.fr/hal-01560180

[44] Michel Buffa, Du Web aux wikis : une histoire des outils collaboratifs, Interstices, 23 mai 2008, https://interstices.info/jcms/c_37151/du-web-aux-wikis-une-histoire-des-outils-collaboratifs

[45] A brief history of CSS until 2016, Bert Bos, Style Activity Lead, W3C, 17 December 2016, https://www.w3.org/Style/CSS20/history.html

[46] The Mobile Web, Tim Berners-Lee, keynote, GSM congress, Barcelone, 2007