



1989 : L'idée du Web

1992 : info.in2p3.fr (4°, 5° ? serveur Web)

2000 : hal.archives.ouvertes.fr

2010 : Vers le Web des données

daniel.charnay@in2p3.fr

Septembre 92, première page en France!



File Edit Options Navigate Annotate Help

Document Title: Centre de Calcul IN2P3

Document URL: http://info.in2p3.fr

Centre de Calcul de l'IN2P3

27 boulevard du 11 novembre 1918
69100 VILLEURBANNE
Tél. (33) 78 93 08 80
Telex 306 906



[CERN](#) - [SLAC](#) - [DESY](#)

De ce site et des versions qui ont suivi il ne reste aucune trace ... Des premiers sites Web que reste-t-il ?
web.archives.org se souvient d'info.cern.ch ou d' info.in2p3.fr à partir de 1997
« Le Web sera le cauchemar des historiens parce qu'il évolue sans cesse sans laisser de trace » Robert Caillau

Daniel Charnay & Wojciech Wojcik, webmasters



NeXT station

Les origines



- ❖ **Vannevar Bush en 1945** décrit un bureau électromécanique fondé sur des index croisés (hyper liens) accédant à des microfilms : Memex. *L'idée est déjà d'aider les chercheurs à organiser et retrouver de l'information !*
- ❖ **Theodor Holm Nelson** sociologue américain, pionnier de l'histoire des technologies de l'information, inventeur de l'hypertexte et du terme. Porteur du projet Xanadu (1965) qui ne verra jamais le jour mais qui contient les principes du Web (réseau d'ordinateurs sur lignes téléphoniques). Trop précoce (1970) trop complexe...
- ❖ **APPLE (1987)** HyperCard et son langage hyperTalk. Première interface graphique pour circuler entre documents (les cartes) mais limitée à une navigation « dans l'ordinateur »

Le fondement : l'hypertexte



- ❖ **Les liens hypertextuels** : liens sémantiques qui permettent de relier une idée à une autre ; utilisation de la souris pour accéder à un lien (Douglas Engelbart, Xerox, la souris, le multifenêtrage, l'interface homme-machine, le tout mis en musique par Apple). La représentation du lien : couleur différente, changement de l'icône véhiculée par la souris (la flèche devient un doigt)
- ❖ **La non linéarité de la lecture** (on clique sur un lien avant d'avoir lu la totalité de la page affichée : plus proche du fonctionnement de la pensée ?) *Une approche de la capacité de l'individu à savoir rechercher. Le danger « perdu dans le cyber espace »*
- ❖ **Le lecteur/auteur** : l'enrichissement d'un texte par l'auteur mais aussi et surtout, par une suite de lecteurs qui déposent des liens ... (Enrichissement d'un article dans Wikipédia). Un hypertexte devient une œuvre collective

La naissance



- ❖ **Tim Berners Lee et Robert Caillau**, informaticiens au CERN, créent le Web en 1992 mais la proposition aux dirigeants du CERN date de 1989 ; elle avait alors suscité un enthousiasme « relatif » ... « formule polie... proposition passionnante :o »
- ❖ **Un support, le réseau Internet** (définition de TCPI/IP et du mot Internet en 1982). Pour le grand public, la notion de Web va se confondre avec la notion d'Internet : expression *rechercher sur Internet plutôt que rechercher sur le Web*. Le Web n'est qu'une application de l'Internet qui en véhicule bien d'autres : par exemple le mail, la visioconférence, etc.
- ❖ **La technologie du Web :**
 - ◆ Un protocole HTTP dialogue de communication entre navigateur et serveur,
 - ◆ Un langage HTML décrit (et structure ?) l'information interprétée par le navigateur (texte, images, autres médias via des plugins)
 - ◆ Un principe d'identification et d'adressage, les URL et URI
 - ◆ Mais aussi et surtout une passerelle avec les applications informatiques, le CGI. Un serveur Web peut distribuer les résultats d'applications près-Web ! (base de données par exemple)
- ❖ **Le premier navigateur graphique** (aussi éditeur) est disponible sur NeXT (CERN, 92), il est suivi par Mosaic (NCSA, 93) qui est multi plate-forme (Unix, Mac, puis Windows), sa descendance sera Netscape puis Firefox

Les raisons du succès



- ❖ **Simple, gratuit, ouvert**
- ❖ **Le CERN** met explicitement la technologie Web dans le domaine public (habitude de la communauté : la CERNlib)
- ❖ **HTTP** est simple (essentiellement 2 verbes GET et POST)
- ❖ **HTML** est simple et permissif ! Pas besoin d'être informaticien, même mal codée la page s'affiche !
- ❖ **L'appropriation est facile** [*view page source*] « ouvre » le code de la page et facilite l'apprentissage

Un espace d'exécution des applications informatiques



- ❖ ***Dès l'origine les formulaires et la passerelle CGI font du Web le lieu où se déroulent le dialogue entre l'utilisateur et un programme informatique***
 - ◆ Exemple, on paramètre sa box ADSL, son imprimante depuis un site Web intégré, on réserve ses billets de trains sur l'application de la SNCF, on fait sa bibliographie sur HAL...).
 - ◆ Le navigateur est devenu l'IHM de toutes les applications informatiques
- ❖ **Les langages du Web facilitent la création des applications**
 - ◆ PHP (Rasmus Ledorf 94)
 - ◆ JavaScript (Netscape 95) né de la nécessité de donner de l'intelligence au navigateur. Il ouvrira plus tard la porte à Ajax, faculté à ne recharger qu'une portion de la page
 - ◆ XML, CSS, XSLT (W3C, 96) séparation des données et de leur présentation (entre autres)

Et pendant ces années ?



❖ ... et jusqu'en 1995, Microsoft ne croit ni au Web ni à l'Internet, il veut lancer « The Microsoft network »

❖ Si.



❖ *Mais il comprendra très vite que le Web devient incontournable, Windows 95 sortira avec Internet Explorer, la guerre des navigateurs peut enfin commencer !*

“The network is the computer”

(John Gage – Sun Microsystems)



- ❖ **Aujourd’hui a-t-on besoin d’autre chose que d’un navigateur ?**
 - ◆ Google propose les principales applications dont on a besoin : traitement de texte, tableur, agenda, mail, traducteur, etc.
 - ◆ On peut stocker de l’information chez Google, ses photos chez Flickr, sécuriser des documents importants chez son banquier
 - ◆ L’outil universel d’accès à internet est le navigateur ! Les « portes d’entrée du Web » deviennent : Google, Facebook, Youtube, eBay, Flickr, Wikipedia, ... Le navigateur fédère l’accès à tous les médias, la plate-forme Web distribue tout les médias (sons, vidéo, webcast, mail, textes de divers formats (PDF)...)
 - ◆ Le navigateur est sur toutes les plate-formes, dont le téléphone portable (Il y a plus d’individu équipé d’un mobile sur la planète que d’individus ayant une connexion Internet)

- ❖ **Le navigateur est-il sur le système d’exploitation ou le système d’exploitation sera-t-il dans le navigateur ? (Google Chrome OS)**

Le retour aux sources



- ❖ Créé pour faciliter l'échange d'informations entre chercheurs, le Web retrouve son âme avec **les archives ouvertes**
- ❖ **Diffuser gratuitement la littérature scientifique**
 - ◆ ArXiv (Paul Ginsparg, physicien, ouvre la première archive ouverte en 1991)
 - ◆ PubmedCentral (Harold Warmus prix Nobel de médecine)
 - ◆ HAL (CCSD, 2001)
 - ◆ ...

Les archives ouvertes



- ❖ **Le Web facilite le libre accès aux résultats de la recherche**
- ❖ **La réappropriation de la communication scientifique devient possible**
 - ◆ Revendication « humaniste » : le libre accès à la connaissance, gratuité, accessibilité
 - ◆ Revendication pragmatique : les budgets documentaires des organismes n'arrivent plus à suivre les tarifs des éditeurs
- ❖ **Communication scientifique directe, le réseau des chercheurs**
 - ◆ Un circuit court et rapide entre chercheurs pour faire connaître leurs résultats
- ❖ **Auto archivage,**
 - ◆ Les chercheurs déposent eux-mêmes leurs
- ❖ **Interopérabilité**
 - ◆ Des protocoles issus du Web (OAI-PMH) vont favoriser l'accès aux bases d'articles scientifiques
- ❖ **Archivage à long terme**

Les archives ouvertes renouent avec une vieille tradition, celle de l'échange direct de correspondances scientifiques entre les savants.

Le WEB apporte l'instantanéité



Les AO, le Web de la science



❖ Visibilité, accessibilité, pérennité

- ◆ L'impact d'une publication est accru de manière substantielle si celle-ci est accessible librement sur le Web «*On line or invisible ?*» S. Lawrence, NEC
- ◆ Dans les domaines des mathématiques, de l'astrophysique, de la physique, avec un échantillonnage de 14 millions d'articles du WOS sur 10 ans, le taux de citation serait de 2.5 à 5 fois supérieur pour les articles en libre accès sur le Web (S. Harnad, 2004)
- ◆ Des études sur les liens figurant en notes de bas de page et sur les références des publications révèlent une inquiétante «dégénérescence»
 - 4% des liens cités dans les articles de 3 journaux scientifiques (Nature, Science et The New England Journal of Medicine) sont inaccessibles 3 mois après leur publication ; ce taux monte à 10% après 15 mois, 13% après 27 mois ...
 - 40% des liens référencés par les articles de deux journaux informatiques sont inaccessibles 4 ans après la publication
- ◆ Les archives ouvertes vont assurer la pérennité, l'archivage à long terme



❖ En France HAL devient l'AO nationale

- ◆ Créée par le CNRS, soutenue par le ministère, l'académie des sciences, elle est utilisé par la quasi-totalité des établissements de recherche ainsi que par de nombreuses universités et grandes écoles
- ◆ 1800 articles et thèses de toutes les disciplines sont déposés chaque mois et plus de 130 000 documents sont accessibles en texte intégral
- ◆ Interconnecté à ArXiv et PubmedCentral, HAL est moissonné par les principaux moteurs de recherche, et, grâce à OAI-PMH aux moteurs spécifiques comme Oaister, Google Scholar, Repec,
- ◆ HAL s'interconnecte avec de nombreuses applications , STAR pour les thèses électroniques, GRAAL pour la gestion de la recherche des universités, l'ANR, etc.

Web sémantique & Web de données



❖ Après un Web de documents, un Web de données

- ◆ http://www.dailymotion.com/video/x8omty_tim-bernerslee-the-next-web-of-open_tech
- ◆ « Donnez-nous vos données brutes » (Tim Berners-Lee)
 - Exemple : diffuser une publications scientifique mais aussi distribuer les données brutes qui ont conduit à cette publication, lier les données entre elles, permettre une « autre » analyse, des travaux sur des sur-ensembles

❖ Un Web sémantique

- ◆ Un Web compréhensible et utilisable par des machines, des logiciels
- ◆ Décrire les données en RDF (accessibles via Sparql)

❖ Pour l'utilisateur

- ◆ Une recherche/navigation par facettes, des recherches floues, en langage naturel
 - *Voir le projet ISIDORE du TGE Adonis sur l'accès unifié aux données*



C'est tout pour aujourd'hui!