



Pour réagir aux articles : courrier@pouirlascience.fr
ou directement sur les pages correspondantes du site www.pouirlascience.fr

✓ DE L'ANALOGIE EN BIOLOGIE

Dans l'article *Les mots de la biologie sont-ils bien choisis ?* (*Pour la Science* n° 399, janvier 2011, http://bit.ly/pls399_pointdevue), Pierre-Henri Gouyon et Marc-André Selosse plaident pour une « vision unifiée du vivant » qui justifierait en biologie l'emploi de termes issus du vocabulaire courant, tels que « gène égoïste » ou « stratégie tricheuse », sans quoi nous encourrions le risque d'une surenchère de mots nouveaux et d'une dissociation maintenue entre les humains et les processus naturels.

Mais la vraie question est de savoir si, entre l'image métaphorique et ce qu'elle illustre, les mécanismes sous-jacents sont les mêmes. Les métaphores biologiques procèdent par analogie. Une analogie est une ressemblance de rapports. Supposons qu'un mécanisme A explique les relations entre des entités du monde humain et que, par ailleurs, les rapports entre des entités du monde naturel ressemblent à ceux qu'entretiennent les entités du monde humain. Le mécanisme A est alors transposé métaphoriquement pour suggérer une explication des relations naturelles, et cette transposition sert de justification pour supposer l'existence d'un mécanisme B similaire à A.

Le problème, c'est que l'analogie ne garantit pas que la transposition soit adéquate, et ne peut alors pas avoir valeur de démonstration. Elle ne peut qu'être suggestive. S'il s'avérait que les mécanismes sous-jacents étaient différents, l'usage d'un mot commun serait pire que l'usage de mots différents, car au nom d'un rapprochement entre l'homme et la nature, nous laisserions le mot commun nous tromper, au moins potentiellement, sur ce qui rapproche vraiment de l'homme. Or nous sommes en droit de douter que ce que nous appelons « stratégie » chez un humain corresponde, dans le processus même désigné par ce nom, à une « stratégie » adaptative d'une population ou d'une espèce.

Enfin, il ne faut pas craindre les mots nouveaux, si ceux-ci sont justifiés. C'est le rôle des sciences que de donner accès rationnellement aux objets et aux mécanismes qui échappent au sens commun, et donc pour lesquels nous n'avons pas de mots disponibles dans la langue vernaculaire. C'est en gagnant de la précision avec des mots nouveaux que l'homme se rattachera de la façon la plus exacte au reste du vivant.
Guillaume Lecoindre, MNHN, Paris

✓ LE NOM DES LANGUES

L'article *L'arrivée des langues indo-européennes en Europe*

(*Pour la Science* n° 400, février 2011, http://bit.ly/pls400_langues) a suscité plusieurs réactions. En voici deux :

– Dans l'arborescence italique de l'arbre de parenté des langues indo-européennes, on trouve une langue inexistante : le valachien. Le terme correct est plutôt « valaque ». Mais même ainsi, cette dénomination est impropre, car il s'agit d'un adjectif désignant des régions géographiques et des populations, et non des langues. Le nom exact de la branche est le roman oriental, qui se ramifie en aroumain, daco-roumain, istro-roumain et mégléno-roumain.

Ion Cepleanu

– Le néerlandais est indiqué sous la dénomination de sa variante dialectale belge, le flamand. N'est-ce pas inexact ?

Elisabeth Hooghe-Peters

→ RÉPONSE DE LA RÉDACTION

Le terme « flamand » est bien celui de l'auteur, Ruth Berger (*Flämisch* en allemand). Par ailleurs, bien que le nombre de néerlandophones soit plus élevé aux Pays-Bas qu'en Belgique, cela ne rend pas l'usage du terme flamand faux pour autant sur le plan linguistique, si l'on admet que le terme flamand désigne toutes les formes dialectales du néerlandais au Sud du Rhin, c'est-à-dire une très grande partie des néerlandophones du monde.

✓ EN ORBITE AUTOUR DE RIEN ?

Selon l'article *Le télescope spatial*

James Webb, un observatoire origami

(*Pour la Science* n° 398, décembre 2010, http://bit.ly/pls398_jwst), cet instrument

sera mis en orbite autour du point de Lagrange L₂. Comment peut-on se mettre en orbite autour d'un point qui n'attire rien ? Ne suffit-il pas de rester en ce point ?

Hugo Roca

→ RÉPONSE DE LA RÉDACTION

Si l'on se place dans le référentiel héliocentrique tournant avec la Terre, les points de Lagrange sont des points d'équilibre entre l'attraction du Soleil, de la Terre, et la force centrifuge. Un objet

placé en l'un de ces points (il en existe cinq) décrit une trajectoire autour du Soleil ayant la même période que la Terre.

Le point de Lagrange L₂, situé à 1,5 million de kilomètres de la Terre dans la direction opposée au Soleil, et son symétrique L₁, à la même distance dans la direction du Soleil, sont des points instables : un corps situé à proximité – et un objet étendu ne peut se trouver en un seul point – finit toujours par s'en éloigner, avec un temps caractéristique de l'ordre du mois. Une sonde placée en L₂ ne peut donc pas tourner sur une orbite stable comme le ferait un satellite en orbite terrestre. Mais il suffit de petites corrections de trajectoires répétées pour se maintenir au voisinage de L₂. Les orbites ainsi décrites (dites de Lissajous) sont des trajectoires complexes, très différentes des ellipses classiques.

✓ INSECTES PHOTONIQUES

Dans l'article *Des insectes à la photonique*, (*Pour la Science* n° 401, mars 2011, http://bit.ly/pls401_photonique), l'auteur attribue les changements de couleur du *Morpho godarti* au changement de l'indice de réfraction de l'air produit par les vapeurs d'acétone ou de trichloréthylène. L'indice de réfraction indiqué pour la vapeur d'acétone est en fait celui de l'acétone liquide. Celui de la vapeur d'acétone est 1,001090, bien peu différent de celui de l'air (1,0003). Comment ce faible changement d'indice peut-il entraîner un changement de couleur ?

Louis Peralta

→ RÉPONSE DE SERGE BERTHIER

Louis Peralta a parfaitement raison. La faible différence d'indice de la vapeur d'acétone et de l'air ne peut expliquer le changement drastique de couleur. Cela montre en fait que le matériau volatil se trouve en phase liquide dans la structure. Tout vient de ce que l'on appelle « vapeur ». Dans le cas présent, et dans les expériences menées en particulier en Hongrie, l'acétone se trouve sous forme de micro/nanogouttelettes qui pénètrent dans la structure et adhèrent à sa surface, formant un film liquide continu, donc d'indice plus élevé. La surface développée de la structure est énorme et permet une condensation efficace des vapeurs. C'est tout l'intérêt de cette technique.

PLS402

QUESTIONS OUVERTES

Faut-il archiver le Web ?

Ce qui est affiché sur la Toile change très vite. Un archivage périodique est donc nécessaire pour en garder la mémoire, avec des difficultés qui ne sont pas seulement techniques.

Gildas ILLIEN

Brewster Kahle, ingénieur américain diplômé du MIT, a inventé le premier système de publication sur l'Internet, WAIS (*Wide Area Information Service*). En 1995, il rend visite à des amis travaillant à Palo Alto sur le grand moteur de recherche de l'époque, *Alta Vista*. Il y découvre que la quasi-totalité du Web d'alors, 16 millions de pages indexées, tient dans l'équivalent de deux distributeurs automatiques de café : le voilà qui enlace, littéralement, tout l'Internet !

Marqué par cette expérience et fasciné par le mythe de la Bibliothèque universelle d'Alexandrie, B. Kahle fonde l'année suivante *Internet Archive*, un organisme à but non lucratif qui se fixe pour objectif d'archiver tout le Web afin de conserver la mémoire numérique du monde et de la rendre accessible à tous. Son site, www.archive.org, basé à San Francisco, permet depuis 1997 de retrouver ce que les bibliothécaires appellent les incunables de l'Internet (par exemple les premières pages du site Web du quotidien *Le Monde*).

Les pouvoirs publics étant toujours un peu plus lents face à l'innovation, il faudra attendre août 2006 pour que le Parlement français vote la loi DADVSI qui institue un dépôt légal des « signes, signaux, écrits, images, sons ou messages de toute nature » publiés dans le cadre d'une communication au public par voie électronique – c'est-à-dire un dépôt légal de l'Internet. Celui-ci est confié à la Bibliothèque nationale de France (BnF), chargée de la collecte des sites français, à l'exception des sites des émetteurs de la radiotélévision qui relèvent, eux, de l'Institut national de l'audiovisuel (INA).

Cependant, les bibliothèques nationales – la BnF, mais aussi ses homologues de Grande-Bretagne, de Suède, du Danemark ou d'Australie – n'ont pas attendu la mise en place d'une législation pour engager des expérimentations visant elles aussi à mettre le Web en boîte à des fins patrimoniales. L'aventure de l'archivage du Web a débuté il y a plus de dix ans. Les ingénieurs pionniers ont d'abord vu le défi technique d'une telle entreprise. Aujourd'hui, alors que notre mémoire du monde numérique commence à faire défaut, beaucoup s'aperçoivent que cet archivage soulève aussi des questions de société : peut-on vraiment tout conserver ? À qui et à quoi un Web archivé pourra-t-il servir ?

Oubli ou mémoire ?

Google n'a que 15 ans, *Wikipedia* 10 ans, *Facebook* à peine 5 ans. Et nous ne savons pas ce qu'il restera en ligne, dans un an ou dix, des blogs et des réseaux sociaux ayant contribué à la propagation des révoltes en Tunisie et ailleurs. Faute de recul, il est difficile de savoir ce qui mérite d'être conservé et ce qui n'a pas été fait pour durer. Cette problématique n'est pas nouvelle, mais est centrale dans tout projet patrimonial. Les contours et les usages du patrimoine « numérique » (c'est-à-dire issu du Web, et non de la numérisation des livres) sont, de fait, encore difficiles à cerner.

Internet est un média éphémère, où l'instantanéité est la règle, mais nous y accumulons une quantité croissante de données numériques : photographies, conversations, enregistrements sonores et filmés, publications sur des blogs, des réseaux sociaux...

Avec le Web 2.0, ces contenus qui constituent des pans entiers de notre vie numérique ne sont pas seulement (voire pas du tout) conservés sur nos disques durs, mais directement sur la Toile.

Le déluge de données est tel que la difficulté n'est plus tant de trouver les informations que de les trier et de supprimer celles n'ayant pas d'utilité. À première vue, ce grand ménage semble se faire tout seul. Les webmasters évaluent la durée de vie moyenne des sites à deux ou trois ans : le Web s'évaporerait ainsi naturellement. Et au fond, l'idée que les données d'hier disparaissent à mesure que de nouvelles strates s'ajoutent serait plutôt rassurante : à quoi bon une mémoire numérique si les marées du Web peuvent, d'elles-mêmes, nous éviter de trier ?

Cathy Marshall, de la Société *Microsoft Research*, a étudié les comportements des Américains confrontés à leur propre mémoire numérique. Beaucoup estiment qu'il n'est pas utile de se préoccuper de leur conservation au motif que « d'autres » (« quelqu'un », *Google*, le Gouvernement fédéral, la Bibliothèque du Congrès...) s'en chargeraient déjà. Elle a même perçu chez certains un « amour du risque », pour qui la disparition volontaire ou accidentelle de leurs données aurait l'effet d'une libération – ce serait une page de sa vie que l'on pourrait enfin tourner : affranchi de son existence numérique, on peut tout recommencer à zéro.

Le principe même d'une mémoire numérique n'est-il pas fondamentalement dangereux ? Les grands moteurs indexeurs comme *Google* assureraient une mémoire totale, capable de conserver vos meilleurs souve-

nirs comme les pires et de ruiner votre réputation. Selon une récente enquête du CRÉDOC (Centre de recherche pour l'étude et l'observation des conditions de vie), les Français seraient préoccupés par la confidentialité et la protection des données sur Internet, et l'opinion se montrerait très favorable au « droit à l'oubli numérique » cher au sénateur président de la CNIL (Commission nationale de l'informatique et des libertés), Alex Türk, et à l'ancienne secrétaire d'État à l'économie numérique, Nathalie Kosciusko-Morizet. En effet, 91 pour cent des personnes interrogées estiment que « les sites Internet devraient permettre à chacun d'effacer simplement les informations personnelles qui ont été communiquées ».

Le spectre de « Big Brother »

Dans cette actualité où la mémoire peut sembler à la fois superflue et dangereuse, on peut donc se demander pourquoi une institution publique, telle la BnF, s'évertue à archiver des échantillons du grand fatras du Web ; serait-elle capable de trouver une juste mesure entre le spectre de *Big Brother* et la perspective tout aussi angoissante d'un monde privé d'histoire ?

Les internautes manquent peut-être de distance historique sur leurs propres pratiques pour se préoccuper de questions longtemps réservées aux chercheurs et aux conservateurs du patrimoine. C. Marshall observe cependant que la « non-politique » de conservation pratiquée par la majorité d'entre nous permet, assez paradoxalement, de retrouver par hasard des données oubliées. On compare souvent cette pratique de la « redécouverte » des fichiers oubliés dans les tréfonds des répertoires à l'expérience magique « d'une boîte à chaussures dans laquelle on retrouverait, pêle-mêle, de vieilles photographies, des coupures de presse, des lettres d'amour ». Une image qui n'est pas sans rappeler, pour les plus anciens,

1. LES MAGASINS NUMÉRIQUES de la Bibliothèque nationale de France, qui contiennent les archives du Web français stockées sur disques et sur bandes magnétiques.



BnF

L'AUTEUR



Gildas ILLIEN est conservateur et chef du Service Dépôt légal numérique à la Bibliothèque nationale de France, à Paris.

l'expérience du furetage dans un grenier familial, voire... dans une bibliothèque.

Alors, garder la trace du Web aurait peut-être et malgré tout quelque chose de bon ? Une histoire tristement ordinaire et contemporaine l'illustre. Symbole de l'industrie dans l'Est de la France, l'entreprise *Vogica*, spécialiste depuis 1976 des cuisines et salles de bain faites sur mesure, a été placée en liquidation judiciaire le 8 novembre 2010. Une quarantaine de magasins ont fermé, laissant un millier de salariés sans emploi. Consulté un mois après, le site www.vogica.fr avait entièrement disparu ; à sa place, figurait un avis de 12 lignes signé du mandataire chargé de la liquidation judiciaire.

Où les anciens salariés de l'entreprise peuvent-ils retrouver la mémoire numérique de leur entreprise et, par exemple, montrer que six mois avant sa fermeture, la rubrique « offres d'emploi » du site proposait encore 28 postes ? Où les chercheurs qui, dans 10 ou 30 ans, écriront l'histoire du déclin du tissu industriel français retrouveront-ils les traces du développement, puis de la chute de cette société ? Où, enfin, les amateurs

de publicités retrouveront-ils les clips qui ont fait la popularité de l'enseigne ?

Toutes ces informations sont conservées à la BnF, dans ses « archives de l'Internet ». L'histoire du site de *Vogica* illustre la problématique fondamentale du patrimoine : ce qui semble anecdotique et sans valeur aujourd'hui peut, un jour ou l'autre, devenir une source précieuse pour l'histoire collective et intéresser nombre de nos concitoyens. Et le Web étant devenu un média de masse, où publier est désormais à la portée de chacun, le patrimoine national ne peut plus se limiter à la production des élites et des spécialistes. Pour la première fois, il est possible de l'étendre aux propos des citoyens les plus ordinaires. Le dépôt légal du Web répondrait-il au défi que représente l'invention d'une mémoire de masse ?

Le dépôt légal, né d'une ordonnance royale de François I^{er} en 1537, a permis de réunir la plus grande part des collections de la BnF. Il oblige tout éditeur, imprimeur, distributeur ou importateur à déposer des exemplaires de chaque document qu'il édite, imprime, produit, distribue ou importe sur le territoire français, à l'un des organismes dépositaires (BnF, INA, etc.). Le dépôt légal s'est étendu à tous les supports et modes de diffusion, jusqu'à la révolution numérique. Ainsi, depuis le vote de la loi de 2006, la BnF archive chaque jour, au moyen de robots moissonneurs, des centaines de sites Web. À chaque connexion, le robot d'archivage s'identifie auprès des serveurs de l'éditeur, déclarant son identité et les raisons légales de sa visite. La BnF archive les contenus récoltés dans de vastes baies de stockage – ses magasins numériques. Des copies des sites de journaux, de revues savantes, de créations artistiques, mais aussi des blogs, des jeux en ligne, des sites publicitaires ou commerciaux côtoient désormais les vénérables collections d'estampes, de cartes et de manuscrits. Et depuis 2008, ces archives sont accessibles au public.

À la différence du dépôt légal usuel, le dépôt légal des sites Web n'implique aucune démarche active de la part des éditeurs ou des webmasters. Les collectes se font automatiquement au moyen d'un robot du nom de *Heritrix*. C'est un logiciel libre, comme

Le circuit de traitement d'une archive Web

Une archive Web se constitue en trois étapes – collecte, conservation et signalement. Les défis à chaque étape sont la volumétrie, l'hétérogénéité et la volatilité des données : l'obtention d'une image fidèle du Web exige un parc de serveurs robustes et capables d'archiver un grand nombre de fichiers très variés dans un temps limité.

Les fichiers sont récoltés au moyen de robots logiciels nommés *crawlers* ou *spiders*, le plus utilisé par les bibliothèques nationales étant *Heritrix*. À partir d'une liste d'adresses URL, ces robots copient tous les contenus qu'ils rencontrent, comme le ferait un internaute surfant sur tous les liens d'un site. Les bases de données, les contenus payants ou accessibles par iden-

tifiant, etc., sont un obstacle, et la moisson se limite souvent à la surface du Web. Le paramétrage des robots et leur surveillance par des professionnels permettent de choisir la fréquence et la profondeur de capture des sites.

Les archives sont stockées dans un format (WARC) qui regroupe les fichiers et les métadonnées associées à la collecte (par exemple pour dater l'archive). La sauvegarde se fait sur disques ou sur bandes, dans des entrepôts numériques bien maintenus (copie sur plusieurs sites par exemple). Pour suivre l'évolution des formats et des logiciels, des campagnes de migration (d'un format de fichier vers un format plus récent) ou d'émulation (par la reconstitu-

tion artificielle de l'environnement initial de publication des sites) sont prévues.

Des techniques d'indexation permettent la recherche et la visualisation dans les archives. L'indexation en plein texte permet de rechercher par mots, comme avec un moteur de recherche en ligne. L'application *Wayback Machine* « remonte le temps » en cherchant, à partir de l'adresse d'un site et d'une frise chronologique cliquable, toutes les anciennes versions du site. D'autres techniques offrent aux chercheurs des traitements plus pointus (fouille de données, de liens, analyses sémantiques de corpus...) et des représentations cartographiques et dynamiques des données.

tous ceux que développe le Consortium international pour la préservation de l'Internet (IIPC), organisation internationale d'une quarantaine d'institutions culturelles qui ont fait alliance avec *Internet Archive* dès 2003 afin de promouvoir et de faciliter l'archivage du Web à l'échelle planétaire.

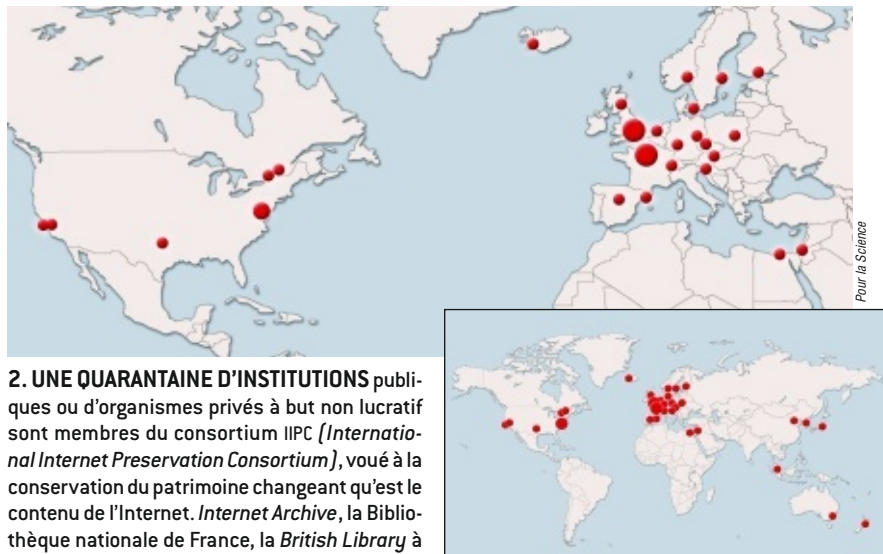
Si tout ou partie du site visité reste inaccessible au robot *Heritrix* pour des raisons techniques (base de données, contenu protégé par mot de passe, formulaire d'accès...) ou commerciales (contenu payant, abonnement...), la loi autorise la BnF à solliciter l'éditeur pour trouver des solutions techniques lui permettant d'améliorer la collecte.

Il est impossible d'être exhaustif

Compte tenu des volumes de données en ligne, la BnF ne peut garantir l'exhaustivité de son archivage. Aussi n'a-t-elle plus l'ambition ni la mission de tout conserver. Elle procède donc par échantillonnage, fondé sur la combinaison de plusieurs types de collectes (collectes automatiques et collectes ciblées, assistées par des bibliothécaires, plus fréquentes et plus profondes). La politique de la BnF est de constituer une collection représentative de l'Internet français tout en archivant certains contenus qui sont dans la continuité de son patrimoine imprimé.

Le service du « dépôt légal numérique » créé à la BnF en 2008 pour organiser et développer cette activité emploie neuf bibliothécaires et informaticiens à temps plein... ainsi qu'une quarantaine de serveurs chargés de la collecte et de l'indexation des données, une sauvegarde étant par ailleurs effectuée à la fois sur disque (pour la diffusion) et sur bande (pour la conservation). Cela représentait fin 2010 près de 200 téraoctets de données (1 téraoctet = 1 000 gigaoctets).

Ce volume peut aujourd'hui paraître modeste, mais c'est l'une des plus grandes collections d'archives Web du monde, après *Internet Archive* (près de deux pétaoctets, soit 2 000 téraoctets). Le régime juridique du dépôt légal numérique français, similaire à celui des pays nordiques, est en effet très favorable : il lui permet de collecter des sites sans avoir à demander l'accord de chaque édi-



2. UNE QUARANTAINE D'INSTITUTIONS publiques ou d'organismes privés à but non lucratif sont membres du consortium IIPC (*International Internet Preservation Consortium*), voué à la conservation du patrimoine changeant qu'est le contenu de l'Internet. *Internet Archive*, la Bibliothèque nationale de France, la *British Library* à Londres, la Bibliothèque du Congrès à Washington figurent parmi ses principaux membres. Il existe aussi d'autres acteurs de l'archivage du Web qui ne font pas partie d'IIPC, par exemple des grandes entreprises qui stockent les contenus successifs de leurs propres sites Web, notamment pour leurs besoins juridiques.

teur ou webmestre. Ce n'est pas le cas dans la plupart des pays de droit anglo-saxon. La *British Library* ou la Bibliothèque du Congrès, par exemple, ne disposent que d'archives sélectionnées et négociées à l'unité. Leurs collections couvrent certaines thématiques ou des événements (les attentats du 11 septembre 2001, l'ouragan Katrina...), mais rien qui ressemble au dépôt légal au sens où on l'entend en France.

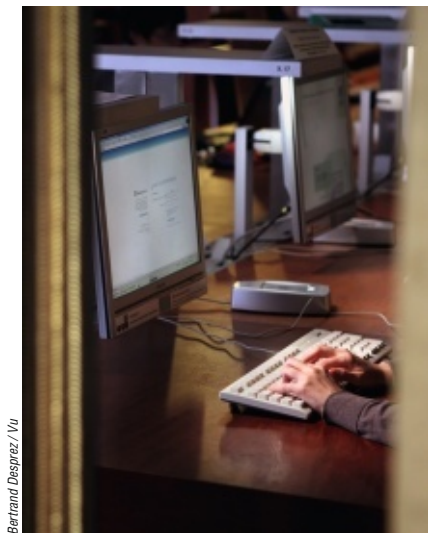
La politique patrimoniale mise en place au fil des ans à la BnF repose sur plusieurs types de collectes, visant à refléter le plus fidèlement possible la société française. Par exemple, la bibliothèque réalise une fois par an un instantané de la totalité des noms de domaines enregistrés sous « .fr » auprès de l'AFNIC (plus de 1,8 million fin 2010, qui ne représenteraient qu'un tiers des sites effectivement enregistrés en France la même année, à comparer aux 192 000 millions de noms de domaines recensés fin 2009 dans le monde). Cette collecte a occupé l'été dernier une trentaine de serveurs durant 12 semaines, et a recueilli près de 24 téraoctets de données, soit 832 millions d'URL.

Pour compléter ces collectes à grande échelle, fidèles à l'esprit du dépôt légal qui n'exclut aucun contenu *a priori*, la BnF col-

lecte aussi plus fréquemment et plus profondément des sites dont les auteurs ou les thématiques s'inscrivent dans la continuité de ses collections.

Chaque matin, le même robot va ainsi « chercher le journal », c'est-à-dire la une en ligne d'une centaine de sites d'actualités qui donnent les nouvelles du jour. Le robot veille à la pluralité des opinions et des sensibilités. Il puise aussi bien dans la presse nationale, régionale, gratuite, que dans des sites d'actualités nés avec le Web, tels *Mediapart* ou *Rue89*. Le chercheur de demain retrouvera ainsi aisément les événements marquants d'une date précise avant de plonger dans des fonds plus spécialisés qui l'attendent déjà. Ces derniers sont le fruit d'un patient travail de veille et de sélection effectué par un réseau de 80 bibliothécaires, souvent en partenariat avec des bibliothèques régionales, des associations ou des laboratoires de recherche.

Depuis 2002, chaque grande campagne électorale mobilise ainsi une équipe de la BnF qui suit jour après jour les faits et gestes numériques des candidats, de leurs opposants et partisans, et veille à leur enregistrement dans les archives portant sur les élections. On y retrouve tracts, affiches,



Bertrand Desprez / Vu

3. L'ACCÈS AUX ARCHIVES DU WEB est souvent strictement encadré en raison des données à caractère personnel qu'elles contiennent. À la BnF, cet accès est ainsi réservé aux chercheurs. En Scandinavie, les archives Web ne sont pas accessibles. En revanche, celles du site privé américain www.archive.org sont disponibles en ligne.

✓ SUR LE WEB

Le consortium IIPC :
<http://netpreserve.org/>

Dépôt légal des sites Web
à la Bibliothèque nationale
de France :

http://www.bnf.fr/fr/professionnels/depot_legal/a_dl_sites_web_mod.html

G. Illien, Le dépôt légal de l'Internet
en pratique : les moissonneurs
du Web, *Bulletin des Bibliothèques
de France*, n° 6, pp. 20-27, 2008,
disponible sur :

<http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004>

Bibliographie établie par la BnF :
http://bibnum.bnf.fr/conservation/bibliographie_dl_web.pdf

CRÉDOC, La diffusion
des technologies de l'information
et de la communication
dans la société française,
décembre 2010
(rapport disponible sur
<http://www.credoc.fr/>).

discours, vidéos et autres documents numériques qui cristallisent les grands enjeux de société d'une époque. Le site de campagne du candidat Jacques Chirac archivé en 2002 (www.chiracaveclafrance.net, disparu depuis), où l'on trouve par exemple son appel aux Français le soir du premier tour de l'élection du 21 avril 2002, y fait figure de pièce de collection...

La BnF suit la vie des mouvements sociaux en ligne et conserve les sites des organisations syndicales, des grandes associations et des grandes causes, mais aussi des entreprises cotées au CAC 40 dont elle archive notamment tous les rapports annuels. Elle s'intéresse aux arts et à l'écriture littéraire. L'une de ses collections les plus fascinantes est faite de journaux personnels en ligne et de sites d'écrivains connus ou non. C'est une sorte de salon littéraire virtuel où l'on découvre les traces des premiers blogueurs (qui étaient des informaticiens) et de leurs suivants : policiers, infirmières, agriculteurs ou caissières qui, un jour, ont préféré le clavier à la plume pour raconter leur vie ou livrer leurs observations. Ces récits de vie numériques, qui côtoient les premiers brouillons et essais en ligne d'écrivains de renom, forment une galerie inédite ; pourquoi l'histoire de la littérature devrait-elle s'arrêter à Casanova, Proust ou Flaubert ?

Le robot est également programmé pour réagir aux urgences et aux grands événements imprévus. Les bibliothécaires suivent l'actualité et sont parfois confrontés à des choix difficiles. Ceux de la BnF ont, par le passé, pris des initiatives de collecte spontanée, jusque dans la rue, dont les historiens leur sont redevables aujourd'hui, comme ce fut le cas pendant l'Occupation ou Mai 68.

Ainsi, en décembre 2010, la BnF décida d'archiver l'intégralité du site controversé *Wikileaks* : l'un des serveurs miroirs étant hébergé chez OVH à Roubaix, le site était juridiquement assujéti au dépôt légal français, fondé depuis cinq siècles sur la relation au territoire national. Il ne s'agissait pas de prendre parti, mais de s'assurer que les « archinautes » du futur auraient les moyens de retrouver trace d'une affaire et d'un débat planétaires, qui seront probablement déterminants pour l'évolution de la régulation des

usages et de la liberté d'expression sur Internet. L'initiative était pertinente puisque, au même moment, le gouvernement américain interdisait l'accès au site *Wikileaks* depuis les agences fédérales américaines, y compris à la Bibliothèque du Congrès.

En prise avec l'actualité

Quelques semaines plus tard, en partenariat avec *Internet Archive*, les robots de la BnF partaient moissonner les sites de Tunisie et d'ailleurs qui ont permis de suivre les révoltes d'Afrique du Nord et du Moyen-Orient. On peut se réjouir, pour une fois, que la France ne soit pas la plus démunie juridiquement et techniquement pour se protéger, face à ce type d'événements, contre l'oubli.

Entre droit à l'oubli et devoir de mémoire numérique, on voit ainsi s'opposer des arguments convaincants de part et d'autre. Idéalement, un archivage qui ne conserverait du Web que sa partie publique, filtrant toute donnée personnelle, constituerait évidemment le compromis le plus acceptable entre la protection des individus et la conservation d'un patrimoine commun. Ce scénario est toutefois impossible à mettre en œuvre dans l'état actuel du droit et des technologies : même si les collectes de la BnF ne représentent qu'une part infime de la totalité des données publiées en ligne, pour rester représentatives elles doivent être effectuées à une échelle suffisamment grande, ce qui implique une large part d'automatisation.

Or identifier et filtrer les données personnelles à cette échelle n'est pas possible, pour des raisons économiques et informatiques, mais aussi parce qu'il est difficile d'apprécier le caractère personnel des informations. Les pratiques d'expression en ligne ont rendu très poreuses les lignes de partage habituelles entre espace public et privé, entre publication et communication interpersonnelle. De ce point de vue, les robots d'archivage ne sauraient faire la part des choses, même s'ils respectent les espaces privés lorsqu'ils sont clairement signalés comme tels (par exemple, le robot n'a pas d'amis sur *Facebook* et ne peut s'introduire dans un réseau social à moins qu'il ne soit déclaré public et visible par tous).

Plusieurs bibliothèques nationales scandinaves (Norvège, Danemark...), pourtant en pointe dans le domaine de l'archivage, font les frais de cette situation : leurs archives du Web restent, pour ces motifs, inaccessibles au public faute de pouvoir en extraire les informations à caractère personnel.

Un accès bien encadré

Un strict encadrement de la consultation des données constitue en revanche une solution pragmatique. C'est *a priori* celle retenue par le législateur français. En l'attente d'un décret qui doit préciser, après consultation de la CNIL, les conditions spécifiques de la consultation des collections issues du dépôt légal de l'Internet, c'est le régime général de communication du dépôt légal qui est appliqué à la BnF depuis l'ouverture de la consultation, en 2008.

Selon le Code du patrimoine, l'auteur d'un site ne peut interdire « la consultation de l'œuvre sur place par des chercheurs dûment accrédités [...] sur des postes individuels de consultation dont l'usage est exclusivement réservé à ces chercheurs ». En pratique, cela signifie que, contrairement aux données accessibles depuis le site www.archive.org, celles de la BnF ne peuvent en aucun cas être mises en ligne. Leur consultation a lieu dans les murs de la BnF et des dispositions pratiques ont été mises en œuvre afin de réserver l'exploitation des données à un usage scientifique ou personnel : notamment, il n'est pas possible d'effectuer de copie numérique des données.

Une autre voie envisagée par la BnF pour améliorer sa mission de dépôt légal du Web est la participation directe des internautes à la sélection des sites à conserver. En 2011, elle projette d'ouvrir un nouveau

service en ligne dédié à la communication avec les internautes en demande de « mémoire ». Chacun pourrait, par exemple, y indiquer l'adresse de son site au moment de sa fermeture ou de sa refonte, afin que la BnF l'archive à temps et le conserve dans son magasin numérique.

Bien sûr, les innovations actuelles (*smartphones*, tablettes numériques, émergence d'un Web sémantique, etc.) nous éloignent des modèles de publication traditionnels et rendront vite obsolètes les technologies employées. Mais c'est déjà une autre ère et un nouveau défi pour les institutions dans leur course contre le temps et l'oubli. En attendant, ce que la BnF et les autres membres du Consortium IIPC auront réussi à sauvegarder de l'âge d'or de l'Internet permettra aux générations futures de retracer les mutations technologiques et culturelles qui auront marqué notre temps. ■



LES RENDEZ-VOUS DU MUSÉUM

AU JARDIN DES PLANTES

2011, Année internationale des Forêts.
Dans le cadre de cet événement, le Muséum vous propose des cycles de conférences et de films.

ENTRÉE LIBRE
DÉTAILS SUR
WWW.JARDINDESPLANTES.NET,
DANS "L'AGENDA DU JARDIN"

LES CONFÉRENCES

Cycle de conférences « La forêt : son rôle, son état, ses habitants » :

Lundi 28 mars : Quelle(s) forêt(s) en France en 2100 ? V. Badeau

Lundi 4 avril : Des forêts et des hommes sous les tropiques, une vieille cohabitation. S. Bahuchet

LES FILMS

Cycle de films « L'usage de la forêt » :

Lundi 11 avril - 18h :
D'une rare crudité - Anim., 2010, 7' 35". E. Davaud, J. Mougel et M. Szymczak
Secrets de plantes : l'if, aux frontières de la vie - Docu., 2010, 52'. J.-L. Bouvret

Cycle de films « Les chemins de l'eau » :

Samedi 16 avril - 15h30 :
Les eaux - Docu., 2008, 15'. K. Kiwanga
Un fleuve humain - Docu., 2006, 92'. S. L'Espérance

Auditorium de la Grande Galerie de l'Évolution
36 rue Geoffroy Saint-Hilaire, Paris 5^e
Métros : Gare d'Austerlitz - Jussieu

LES COURS PUBLICS, LES JEUDIS À 18H

Cycle « Anatomie comparée » Jean-Pierre Gasc

Jeudi 31 mars : Sélection et optimalité
Jeudi 7 avril : La forme et l'information

Grand Amphithéâtre du Muséum
57 rue Cuvier, Paris 5^e - Métro : Jussieu



Dinosaures : on a retrouvé leur sang

Mary Schweitzer



**La préservation
de matériaux organiques
durant plusieurs dizaines
de millions d'années passait
pour impossible.
Pourtant, les paléontologues
en trouvent de plus en plus
dans des os de dinosaures.**

De petites boules rouges ! Quand un collègue me les montre sous le microscope, je refuse d'y croire. Ces minuscules objets se trouvent dans ce qui ressemble à un capillaire serpentant au milieu de la masse jaunâtre de l'os. Chacune est dotée d'un centre foncé (*voir l'encadré pages 32 et 33*). Un noyau cellulaire ? Je sais que, hormis les mammifères, tous les vertébrés – dont les dinosaures – ont des cellules sanguines à noyau. Ces petites sphères rouges dotées d'un centre sombre ressemblent donc à des globules rouges. Mais la lamelle d'os que nous venons de glisser sous l'objectif du microscope provient d'un *Tyrannosaurus rex* trouvé au Montana dans les Rocheuses. Or cet animal est mort il y a 67 millions d'années, et chacun sait que la matière organique ne résiste jamais autant de temps...

Depuis plus de 300 ans, les paléontologues partent du principe que la taille et la forme des os sont les seuls indices livrés par les fossiles. On pensait qu'au cours de la fossilisation, les molécules organi-

ques constituant les tissus mous étaient remplacées par des minéraux provenant du sol, et que seuls les minéraux des os subsistaient. Cette après-midi de 1992, ce que je voyais de mes propres yeux suggérait pourtant l'existence d'exceptions. Mais je me disais aussi que ces petites boules ressemblant à des globules rouges à noyau résultaient peut-être de quelque processus géologique inconnu.

À l'époque, tout juste diplômée de l'Université d'État du Montana, je n'étais pas une paléontologue confirmée. Parce que je demandais l'avis de divers enseignants et thésards, Jack Horner, le conservateur du Département de paléontologie du Muséum des Rocheuses, à Bozeman, eut vent de mes petites sphères rouges. Il me pria d'y jeter lui-même un coup d'œil, et je le vis alors, le front plissé par l'attention, regarder dans le microscope pendant ce qui me sembla des heures. Puis il me demanda ce que j'en pensais. Je lui répondis que ces structures avaient la taille, la forme et la couleur de globules rouges et