



Une histoire politique de l'archivage du web :

→ LE CONSORTIUM INTERNATIONAL POUR LA PRÉSERVATION DE L'INTERNET

GILDAS ILLIEN

Bibliothèque nationale de France
gildas.illien@bnf.fr

Gildas Illien est conservateur en chef au service Dépôt légal numérique de la Bibliothèque nationale de France. Depuis 2007, il a été également responsable technique et trésorier du Consortium international pour la préservation de l'internet (IIPC). Diplômé de l'Institut d'études politiques de Paris et titulaire d'un master en communication de l'université McGill (Montréal), il a auparavant exercé à la bibliothèque universitaire de Paris 8, fait partie de l'équipe de préfiguration de la bibliothèque de l'Institut national d'histoire de l'art et dirigé les médiathèques de l'Institut français de Vienne et du Centre culturel français d'Oslo.

Le web a vingt ou trente ans de moins que la plupart des bibliothécaires, et peu nombreux sont ceux qui se soucient d'en garder la mémoire. En France, comme dans la plupart des pays, le dépôt légal de l'internet est de création récente¹, et la responsabilité juridique de sa conservation incombe principalement à la Bibliothèque nationale. C'est un champ d'expertise isolé, dont la technicité rebute souvent les professionnels. Pourtant, à l'échelle mondiale, le projet d'archiver le web soulève des questions indéniablement politiques. Les bibliothèques nationales sont héritières du concept d'État-nation, qui présuppose une certaine unité entre une culture, une langue, un territoire, des supports de publication. Comment peuvent-elles prétendre conserver un média dont les contenus sont massivement dématérialisés et partagés hors des frontières de l'État? De quel droit? Et à quel coût? Quelles publications doivent-elles sauver en priorité? Selon quels critères de sélection? L'immensité du réseau conduira-t-elle demain les institutions de mémoire des grandes puissances à se partager, façon Yalta, la conservation de tous les savoirs nés numériques du monde? Qu'advient-il du patrimoine immatériel des régions moins favori-

sées? Parce que le web appartient à tout le monde et est devenu la source principale de la mémoire collective de ce début du XXI^e siècle, il est essentiel de mieux faire connaître les dispositifs conservatoires en cours de construction, ainsi que les organisations qui les portent.

Dans la jeune galaxie de la préservation numérique, la coopération internationale s'organise dans les marges, voire tout à fait en dehors des principales associations et conférences de bibliothécaires. Depuis quelques années, on assiste à l'émergence de groupements interlopes et relativement confidentiels. Qu'ont-ils de particulier? D'abord, leurs méthodes de travail sont issues de la culture et des pratiques du web. Ensuite, ces communautés favorisent la rencontre de professionnels qui, jusqu'ici, se parlaient assez peu : bibliothécaires, archivistes, ingénieurs, entrepreneurs, chercheurs. Enfin, on pourrait dire que ce sont des organisations de *technocrates*, dans la mesure où leur création résulte au départ d'un problème technique nécessitant la mobilisation d'experts. Leurs échanges portent d'abord sur les données, les logiciels, les normes. Ce n'est que dans un second temps que leur activité est susceptible de provoquer l'intervention du politique. C'est justement cette question qui nous intéresse : dans quelles circonstances un problème technique devient-il politique?

On a choisi d'explorer cette problématique à partir de l'exemple du consortium international pour la préservation de l'internet (International

1. Pour une présentation de l'organisation du dépôt légal du web à la BnF, voir l'article de Gildas Illien, « Le dépôt légal de l'internet en pratique : les moissonneurs du web », *BBF*, 2008, n° 6, p. 20-27. Disponible en ligne : <http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004>



Carte des membres de l'IIPC (janvier 2010).

Internet Preservation Consortium, IIPC)², qui regroupe aujourd'hui une quarantaine d'institutions. Il est en effet représentatif de ces nouvelles organisations qui interrogent la capacité des bibliothèques à transformer un fait technique en enjeu de politique publique.

L'objectif de cet article est à la fois de dresser l'état des forces internationales engagées dans ce secteur et d'identifier de bonnes pratiques de gouvernance et des méthodes de travail innovantes. On verra ainsi comment le numérique active de nouveaux leviers de coopération et d'influence, qui nécessitent en retour le développement d'organisations et de compétences *ad hoc*. En toile de fond du récit de l'histoire de l'IIPC, on s'intéressera à trois activités qui découlent de ses missions fondamentales : le développement logiciel

(la technique), la communication et le lobbying (la politique), la création collaborative de contenus (les collections).

Aux origines du consortium IIPC : s'unir pour moissonner le monde

À la fin des années quatre-vingt-dix, l'archivage des sites web, pourtant bien moins nombreux et moins complexes qu'aujourd'hui, constituait déjà un défi si considérable qu'une seule institution, fût-elle la fondation pionnière Internet Archive³, ne pouvait le relever seule. Les ressources à mobiliser étaient importantes, alors qu'il n'existait aucun débouché commer-

cial. La révolution qu'allait jouer le web dans la transformation des pratiques culturelles et sociales n'allait pas de soi non plus. Trouver des financements publics pour s'engager dans cette voie a nécessité l'intervention de personnalités visionnaires dotées d'une grande force de conviction. Ces personnalités, comme Brewster Kahle, fondateur d'Internet Archive aux États-Unis, ou Catherine Lupovici⁴ et Julien Masanès⁵ en France, avaient un profil atypique, marqué à la fois par une grande maîtrise technique et une vision patri-

2. Site officiel du consortium IIPC : www.netpreserve.org

3. www.archive.org
Internet Archive a commencé à archiver le web en 1997.

4. Catherine Lupovici, « Web archives long term access and interoperability: the International Internet Preservation Consortium activity », dans *Actes du 71^e Congrès de la Fédération internationale des associations de bibliothécaires et d'institutions (IFLA)*, Oslo, Norvège, 2005. En ligne : <http://archive.ifla.org/IV/ifla71/papers/194e-Lupovici.pdf>

5. Julien Masanès, « Towards continuous Web archiving: first results and an agenda for the future », dans *D-Lib Magazine*, 2002, vol. 8, n° 12. En ligne : www.dlib.org/dlib/december02/masanes/12masanes.html

moniale résolument futuriste. C'est grâce à leurs efforts de dissémination qu'une première prise de conscience internationale a été possible. Celle-ci s'est concrétisée quelques années plus tard lorsqu'une dizaine de bibliothèques nationales réparties entre l'Europe (en particulier les pays nordiques, la France et la Grande-Bretagne), l'Amérique du Nord et l'Australie ont choisi de s'associer à Internet Archive pour fonder le consortium.

Pourquoi cette alliance ? Pour ces bibliothèques *nationales*, l'archivage du web présentait une particularité inédite au regard de traditions structurées par les législations et les cultures de chaque pays : il pouvait difficilement se concevoir en dehors du reste du monde. L'internet était et demeure un média international, tant du point de vue de sa gouvernance, de son architecture, que de ses contenus. Tout partage documentaire du web allait de fait impliquer une concertation sur les frontières et les responsabilités, concertation forcément plus complexe que lorsqu'on pouvait s'en tenir aux critères relativement simples du territoire, de la nationalité ou de la langue. Car, si l'on conçoit aisément que le site de l'Élysée est un site français et celui de la Maison-Blanche un site américain, la chose se complique quand on pense aux archives de Twitter, de YouTube ou de Wikipedia. La nature même de ce gigantesque fonds nécessitait donc, dès le départ, une vision d'interopérabilité entre collections : il eût été irresponsable de ne pas se projeter dans l'avenir, même lointain, où un citoyen serait en droit de disposer d'un accès unique au patrimoine du web, nécessairement planétaire, plutôt qu'à ses pièces détachées confinées dans des silos nationaux. La solution la plus simple, qui fut d'ailleurs proposée par Brewster Kahle en 2003, aurait peut-être été la construction d'une seule et unique archive mondiale alimentée par toutes les bibliothèques nationales. Cependant, du fait des conventions internationales comme des législations nationales sur le droit d'auteur et sur la protection des données personnelles, ce projet ne put voir le jour. En effet, la plupart des législations nationales imposent des restrictions de consultation aux archives

de l'internet. Pour pouvoir les communiquer en ligne, il est généralement nécessaire d'obtenir l'accord préalable des éditeurs – une démarche qui n'est pas possible pour des collections qui se chiffrent en millions de sites⁶.

La solution qui se dégagea fut une solution de compromis. Chaque institution allait archiver de son côté, mais en concertation étroite avec les autres, afin d'éviter la dispersion des efforts, la redondance des contenus et surtout afin de garantir la possibilité de fusionner ou de relier un jour entre elles toutes les collections nationales. Il faudra probablement attendre des décennies pour que ce rêve se réalise, s'il se réalise jamais, mais la gestion du patrimoine implique forcément de faire des paris et de prendre des risques sur le long terme. À l'heure où l'on s'évertue à numériser en nombre les fonds d'imprimés des bibliothèques pour les diffuser en ligne, il reste évidemment paradoxal de territorialiser des segments de l'internet qui ne sont le plus souvent consultables qu'entre les murs des institutions, mais les bibliothèques nationales ne peuvent en aucun cas s'affranchir du droit.

Ces choix conduisirent rapidement à placer la technique au centre des préoccupations. L'urgence était de disposer de robots capables de collecter automatiquement de très grands volumes de sites. Par ailleurs, les pionniers de l'IIPC comptaient parmi eux des professionnels expérimentés, qui avaient œuvré à la réalisation de catalogues collectifs. Ces derniers comprirent très vite que l'interopérabilité future des collections nécessitait l'établissement en amont d'une communauté de formats, de normes et de logiciels, facilitant la production de données homogènes par les institutions. Les entreprises et les logiciels propriétaires n'étaient pas légion dans

6. Seuls Internet Archive et la Bibliothèque nationale et universitaire d'Islande prennent le risque juridique de diffuser leurs archives en ligne sans permission préalable des éditeurs. D'autres institutions, comme la British Library, proposent la consultation en ligne d'un nombre limité de publications archivées pour lesquelles elles ont obtenu l'accord individuel de chaque éditeur de site.

ce secteur. Un appel d'offres pour le développement d'un robot répondant à leurs besoins, lancé conjointement par les trois poids lourds du groupe, la British Library, la Bibliothèque du Congrès et la BnF, resta infructueux. Elles durent se rendre à l'évidence : le marché n'était pas mûr. Elles firent alors le choix de s'engager dans la voie du développement *open source*, à partir des nombreuses briques déjà conçues par Internet Archive. C'est la raison historique pour laquelle les technologies aujourd'hui utilisées par toutes les bibliothèques moissonneuses sont des logiciels libres, propices aux développements collaboratifs qui allaient marquer les premières années du consortium.

Historiquement conçu et construit en réseau, c'était donc en réseau qu'internet devait être conservé. Les pionniers étaient isolés et avaient besoin d'un forum au sein duquel partager leurs expériences et leur foi en « *l'importance d'une collaboration internationale pour assurer la préservation des contenus de l'internet pour les générations futures* » – le credo de l'IIPC, qui figure en préambule de son accord fondateur. C'est fondamentalement la nécessité d'une *communauté* dédiée à l'archivage du web qui a motivé la création du consortium en juillet 2003.

Une organisation virtuelle ?

Lors de sa fondation, le consortium s'est fixé trois missions fondamentales (missions techniques, politiques, documentaires) et quatre activités principales, qui figurent dans l'accord signé par ses membres (voir encadré). On verra que ces missions et activités ont toutes été explorées au cours des sept dernières années, mais plutôt par phases que simultanément. L'interprétation que l'on propose de l'histoire de l'IIPC est en effet celle d'une évolution progressive de la technique vers la politique, une dynamique que l'on retrouve dans toute forme d'innovation.

L'adhésion à l'IIPC s'est toujours faite par cooptation, les institutions candidates devant justifier de réalisa-

tions ou de projets significatifs dans le domaine de l'archivage de l'internet, s'acquitter d'une cotisation annuelle (de 2 000 à 8 000 euros, selon l'importance de leur budget) et accepter les termes de l'accord signé pour trois ans par toutes les parties. Celui-ci a été renouvelé à trois reprises depuis 2003, chacune de ces périodes de trois ans correspondant, on le verra, à un cycle d'évolution.

La gouvernance du consortium a en effet évolué au fil des ans afin de s'adapter à l'effectif et aux aspirations de ses membres. Ces évolutions reflètent un triple mouvement : déconcentration, professionnalisation, démocratisation. Le pilotage était initialement assuré exclusivement par la BnF, mandatée à ces fins par le comité de pilotage composé des membres fondateurs. Puis les tâches ont été progressivement réparties entre plusieurs institutions : la Bibliothèque du Congrès a pris en charge la communication de l'IIPC en 2007, tandis qu'Internet Archive s'est vu confier son pilotage technique à partir de 2010. Cette répartition a favorisé une spécialisation des tâches au sein de l'équipe qui assure la gestion opérationnelle du consortium. La présidence du consortium est devenue tournante et annuelle à partir de 2007, la France, l'Islande, le Canada, Singapour et la Grande-Bretagne se succédant à ce poste. À partir de 2010, la douzaine de représentants, inamovibles, qui siégeaient au comité de pilotage furent élus et renouvelés par tiers par l'assemblée générale. Celle-ci, instituée en 2007, n'avait jusqu'alors aucun pouvoir décisionnel.

Le consortium est une organisation quasiment virtuelle, dont les membres ne se rencontrent en personne qu'une ou deux fois par an. Il ne dispose pas de siège, ni de personnel salarié permanent. Son fonctionnement repose entièrement sur le volontariat et l'engagement collaboratif des membres. Ses ressources financières proviennent des cotisations annuelles, qui couvrent les dépenses associées à l'organisation de réunions et conférences, à la maintenance du site web, au financement de développements informatiques, d'études, et à la prise en charge de certains frais de mission.

Le comité de pilotage (*Steering Committee*) définit la stratégie de l'IIPC. Il vote les décisions importantes, notamment budgétaires. Il se réunit deux ou trois fois par an et, depuis peu, organise tous les deux mois une conférence téléphonique et des votes électroniques. Quatre personnes, appelées officiers de l'IIPC, occupent des fonctions de coordination spécifiques : le président (*Chair*), élu chaque année par le comité de pilotage, joue un rôle d'impulsion et de représentation pour l'ensemble de la communauté, dont il préside toutes les rencontres ; le responsable de la communication (*Communication Officer*), désigné pour 3 ans par le comité

de pilotage, coordonne la communication (site web, liaisons internes, organisation des réunions et conférences) ; le responsable des programmes (*Program Officer*) est également désigné pour 3 ans par le comité de pilotage. Il coordonne la stratégie technique et les activités des groupes de travail ; le trésorier (*Treasurer*), désigné dans les mêmes conditions, assure la coordination administrative et financière du consortium.

Trois groupes de travail permanents (*working groups*) et des groupes d'intérêt (*task forces*) complètent ce dispositif. Ils constituent les espaces de rencontre et de collaboration privilégiés. Les groupes de travail permanents, dont chacun est copiloté par un binôme issu de deux institutions différentes, sont organisés autour de trois axes : collecte, préservation, accès. Les pilotes de ces groupes participent au comité de coordination du consortium, sorte de bureau exécutif où siègent également le président et les trois officiers. Les groupes d'intérêt sont créés en fonction de l'agenda de la communauté. Par exemple, un groupe d'intérêt intitulé *Advocacy and Outreach Task Force* travaille actuellement à l'amélioration de la visibilité politique de l'IIPC auprès de l'industrie du web et des pouvoirs publics.

Chaque groupe est libre d'organiser son travail comme il l'entend, en utilisant notamment la téléconférence, indispensable compte tenu de l'éparpillement géographique des membres sur quatre continents. Le décalage horaire est évidemment une difficulté lorsqu'il s'agit de réunir, par exemple, des collègues installés à San Francisco, Washington, Paris, Tokyo et Wellington : il n'est pas rare de se réunir virtuellement en pyjama depuis son domicile. C'est pourquoi, malgré les coupes budgétaires qui affectent tous les établissements, la réunion de l'assemblée générale, accueillie une fois l'an dans une ville chaque fois différente, est l'occasion privilégiée de rencontres en face-à-face indispensables à la cohésion d'un groupe qui se nourrit des relations personnelles qui se tissent au fil du temps.

Le travail au quotidien au sein du consortium doit s'accommoder des contraintes propres à toute forme de

Les missions fondamentales du consortium IIPC

- Travailler en collaboration, dans le cadre législatif de leurs pays respectifs, pour identifier, développer et faciliter la mise en œuvre de solutions permettant de sélectionner, de collecter et de préserver les contenus de l'internet et d'en assurer l'accessibilité.
- Faciliter la couverture internationale des collections d'archives de contenus de l'internet, en conformité avec leurs cadres législatifs nationaux et en accord avec leurs politiques respectives de développement des collections nationales.
- Plaider vigoureusement au niveau international en faveur d'initiatives et de lois encourageant la collecte, la préservation et l'accès aux contenus de l'internet.

Afin de réaliser ces objectifs, le consortium s'engage à :

- offrir un forum pour le partage des connaissances sur l'archivage des contenus de l'internet ;
- développer et promouvoir des normes pour la collecte, la préservation et l'accès à long terme aux contenus de l'internet ;
- favoriser le développement de logiciels et d'outils appropriés et interopérables, de préférence sous licence libre (*open source*) ;
- améliorer la sensibilisation aux questions liées à la préservation des contenus de l'internet et aux initiatives associées, notamment par le biais de conférences, d'ateliers, de formations, de publications.

coopération internationale. L'année budgétaire n'est pas la même d'un pays à l'autre, ce qui complique la gestion administrative. Les institutions qui ont les moyens de financer les déplacements réguliers de leurs représentants jouent un rôle décisif. Celles qui sont brutalement frappées par la RGPP⁷ locale suivent les projets plus mollement et finalement plus du tout. L'hiver des uns est l'été des autres, rendant difficile la collaboration pendant les grandes vacances. La question de la langue de travail est évidemment centrale : la maîtrise de l'anglais (et de ses différents accents : américain, écossais, australien... mais aussi espagnol, allemand ou japonais) est indispensable à la survie au sein du groupe. Le français est peu parlé, sinon avec les Québécois et les Suisses. L'organisation des réunions et des événements sociaux requiert un tour de force programmatique et culinaire pour tenir compte des régimes et des horaires alimentaires propres aux différentes cultures et religions. Certains clichés inavouables se vérifient en pratique. Les Américains, qui marchent au café clair, sont plutôt concis et efficaces de bon matin. Les Français sont très bavards et aiment travailler tard autour d'un bon repas. L'ingénieur anglais est particulièrement inventif au pub. Le collègue japonais ne dit jamais non (culturellement, cela lui est impossible), ce qui oblige à formuler les propositions de manière assez diplomatique lorsqu'on doit procéder à un vote. Cependant, les effets secondaires associés au décalage horaire et aux temps de vol pèsent sur tous et finissent par aplanir les différences. C'est souvent dans des moments d'épuisement collectif que les décisions les plus pragmatiques et les plus tranchées sont prises.

Ce tableau peut sembler anecdotique. Les facteurs culturels et géographiques jouent pourtant un rôle aussi important que les compétences et les affinités professionnelles dans le succès ou l'échec de la plupart des projets – on aurait tort de les minorer.

L'ère des ingénieurs

Après cette présentation du cadre général de fonctionnement de l'IIPC, nous pouvons en survoler l'histoire. La première phase du consortium (2003-2006) a été celle de l'innovation technologique. À cette époque, le consortium ne compte que douze membres, et ses réunions mobilisent au mieux une trentaine de personnes. Il s'agit principalement d'ingénieurs férus du web, chaperonnés par des décideurs au profil technico-stratégique, qui dessinent la trajectoire et veillent sur les ressources budgétaires ; en revanche, peu de bibliothécaires sont impliqués. C'est l'âge d'or du développement logiciel, où sont posés les fondements technologiques de l'archivage d'un web encore assez simple, principalement en HTML. On travaille alors principalement à consolider l'architecture de logiciels, tels le populaire robot moissonneur Heritrix⁸, adopté par la BnF pour ses collectes du dépôt légal, ou encore la Wayback Machine⁹, la « machine à remonter le temps », qui permet l'indexation et la recherche par adresse URL. Dans le même temps, des groupes de travail aux noms mystérieux (« *Framework* », « *Metrics* », « *Deep Web* ») planchent sur des sujets pointus et prospectifs qui préfigurent des questions devenues centrales depuis, qu'on pense à l'archivage du web profond ou à la difficulté de mesurer de manière uniforme les contenus et les usages des archives.

Les pionniers commencent à moissonner la Toile, généralement à titre expérimental, et saturent, dans l'euphorie des commencements, leurs premiers serveurs de test. Internet Archive, installée dans une petite maison en bois du parc du Presidio, à San Francisco, accueille en stage de jeunes ingénieurs fraîchement recrutés par les BN d'Islande, du Danemark, de France ou d'Australie. Ceux-ci reviennent chez eux avec des photos où on les voit boire des sodas et manger des pizzas tout en scrutant joyeusement des lignes de code et d'URL sur des écrans. Dans une ancienne mine

du Cercle polaire, à Mo i Rana, les Norvégiens installent leur première ferme de serveurs et partent à l'assaut de leur domaine national, le *.no*. En Islande, un ingénieur de 25 ans capture et indexe à lui seul tout le web national, mais ne fait pas cela à temps plein. On apprend sur le tas, on parle de données plutôt que de collections. Les choses se font en masse et à la louche. Les partenaires de l'IIPC sont peu nombreux à proposer une consultation publique de ce qui s'apparente encore à une boîte noire. L'urgence est alors de collecter, l'accès et la conservation de long terme ne sont pas identifiés comme des besoins immédiats. Si bien qu'il n'est pas rare de perdre ou de détruire des données qui, faute de loi, ne sont pas encore devenues inaliénables. Cette époque, profondément sympathique et créative, signe la rencontre du troisième type entre les cadres de bibliothèques nationales multacentraires et des ingénieurs fous. Cette rencontre focalisée, on l'a vu, sur les logiciels libres, va durablement marquer la culture de l'organisation, qui conservera un style résolument décontracté et pragmatique ainsi que l'habitude du travail collaboratif à distance.

À la conquête du monde

La seconde phase de l'IIPC (2007-2009) conduit à l'élargissement du consortium. De nombreux pays mais aussi d'autres types d'organisations (archives nationales et audiovisuelles, comme l'INA ; bibliothèques universitaires, comme celle de Harvard) frappent à la porte. Les membres du comité de pilotage s'interrogent : prendront-ils le risque d'accueillir des débutants et d'enrayer la dynamique d'innovation ? Ils choisissent d'étendre le réseau afin d'atteindre la taille critique qui lui fait défaut pour que la cause de l'archivage du web soit défendue par une communauté audible. C'est un choix décisif, qui marque les débuts de la « politisation » de l'IIPC, et en particulier son souci d'acquiescer une vraie stature internationale. La Bibliothèque du Congrès prend les rênes de la communication et conduit de main de maître une campagne de recru-

7. Révision générale des politiques publiques.

8. <http://crawler.archive.org>

9. www.waybackmachine.org



Réunion de l'IIPC à Vienne en 2010

tement et de sensibilisation dans de nombreuses régions du monde, y compris au Sud. Les motivations de l'élargissement ne sont en effet pas seulement guidées par l'instinct de survie ou de rayonnement de l'organisation. Il s'agit aussi de sensibiliser les pays émergents aux enjeux du patrimoine numérique. Le choix de logiciels *open source* utilisables gratuitement par tous répondait à cette même préoccupation : faire financer par les institutions occidentales des outils que les moins nantis pourraient réutiliser pour archiver leur propre patrimoine.

Le consortium ne réussit cependant pas à recruter dans les pays en développement. Conserver le web n'est pas une priorité quand on manque de bibliothèques, de livres, d'ordinateurs et d'électricité. L'IIPC s'ouvre néanmoins largement à l'Asie, avec l'arrivée des BN du Japon, de Singapour, de Corée du Sud, de Nouvelle-Zélande et même de Chine. La participation européenne se renforce aussi, que ce soit en Allemagne, en Pologne, en Espagne, ou en Slovaquie, et avec

la participation inédite d'organismes de recherche et de développement (la Fondation European Archive, devenue depuis Internet Memory Foundation¹⁰), et de « jeunes pousses » (comme la société britannique Hanzo). Au terme de cette campagne, l'IIPC triple en deux ans le nombre de ses adhérents. Sa liste de diffusion compte plus de 250 destinataires dans 39 institutions. Son modèle de fonctionnement est imité par d'autres organisations comme le NDIIPP (National Digital Information Infrastructure and Preservation Program) qui coordonne les actions nationales pour la préservation du numérique aux États-Unis et compte aujourd'hui plus d'une centaine d'institutions membres.

Les cotisations abondent un budget qui se chiffre désormais en centaines de milliers d'euros que l'on n'arrive pas toujours à dépenser faute d'institutions et d'ingénieurs pour porter les projets informatiques.

10. <http://internetmemory.org>

Ce mouvement d'expansion est indubitablement le fruit d'une bonne campagne de marketing. Mais il correspond également à des évolutions plus profondes au sein des institutions. Les archives du web commencent en effet à y perdre leur statut expérimental et à poser des problèmes classiques de fonctionnement et de pérennisation. En 2006, en France et en Finlande, le dépôt légal du web est devenu une mission officielle, inscrite dans la loi. Un petit nombre d'archives s'ouvrent même au public : c'est par exemple en 2008 que les archives de l'internet français sont déployées dans les salles de lecture de la BnF¹¹. Les données collectées se chiffrent rapidement en dizaines, puis en centaines de téraoctets. Leur masse leur

11. Sara Aubry, « Les archives de l'internet : un nouveau service de la BnF », dans le dossier « Pleins feux sur la veille », *Documentaliste – Sciences de l'information*, 2008, vol. 45, n° 4, p. 12-13. En ligne : www.cairn.info/revue-documentaliste-sciences-de-l-information-2008-4.htm

donne de la visibilité et soulève des problèmes de coûts. Ce passage progressif de l'expérimentation à ce que les Anglo-Saxons appellent « *business as usual* » conduit à mobiliser des équipes peu nombreuses sur des opérations de traitement et de maintenance qui requièrent la formalisation de procédures, l'établissement de contrats de service, la fixation d'objectifs volumétriques, au détriment de la recherche et du développement. Pour répondre à ces nouvelles préoccupations, les bibliothèques nationales qui siègent au comité de pilotage y désignent des professionnels de plus en plus gradés, qui ne sont plus nécessairement des directeurs de l'informatique ou des spécialistes de la préservation numérique : des gestionnaires et des conservateurs de haut rang rejoignent le groupe, apportant avec eux des méthodes de travail qui diffèrent de celles de l'ère précédente. La comptabilité analytique, les « *business plans* » et la gestion de projet à l'anglo-saxonne font leur apparition.

En Europe, la bibliothèque numérique Europeana se construit en réponse à Google et à ses projets et capte l'attention des politiques et des médias. Le dépôt légal numérique grandit discrètement à l'ombre de la numérisation. Pourtant, la Conférence des directeurs de bibliothèques nationales et la Commission européenne identifient, encore timidement, l'archivage du web comme un enjeu potentiellement stratégique. Un premier projet européen (Living Web Archive, LIWA)¹² est financé par la Commission à partir de 2008. Des liens se tissent également entre les membres de l'IIPC et les communautés de la préservation numérique, comme celles que rassemble la conférence iPRES¹³ : alors que s'engage la construction d'entrepôts numériques comme Spar (à la BnF) ou Rosetta (en Nouvelle-Zélande), destinés à accueillir d'importants volumes d'archives du web, cette coopération s'impose. Elle permet de faire reconnaître

les spécificités de ces nouvelles collections, qui suscitaient jusque-là une certaine méfiance en raison de leur volumétrie monstre et de la diversité de leurs formats de fichiers. Les travaux de normalisation orchestrés par l'IIPC contribuent pour beaucoup à leur reconnaissance institutionnelle, en particulier la normalisation en 2009 par l'ISO du format WARC¹⁴, le format conteneur des archives du web.

La conjonction de ces évolutions (plus de membres, aux profils plus variés, et dont un certain nombre sont désormais tenus par leurs législations nationales à une obligation de résultat) oblige à repenser les modalités de travail au sein de l'IIPC. Aux objectifs de collecte s'ajoutent désormais de manière plus tangible ceux de la consultation et de la préservation. C'est à cette époque que ces trois problématiques, qui recoupent celles du circuit du document en bibliothèque, sont retenues pour coordonner durablement l'activité des groupes de travail permanents, rendant l'organisation plus lisible par des non-techniciens. Elles posent aussi la question de l'intégration des archives du web dans des flux de traitement existants : peut-on les cataloguer, les indexer, les conserver selon des procédures connues des services et des personnels en place ?

Le consortium se heurte alors aux différences d'attentes et de styles qui opposent les ingénieurs, conscients que les technologies inventées quatre ans plus tôt ne suivent pas les évolutions du web, et les décideurs, impatients de stabiliser des modèles de production. Ces derniers trouvent difficilement leur place dans un environnement encore très marqué par l'informatique et la modélisation. Le consortium s'attache alors à maintenir le lien entre ces communautés. À Paris, Canberra, Ottawa et, plus tard, à Singapour, il organise pendant quatre années consécutives cinq jours de conférences, ateliers, formations et réunions dans le but de

brasser les métiers, les compétences, les cultures, les pratiques. Une centaine de personnes assistent à chaque fois à ces rencontres. Il est devenu nécessaire d'affréter un ou deux cars pour les transporter ; on produit les premières mallettes et les premières clés USB marquées du logo de l'IIPC. La réception d'ouverture de l'assemblée générale de 2008 a lieu dans le spectaculaire hall d'honneur de la BN d'Australie. Elle s'ouvre par la traditionnelle danse du kangourou, qui précède l'allocution d'un chef maori. À la surprise générale, celui-ci explique, en substance, que l'archivage du web s'inscrit dans le sillage des initiatives de l'Unesco pour sauvegarder le patrimoine immatériel des civilisations en voie d'extinction.

Ce temps d'expansion et d'intense dissémination aussi bien à l'extérieur qu'à l'intérieur de la communauté est ainsi le moment où la technique bascule vers le politique. Pendant ces années, l'organisation a perdu en rapidité et en agilité informatique ce qu'elle a gagné en visibilité politique. Paradoxalement, alors que les outils qu'elle a développés sont adoptés par un nombre croissant d'institutions, ils sont en perte de vitesse car ils n'ont pas su évoluer aussi vite que le web : une grande part du patrimoine échappe désormais aux robots.

Des données devenues collections : vers une géopolitique du patrimoine numérique ?

L'année 2010 est scellée par la signature d'un troisième accord et un renouvellement important des instances de pilotage du consortium. Les difficultés du développement logiciel sont reconnues mais pas franchement résolues. Certains voudraient renouer avec l'âge d'or informatique et donner aux ingénieurs des bibliothèques les moyens de travailler davantage ensemble. D'autres, frappés de plein fouet par les réductions budgétaires, considèrent au contraire que les défis technologiques de l'archivage ne peuvent plus être résolus de cette manière. L'amélioration de la qualité des

12. <http://liwa-project.eu>

13. iPRES : International Conference on Preservation of Digital Objects. Existe depuis 2003.

14. *Information and documentation – WARC file format*, ISO 28500:2009, ISO TC46/SC4, 2009.

captures du web et la maintenance des applications vont probablement nécessiter que l'on se tourne vers des organismes extérieurs entièrement dédiés à la recherche et au développement, voire vers l'industrie informatique et les grands éditeurs de sites. Ce sont les questions politiques qui prennent aujourd'hui une place prépondérante dans les discussions. Elles s'orientent de plus en plus vers des problématiques de contenus et de responsabilité que l'on pourrait qualifier de *géo-politique patrimoniale*.

Dès 2009, huit bibliothèques nationales européennes avaient réalisé conjointement la collecte des sites web de la campagne des élections européennes, tandis qu'une coalition de bibliothèques américaines avait saugrenu dans l'urgence des centaines de sites documentant les huit années de pouvoir de l'administration Bush, sites que la NARA (National Archives and Records Administration)¹⁵ avait renoncé à archiver. Dans le même temps, l'IIPC avait lancé un projet pilote sur trois ans visant à expérimenter l'interopérabilité des collections nationales : un essai d'archivage collaboratif en grandeur réelle des sites des Jeux olympiques de Londres de 2012. Mais, début 2010, l'histoire du web semble s'accélérer, poussant les institutions à élargir sans plus attendre les frontières de leurs interventions patrimoniales. Tous les observateurs s'accordent à reconnaître que le jeu des relations internationales est profondément transformé par l'internet. Les sites d'interactivité et d'échanges identifiés sous le sigle 2.0 apparaissent de plus en plus clairement comme des catalyseurs des aspirations, revendications et mobilisations politiques des sociétés civiles. L'administration américaine prend acte de la puissance politique du web. Dans un discours prononcé en janvier de la même année, la secrétaire d'État Hillary Clinton déclare que la défense et la promotion de l'internet dans le monde sont l'axe prioritaire de la nouvelle doctrine diplomatique de Washington. Au même moment, la Bibliothèque du Congrès, la BnF et Internet Archive réalisent

ensemble la collecte d'urgence d'un ensemble de sites relatifs au séisme en Haïti. Un an plus tard, elles renouvellent cette coopération spontanée, d'abord pour archiver les sites de Wikileaks, puis, très récemment, à l'occasion de la révolution du Jasmin en Tunisie et dans le reste de l'Afrique du Nord. Au risque de s'écarter de leurs missions initiales, elles laissent leurs robots s'aventurer dans des zones grises, sans territoire fixe. Car les bibliothèques du consortium ne peuvent plus ignorer des événements et des contenus numériques particulièrement volatils documentant une future histoire du monde qui n'est pas réductible à la somme de leurs histoires nationales.

Plus nombreux à être associés à leur sélection et à leur valorisation, les bibliothécaires finissent ainsi par s'intéresser aux archives du web. On recense les premiers cas de disparition critique de contenus de l'internet et les premières demandes de particuliers, d'entreprises et même de partis politiques qui ont perdu leurs données. L'émergence d'un public potentiel pour ce nouveau fonds qui dépasse celui des chercheurs universitaires stimule les travaux documentaires et scientifiques. Pour répondre à ces besoins, les professionnels commencent à formaliser des politiques documentaires d'un genre nouveau¹⁶. Parce que l'archive du web devient plus intelligible et plus respectable en tant qu'objet bibliothéconomique, on commence à la regarder comme une *collection* à part entière. Ironiquement, sept ans après sa fondation, la politique du consortium s'empare enfin des questions documentaires qui avaient motivé sa création.

16. Voir, par exemple, la contribution de Gildas Illien et Clément Oury, « Quelle politique documentaire pour l'archivage des sites internet ? », in *Les collections électroniques, une politique documentaire en mouvement*, sous la dir. de Pierre Carbone et François Cavalier, Paris, Éditions du Cercle de la librairie, 2009, p. 157-178.

Conclusion

Au terme de ce récit, que retenir de l'expérience de l'IIPC ? La sociologie des organisations nous enseigne que les intentions initiales d'une organisation se réalisent rarement. En réalité, ce qu'elles produisent est le fruit d'une conjonction de facteurs internes (valeurs, croyances et interactions individuelles des acteurs) et externes (hasards, accidents, opportunités). La brève histoire de l'IIPC confirme cette analyse. Ce qui est néanmoins remarquable, et à porter au crédit des personnalités à l'origine de cette aventure, c'est qu'elles avaient dès le départ identifié les grands sujets mais aussi les points de tension qui allaient déterminer toute sa dynamique d'innovation. Essayons d'en tirer quelques enseignements utiles pour le développement de la coopération internationale dans le domaine numérique.

Dans le domaine des bibliothèques numériques, le développement logiciel est le nerf de la guerre

C'est peut-être une évidence, mais il est bon de la rappeler : ce n'est pas parce qu'il est gratuit qu'un logiciel libre ne coûte pas cher. Dans les activités du patrimoine, vouées par définition à fonctionner sur un temps long, il est encore plus dangereux qu'ailleurs d'être tributaire de logiciels propriétaires. Néanmoins, le recours à l'*open source* implique de disposer de moyens informatiques conséquents, et d'accepter en outre de contribuer ponctuellement à des développements qui ne répondent pas aux priorités immédiates de l'établissement. À la place du paiement de licences, il faut impérativement prévoir du temps de travail d'ingénieur dédié au développement (distinct du temps de travail consacré aux opérations de production), ainsi que de certains déplacements à l'étranger. La coopération internationale offre des possibilités de mutualisation des charges de l'*open source* qui sont réellement intéressantes. Mais, pour que ce calcul soit rentable, il faut que le développement s'appuie sur une communauté d'utilisateurs assez

15. www.archives.org

nombreuse et assez active. C'est un peu toute la différence entre la poignée de développeurs qui a inventé le robot Heritrix et l'immense communauté qui maintient le navigateur Firefox.

La mixité et l'organisation interne des communautés numériques déterminent leurs chances de réussite

Pour que les bibliothèques s'emparent durablement d'un défi technique et se donnent les moyens de le relever, il est nécessaire que des décideurs, des bibliothécaires et des usagers se mobilisent dès le départ au côté des experts techniques afin que les problèmes soulevés fassent sens pour l'ensemble des acteurs potentiellement impactés par l'innovation – et susceptibles de la mettre en œuvre ensuite. Sans cette alliance objective et cet accompagnement métier de l'expertise, la portée politique de l'innovation technologique ne sera pas comprise assez tôt. La construction

d'une relation de confiance entre ces familles professionnelles est toutefois la chose la plus difficile à réaliser. Si la structure de l'organisation enferme les experts, elle les étouffera et, avec eux, toutes les forces innovantes. Si, au contraire, elle les laisse inventer en orbite, trop à l'écart des contraintes documentaires, humaines et économiques, l'innovation ne trouvera pas non plus le chemin de la production.

La question de la taille et du champ de l'organisation est déterminante

Trop petite, confinée à un seul aspect du problème, ou à une seule région du monde, une organisation n'aura ni l'audience ni la taille critique nécessaires à sa survie. Trop grande, trop ambitieuse, elle aura à embrasser trop de problèmes pour en résoudre aucun correctement. De ce point de vue, la culture professionnelle des Américains, moins holistique et plus pragmatique que celle des Français, peut être une source d'inspiration :

plutôt que de s'attaquer en permanence à tous les problèmes dans leur globalité sans en avoir les moyens, il vaut mieux couper le travail en plus petits morceaux (les «*work packages*»), les répartir entre acteurs autonomes, et phaser leur réalisation dans le temps, à raison d'un ou deux grands objectifs par an. De la même façon, on préférera des organisations souples et évolutives, affranchies de contraintes administratives ou salariales potentiellement paralysantes dans le cas où le besoin de s'adapter rapidement à un environnement en mutation se ferait sentir. De ce point de vue, le modèle de l'IIPC, qui conduit à revisiter les modalités de gouvernance tous les trois ans, est intéressant, puisqu'il permet d'envisager assez facilement toutes sortes d'évolutions. Y compris sa propre dissolution, le jour où l'archivage du web aura été banalisé et ne justifiera plus l'existence d'une entité distincte des autres organisations et associations de bibliothèques. Une organisation n'est pas une fin en soi. ●

Février 2011