

La préservation des connaissances : instrumentation de contenus et interprétation des vues documentaires

Bruno Bachimont (1)
bruno.bachimont@utc.fr

Stéphane Crozat (2)
stephane.crozat@utc.fr

Université de Technologie de Compiègne, UMR CNRS Heudiasyc (1)

Université de Technologie de Compiègne, Unité Recherche Action Ingénierie des contenus et des savoirs (2)

Mots-clés : préservation, connaissance, gestion, OAIS, chaîne éditoriale, herméneutique, philologie

Keywords: preservation, knowledge, management, OAIS, editorial chain, hermeneutics, philology.

Résumé : La gestion des connaissances tente d'explicitier la connaissance pour la rendre capitalisable sous les formes de documents. La préservation des contenus vise à maintenir l'intégrité et l'authenticité des contenus au cours du temps et de la longue durée. Or, la gestion des connaissances doit prendre en compte la transmission dans le temps des contenus qui capitalisent la connaissance, et la préservation ne s'effectue qu'en entretenant les connaissances associées aux contenus. L'approche proposée ici consiste à croiser ces deux points de vue pour la préservation des connaissances.

Abstract : knowledge management aims at explicating knowledge in order to make it usable by means of documents. Content preservation aims at ensuring content integrity and authenticity through time. While knowledge management should take into account document transmission, content preservation relies on knowledge and its maintenance through time. The proposed approach consists in merging these two points of view for knowledge preservation.

1 Introduction de l'article

Alors que l'on parle habituellement de gestion des connaissances, on évoque plutôt la préservation des contenus. En effet, la connaissance serait cet actif immatériel qu'il faudrait seulement gérer, alors que les contenus sont des objets matériels dont il faut assurer l'intégrité physique. Mais les connaissances s'expriment à travers des inscriptions qui les matérialisent, et les contenus s'interprètent par des connaissances qui les actualisent. Les opposer est donc artificiel, les traiter séparément stérile.

Or, ces deux domaines restent étonnement indépendants, la gestion des connaissances ignorant la longue durée, la préservation la connaissance. La proposition de cet article est de les articuler, chacun pouvant apporter à l'autre ce qui lui manque. La préservation nous apprend que la connaissance ne s'entretient que par son usage et sa mise en pratique : l'usage permet la conservation, non l'inverse. La gestion nous apprend qu'il faut expliciter les différents savoirs dont on dispose pour permettre leur formulation et transmission.

Nous proposons donc une approche de la préservation par l'accès où les connaissances sont capitalisées, préservées et transmises à travers une pratique entretenue et permanente, reposant sur une instrumentation des contenus. Dans le contexte numérique contemporain, les contenus matérialisant les connaissances sont en effet soumis à des transformations permanentes, que ce soit pour les lire ou les exploiter, si bien que tout accès devient dès lors une transformation des contenus. La préservation par l'usage devient une herméneutique des avatars documentaires créés par l'usage.

2 Une approche patrimoniale pour la capitalisation

2.1 Gestion des connaissances et inscription documentaire

La gestion des connaissances se définit essentiellement aujourd'hui comme le fait d'explicitier, rassembler et diffuser les savoirs et savoir faire d'une organisation ou d'une communauté. Ses principales difficultés relèvent alors de la difficulté de formuler et formaliser ces savoirs et savoir faire pour pouvoir les collecter, et d'autre part de les adapter et les convertir pour les diffuser et partager. Formulée ainsi, la problématique se pose essentiellement en terme de transcription, de l'implicite à l'explicite, de l'informel au formel, du pratique au théorique, du spontané au rationalisé. Cela étant, la gestion des connaissances bute sur deux difficultés :

- L'ancrage documentaire des contenus rassemblés et formulés
- La gestion dans le temps de la vie de ces contenus, pour préserver tant leur intégrité que leur intelligibilité.

Le premier problème correspond au fait de pouvoir instrumenter les contenus de manière à permettre leur exploration, comparaison et exploitation. En particulier, il est nécessaire de pouvoir mettre en résonance des contenus différents mobilisant des formats hétérogènes comme le texte, la vidéo, le son, etc. Une problématique générale du dossier de la connaissance émerge, où le lecteur / utilisateur se trouve confronté à une multiplicité de contenus devant lesquels il doit être capable d'abstraire une connaissance et de la mobiliser.

Le second problème correspond au fait que le numérique est un support par nature instable permettant la transformation et la manipulation des contenus. Comme en témoigne les ressources bureautiques, un même fichier se donne à lire de manière différente selon les traitements de texte et outils utilisés. La question est alors de savoir en quoi un contenu vu différemment selon différentes vues produites par différents outils reste le même ; en quoi peut-il garder son authenticité à travers les accès à chaque fois modifiés qu'on peut en avoir. De même, les outils numériques permettent de généraliser une logique du emploi, de la réutilisation, de la ré-éditorialisation. Nombre de contenus sont originaux par leur intention, moins par leur composition qui peut reprendre force d'éléments préexistants. Un nouveau problème émerge ainsi de suivre la généalogie des contenus à travers leur leur mutation documentaire. La problématique est donc la suivante:

- d'une part, on a différents contenus renvoyant à une synthèse à construire et abstraire;
- d'autre part, on a différentes vues de « mêmes » contenus qu'il faut confronter et discuter pour déterminer et étudier l'authenticité et l'identité des contenus.

Le parti pris de cet article est de mettre en avant l'utilisation et la réutilisation des contenus comme approche générale de la capitalisation et de la préservation. D'une part on ne capitalise que ce dont on a besoin et l'on ne préserve que ce que l'on utilise ; et d'autre part l'accès aux contenus pour leur usage est une garantie de leur gestion continue et de leur préservation.

Ainsi, il faut prendre acte et non le regretter que le numérique confère une mutabilité intrinsèque aux contenus. Le problème n'est plus de conserver un contenu identique à lui-même, mais de le suivre et le reconnaître à travers ses différents avatars et représentations qu'il adopte au cours du temps. L'utiliser, c'est donc le transformer, le laisser identique à lui-même, c'est l'oublier. Cette mutabilité se retrouve tant sur le plan de la forme que du fond :

- un contenu numérique doit posséder une lisibilité technique car il repose sur une instrumentation donnée pour être lu ;
- un contenu en général doit reposer sur une lisibilité culturelle car il s'inscrit dans un système de normes et de codes pour être compris et interprété.

Par conséquent, gérer la connaissance, capitaliser pour la transmettre et la réutiliser, sont des tâches qui se définissent sur le fond de cette mutabilité. Pour se souvenir, il faut transformer, pour comprendre il faut interpréter. La capitalisation des connaissances doit donc s'effectuer comme une herméneutique des avatars documentaires créés par l'usage.

2.2 La gestion patrimoniale des contenus

L'approche est ici de considérer ce qui a fait le succès de la transmission des connaissances, pratiques ou théoriques, dans le temps, voire la longue durée, pour l'adapter à la gestion des connaissances et de leur

inscription documentaire. Le problème abordé sous l'angle de la transmission peut se résumer de la manière suivante :

- Des contenus documentaires expriment certaines connaissances, savoir faire ou pratiques ;
- Ces contenus, au moment de leur production, sont encore plongés dans un environnement, une tradition de lecture et d'interprétation, qui permet de recouvrer leur signification et d'accéder à leur mise en œuvre, pratique ou théorique.
- Avec le temps cependant, ces contenus se décontextualisent progressivement : un fossé d'intelligibilité se creuse si bien qu'ils deviennent illisibles et incompréhensibles.

Afin d'éviter le creusement de ce fossé, il faut donc que le contexte d'interprétation des contenus soit maintenu au fil du temps. Pour cela il faut que les contenus soient utilisés (lu, annotés, rééditorialisé, etc.) et que les modalités de cet usage soit contrôlé par un dispositif qui permettent et maintiennent les conditions de l'usage dans le temps. La solution élaborée par la tradition occidentale repose sur les éléments suivants :

- Une conservation des supports physiques des contenus. C'est le rôle traditionnel des bibliothèques, qui rassemblent, inventorient et préservent les contenus dont elles ont la charge.
- Une interprétation permanente et dynamique des contenus, qui maintient l'intelligibilité de ces derniers, au besoin en les enrichissant de commentaires, gloses, explicitations, venant faciliter la recontextualisation des contenus dans l'environnement social et intellectuel du moment. C'est traditionnellement le rôle de l'université.
- Quand les contenus ont une dimension performative, comme les outils techniques ou les instruments de musique, on ajoute un « conservatoire » qui a pour but de répéter et transmettre les pratiques associées aux instruments : elles répliquent en leur sein la dichotomie ci-dessus, en conservant les outils et en transmettant les savoir faire. Ce sont les conservatoires des arts et métiers par exemple, ou les conservatoires de musique.

On voit ainsi que la capitalisation au cours du temps repose sur deux piliers fondamentaux : la conservation des contenus et le maintien d'une tradition d'usage et de lecture. Seul l'usage répété et systématique évite la rupture de la transmission du savoir. La conservation peut être plus ou moins instrumentée, de manière à permettre un accès décontextualisé et autoriser ainsi des relectures dans des contextes différents de l'origine, mais il reste toujours la nécessité de s'inscrire dans un cadre partagé par les compétences du lecteur d'une part et le contexte du document d'autre part. Si par exemple, on peut ajouter des glossaires pour expliquer des mots anciens, on supporte ainsi la perte de savoir, mais on repose sur la connaissance de la langue commune. La question est à présent de savoir comment instrumenter l'usage documentaire pour faciliter la préservation des connaissances.

3 Gestion patrimoniale des contenus et gestion documentaire

L'enjeu est de proposer une approche de la gestion documentaire des connaissances reprenant ces clefs du succès de la transmission patrimoniale des contenus. Cela pose plusieurs questions :

- Identifier les rôles et les actants : que sont les contenus, les interprètes, les institutions ? On s'aperçoit que, souvent, la capitalisation des connaissances constitue bien une sorte de bibliothèque, mais laisse ses documents dormants, sans une université venant constamment lire et actualiser le contenu.

Instrumenter les rôles et les contenus : comment traduire de manière éditoriale la conservation des contenus de manière à rendre possible la dynamique interprétative et à intégrer dans la conservation et la transmission le résultat de ces interprétations ?

L'idée est alors de constituer une éditorialisation des contenus et de la glose comme la condition de la gestion documentaire, de manière à conserver son intelligibilité. Les propositions que l'on peut faire sont les suivantes :

- La gestion des contenus : pour une philologie documentaire et numérique. L'objectif est de revenir sur l'identité d'un contenu, son authenticité et intégrité, pour définir quelles sont ses différentes versions, variations et transformations. Ce problème est particulièrement aigu dans le contexte numérique où le simple fait de lire un contenu à l'écran est d'emblée la transformation de la ressource conservée, où les reproductions se réalisent instantanément (copier/coller) et où un même contenu physique (un fichier par exemple) peut être mobilisé au sein de document différents (une image, une définition, etc.). La solution générale au problème de l'identité du contenu est le maintien d'une généalogie qui permette de

connaître les liens entre une instance physique d'un contenu et ses parents (pères, frères, cousins, clones, etc.).

- L'interprétation des contenus : pour une herméneutique documentaire. L'objectif est de préciser comment constituer un appareil critique pour l'explicitation sédimentée d'un contenu, qui vient enrichir sa consultation. Le numérique permet une approche nouvelle de cette sédimentation, dans la mesure où l'ajout, le retrait, le référence, la variation devient une modification *intégrée* au contenu - en fait elle devient le contenu lui-même et non seulement une glose au sens d'une marque « en marge ». La solution générale au problème de l'interprétation du contenu est l'enregistrement explicite des différentes couches de modification qui viennent enrichir un contenu au cours du temps et des usages.

Puisque l'approche est de *transformer* les contenus pour les conserver, transformation matérielle pour préserver leur accessibilité physique (changer de format, de support, etc.), et de les *enrichir* pour préserver leur accessibilité intellectuelle, à l'instar des contenus plus anciens comme ceux de la tradition manuscrite, il convient d'élaborer une philologie et herméneutique numérique pour rendre possible et gérer cette *transmanence* des contenus, la persistance patrimoniale du contenu à travers son évolution dans le temps.

4 Instrumentation éditoriale des contenus pour leur transmission

Nous avons posé comme stratégie de préservation une démarche fondée sur l'accès aux contenus, combiné à une gestion de leur identité et de leurs enrichissements. Dans le contexte du documentaire numérique, nous proposons de reformuler cela comme la mise au point d'un dispositif permettant l'instrumentation éditoriale - nous pourrions ajouter continue - des contenus.

L'hypothèse est que les usagers accèdent aux contenus pour les utiliser, et non pour les préserver, mais que cet usage est une source d'information riche pour la préservation. Or, utiliser un contenu, dans un contexte numérique signifie toujours le transformer (même pour le lire, surtout pour l'annoter ou le réutiliser). Ce que le numérique rend possible - voire impose - le couple homme-machine ne le gère pas forcément. Ainsi nous copions, référençons, commentons à loisir, mais ces transformations ne sont que rarement exploitables a posteriori à des fins de préservation : l'on perd les liens entre les copies, l'on ne sait plus quel fragments est le dérivé de l'autre, etc. Ce qui nous incite à une gestion de l'instantané (le Web et ce qu'il me propose dans l'état où il me le propose maintenant) plutôt qu'à une gestion documentaire rigoureuse (tel contenu dans cette version, relié à tel autre contenu dans telle version, etc.). Notre proposition est donc de proposer aux usagers un dispositif éditorial reposant sur deux facettes :

- Il est l'instrument de l'usage au quotidien : il permet d'écrire, lire, annoter, commenter, compléter, fragmenter, réutiliser, augmenter, enrichir, etc.
- Il est l'instrument de la préservation dans le temps : il enregistre les actes des usagers et maintient la généalogie des contenus (fonction philologique) et l'actualisation des contenus (fonction herméneutique et d'accessibilité physique.)

L'approche présentée ici s'appuie sur plusieurs éléments :

- La chaîne éditoriale Scenari, chaîne permettant de créer des contenus multimédia structurés selon une modélisation documentaire donnée.
- La norme OAIS (Open Archive Information System) qui propose une organisation des informations et des processus pour gérer sur le temps long l'évolution d'une archive, c'est-à-dire de documents dont on veut conserver l'intelligibilité.
- L'approche Cyclops, élaborée pour la gestion patrimoniale de contenus artistiques et musicaux. Cette approche, fondée sur la préservation par l'accès, propose aux détenteurs de contenus de les déclarer et définir

5 Standards techniques et méthodes

5.1 La chaîne éditoriale Scenari

Scenari est un outil intégré de conception et de déploiement de chaînes éditoriales XML [1]. Il se compose de SCENARIBuilder et de SCENARICHAIN. SCENARIBuilder est un outil permettant de créer des *modèles*

documentaires, c'est à dire des règles et programmes fixant la structure d'une famille de documents (schéma XML, règles de validation, etc.) et son comportement (logique d'édition, formats de publication, etc.). SCENARIchain est l'environnement permettant d'exécuter les modèles créés avec SCENARIBuilder, il propose principalement un éditeur XML WYSIWYM¹, un gestionnaire d'intégrité et des interfaces de publication (HTML, Open Document, etc.).

Le principe de la chaîne éditoriale XML consiste à proposer un environnement permettant la création des contenus selon un langage XML qui représente la structure documentaire logique, plutôt - comme dans les outils bureautique - qu'un format de mise en forme. Dans un second temps l'outil permet de transformer ces contenus XML (dit canoniques) en des vues lisibles suivant des format comme HTML pour le Web ou Open Document pour l'impression. Cette approche modifie assez profondément le rapport au document numérique, en ouvrant en particulier les voies du polymorphisme et de la rééditorialisation *sans recopie*.

Le *polymorphisme* consiste en la possibilité technique de disposer d'une *source unique* (*single sourcing* en anglais) de contenu et de la transformer à volonté selon les supports et mises en formes désirés. Le polymorphisme est un possible technologique qui reste limité dans la pratique : en effet il est rare que l'on souhaite présenter exactement la même information sous deux supports différents pour deux usages différents. Une nouvelle publication implique généralement la sélection du contenu (telle partie en plus, telle partie en moins), sa réorganisation (telle partie avant telle autre), sa remise en contexte (introduction, conclusions, transitions), etc.

L'idée est alors de profiter du découpage logique du contenu formalisé selon un langage XML métier, pour appliquer des césures physiques (découpage de fichiers XML et utilisation de liens par référence). Il devient alors possible de partager de mêmes fragments documentaires entre plusieurs documents, ce qui permet la réutilisation sans recopie. On appelle *ré-éditorialisation* (le terme anglais de *repurposing* étant encore plus adéquat) la remise en contexte de fragments issus d'un fonds documentaire, par leur ré-agencement au sein d'un nouveau document, leur augmentation par une création de contenus spécifiques et leur publication sur un nouveau support et/ou pour un nouveau public.

La chaîne éditoriale XML est donc un outil pour l'usage (il permet de créer, modifier, réutiliser, publier) et pour la préservation (il assure l'accessibilité technique par le langage XML technologiquement indépendant, il permet la réutilisation sans recopie, il facilite la mémorisation des différentiels entre versions, etc.).

5.2 Open Archive Information System

OAIS est une norme internationale issue du monde de l'astronomie qui a vocation à gérer les archives, en particulier numérique. Signifiant *Open Archive Information System*, OAIS n'a pas pour vocation de décrire les opération technique de l'archivage ni de prescrire des processus particuliers, mais de fixer un cadre, une terminologie et un référentiel pour la gestion des archives. Elle propose un modèle d'information et un modèle de processus. Le modèle d'information repose, entre autres choses, sur l'idée fondamentale qu'un objet numérique n'est lisible que si l'on possède une information sur la manière de le lire. OAIS appelle cette information la *representation information*. Par exemple, une suite de caractères aura pour information de représentation le fait qu'elle est exprimée en ASCII. Cette information a elle-même besoin d'être exprimée de manière numérique, et donc, en tant qu'information numérique, elle possède donc sa propre information de représentation. Par ailleurs, le modèle de processus prescrit que l'archive, comme institution, doit en permanence surveiller l'environnement pour détecter quand un écart, un fossé, se creuse entre ce que l'archive contient et ce que son environnement peut comprendre. En effet, une archive n'est pas gardée pour elle-même, mais pour une communauté désignée qui pratique la connaissance et les savoirs lui permettant de se saisir de l'archive. Quand un écart s'installe entre l'information numérique, son information de représentation et la communauté désignée, c'est qu'il faut enrichir l'archive de nouvelles informations de représentation.

A la base de nombreux projets de préservation, OAIS est une norme étonnement ouverte qui permet d'articuler l'usage de l'archive à sa conservation. Cependant, si elle intègre bien le fait de devoir surveiller l'exploitation par la communauté désignée et sa capacité de compréhension, elle ne prévoit pas d'articulation particulière pour intégrer de manière régulière les connaissances produites par cette communauté dans son utilisation de l'archive, faisant naturellement évoluer l'archive par son usage même. OAIS en reste encore à une vision centralisée, où l'archiviste contrôle et surveille l'usage pour faire évoluer le contenu, au lieu de suivre l'évolution du contenu par l'usage. C'est pourquoi nous proposons une méthodologie pour l'archivage reposant sur la préservation par l'accès, l'enjeu étant de pouvoir enrichir l'archive et la connaissance par le produit de son usage, conservant ainsi son actualité et son intelligibilité. Cette approche a été étudiée dans les arts médiatiques (voir section suivante).

¹What you see is what you mean.

6 Terrains

6.1 Contenus artistiques : préservation des arts médiatiques

Les arts médiatiques proposent un cadre particulièrement stimulant pour la capitalisation et la gestion des connaissances dans la mesure où les objets conservés sont soumis à une double obsolescence technologique et culturelle. Exploitant des dispositifs hétérogènes, faisant appel à différents dispositifs et techniques, les oeuvres médiatiques se conservent difficilement, le maintien de leur identité interdisant la possibilité de les rejouer ou de les réactiver et donc d'exprimer leur intelligibilité. De nombreuses initiatives préconisent de revenir aux propriétés significatives, ou les invariants de l'oeuvre, dont on conserverait la description pour la reproduire plus tard en respectant ces descriptions. Autrement dit, on n'interdit pas l'usage au profit de l'identité, mais on accepte de réinventer l'oeuvre dans sa totalité pour lui permettre de rester vivante, au prix de transformations à qui l'on impose de respecter certains invariants.

6.2 Expertises d'entreprises : capitalisation des expériences

L'UTC et la Caisse régionale de Crédit Agricole Brie Picardie collaborent dans le cadre du projet de recherche CAP-XP visant à créer une plate-forme d'échange de connaissances métier et d'expériences pour le réseau des commerciaux de la banque privée.

Le projet a mis à disposition d'un réseau d'une trentaine de commerciaux et directeurs d'agence un système éditorial leur permettant de consigner des comptes rendus d'expérience (une action de vente réussie, une action de conseil, etc.) et des éclaircissements techniques directement liés à leur métier (modification de loi, tendance économique, etc.). Le système permet (notamment) :

- l'édition XML des contenus en ligne (des contenus de type « histoire » et « note technique »)
- la publication pour la consultation Web directe, sous forme de fiches imprimables et sous forme de diaporamas
- la constitution de dossier regroupant des histoires ou notes techniques
- la « promotion » des histoires et notes techniques en (respectivement) cas et fiches techniques. Cette promotion signifie que les contenus sont copiés dans une nouvelle version qui va pouvoir être enrichie par la communauté, avec pour objectif d'obtenir des versions officielles finalisées (validées par la communauté d'utilisateurs)

Ce dispositif rend possible deux axes perpendiculaires de sédimentation d'usage pour passer de la captation à la capitalisation : un axe vertical, le *raffinement* progressif et un axe horizontal, la réutilisation.

À la création du contenu, celui-ci est plutôt brut (selon le temps que le contributeur aura pu passer à l'élaborer) et plutôt personnel (selon la prise de recul du contributeur). Au fur et à mesure des usages de ce contenu par les autres contributeurs (commentaires, corrections, co-élaboration lorsque les contenus sont promus au niveau de cas ou fiches) le contenu est amélioré pour devenir un contenu d'intérêt général au niveau de la communauté (tout les contenus n'auront bien entendu pas cet avenir).

Sur un axe complémentaires, et quelque soit le niveau de leur raffinement, les contenus peuvent être mobilisés au sein de différents dossiers, pour l'usage quotidien des membres de la communauté : fiche pour soi en vue d'un rendez-vous, dossier technique pour un client, diaporama de présentation pour les collaborateurs de l'agence, etc. Ce second axe est à la fois le moteur du premier (on améliore car on a un besoin) et son consommateur (on peut utiliser les contenus car ils sont de qualité). Le système a été installé en juin 2009 et donnera ses premiers résultats mesurables en septembre 2009, le colloque sera l'occasion de les présenter, et en particulier de valider l'hypothèse du couplage amélioration-utilisation.

7 Conclusion

Les approches proposées sont complémentaires, l'une apportant la méthodologie patrimoniale, l'autre l'approche documentaire, la dernière un environnement normatif. L'objectif consiste alors à proposer un modèle documentaire permettant l'édition structurée des différentes interprétations d'un contenu, intégrées avec sa version courante ainsi que ses versions antérieures. Dans tous les cas, il faut documenter tant la forme que le fond et construire une perspective temporelle permettant au lecteur d'accéder au contenu (la dernière version) et

DRAFT

d'adopter s'il le désire une lecture critique qu'il peut approfondir aussi loin qu'il le veut grâce au modèle documentaire proposé.

Bibliographie

- [1] Bachimont, B. (2007). Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents. — Hermès, Paris.
- [2] CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (2002) Reference Model for an Open Archival Information System (OAIS). Washington.
- [3] S. Crozat, Scenari la chaîne éditoriale libre, Eyrolles, Accès Libre, 2007