

Chapitre 1

Archivage audiovisuel et numérique: les enjeux de la longue durée

L'archivage connaît depuis peu un regain d'intérêt remarquable alors que cette question ne déborde traditionnellement pas des communautés d'experts et spécialistes où elle est traitée. Mais la mutation numérique généralisée que nous vivons permet de poser dans le temps court de la gestion des collections des questions habituellement abordées dans le temps long de la mémoire et de ses archives. Alors que les producteurs et consommateurs de contenus partageaient jusqu'ici un même cadre technologique reportant au delà de leurs préoccupations les problèmes relevant de la préservation et de la transmission de leurs contenus, la succession débridée des formats et outils numériques dans un rythme sans cesse croissant amène les usagers du numérique à envisager non seulement la communication des documents dans l'espace de l'échange, lié à leur activité sociale et économique, mais aussi la pérennisation dans le temps de la transmission de ces mêmes contenus. La question temporelle entre dans la quotidienneté de l'utilisateur du numérique, qu'il soit amateur, spécialisé ou professionnel : que faire de mes anciennes vidéos ? Quel nouveau format adopter ? Peut-on faire confiance aux nouveaux supports ? Comment être sûr de ne rien perdre ?

Cette nouvelle préoccupation découle pour une large part des conditions techniques mises en place par la mutation numérique. Mais les difficultés rencontrées, les réponses envisagées, les concepts mobilisés, n'en sont pas pour autant toujours inédits. En effet, le rapport à la mémoire qu'instaure le numérique, s'il fait rupture avec la tradition récente de l'imprimé et de ses collections, n'en appelle pas moins à des continuités sur le temps beaucoup plus long de l'histoire de l'écriture et de la mémoire humaine. Si l'imprimé nous a habitué un temps avec des

contenus dont l'origine, attestée par les pratiques éditoriales, n'appelle que la conservation à l'identique, dans son intégrité physique, du contenu, le numérique renoue avec des difficultés familières aux codicologues et philologues des manuscrits antiques et médiévaux, tant par l'origine incertaine des contenus que leur variabilité que la mutabilité du numérique leur permet d'avoir. Confronté à des contenus sans cesse variables, toujours transformés par rapport à un original qu'on ne peut alors que supputer, l'usager du numérique peut s'inspirer tant des *scriptoria* monastiques que des premières imprimeries-librairies de la Renaissance. Inspiration n'est pas raison, il faudra faire la part entre la métaphore et le modèle : le numérique ne trouvera pas son intelligibilité dans une projection, trop hâtive, des modèles issus des traditions manuscrites, médiévales ou antiques. Mais la compréhension des problèmes de jadis devraient permettre d'établir des voies de réflexion que l'on pensait révolues dans l'extase béate dans laquelle l'exception de l'imprimé nous avait plongés.

Le problème des contenus numériques

Les contenus numériques, par beaucoup d'aspects, posent des difficultés que l'on avait déjà rencontrées avec des médiums¹ précédents. En particulier, les médiums audiovisuels, et de manière générale les médiums que nous aimerions appeler ici « médiums technologiques ». Précisons cette notion.

Les médiums technologiques

Les dimensions documentaires

Tout document est inscrit sur un support. L'articulation du contenu au support sur lequel il est inscrit varie en fonction de la technologie mise en œuvre pour l'élaboration du support et le procédé d'inscription utilisé. Cette articulation détermine le type du document et ses propriétés. Plusieurs dimensions permettent de décrire cette articulation [BAC 07] : tout d'abord un dimension de préservation ou d'enregistrement, destinée à donner la persistance et la pérennité au contenu ; en suite une dimension des restitution ou d'appropriation, destinée à permettre l'accès à

¹ La langue française nous gratifie d'une complexité insoupçonnée avec le terme de « médium » dont le pluriel n'a pas la même signification que le singulier. Si le médium peut signifier le support physique, le véhicule matériel, d'une transmission ou d'une inscription, les médias désignent le moyen de diffusion de masse élaboré à partir des médiums. Nous réserverons le terme de médium et son pluriel « médiums » pour qualifier les supports physiques, et le terme de média et son pluriel « médias » pour qualifier les moyens économiquement et industriellement organisés de diffusion, comme la télévision ou la presse.

l'information conservée. Enfin, un dispositif de restitution ou encore de publication permet passer de l'enregistrement persistant à une vue restituée ou publiée :

- Dimension de préservation

- Le support d'enregistrement du document

Un document est l'inscription matérielle d'un contenu. Le support d'enregistrement est par conséquent le support d'inscription, le support que l'on charge de préserver et conserver le contenu inscrit. Par exemple, un livre a pour support d'enregistrement le papier, un document audiovisuel le support d'une bande magnétique vidéo, ou le support d'un film argentique, le document numérique enfin la mémoire adressable d'un support informatique.

- La forme d'enregistrement

Le support d'inscription permet de consigner le contenu dans une certaine forme. La forme d'enregistrement est la forme sous laquelle le contenu est inscrit sur le support d'enregistrement. Pour un livre, c'est la typographie d'un répertoire alphabétique, pour un document audiovisuel, le signal magnétique sur le support vidéo, ou enfin le code numérique binaire des documents numériques.

- Dimension de restitution

- Le support d'appropriation du document

Le document ne préserve et ne conserve un contenu que pour le rendre accessible et public, puisqu'il résulte d'un acte de publication. Un document doit par conséquent pouvoir être lu, consulté, visualisé sur un support le permettant. Ce support est celui où un utilisateur peut faire sien le contenu, se l'approprier. Un support d'appropriation est l'écran de télévision, le papier du livre, un haut-parleur, etc.

- La forme physique d'appropriation du document

Le support physique présente le contenu. Pour cela, ce dernier doit être présenté dans une forme physique compatible avec le support physique, pour que le contenu puisse être appréhendé par l'utilisateur. Par exemple, à partir du signal magnétique (forme d'enregistrement) de la cassette vidéo (support d'enregistrement), la télévision (support d'appropriation) reconstruit un signal visuel (les points de couleurs de l'écran) regardé par le spectateur.

- La forme sémiotique d'appropriation du document

La représentation affichée sur le support d'appropriation respecte une structure ou une forme telle qu'elle est directement perceptible par l'utilisateur. Cette forme est donc la forme d'appropriation permettant à un utilisateur de s'appropriier le contenu. La forme d'appropriation correspond à une forme directement interprétable, dans la mesure où elle appartient à un registre sémiotique dont l'utilisateur a déjà fait l'apprentissage culturel ou scolaire. Lorsqu'un document mobilise plusieurs formes sémiotiques d'appropriation, on dira que le document est multimédia. Par exemple, l'audiovisuel est multimédia car il mobilise l'image, la musique, le bruit et la parole. L'image peut elle-même être multimédia si elle comporte des textes et des structures iconiques.

- La modalité d'appropriation

La forme sémiotique d'appropriation s'adresse à une ou des modalités perceptives. Quand il y a plusieurs modalités d'appropriation, le document est multimodal. Par exemple, l'audiovisuel est multimodal car il s'adresse à la vue et à l'ouïe.

- Le dispositif de restitution - publication

Le dispositif de restitution permet de passer du contenu enregistré au contenu publié et ainsi rendu accessible et perceptible pour sa lecture. Il donne lieu à des vues publiées, vues donnant le contenu perceptible prêt et idoine pour l'interprétation.

Ces dimensions permettent de distinguer les « médiums perceptibles » et les « médiums technologiques » :

- Les médiums perceptibles ont leurs support d'enregistrement et support d'appropriation confondus : le support sur lequel on lit est celui que l'on range et stocke ; de même, leurs formes d'enregistrement et d'appropriation sont confondues : ce que l'on lit est bien ce que l'on a inscrit sur le support. Cela signifie que l'on peut accéder au contenu indépendamment de toute médiation technique ou mécanique. Le contenu est ainsi inhérent au support et, si on prend soin du support, alors le contenu prendra soin de lui-même dans la mesure où ses conditions de conservation sont celles de son support.
- Les médiums technologiques ont un support d'enregistrement distinct et séparé du support de restitution. Ce que l'on conserve n'est pas ce que

l'on consulte, car il est nécessaire de recourir à une médiation technique et mécanique pour accéder au contenu. Ainsi, il ne suffit pas de conserver le support d'enregistrement pour accéder au contenu, il faut également conserver les conditions d'accès technique à l'information enregistrée pour la restituer dans une forme immédiatement consultable et interprétable.

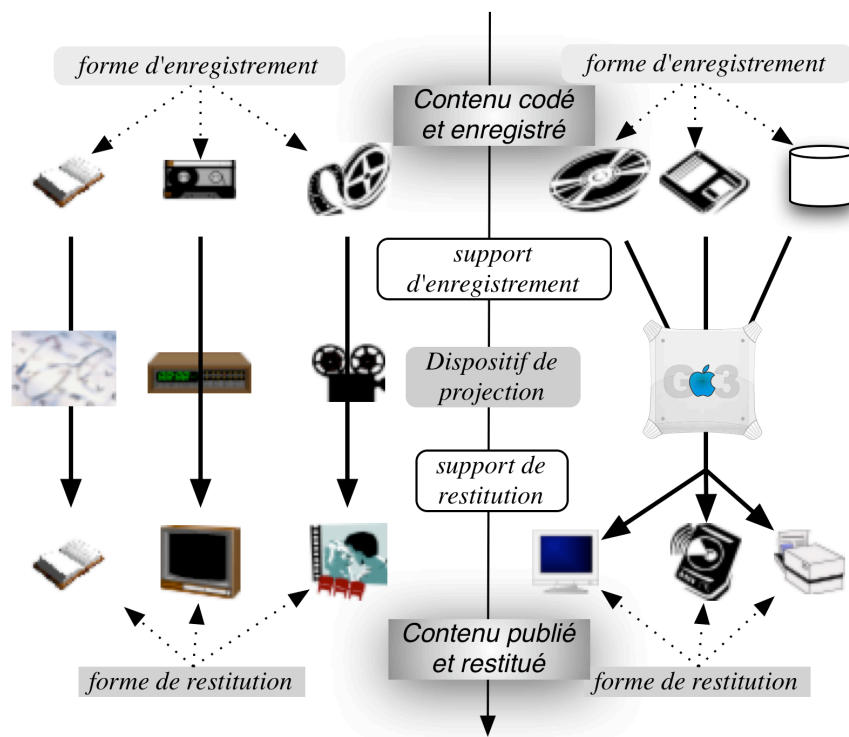


Figure 1. Les dimensions documentaires

Lisibilités technique et culturelle, fossés d'intelligibilité et d'obsolescence

La préservation documentaire repose ainsi sur la préservation de deux possibilités fondamentales d'accès au contenu (cf. figure suivante) :

- La *lisibilité technique*, qui permet d'accéder à l'information enregistrée. Elle repose sur la nécessité de conserver la connaissance technique et l'instrumentation associée pour exploiter l'information conservée. Elle se heurte au *fossé d'obsolescence* selon lequel les

techniques d'accès au codage conservé sortent progressivement de l'état de l'art courant des techniques.

- La *lisibilité culturelle*, qui permet d'accéder à l'information restituée. Cette information est immédiatement perceptible et interprétable dans la mesure où elle n'a plus besoin de médiation technique pour être consultée. Mais elle repose sur une médiation culturelle, nécessitant la plupart du temps apprentissage et éducation, pour interpréter, comprendre et exploiter le contenu consulté. La lisibilité culturelle se heurte au *fossé d'intelligibilité* selon lequel les codes sémiotiques et culturels d'un contenu deviennent de plus en plus étrangers à la culture du moment si bien que le contenu, tout perceptible qu'il soit, devient inintelligible puisqu'on ne comprend pas ce qu'il signifie, à quelle question il renvoie, à quelle interprétation il se prête.

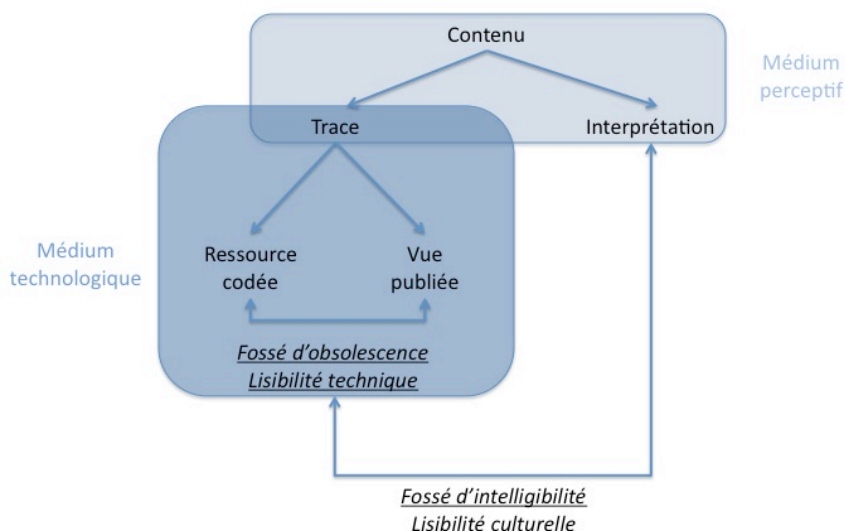


Figure 2. Les médiums perceptibles doivent surmonter le fossé d'intelligibilité et assurer la lisibilité culturelle. Les médiums technologiques doivent en plus surmonter le fossé d'obsolescence et assurer la lisibilité technique.

La complexification au cours du temps des supports documentaires a pour conséquence de rendre de plus en plus ardue la tâche de préservation de la lisibilité

technique et culturelle. En effet, pour les médiums perceptibles, il suffit de s'assurer de la lisibilité culturelle. Ainsi faut-il avoir des spécialistes de l'interprétation du contenu, qui peuvent pour une large part d'entre eux rester ignorants des propriétés techniques du support d'enregistrement et de restitution. Beaucoup de spécialistes du latin médiéval ne sont pas pour autant des codicologues s'intéressant aux aspects matériels des livres, aux conditions de leur fabrication et de la préservation physique de ces derniers. Pour les médiums technologiques, il faut en outre conserver la lisibilité technique. Mais cette dernière peut être redoutablement complexe, car on n'est plus dès lors dans la conservation, mais dans la reconstruction, voire dans la réinvention du contenu à partir de ce qui a persisté. On peut mieux le réaliser quand on envisage le passage de l'audiovisuel au numérique.

L'audiovisuel a en effet eu pour particularité d'introduire une séparation entre le support d'enregistrement et le support de restitution : la forme de restitution étant temporelle, le support d'enregistrement étant nécessairement spatial, il est impossible de les associer directement. Il faut avoir une forme d'enregistrement qui soit spatiale pour être associée au support d'enregistrement et ainsi participer à sa pérennité, et avoir une forme de restitution temporelle qui soit reconstruite par l'enregistrement. Le dispositif technique de lecture permet de passer de *l'organisation spatiale* de l'information enregistrement au *déroulement temporel* du contenu. On en arrive ainsi à l'essence même de l'enregistrement, qui consiste à coder spatialement un contenu pour le reproduire ensuite en une séquence temporelle. Les médiums perceptibles font l'économie d'un codage technique car l'apprentissage culturel de la lecture permet de passer du codage spatial de l'écriture au déroulement temporel de la lecture. Pour les médiums technologiques, la forme de codage ne se prête pas à une lecture humaine mais doit en passer par une reconstruction d'une forme lisible. Cette forme lisible est temporelle dans le cas de l'audiovisuel, qui doit restituer une forme animée dans le temps. Mais elle peut être spatiale quand elle restitue une forme directement lisible via une médiation ou lecture culturelle qui temporalise le contenu (par exemple, pour une image ou un texte numériques). Dans tous les cas, la lecture est temporelle, mais sa temporalité peut être librement reconstruite par la lecture du contenu décodé, ou imposée par le décodage comme dans le cas de l'audiovisuel.

De la consultation à la ré-invention du contenu conservé

Mais la question qui vient est de déterminer la relation existant entre le contenu décodé et le code conservé. Quelle fidélité peut-on établir ? Comment même définir une telle notion ? Dans la plupart des cas, le dispositif de lecture est uniquement conçu pour décoder le contenu enregistré et restituer une forme perceptible. Par conséquent, le contenu restitué est la conséquence directe, causale et donc fidèle du contenu enregistré. Trivialement dit, la seule chose que sait faire mon magnétoscope est de lire le contenu de mes cassettes VHS pour les afficher sur l'écran de ma

télévision. Par conséquent, ce que je vois est par définition ce qui est sur la cassette, et donc je le crois. Mais qu'en est-il quand la reconstruction repose sur un programme qui peut utiliser non seulement l'information enregistrée et conservée, mais également bien d'autres informations ? Quand la médiation technique n'est plus dédiée au type de codage permettant la conservation de l'information mais devient calculatoire et universelle comme l'est tout ordinateur, quel rapport établir entre ce que l'on consulte et ce qui est conservé ?

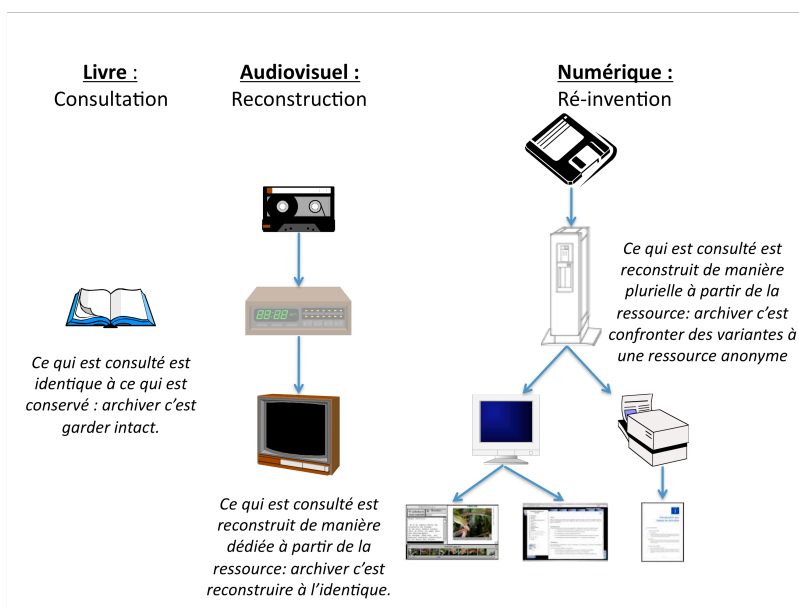


Figure 3. La déconstruction du document : avec le livre, ce qui est lu est ce qui est conservé ; dans l'audiovisuel, ce qui est vu est ce qui est reconstruit par un mécanisme dédié au codage de l'enregistrement ; dans le numérique, ce qui est consulté est ré-inventé par les possibilités calculatoires de l'outil de lecture.

Comme le montre le schéma ci-contre, on a bien une dématérialisation progressive et croissante du document qui transforme la consultation en une reconstruction (e.g. audiovisuel analogique) pour devenir à présent une ré-invention avec le numérique.

Numérique et ré-invention

La question qui se pose est l'importance qu'il faut donner à cette ré-invention et si malgré tout on ne peut, comme pour l'audiovisuel, reposer sur des contraintes propres à la ressource pour contrôler les publications qui seront effectuées. Le

numérique a ceci de particulier que les reconstructions peuvent être parfaitement arbitraire par rapport au contenu qui a été enregistré. Contrairement à un magnétoscope qui interprète toujours de la même manière une cassette VHS, un programme peut reconstruire le contenu en fonction de paramètres qui dépendent autant de son environnement que de la ressource codée. Pour bien le réaliser, il faut se souvenir des propriétés du numérique, que nous voudrions qualifier ici de la manière suivante :

- Le numérique entre dé-sémantisation et interprétation.

Cela signifie que le numérique est un pur code sans signification ; il est indépendant de la manière dont il sera implémenté matériellement d'une part, et interprété sémantiquement d'autre part (voir figure suivante). Ainsi un même train binaire peut être lu comme un son ou comme une vidéo. Toutes implémentations ou interprétations seront donc libres et arbitraires par rapport à la ressource binaire qui n'introduit donc aucune contrainte quant à sa mobilisation.

- Le numérique entre fragmentation et recombinaison.

Cela signifie qu'un contenu numérique est rapporté à des unités vides de sens qui le fragmentent en autant d'entités arbitrairement manipulables par des programmes. La manipulation n'ayant pour seules limites celles du calcul, la transformation d'une ressource n'est pas contrainte par la nature de la ressource elle-même ni par ce qu'elle représente.

Ces deux propriétés font système : parce que le numérique dé-sémantise, les unités obtenues par la discrétisation sont librement manipulables. De même, c'est le fait de discrétiser en symboles binaires reliés au contenu d'origine au moyen d'un code conventionnel qui assure le caractère non sémantique de la ressource.

On comprend dès lors que, en ce qui concerne les documents, la numérisation a pour conséquence de permettre des reconstructions arbitraires à partir d'une ressource elle-même anonyme et a-sémantique car binaire. On est donc confronté à des vues publiées que seule notre connaissance des conditions globales de la numérisation ou production numérique, de la publication à partir de la ressource, permet de comparer et articuler, critiquer et hiérarchiser sans que l'on puisse s'appuyer tant sur la ressource que sur les vues elles-mêmes.

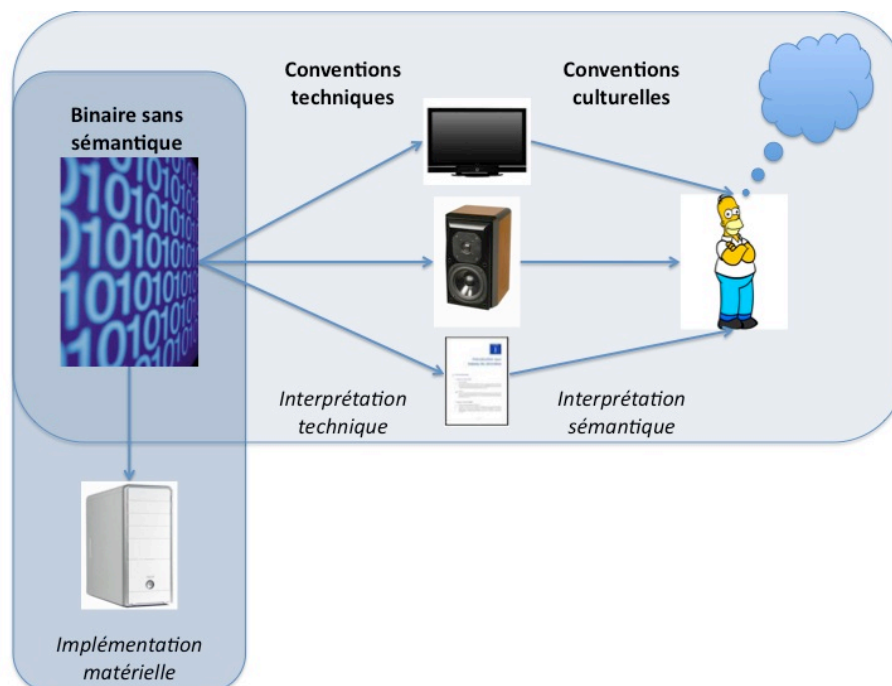


Figure 4. Le code binaire ne possède aucune sémantique par lui-même et la reçoit par des conventions extérieures : il peut devenir image, son ou texte selon les codes adoptés.

Le numérique : un nouveau régime de la mémoire

Le dilemme archival

Le numérique introduit par conséquent un nouveau rapport au document et à sa conservation. Car la conservation n'est plus seulement le fait de garantir l'intégrité et l'authenticité d'une trace matérielle, mais aussi le fait de permettre son exploitation technique dans le but de reconstruire le contenu. Or, préservation et reconstruction ne sont pas toujours compatibles, notamment quand le fait de garantir la possibilité de la reconstruction entraîne la transformation du codage enregistré et conservé. C'est le dilemme archival du numérique :

- D'une part, il faut garantir l'accessibilité technique du codage enregistré et donc le faire évoluer pour le garder compatible avec l'environnement technologique courant. Mais dans ce cas, on abandonne l'identité à soi physique du codage puisqu'il doit évoluer, et on compromet l'intégrité

et l'authenticité puisqu'il faut garantir que toute évolution reste bien fidèle au contenu original.

- D'autre part, il faut conserver le contenu dans son intégrité et authenticité. Intégrité signifie que le contenu reste identique à lui-même au cours du temps : il conserve tous ses parties et attributs physiques et matériels permettant de le constituer et de le déclarer complet et conforme à sa définition et identité. Authenticité est le fait que le codage est bien ce qu'il prétend être et qu'il est conforme au contenu dont il est le codage. Assurer l'intégrité et l'authenticité implique une conservation qui n'altère ni ne transforme le contenu codé. Mais cela peut compromettre l'accessibilité technique puisque le contenu n'est plus obligatoirement conforme aux techniques d'accès du moment et peut être frappé d'obsolescence technologique outre la corruption physique inhérente à tout support matériel.

Le dilemme archival est que l'archiviste des mediums numériques est forcément le compromis et le renoncement à l'une de ses tâches : s'il assure la meilleure préservation possible au sens de préserver l'identité à soi du contenu, il compromet l'accessibilité ; s'il garantit une accessibilité permanente, il compromet la préservation du contenu.

L'archiviste de l'audiovisuel et du numérique doit donc construire un cadre théorique et méthodologique permettant de résoudre ce dilemme et d'assurer la préservation de contenu qui par nature sont à ré-inventer et pas seulement à conserver.

L'enjeu du numérique : la mutabilité et la variance

Le numérique introduit en effet une modification profonde dans notre rapport au contenu dans la mesure où il met clairement en relief le caractère provisoire et temporaire de l'identité d'un contenu et de sa définition. Le numérique nous amène à renouer avec une économie de la variante où un contenu n'existe qu'à travers des exemplaires qui donnent des versions différentes mais voisines sans qu'aucune ne soit la référence canonique permettant de juger les autres. Seules les variations entre ces versions permettent l'élaboration d'une critique du contenu et l'établissement d'une identité du contenu, identité putative et non héritée, proposée et non trouvée, critiquable et non dogmatique, argumentée et non démontrée, provisoire et temporaire et non définitive.

Cette situation n'est pas si nouvelle qu'on ne puisse lui trouver des antécédents. Au contraire, le numérique nous entraîne dans une proximité improbable avec la

culture des manuscrits, faisant de l'ère de l'imprimé une parenthèse exceptionnelle où, par delà la multiplicité des variantes, des versions canoniques s'imposaient comme référence pour la critique et la référence pour les autres. Mais le caractère conventionnel de ces versions canoniques, comme on va le voir dans la suite, permet de dégager une communauté de problème sous-jacente avec l'ère du manuscrit hier ou l'ère du numérique aujourd'hui, communauté de problème car en fait inhérente à la notion même de contenu.

Contenus et mutations

Revenons à la notion de contenu et tâchons de préciser ce que le numérique implique quant à sa définition. Prenons les points de repères suivants :

- Un *contenu* se définit comme une forme d'expression s'incarnant sur un support matériel. Par forme d'expression il faut comprendre toute forme émanant d'une intention de communication et pouvant être reconnue comme telle. Par support matériel, il faut comprendre le véhicule physique permettant de transmettre le contenu. Ainsi la parole repose sur l'onde sonore comme véhicule matériel, onde que l'on peut physiquement enregistrer par exemple, et une forme d'expression qui est la structure du véhicule matériel sonore qui permet de distinguer cette parole du bruit et de la comprendre et interpréter.
- Une inscription est un contenu fixé sur un support matériel. Alors que le contenu peut être évanescent, comme par exemple la parole, l'inscription est pérenne et profite de la permanence d'un support matériel pour acquérir une persistance temporelle. Tout support matériel ne peut fixer un contenu, seuls les supports spatiaux et statiques permettent de le faire. C'est en ce sens que tout support d'enregistrement est nécessairement spatial.
- Enfin, un *document* est une inscription prise dans un contexte documentaire de production et de réception. Un document est donc une inscription instituée comme document, dont on reconnaît le statut intentionnel et qui est interprété et compris en fonction d'autres documents associés à sa production, transmission et réception.

La question qui se pose à présent est de déterminer l'identité d'un document. On s'aperçoit, après une réflexion rapide, que la notion n'est pas si simple. Prenons deux exemples pour fixer les idées. Si je considère *Madame Bovary* de Flaubert, je considère qu'il s'agit de la même œuvre qui est matérialisée par mon exemplaire de la Pléiade ou dans la collection *Librio*. En effet, je retrouve le même texte, le même chapitrage, et la même structuration en paragraphe. Mais je considère comme non

pertinentes les différences de polices de caractères entre une collection et une autre par exemple, ainsi que la pagination. Par ailleurs, si je considère un recueil de poèmes comme *Alcools* d'Appolinaire, la pagination est importante, la taille des lignes ainsi que la forme graphique en général des poèmes. On voit donc que les critères d'identité de contenu d'une expression matérielle à une autre sont variables, et que les invariants, ce qui doit être préservé d'une expression à une autre pour qu'on puisse dire qu'elles expriment la même chose, sont variables en fonction du type de contenu. Aussi peut-on distinguer plusieurs niveaux dans l'identité d'un contenu :

- L'identité numérique d'un contenu : il s'agit d'un objet matériel particulier, un exemplaire donné, voire unique.
- L'identité matérielle d'un contenu : il s'agit de la forme matérielle d'un contenu que l'on peut retrouver de manière identique d'un contenu à un autre, comme entre les exemplaires d'un même ouvrage dans une même collection (production de série).
- L'identité expressive d'un contenu : il s'agit du contenu tel qu'il est reconnu le même indépendamment des différentes matérielles que l'on peut constater. C'est *Madame Bovary* reproduit dans plusieurs collections différentes.
- L'identité sémantique d'un contenu : il s'agit du sens du contenu. Résultant de son interprétation, l'identité sémantique du contenu résulte de sa confrontation avec d'autres documents dans le cadre d'un contexte interprétatif. C'est à ce niveau qu'on dira que l'ennui dans *Madame Bovary* a telle ou telle signification, et s'exprime de telle ou telle manière.

La difficulté est de savoir à quel niveau doit s'exercer la préservation. Quand on a affaire à des mediums perceptibles, on sait que la préservation d'un exemplaire matériel permet de conserver l'accès au contenu. Aussi définit-on souvent la préservation comme la conservation de l'identité numérique de contenu, dont on maintient l'intégrité physique, l'identité à soi du contenu et la connaissance de son origine permettant d'établir son authenticité.

Quand on a affaire à des mediums technologiques, la préservation ne peut concerner l'identité numérique des contenus. Elle doit concerner l'identité expressive du contenu, c'est-à-dire la capacité de reproduire ou ré-inventer le contenu en préservant ses invariants constitutifs de son identité. Quand une tradition culturelle permet de dégager de tels invariants, n'importe quelle approche peut être utilisée dès l'instant que ces invariants sont respectés. Mais de

telles traditions peuvent manquer et souvent il n'existe pas de consensus pour préserver l'identité expressive. On se reporte alors sur l'identité matérielle : on reproduit pour retrouver le « même » contenu. Par exemple, le changement de format pour un contenu audiovisuel doit respecter l'identité expressive, pour que l'on voie la même chose. C'est dire que les débats actuels sur la haute définition ouvrent de nouvelles et difficiles questions pour l'archiviste.

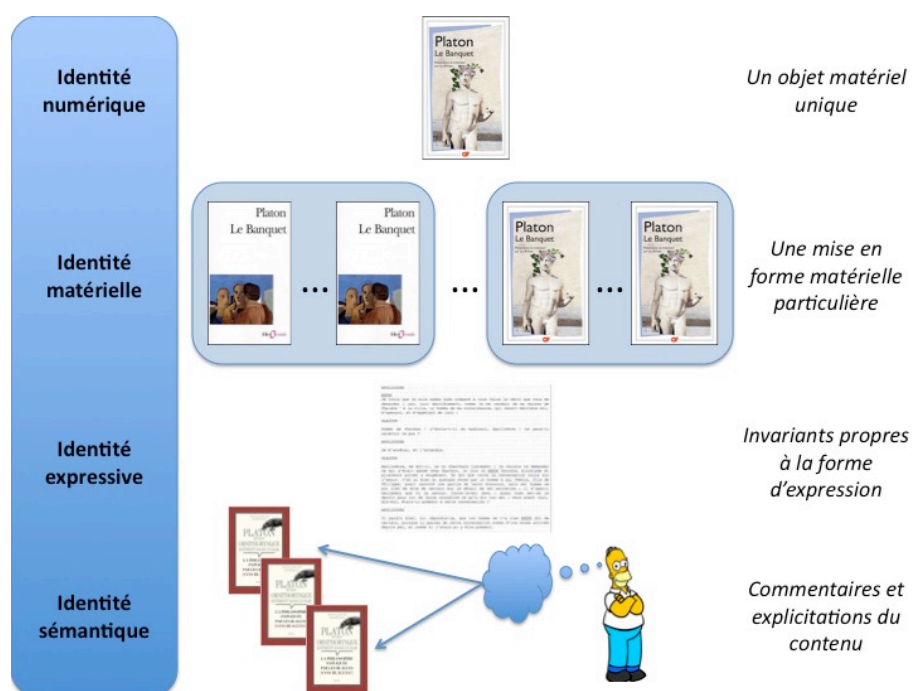


Figure 5. Un document possède plusieurs identités selon le point de vue adopté : il est unique ou interchangeable selon le niveau d'identité choisi.

Avec le numérique, la situation devient encore plus complexe :

- Le numérique consiste en ressources que l'on peut exploiter de multiples manières selon le dispositif de lecture que l'on emploie pour reconstruire le contenu, qui se présente alors de multiples manières. On a donc autant de variantes que de reconstructions possibles.
- La ressource numérique est en tant que telle inaccessible directement sinon par la médiation de la reconstruction : il n'est pas possible de

vérifier que la reconstruction est fidèle à la ressource, car il faudrait pour cela accéder à la ressource via une autre reconstruction.

On a donc des variantes dont l'origine, la ressource à la base de leur reconstruction est inaccessible. Il ne reste donc que l'examen des conditions de la reconstruction, la confrontation des expressions reconstruites entre elles, qui permettent l'établissement de l'intégrité et de l'authenticité des contenus. Il nous faut donc un modèle général permettant de comparer les contenus dont l'origine demeure inaccessible et qui permettent de confronter les variantes ensemble. De plus il faut que ce modèle reprenne à son compte les contraintes inhérentes aux médiums technologiques.

Des approches de la transmission

L'idée à la base de notre modèle conceptuel est que la préservation ne repose pas sur la conservation passive des contenus qu'il faudrait garder intacts, mais qu'elle repose au contraire sur la pratique active de l'interprétation et de l'exploitation. Autrement dit, on ne préserve bien que ce dont on se sert.

Cette idée est soutenue par 3 repères que l'on peut mobiliser ici : l'héritage de la culture antique, le patrimoine musicale dit « classique » et enfin la solution occidentale à la préservation de la culture savante.

L'héritage de la culture antique

On sait que les manuscrits les plus anciens de l'œuvre d'Aristote dont on dispose remonte au 9^e siècle après J.-C. Autrement dit, on ne connaît l'une des œuvres les plus importantes pour la constitution de la culture occidentale qu'à travers des contenus postérieurs de plus d'un millénaire après la disparition de leur auteur présumé. Deux questions se posent dès lors : comment se fait-il qu'on ait encore des versions de cette œuvre d'une part, et comment pouvons-nous y encore comprendre quelque chose d'autre part ? La réponse pour ces deux questions est la même : c'est qu'en fait, dans toute l'histoire de l'occident, on n'a quasiment pas arrêté de lire et commenter Aristote. Que ce soit à travers les copies qui en furent faites et les commentaires qui en furent élaborés, l'œuvre d'Aristote a toujours fait partie de l'horizon intellectuel de nos aïeux. Non que ces derniers aient voulu nous le léguer, mais bien plutôt qu'ils aient voulu le lire pour eux et se le rendre intelligible dans l'horizon de leurs préoccupations et questions. Bref, Aristote fut copié, ce qui explique qu'on en a des exemplaires nouveaux même 1000 ans après la mort d'Aristote. Par ailleurs, comme il fut toujours commenté, tout lecteur a toujours accédé au texte à travers la lecture et le commentaire de ses prédécesseurs. Encore

aujourd'hui, le commentaire critique (pour établir le texte de référence) ou érudit (pour proposer une interprétation) se construit en s'articulant à l'héritage de ces copies et études. Nous lisons à travers les commentaires de nos maîtres de l'Université, qui eux mêmes s'appuient sur la critique allemande du 19^e siècle, qui elle-même reprenait Saint Thomas, lui-même mobilisant Averroès, Alexandre d'Aphrodise, etc. C'est toute une chaîne de lecture et de lecteur dialoguant entre eux à travers les siècles qui permet la transmission et la préservation.

Ce que nous montre cet exemple, c'est que la préservation est la conséquence de l'interprétation et l'usage, et non l'inverse.

L'héritage musical

La musique présente une difficulté qui nous est désormais familière : expression temporelle et dynamique, elle ne peut s'inscrire directement car aucun support spatial ne permet de fixer un contenu temporel. Il faut un codage dont la spatialité, fixée sur un support, permet de programmer le déroulement temporel exécuté par un mécanisme reproduisant la forme temporelle. Par conséquent, tant que de tels mécanismes n'existent pas, il est impossible de conserver la musique. Comment alors la transmettre ? La solution tient en trois éléments :

- Des partitions qui décrivent non la musique elle-même mais la manière de l'exécuter et de la reproduire sur un instrument.
- Des instruments ou lutherie dont on garde le savoir faire pour leur fabrication :
- Des conservatoires où l'on entretient la pratique musicale, faisant la rencontre entre la partition et l'instrument.

La solution musicale est donc très originale : à défaut du contenu lui-même, on en garde un invariant de description, celle de son exécution, mais qui doit, pour fonctionner, être adossé à deux savoir faire complémentaires : la fabrication des instruments et la pratique instrumentale.

Là encore, la contenu ne se préserve qu'à travers l'entretien d'une double pratique qui dépasse l'impossibilité de conserver le contenu lui-même. La pratique de la mémoire est donc l'outil de sa préservation.

L'héritage culturel occidental

1537, François 1^{er} fonde la librairie royale par l'édit de Montpellier. Institution de mémoire permettant de conserver les œuvres du génie national, la librairie royale est créée dans le mouvement que celui qui présida à la fondation en 1530 du Collège des lecteurs royaux, le futur Collège de France. Derrière ces deux fondations, le même esprit humaniste, celui de Guillaume Budé (Ref) : la conservation des œuvres n'est rien sans l'entretien de la pratique de la lecture et de la transmission de cette pratique. De manière synthétique, notre approche occidentale de la mémoire a toujours consisté dans cette double complémentarité entre d'un côté la bibliothèque qui conserve les œuvres et l'université qui entretient la lecture et son sens (la recherche) et assure sa transmission (l'enseignement)².

Encore ici, on retrouve cette transmission par l'usage et l'interprétation, la conservation des objets n'étant que la conséquence et non la cause de la lecture.

La mémoire

Ces exemples suggèrent une conception de la mémoire bien particulière. La mémoire n'est pas un stock que l'on constitue et que l'on transmet, mais une capacité que l'on entretient. L'archive n'est donc pas la mémoire, elle n'en est au plus que la condition. Pour qu'il y ait mémoire, il faut passer de la puissance à l'acte, il faut qu'il y ait un exercice de mémoire, un exercice de la mémoire. La mémoire est donc la tension entre le présent et ses préoccupations et les ressources conservées ; cette tension se résout en individualisant un souvenir, en faisant sourdre un souvenir hors des ressources dans la forme suggérée par l'horizon du présent.

Cette conception de la mémoire, suggérée notamment par l'histoire de la culture et de la mémoire [CAR 00, CAR 02] s'extériorise littéralement dans notre conception de la mémoire numérique. En effet, la ressource numérique ne peut être la mémoire car, numérique, elle est inaccessible directement et nécessite la médiation d'une lecture technique. La ressource est donc bien une condition et non l'objet de la mémoire. Par ailleurs, on a des dispositifs de publication qui, à partir de la ressource, proposent des vues du contenu. Ces vues sont diverses et répondent à différentes préoccupations, que ce soit l'authenticité du contenu vis-à-vis des publications d'une certaine époque (rejouer les jeux vidéos d'arcade dans les conditions des années 70-80), ou l'adéquation vis-à-vis des enjeux de lecture de

² C'est souvent le même acte qui préside à la recherche et à son enseignement, les conditions d'exposition et de transmission étant le moteur et l'aiguillon de la recherche, sauf si l'on se fait une vision paresseuse de l'enseignement, ignorant sa force questionnante, et si l'on a une vision dogmatique de la recherche accumulant les faits et au lieu d'édifier les structures de la compréhension.

l'époque courante. L'enjeu est alors de savoir définir les règles permettant de distinguer, structurer et articuler les différentes vues reconstruites.

La conception de l'archive est donc très proche que ce que Paul Veyne a pu dire de l'histoire en la qualifiant de « roman vrai » [VEY 71]. La mémoire reconstruite par l'archive est bien une fiction, et il faut faire son deuil de l'authenticité simplement établie par une identité avec un état passé. Mais toute fiction n'est pas bonne pour une fonction de mémoire, et des critères de véracité doivent être établies. C'est l'enjeu d'une archivistique audiovisuelle et numérique, et des travaux scientifiques qui en dépendent ou en découlent.

Nouveaux champs de recherche

Ce sont donc de nouveaux champs de recherche qu'il faut aborder pour une archivistique qui prendrait au sérieux les médiums technologiques. Le problème peut en effet être défini de la manière suivante, au vu de nos réflexions antérieures :

- On a des ressources audiovisuelles ou numériques inaccessibles sans médiation technique qui ne sont plus par nature fiable ou fidèle, mais qui peuvent varier arbitrairement.
- Ces médiations techniques peuvent reconstruire autant de vues que nécessaires des ressources, donnant ainsi autant de variantes possibles de mêmes ressources. De plus, la ressource numérique en elle-même n'ayant pas de sémantique propre (cd. *supra*), la signification du contenu ne provient que de sa reconstruction technique et de son interprétation culturelle.
- La question est alors de définir comment comparer et articuler ces différentes versions, et d'en donner une interprétation :
 - Y-a-t-il une version de référence dont les autres seraient dérivées ?
 - Y-a-t-il des critères permettant de distinguer les vues fidèles des vues qui seraient des falsifications ou simplement fautives ?
 - Comment peut-on comparer différentes variantes : comment savoir que des publications différentes sont des variantes d'un même contenu ? Qui est la variante de qui, et de quoi ?

Prenons un exemple simple : un programme conçu pour être diffusé à la télévision, sur un site Web et pour des téléphones portables donnera lieu à différentes publications adaptées à chacun de ces trois supports. Comment les comparer ? Faut-il le faire ? Sont-ils les variantes d'un même programme idéal qu'il faudrait reconstruire abstraitement ? Sont-ce trois programmes indépendants ? De même, entre un programme diffusé sur ces chaînes de télévision, les extraits publiés sur des sites comme YouTube ou Dailymotion, qui est l'original de qui ? Dans le contexte actuel, on voit bien qu'il y a une réponse à cette question ; mais sera-ce toujours la même ? A quand les émissions de télévision sur les extraits les plus populaires de Youtube ?

On s'aperçoit que se formalisent deux problèmes principaux qui peuvent emprunter un vocabulaire très classique pour être explicités :

- Le problème d'une philologie numérique

La philologie est l'art d'établir les textes, c'est-à-dire des versions de référence. Discipline pratiquée par les spécialistes des textes anciens, confrontés à différentes versions qu'un aucun original ne vient départager, elle mobilise conjectures, hypothèses et réfutations pour proposer des versions canoniques sur lesquelles fonder l'interprétation.

- Le problème d'une herméneutique numérique

L'herméneutique est l'art d'interpréter les textes, c'est-à-dire d'en proposer une compréhension, une lecture, une signification. L'interprétation commence quand on a établi l'identité expressive du contenu ; elle commence là où la philologie s'arrête. En fait, les deux travaillent de conserve car c'est l'hypothèse d'un sens à donner qui permet de faire travailler les hypothèses philologiques.

L'enjeu aujourd'hui est de construire ces deux disciplines. Localement, au cas par cas, on sait établir des versions, on sait entreprendre des interprétations. Mais il faut bien avouer que pour le moment aucuns universaux de méthode n'émergent. Ces disciplines sont encore dans les limbes, et l'archivistique numérique en est encore à débroussailler le fossé d'obsolescence pour assurer la préservation des contenus numériques.

Un cadre méthodologique

L'archivage possède ses normes et ses méthodologies. Pour l'archivage numérique, la norme OAIS [CCSDS 02] est sans nul doute la plus pertinente et la

plus efficace pour entreprendre la préservation de contenu. Conçue par et pour la communauté des sciences astronomiques, cette norme s'avère beaucoup plus générale que son cadre initial de conception et elle tend à devenir le cadre imposé pour toute archive numérique. Car il s'agit d'un cadre : pas d'outils informatiques proposés par OAIS, seulement des modèles, volontairement très généraux pour ne pas imposer des choix techniques qui risqueraient d'être adaptés seulement pour un temps et pour une partie des problèmes traités.

OAIS propose un modèle d'information et un modèle de processus. Le modèle d'information repose sur l'idée que la donnée numérique n'a en soi aucune sémantique propre et qu'il faut donc lui associer une information permettant de la relire. On distingue donc la « donnée » proprement dite, ce que nous avons appelé jusqu'à présent la ressource, et l'« information de représentation » qui correspond à ce qu'il faut savoir pour exploiter la donnée. La notion de ce qu'il faut savoir dépend d'une part de la communauté à qui cette donnée est destinée, et d'autre part de ce qu'elle sait à un moment donné. OAIS parle de « communauté désignée ».

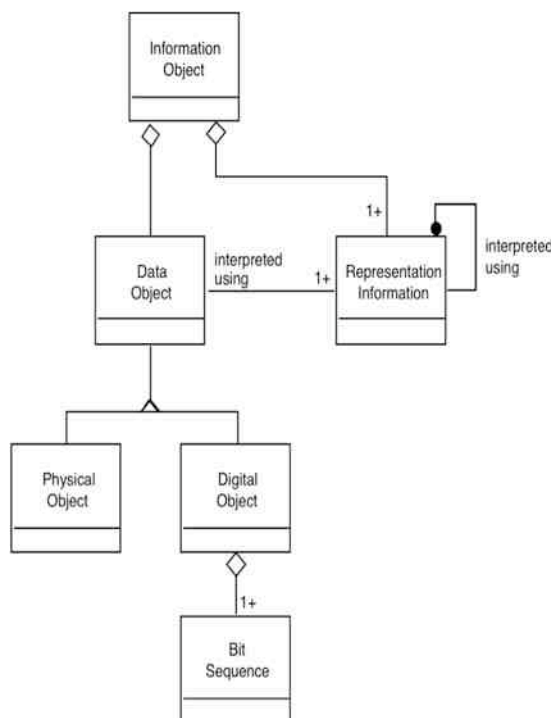


Figure 6. Le modèle d'information d'OAIS (en notation UML) prévoit qu'une information consiste en une donnée et une information de représentation, elle-même consistant en une donnée qu'il faut également décrire. Le modèle est donc récursif et évolutif [CCSDS 02].

Ce peut être par exemple son format, le type de codage associé, etc. Ainsi, à un moment donné, on sait qu'une suite binaire correspond à de l'ASCII. Il suffit de consigner dans l'information de représentation que c'est de l'ASCII, sans préciser davantage ce codage, car la communauté désignée connaît parfaitement ASCII. Mais si au cours du temps la communauté perd la connaissance de ce codage, il faudra enrichir l'information de représentation de cette connaissance. De même, si un mode d'emploi en anglais est associé à une donnée, et que l'anglais devienne une langue morte, il faudra intégrer une grammaire et un dictionnaire de cette langue.

Par conséquent, l'archive OAIS n'est pas établie une fois pour toute et elle évolue au cours du temps : l'archivage est un processus. On a donc des modules de processus qui indiquent comment gérer et administrer l'archive. Fondamentalement, ces processus ont pour but de surveiller l'écart existant entre l'état de l'archive et celui de la communauté désignée pour définir les différences entre la connaissance supposée connue par l'archive et celle effectivement détenue par la communauté. Dès qu'une telle différence est détectée, l'archive est enrichie. Le processus de l'archive selon OAIS repose sur trois objets associés à trois étapes clés :

- L'alimentation de l'archive, où les détenteurs de contenus confient ou remettent à l'archive leurs documents. On a affaire à des SIP (Submission Information Package).
- L'entretien de l'archive, où l'archiviste fait évoluer l'archive en fonction des nécessités du moment. On a affaire à des AIP (Archive Information Package).
- L'exploitation de l'archive, où l'archiviste délivre auprès de la communauté désignée les contenus archivés dans la forme adaptée à ses besoins. On a affaire à des DIP (Delivery Information Package).

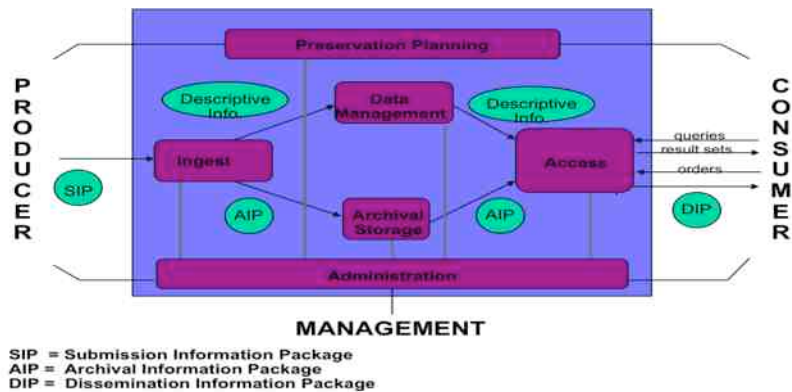


Figure 7. *Les processus prévus par la norme OAIS s'organisent selon les flux d'information et de contrôle de cette figure [CCSDS 02].*

OAIS est donc incontestablement une approche qui correspond à ce que nous avons appris dans notre examen des contenus numériques. Cependant, elle présente quelques défauts ou plutôt incomplétudes :

- Si l'archive est enrichie, elle l'est par le moyen d'un pilotage assuré par le gestionnaire de l'archive. Mais cette vision centralisée n'est pas adaptée à la réalité des contenus ni attestée par l'histoire de l'archive. L'enrichissement vient en fait des utilisateurs de l'archive, de la communauté désignée, et non des archivistes ; de même, les écarts sont établis par les chercheurs et membres de cette communauté, et non par les archivistes eux-mêmes, à moins d'en faire des membres de la communauté, ce qui est toujours possible en principe.
- De même, l'archive évolue sans intégrer les apports des usagers et leur lecture. Or, comme on l'a vu, c'est l'usage et l'interprétation actuels qui permettent d'apporter l'intelligibilité au contenu et de combler le fossé séparant la forme culturelle conservée de la forme sous laquelle il serait compréhensible. Autrement dit, il faut que les DIP puissent devenir des SIP après leur exploitation par les usagers de l'archive.

Ces remarques n'invalident en rien le modèle OAIS mais indiquent plutôt comment le mettre en œuvre. Ainsi, un projet d'archivage doit s'inscrire dans le cadre méthodologique OAIS mais prendre en compte pour sa mise en œuvre le modèle que nous allons à présent expliciter.

Des approches de la préservation

Plusieurs approches émergent alors pour la préservation des médiums technologiques. Il est désormais d'usage de distinguer les approches suivantes [THI 02, BORG 06, GLA 07] :

- L'approche muséologique

Cette approche consiste à conserver les contenus tels quels ainsi que leurs outils de lecture. On reproduit ainsi non seulement l'information à conserver mais aussi les conditions de lecture propres à une époque et à un contenu. Valable pour un volume faible de contenus, cette approche se heurte en outre à la difficulté de maintenir en fonctionnement des outils obsolètes. Mais elle reste utile pour se rendre compte des conditions de jeu pour les jeux d'arcade par

exemple, ou encore la lecture de contenus anciens comme les premières versions Word.

- La migration

La migration consiste à faire évoluer le format technique des contenus pour les garder compatibles et adaptés aux outils de lecture disponibles dans l'environnement technologique du moment. Coûteuse puisqu'il faut l'appliquer à tous les contenus, cette stratégie est aussi la plus simple à mettre en œuvre. De plus, elle permet de faire bénéficier les contenus migrés des perfectionnements liés aux formats et à leurs lecteurs.

- L'émulation

Plutôt que de faire évoluer les contenus, on simule sur les environnements du moment les outils de lecture des formats anciens. Solution séduisante en théorie, car on garde les contenus identiques à eux mêmes sans avoir à la faire évoluer, elle est fragile car une émulation n'est jamais parfaite ni efficace. De plus, il faut suivre l'évolution des outils de lecture par une cascade d'émulateurs, ce qui induit une complexité technologique coûteuse et inefficace. Depuis quelques temps, on pense préserver les contenus en les émulant sur une machine virtuelle, la tâche n'étant alors que d'implémenter la machine virtuelle sur l'environnement cible [LOR 02]. Possédant de fervents défenseurs comme [ROT 99] car permettant un archivage à l'identique respectant intégrité et authenticité, sans avoir à choisir pour la postérité ce qu'il faut retenir d'un contenu pour le faire migrer ou le reproduire, cette approche connaît un surcroît d'intérêt ces derniers temps grâce à l'approche virtuelle.

- La description

Cette dernière approche consiste à ne pas conserver les contenus enregistrés dans la mesure où ils sont partiels, incomplets ou mal définis (formats ad hoc, connaissances imparfaites de ces derniers, etc.). Il vaut mieux par conséquent conserver une description des contenus qui permettent de les reproduire. La description peut porter sur les points essentiels à reproduire, l'intention de l'auteur à respecter (par exemple [DEP 03]), l'apparence graphique, etc.

Aucune de ces stratégies ou approches n'est clairement meilleure que les autres et en particulier aucune ne peut prétendre assumer les avantages et propriétés de toutes les autres. Il reste donc qu'il faut construire une stratégie à chaque fois nouvelle, à partir de ces 4 approches, en fonction des fonds concernées et des objectifs fixés à l'archivage.

La complexité des médiums technologiques impliquent néanmoins que les approches par descriptions vont prendre une ampleur certaine et que les recherches sur les moyens et standards de description seront sans nul doute des enjeux importants des prochaines années. Dans la prolongation de l'approche musicale rappelée plus haut, il s'agit de définir les langages permettant de noter les invariants devant piloter et contrôler la reconstruction ou ré-invention des contenus.

Un modèle résultant

Le modèle qui en résulte tient alors dans les thèses suivantes :

- La préservation s'effectue par l'usage et l'interprétation des contenus qui entraîne alors leur conservation.
- La conservation simple des contenus est inutile :
 - Elle n'assure pas la lisibilité culturelle dans la mesure où elle ne permet pas de surmonter le fossé d'intelligibilité. C'est l'appétence au contenu, entretenue par l'animation et l'alimentation des communautés savantes et culturelles, qui suscite la succession des lectures et des commentaires comblant le fossé d'intelligibilité et préparant la lecture des générations futures.
 - Elle n'assure pas la lisibilité technique dans la mesure où elle ne permet de surmonter le fossé d'obsolescence. C'est la pratique régulière des contenus qui permet d'appréhender la progressive obsolescence des contenus et de les adapter aux conditions technologiques nouvelles.
- La détermination de ce qu'il faut conserver provient de la définition des invariants constituant l'identité d'une œuvre. Cette définition est toujours provisoire et temporaire car elle relève de la convention et tradition culturelle. C'est pourquoi la conservation ne peut être faite une fois pour toute et doit toujours s'associer à une recherche qui remet en cause ces invariants et reconstruit le modèle du contenu. Autrement dit, l'archivage et ses choix appartient à la communauté savante et culturelle, et ne peut se rapporter à des choix fixés par défaut (on garde tout) ou dogmatiquement (on fait le tri). L'archivage relève donc d'une politique de la mémoire qui repose sur un débat public avec les communautés concernées selon leur régime propre de compétence. Ce

débat sera largement contradictoire et polémique, la mémoire véhiculant l'identité et le rapport au présent.

- La conservation n'est pas une préservation de l'intégrité physique des contenus ni de leur identité à soi, mais une ré-invention permanente des contenus à partir des éléments conservés. L'enjeu est de conserver une identité des contenus à travers la transformation de leur ressource et la variabilité de leurs vues re-inventées. Au lieu d'une permanence garantissant l'identité à soi du contenu, il convient de parler d'une transmanence où le contenu évolue pour rester lui-même, où son identité n'est plus par coïncidence avec soi mais par évolution.
- La préservation doit adopter une vision organique de la mémoire, où le contenu évolue, se transforme, s'adapte pour se maintenir et se conserver.

Conclusion

La préservation numérique est une ré-invention de l'archive à partir des ressources conservées. Que ce soit une ré-invention découle du fait l'accès au contenu, i.e. sa publication, s'effectue via une reconstruction calculée dont la nature peut être arbitraire. Par conséquent l'archivage ne peut plus seulement se penser à partir de l'origine de l'archive qu'il faut garder en l'état mais en fonction de l'exploitation qui en sera faite, même si cette exploitation se fixe pour tâche de rétablir l'authenticité et l'intégrité du contenu. Evolution permanente, l'archivage ne peut se penser comme gardien d'un contenu mais comme témoin d'une transformation permanente d'une identité qu'il faut continument faire évoluer.

Références

- [BAC 07] BACHIMONT, B. (2007). Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents. — Hermès, Paris.
- [BOR 06] BORGHOFF, U. M., RÓDIG, P., SCHEFFCYK, J. & SCHMITZ, L. (2006) Long-Term Preservation of Digital Documents: Principles and Practices, Berlin, Springer.
- [CAR 02a] CARRUTHERS, M. (2002). Le livre de la mémoire : la mémoire dans la culture médiévale. — Macula, Paris.
- [CAR 02b] CARRUTHERS, M. (2002). Machina Memorialis : méditation, rhétorique et fabrication des images au Moyen Âge. — Gallimard, Paris.

- [CCSDS 02] CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (2002) Reference Model for an Open Archival Information System (OAIS). Washington, http://www.archivistes.org/IMG/projet_norme_oais_version_francaise.pdf.
- [CLE 08] CLECH-CHARTON, S. L. (2008) Guillaume Budé ; l'humaniste et le prince, Paris, Riveneuve - éditions.
- [DEP 03] DEPOCAS, A., IPPOLITO, J. & JONES, C. (Eds.) (2003) L'approche des médias variables : la permanence par le changement, Montréal, New York, Fondation Daniel Langlois et Guggenheim Museum Publications.
- [GLA 07] GLADNEY, H. M. (2007). Preserving Digital Information. — Springer, Berlin.
- [LOR 02] LORIE, R. (2002) The UVC : a Method for Preserving Digital Document - Proof of Concept. IN VAN DIESEN, R. J. (Ed. IBM / KB Long-Term Preservation Study). IBM / Koninklijke Bibliotheek. http://www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf
- [RO 99] ROTHENBERG, J. (1999) Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Washington, D.D., Council on Library and Information Resources.
- [THI 02] THIBODEAU, K. (2002) Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. The State of Digital Preservation: An International Perspective. Washington D.C., Council on Library and Information Resources.
- [VEY 77] VEYNE, P. (1971). Comment on écrit l'histoire. — Seuil, Paris.