

# L'Univers numérique

## Pour une préservation à long terme du patrimoine numérique

Dans notre série qui accompagne l'Année européenne du patrimoine culturel, le présent article soulève un problème fondamental de la culture actuelle du savoir : la conservation du patrimoine numérique. Alors que bibliothèques et archives de toutes sortes (parchemin, papier, cinéma, photographie, ...) ont tout doucement commencé à numériser leurs fonds, seuls quelques spécialistes semblent se préoccuper dès à présent du long terme.

Dans le film *Rogue One, A Star Wars Story* (Gareth Edwards, 2016), l'intrigue toute entière repose sur la récupération par l'Alliance rebelle (le côté lumineux de la Force) des plans de construction de l'Etoile de la Mort, qui se trouvent dans les archives de l'Empire (le côté obscur de la Force). On vous épargnera les 1001 détails de la saga *Star Wars*, mais sachez que c'est le vol de *données numériques* concernant la construction de l'Etoile de la Mort qui permettra aux rebelles de combattre l'Empire et Dark Vador dans *Star Wars: Episode IV – A New Hope*, le premier film tourné par George Lucas en 1977. L'histoire du vol de ces données, qui se situe chronologiquement avant le premier *Star Wars*, est ainsi en quelque sorte à la source de tous les récits qui suivront.

L'ingénieur Garen Erso (Mads Mikkelsen) a été contraint de construire l'Etoile de la Mort, mais il a introduit dans les plans (et peut-être aussi dans le système d'archivage car les données semblent mal protégées) une faille, avec l'idée que des années plus tard sa fille et les rebelles pourront récupérer ces données et s'en servir pour détruire l'Etoile. Pour cela, il a dû faire confiance à la politique de gestion

des archives de l'Empire et en effet, à travers toute la saga *Star Wars*, il n'est jamais question de corruption de fichiers ou de problèmes d'interopérabilité, ce qui suppose que les archivistes de la galaxie ont mis en place de bonnes stratégies de gestion des données. Imaginez qu'en essayant d'ouvrir le jeu de données, l'ordinateur des rebelles ait affiché un message d'erreur ou de corruption du fichier ! L'histoire de Leia, Han, Luke et tout le destin de la galaxie en auraient été changés. Mine de rien, le film touche là à la nature pérenne des archives qui dépasse le cadre d'une vie humaine. L'archiviste a pour objectif l'élaboration de stratégies permettant de garantir la transmission aux générations futures de données et des informations nécessaires pour les décrypter. Une décision prise à un certain moment peut dès lors avoir des conséquences des années, voire des décennies, plus tard.

### Le patrimoine numérique

Le patrimoine numérisé ou nativement numérique remplit deux fonctions : d'une part la conservation pérenne des données et de l'autre l'accès facile et rapide des utilisateurs à ces mêmes données. L'association britannique Digital Preservation Coalition (DPC)<sup>1</sup> définit ainsi la préservation numérique : « a series of managed activities necessary to ensure continued access to digital materials for as long as necessary ». La préservation du patrimoine numérique existe parallèlement à celle du patrimoine analogique (ou physique), mais elle ne peut pas être

Alessandra Luciano

---

**Le patrimoine numérisé ou nativement numérique remplit deux fonctions: d'une part la conservation pérenne des données et de l'autre l'accès facile et rapide des utilisateurs à ces mêmes données.**

---

Alessandra Luciano est la responsable des archives films et tv au Centre national de l'audiovisuel à Dudelange. Elle est diplômée en théories et philosophies du cinéma de la Columbia University ainsi qu'en préservation et présentation de l'image animée, avec spécialisation en image d'archives de l'Université d'Amsterdam. Ses intérêts de recherche portent sur l'histoire et le développement technologique du cinéma, les images d'archives et la culture populaire.

---

**Toutes les institutions culturelles qui ont pour mission la préservation à long terme d'un patrimoine font face aujourd'hui à un déluge de données numériques.**

---

traitée, gérée et abordée comme celle des supports analogiques. Même si quelques tâches sont similaires, la préservation des données numériques nécessite un savoir-faire complémentaire, d'autres actions à mener et une approche, voire même une philosophie différentes.

### Les nouveaux défis du numérique

Si le numérique offre des avantages comme l'accès rapide aux documents et à l'information, il est aussi créateur de nouveaux défis. Toutes les institutions culturelles qui ont pour mission la préservation à long terme d'un patrimoine font face aujourd'hui à un déluge de données numériques qui exigent des décisions en termes de politique de collecte, de gestion de l'espace de stockage et d'organisation du travail. La quantité des documents numériques ne cesse en effet d'augmenter.

Outre la masse des données, leur sécurisation constitue l'un des grands enjeux des archives numériques. De par leur nature, les données numériques risquent à tout moment d'être (volontairement ou non) effacées ou corrompues. Alors qu'il est possible de conserver du papier ou de la pellicule sans trop s'en préoccuper si ces supports sont stockés dans des conditions optimales, les fichiers numériques exigent une surveillance constante.

La préservation de fichiers numériques implique ainsi des problématiques supplémentaires qui concernent le stockage, la sécurité mais aussi la standardisation des données à préserver. A ces facteurs s'ajoutent la propriété intellectuelle et l'obsolescence technologique des formats numériques et des outils nécessaires pour lire les données. Contrairement aux supports physiques visibles à l'œil nu, un fichier numérique n'est en effet lisible qu'à travers une interface, à l'aide d'un ordinateur, qui suit des lignes de commandes afin d'accéder au contenu. Il est impossible de déduire la condition du document et sa conformité aux stratégies mises en place par l'archive, sans l'ouvrir grâce à un logiciel.

### La bonne gestion des archives numériques

La standardisation des formats numériques et leur conservation à long terme sont des questions qui sont loin d'être résolues. Il existe plusieurs actions à entreprendre pour garantir une bonne gestion des archives. Nous allons nous en tenir ici à deux d'entre elles, parmi les plus essentielles : la bonne connaissance des formats numériques et la migration. Ces deux étapes fonctionnent en complément, car « the migration of data files can be more difficult to

carry out if the authenticity and integrity of the files are not guaranteed »<sup>2</sup>.

Il est d'abord primordial de comprendre les données qui sont transmises aux archives. Cela implique de documenter l'environnement dans lequel les données ont été créées ainsi que les aspects techniques (l'ensemble formant les métadonnées techniques). Il est possible de traiter cette partie du travail de plusieurs façons. Il existe des règles de catalogage ; la façon dont l'information est structurée est importante dans la mesure où cette information sera interprétée par la suite par des outils informatiques. Il faut également comprendre la structure physique des fichiers numériques afin de savoir quelles étapes entreprendre pour leur conservation. Cela peut être difficile en particulier dans le cas des images animées, car il n'existe pas de format normalisé et la structure d'une vidéo numérique peut varier, même si elle a le même nom de conteneur<sup>3</sup> et le même codec<sup>4</sup>. Ceci peut devenir un problème lors d'une conversion ou d'une migration car les fichiers peuvent se comporter différemment et afficher des anomalies bien qu'ils aient le même nom.

Dans les archives, on distingue par ailleurs entre un format d'exploitation et de préservation. La version qu'on appelle souvent master, donc le fichier de préservation, doit contenir le plus d'informations possible, ne pas être comprimée, devrait idéalement être dans un format non-propriétaire<sup>5</sup> et surtout doit être stockée dans plusieurs endroits. La duplication des archives, ou *mirroring*, est une bonne pratique qui permet en cas de catastrophe (naturelle ou pas) de récupérer le contenu des archives qui sont dupliquées dans un autre lieu. La version d'exploitation peut par contre être dans des formats différents, éventuellement comprimés, car elle ne sert qu'à la consultation et à l'utilisation et doit

StorageTek, conservation du support numérique,  
Photo: Romain Girtgen © CNA



donc en premier lieu satisfaire le besoin des clients et utilisateurs.

L'un des principaux défis pour la préservation numérique est la rapidité avec laquelle les technologies évoluent, obligeant les archives à constamment s'adapter. C'est un moyen pour les fournisseurs de s'assurer que leurs clients continueront à acheter et à investir dans de nouveaux produits. Si l'archive n'est pas compatible avec les technologies les plus récentes, elle ne pourra plus lire certains fichiers qui seront alors irrémédiablement perdus. Sans constante mise à jour (tous les cinq à dix ans) du stockage numérique et du système d'exploitation, l'infrastructure et le contenu deviendront obsolètes. Cette obsolescence programmée fait de la migration des données une nécessité. La migration est une des stratégies principales pour la sauvegarde des archives. Cette opération permet de migrer sans perte le contenu des fichiers vers un nouveau format en assurant que les fichiers restent accessibles et lisibles par des systèmes d'exploitation.

L'évolution constante et rapide des technologies change aussi les exigences des utilisateurs. La migration est donc aussi une nécessité si les archives veulent continuer à fournir aux utilisateurs les formats qu'ils désirent. Mais si la migration est une tâche complexe, elle a aussi l'avantage d'obliger les archives à surveiller constamment l'état et la pérennisation de ses données. Une archive a plus de chances de détecter une anomalie dans ses données et donc d'y remédier si elle est obligée de lire régulièrement ses fichiers numériques.

La vérification des fichiers et la migration sont ainsi complémentaires dans le workflow d'une archive. Car la migration de milliers de fichiers est plus facile à réaliser si l'intégrité des fichiers est garantie et fiable. La notion de fiabilité signifie qu'on réduit le risque de changer l'information manipulée et qu'on assure une prévisibilité maximale dans ce qui se passera lors de futures migrations.

## Conclusion

S'ils veulent garantir l'accès aux données non altérées et non corrompues à de multiples générations d'utilisateurs, les responsables des archives numériques doivent anticiper continuellement les changements et nouveaux standards technologiques à venir, ce qui exige une évolution permanente de leur savoir-faire et donc un investissement considérable (en temps et en argent) dans leur formation continue ainsi qu'un contact permanent avec les autres acteurs du secteur. De leur côté, les responsables politiques doivent s'investir dans la conservation pérenne du

patrimoine national en définissant des approches et une philosophie communes, en standardisant dans la mesure du possible les technologies mises en œuvre, tout en veillant à l'interopérabilité nationale et internationale des standards mis en place et en étant conscient des défis, du coût et des enjeux de la conservation pérenne du patrimoine numérique. ♦

- 1 <https://www.dpconline.org>, consulté le 17.02.2018
- 2 PREFORMA handbook (<http://www.preforma-project.eu/download.html>, consulté le 16.02.2018). Preforma est un projet européen, dont l'objectif est le développement d'outils open source à utiliser dans la standardisation et la validation des formats de fichiers numériques qui doivent être conservés à long terme dans les archives. Le CNA collabore au projet Preforma et l'initiative « No Time to Wait » qui en découle.
- 3 Les conteneurs sont utilisés pour stocker des objets numériques sous une forme organisée qui suit des règles d'accès spécifiques. (Wikipedia, article « conteneur informatique »)
- 4 Un codec est un dispositif matériel ou logiciel permettant de mettre en œuvre l'encodage ou le décodage d'un flux de données numérique, en vue d'une transmission ou d'un stockage. Certains codec intègrent également une fonction de compression ou encore de chiffrement des données. (Wikipedia, article « codec »)
- 5 Un format non propriétaire, ou open source, est un format de données dont les codes sont mis à disposition du public et dont l'utilisation n'est pas restreinte par des brevets. Les utilisateurs ne sont dès lors pas dépendants d'une société (qui peut faire faillite, décider de stopper la production de son produit ou profiter d'une situation de monopole).

*StorageTek*, conservation du support numérique,  
Photo: Romain Girtgen © CNA

