

Module 7C - Section 6 : **S'informer**

Édouard Vasseur @AIAF - PIAF

VF 02/12/2024

Table des matières

Objectifs	3
Introduction	5
1. Les organisations spécialisées dans la préservation numérique	6
1.1. Digital Preservation Coalition (DPC)	6
1.1. Objectifs.....	6
1.2. Les ressources utiles fournies par la DPC.....	7
1.2. Le groupe PIN	7
2.1. Objectifs.....	7
2.2. La Cellule nationale de veille du groupe PIN	8
2.3. Les 4 axes de travail de la Cellule nationale de veille de PIN	8
1.3. Les autres associations nationales	8
1.4. Open Preservation Foundation (OPF)	9
2. Revues et conférences internationales	10
2.1. Les revues	10
2.2. Les conférences internationales	10
3. Autres ressources en ligne	12
3.1. Community Owned digital Preservation Tool Register (COPTR)	12
1.1. COPTR est un wiki	12
1.2. Les fiches du COPTR.....	12
3.2. Wikidata for digital preservation	12
3.3. Carte heuristique des outils de préparation des versements	13
Conclusion	14

Objectifs



Description du module :

La préservation des documents d'archives sur support numérique – ce que les Québécois nomment documents technologiques – constitue désormais un enjeu quotidien des archivistes. L'archiviste dispose désormais d'un important panorama de normes, de standards, d'outils et de retours d'expérience pour lui permettre d'appréhender les documents d'archives sur support numérique et envisager leur préservation dans le temps.

Le but du module est de :

- aider à évaluer la situation en matière de préservation des documents d'archives sur support numérique ;
- permettre de concevoir et mettre en œuvre une politique de préservation des documents d'archives sur support numérique.

L'apprenant doit être en mesure de :

- appréhender les spécificités en matière de préservation des documents d'archives sur support numérique ;
- dresser un état des lieux d'ensembles de documents d'archives sur support numérique ;
- définir et mettre en œuvre une politique de préservation des documents d'archives sur support numérique ;
- concevoir, mettre en œuvre et administrer un environnement permettant de gérer dans le temps les documents d'archives sur support numérique, quels que soient les moyens disponibles ;
- appréhender les différentes catégories de formats de fichiers numériques ;
- savoir comment aller plus loin dans la réflexion.

Positionnement :

Ce module s'inscrit naturellement dans la chaîne archivistique. S'il se concentre sur les questions de planification de la préservation, de mise en œuvre de la préservation et de stockage des documents d'archives sur support numérique, il fournit également des éléments à prendre en compte lors de la mise en place de politiques et procédures de gouvernance de l'information et de gestion de l'archivage/gestion des documents d'activité/gestion des documents institutionnels/records management, de collecte de documents d'archives définitifs et d'accès à ceux-ci.

Il ne s'intéresse en revanche pas à la numérisation de documents d'archives sur support physique ou d'enregistrements sonores et audiovisuels sur support analogique, sauf dans le cas où l'opération de numérisation vise à substituer la version du document sur support numérique à celle sur support physique ou analogique.

Point sur le vocabulaire employé :

- Le terme “préservation” est entendu comme recouvrant « les fonctions de conservation préventive et matérielle » [Direction des Archives de France, Dictionnaire de terminologie archivistique, 2002] ;
- Sont distingués :
 - **les documents d’archives sur support physique**, où l’information est directement accessible à l’œil humain ou ne nécessite, pour le devenir, que l’emploi d’un appareil optique (projecteur) permettant de faciliter son agrandissement
 - **les documents d’archives sur support analogique**, où l’information, pour être intelligible, a absolument besoin de la médiation d’un appareil pour permettre à l’utilisateur de prendre connaissance de l’information (projecteur, lecteur, etc.) ;
 - **les documents d’archives sur support numérique**, qu’ils aient été directement produits avec des outils numériques ou soient le produit de la numérisation de documents d’archives sur support physique ou analogique. L’information, pour être intelligible, a absolument besoin de la médiation d’un environnement matériel et logiciel pour permettre à l’utilisateur de prendre connaissance de l’information ;
- “Document d’archives” est l’expression utilisée pour identifier toute information sur un support qui a besoin d’être prise en charge et conservée, soit pour sa valeur de preuve, soit pour sa valeur informationnelle, soit pour sa valeur patrimoniale ou de recherche. En fonction du contexte, l’expression pourra concerner des documents, des records ou des archives au sens anglo-saxon des termes ;
- “Service d’archives” est l’expression utilisée pour désigner toute structure ou organisme souhaitant mettre en place une politique de préservation de documents d’archives sur support numérique. Ce service d’archives peut être
 - interne à une organisation productrice et en charge de la gouvernance de l’information et de la gestion de l’archivage/gestion des documents d’activité/records management ou de la gestion d’archives intermédiaires ;
 - externe à une organisation productrice, soit qu’il s’agisse d’un prestataire de tiers archivage, soit d’un service d’archives définitif.

Les notions abordées dans ce module peuvent être complétées par :

- le module 9 - Section 2 : Numériser les documents qui présente les techniques de base de transfert de support vers le numérique
- le module 5 Gestion et traitement des archives courantes et intermédiaires

Il est vivement conseillé de prendre connaissance du module 7B Gestion des documents numériques au stade courant avant d’entamer la lecture du module présentement proposé. Certaines notions de base, activées ici, sont exposées plus longuement dans ce premier module.

Le glossaire du PIAF doit être consulté pour les définitions des termes spécifiques.

Introduction

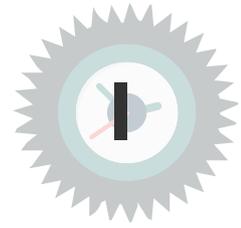


La préservation numérique est encore une pratique professionnelle récente et en pleine évolution, comme le sont d'ailleurs l'information numérique et les documents d'archives sur support numérique.

Aussi tout professionnel intéressé par les questions de préservation numérique ou en charge de fonds d'archives sur support numérique doit se tenir au courant des nouvelles réflexions, des nouveaux outils et des nouvelles expérimentations qui sont menées par les universitaires et les professionnels, dans le monde francophone ou à l'extérieur de celui-ci.

Cette section du module de cours a pour ambition de guider les apprenants et de leur fournir les moyens de continuer à se documenter, pour mettre à jour leur politique de planification de la préservation.

1. Les organisations spécialisées dans la préservation numérique



Introduction

Les associations professionnelles « historiques » (comme le Conseil international des archives, l'Association des archivistes français, l'Association des archivistes du Québec, l'Association des archivistes suisses) **constituent traditionnellement des lieux d'échanges**. La préservation numérique fait partie des sujets qui y sont débattus, parmi tous les sujets de préoccupations des archivistes.

À côté de ces associations professionnelles, **des associations ou des structures dédiées à la préservation numérique se sont fondées depuis le début des années 2000**, regroupant bien souvent archivistes, bibliothécaires, archéologues, professionnels de musées, universitaires et toutes les personnes intéressées par la préservation des documents d'archives et autres objets sur support numérique.

Ces associations ont des objectifs variés :

- mise en commun des pratiques professionnelles ;
- formation initiale et continue ;
- développement et maintenance d'outils.

Si la plupart utilisent la langue anglaise pour leurs travaux, certaines regroupent des professionnels francophones.

1.1. Digital Preservation Coalition (DPC)

Digital Preservation Coalition (DPC – <https://www.dpconline.org/>) est sans doute la structure la plus connue et celle qui met à disposition des professionnels de la préservation numérique le plus de ressources.

1.1. Objectifs

Fondée en 2002 par des professionnels du Royaume-Uni et d'Irlande, DPC a pour mission d'assister ses membres dans leur mission de préservation à long terme des objets numériques, en les aidant à affronter les défis stratégiques, culturels et technologiques posés par ceux-ci.

Le dernier plan stratégique qui couvre la période 2018-2022 donne une bonne idée des objectifs de l'organisation, qui comprennent notamment les axes suivants :

- développer un plaidoyer adapté aux enjeux et défis de la préservation numérique ;
- permettre aux membres de l'organisation de développer les échanges entre eux ;
- identifier et développer les bonnes pratiques ;
- aider les membres de l'organisation à mettre en œuvre des organisations appropriées et durables leur permettant d'affronter les défis de la préservation numérique, etc.

1.2. Les ressources utiles fournies par la DPC

Aux personnes intéressées par la préservation numérique, même si elles n'appartiennent pas à des organisations membres de DPC, l'organisation fournit un certain nombre de ressources utiles – mais en anglais –, régulièrement actualisées par ses membres :

- **le manuel de préservation numérique (Digital Preservation Handbook)**, publié pour la première fois en 2001 et régulièrement actualisé, constitue une ressource utile. Régulièrement mis à jour, traduit en français, il donne accès à des cas d'usages, des retours d'expériences, des présentations et des articles sur la préservation numérique en général, et sur des points particuliers comme la criminalistique numérique (digital forensics), la sécurité, le recours à des services d'infonuagique (cloud computing), etc.
- **des rapports sur des problématiques spécifiques** – notamment sur les questions liées aux différents formats de fichiers – sont rédigés, mis à jour régulièrement et publiés sur le site de DPC, dans la rubrique des *Technology Watch Publications*. Composés généralement d'une cinquantaine à une centaine de pages, ils couvrent des sujets aussi variés que les objets audiovisuels, les messages électroniques, le PDF/A, les réseaux sociaux, le modèle de référence OAIS, les ebooks. Ils sont publiés au rythme d'un ou deux rapports par an ;
- **des notes techniques**, qui donnent en quelques pages (généralement autour de 10) une synthèse des connaissances sur des questions de préservation numérique. Les sujets sont variés et couvrent les formats de fichiers (tableurs, messageries, 3D, etc.) et des activités de préservation (accès par exemple).

C'est également DPC qui a conçu la grille d'évaluation rapide qui a été présentée dans la section 2 de ce module.



Complément

Pour aller plus loin : <https://www.dpconline.org/docs/digital-preservation/training-modules/getting-started-with-digital-preservation/1679-digital-asset-registers-getting-started/file>.

1.2. Le groupe PIN

La France dispose avec l'association Aristote, et plus particulièrement son groupe Pérennisation de l'information numérique (PIN) d'un équivalent à DPC (<https://www.association-aristote.fr/groupe-pin/>).

Ce groupe rassemble les principales organisations françaises intéressées par le sujet, notamment

- le Service interministériel des Archives de France (SIAF),
- les Archives nationales de France,
- la Bibliothèque nationale de France (BnF),
- le Centre national de la recherche scientifique (CNRS),
- le Centre national d'études spatiales (CNES),
- l'Institut national de l'audiovisuel (INA) ou
- l'Institut national de recherches en sciences et technologies du numérique (INRIA).

1.2.1. Objectifs

Le groupe PIN a pour vocation de constituer un lieu d'échange dans le domaine de la préservation numérique. Il organise principalement des réunions plénières où sont partagées expériences et connaissances, ainsi que des sessions de formation dédiées à la question de la préservation numérique.

1.2.2. La Cellule nationale de veille du groupe PIN

En son sein a été créée en 2019 une Cellule nationale de veille (<https://www.association-aristote.fr/cellule-format/>) qui s'intéresse plus particulièrement aux questions liées aux objets numériques, à leurs formats et à leur préservation. Regroupant une dizaine d'institutions publiques, cette cellule a pour objectifs de mutualiser savoir-faire et expertises, de produire des outils pouvant servir de recommandations, de mutualiser la veille et de constituer une passerelle entre les mondes francophone et anglophone.

1.2.3. Les 4 axes de travail de la Cellule nationale de veille de PIN

À ce jour, la cellule a développé quatre axes de travail principaux :

- **un sous-groupe « Expertises »** qui assure un recensement des expertises existantes dans les organisations, tant françaises qu'étrangères, en matière de préservation numérique en général et de formats de fichiers en particulier ;
- **le sous-groupe « Connaissance des formats »** qui réfléchit à la définition d'une politique de formats, ainsi qu'à l'analyse et l'étude de formats de fichiers spécifiques. Il travaille actuellement sur les formats d'images numériques (TIFF) et sur les formats associés aux messageries électroniques (EML, PST). Il a également conçu une grille permettant aux institutions de recenser les formats de fichiers représentés dans leurs collections, et défini neuf critères pour élaborer/établir une politique de formats ;
- **le sous-groupe « Outils et corpus »** qui identifie et recense les outils utiles, en se basant sur le registre COPTR et Wikidata (voir le chapitre 3 plus bas) et qui a également pour objectif d'identifier et de mettre à disposition des corpus de fichiers électroniques en français ;
- **le sous-groupe « Traductions »**, qui procède à des traductions en français de documents de référence publiés en langue anglaise. C'est ce groupe qui a traduit le manuel de préservation de DPC et qui travaille actuellement à la traduction de rapports et de notes techniques de DPC.

Toutes les ressources produites par le groupe PIN et la cellule nationale de veille sur les formats (grilles, traductions, webinaires) sont publiées sur le site de l'association Aristote et sa chaîne YouTube.



Pour aller plus loin : <https://www.association-aristote.fr/>¹

1.3. Les autres associations nationales

Des associations existent également dans d'autres pays, avec des objectifs relativement similaires à ceux des deux associations présentées précédemment :

- aux États-Unis, la *National Digital Stewardship Alliance (NDSA)*, fondée en 2010 et dont le secrétariat est assuré par la Bibliothèque du Congrès, qui a notamment produit les Niveaux de préservation présentés dans la section 2 de ce module ;
- en Allemagne, **le réseau nestor**, officiellement fondé en 2009, dont le secrétariat est assuré par la *Deutsche Nationalbibliothek (NBK)*. Le réseau nestor s'est principalement intéressé aux questions de certification évoquées dans le chapitre 4 de la section 4 de ce module, mais il comprend également des groupes de travail sur les normes, la documentation, l'émulation, la reconnaissance des formats, les formats audiovisuels, l'archivage personnel ou auto-archivage (archivage des documents d'archives privés produits et conservés par des particuliers). Ses ressources sont en allemand ;

¹ <https://www.google.com/url?q=https://www.association-aristote.fr/&sa=D&source=editors&ust=1733335186019344&usg=AOvVaw1WOEnvNe9dr9gZ1UCwNUpt>

- aux Pays-Bas, le *Dutch Digital Heritage Network*, fondé en 2014. Ses ressources sont en néerlandais.

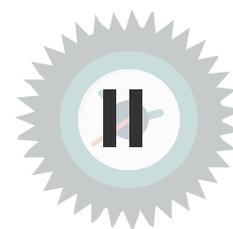
1.4. Open Preservation Foundation (OPF)

L'*Open Preservation Foundation* (OPF) est une organisation à but non commercial de droit anglais issue du projet *Preservation and Long-term Access through Networked Services* (Planets), financée par l'Union européenne, et initiée en 2010.

Cette fondation a pour but de développer et de fournir des outils utiles à la préservation numérique. C'est OPF qui conçoit, développe et maintient des outils libres d'identification de formats, de validation de formats et d'extraction de métadonnées comme JHOVE, VeraPDF, jpylyzer ou fido.

Le blog hébergé par l'OPF est une source utile d'informations sur la préservation numérique en général et sur l'utilisation des outils développés par OPF en particulier.

2. Revues et conférences internationales



Introduction

Une intense activité d'échanges internationaux vient compléter l'offre fournie par les associations professionnelles. Revues et actes des conférences internationales sur le sujet fournissent de nombreuses informations utiles à tout archiviste s'intéressant à la préservation numérique. Comme souvent, ces ressources sont cependant majoritairement disponibles en langue anglaise.

2.1. Les revues

Peu de revues sont malheureusement consacrées à la préservation numérique.

Une des rares exceptions est l'*International Journal of Digital Curation* (IJDC) qui rassemble des articles traitant de sujets stratégiques, tactiques ou opérationnels dédiés à la préservation numérique.

Le milieu universitaire et ses problématiques – préservation des données de la recherche, auto-archivage – y sont prédominants, mais les archivistes intéressés par la préservation numérique y trouveront des articles en libre accès susceptibles de les intéresser.

2.2. Les conférences internationales

La majeure partie des ressources consacrées à la préservation numérique est constituée des actes des conférences internationales qui se tiennent à des intervalles réguliers (généralement annuels) sur le sujet.

Les articles, comptes rendus des ateliers et résumés d'affiches présentés à l'occasion de ces conférences constituent une source unique de réflexion sur la préservation numérique ainsi que sur les grandes problématiques traitées par les universités et les institutions de conservation du patrimoine.



On notera que le coût d'inscription et de participation à ces conférences en limite l'accès à des institutions suffisamment dotées financièrement.

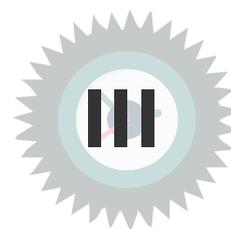
Trois conférences internationales dominent :

- **International conference on digital PREServation** (iPRES) est une conférence annuelle qui réunit depuis 2004 les spécialistes de la préservation numérique. Actualité des normes, retours d'expériences, ateliers pratiques, séances d'affiches rassemblent les professionnels du monde entier, et s'y retrouvent naturellement les experts des principales organisations spécialisées, DPC, le groupe PIN, nestor, NDSA, OPF, etc. Les actes de ces conférences, publiés en ligne, sont particulièrement riches : <https://ipres-conference.org/>
- **International Digital Curation Conference (IDCC)** est une conférence annuelle organisée par le Digital Curation Center (DCC) britannique depuis 2005. Comme l'IJDC, elle traite plus particulièrement des données numériques de la recherche et des problèmes de préservation

qu'elles posent. Les présentations effectuées lors de ces conférences sont disponibles sur le site de DCC, via le programme de chaque conférence, accessible depuis cette page : <https://www.dcc.ac.uk/events/idcc> ;

- ***Preservation and Archiving Special Interest Group (PASIG)*** est un groupe qui organise également depuis 2008 des conférences annuelles sur la préservation numérique. Plus informelles que iPRES, elles ont également un volet plus technique. Les présentations effectuées à l'occasion de ces conférences sont accessibles à partir du site de chacune d'entre elles : <https://www.preservationandarchivingsig.org/resources/past-events>

3. Autres ressources en ligne



Introduction

Trois autres ressources disponibles en ligne sont d'une grande utilité pour tous ceux qui s'intéressent à la préservation numérique :

- le Community Owned digital Preservation Tool Register (COPTR) ;
- Wikidata ;
- la carte heuristique des outils de préparation des versements.

3.1. Community Owned digital Preservation Tool Register (COPTR)

Disponible à l'adresse suivante : https://coptr.digipres.org/index.php/Main_Page

3.1.1. COPTR est un wiki

COPTR est une ressource en ligne sous forme de wiki qui a pour but de fournir un registre des outils utiles pour mener la préservation numérique. COPTR permet également de décrire des processus de préservation numérique comme la prise en charge des fichiers d'imagerie tridimensionnelle, par exemple.

3.1.2. Les fiches du COPTR

Chaque outil dispose d'une fiche particulière qui décrit sa finalité (ex. accès, calcul d'empreinte, émulation, extraction de métadonnées, identification de formats, validation de formats), **les catégories d'objets qu'il prend en charge** (ex. images tridimensionnelles, audiovisuel, bases de données, images, logiciels, sites web, tableurs, etc.), **les formats de fichiers qu'il est capable de traiter**.

Chaque fiche permet également de fournir toute ressource utile sur l'outil, de signaler où trouver davantage d'informations sur lui et de rassembler les retours d'expériences des utilisateurs.

Il est possible de faire des recherches par outil, par fonction, par catégorie d'objet ou par format de fichiers.

Alimenté à l'origine par des organisations comme le Digital Curation Center britannique, le NDSA ou OPF, **COPTR est désormais accessible à toute personne souhaitant contribuer à l'alimentation du registre**.

3.2. Wikidata for digital preservation

Disponible à l'adresse suivante : <https://wikidp.org/>

Wikidata for digital preservation est un portail centralisé permettant d'accéder sous une forme structurée lisible par des machines à des informations relatives aux métadonnées, aux logiciels, aux formats de fichiers, aux outils permettant d'identifier et de préserver les objets numériques dans le

temps. Le fait que ces informations soient accessibles sous une forme lisible par machine permet aux spécialistes de la préservation numérique d'automatiser un certain nombre d'actions à partir de ces informations. La forme retenue pour la structuration des informations permet également des interrogations plus flexibles.



À titre d'exemple, les informations relatives aux formats de fichiers dans ce portail comprennent des informations sur les créateurs de ces formats, sur la manière dont ces formats sont liés à d'autres formats, sur leur type MIME, sur l'identification dans le registre PRONOM, sur les « signatures » de ces formats de fichiers, etc.

3.3. Carte heuristique des outils de préparation des versements

Disponible à l'adresse suivante : <https://framindmap.org/c/maps/433390/public>

Cette carte heuristique, conçue par l'archiviste français Julien Benedetti, recense des outils ouverts (open source), destinés à aider les archivistes souhaitant préparer le versement de documents d'archives sur support numérique, pour les actions suivantes :

- **contrôle de l'intégrité :**
 - outils de calcul d'empreinte ;
 - outils de copie contrôlée ;
- **classement et tri :**
 - exploration d'arborescence ;
 - renommage de fichiers ;
 - dédoublonnage ;
 - réorganisation de fonds d'archives ;
- **préservation :**
 - identification et validation de format ;
 - extraction de métadonnées ;
 - conversion de formats de fichiers.

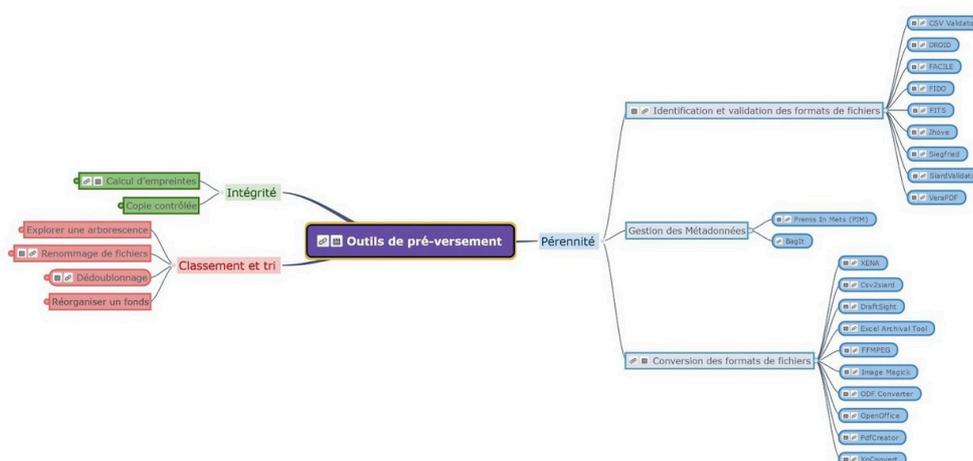
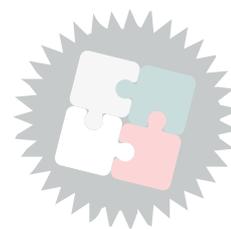


Fig. 1 : carte heuristique des outils de préparation des versements (source : <https://framindmap.org/c/maps/433390/public> ; auteur : Julien Benedetti)

Chaque outil est décrit dans une brève note et est disponible via un lien vers son entrepôt de référence.

Conclusion



Le module est terminé, nous espérons que celui-ci vous a été utile pour mieux appréhender la préservation des documents numériques. Maintenant, c'est à vous d'expérimenter et de partager avec la communauté des archivistes et des spécialistes de la préservation numériques vos expériences et les connaissances que vous aurez acquises.