

La cellule Formats

Pour une expertise nationale sur les formats de fichiers

Lorène Béchard & Mélanie Rebours

#PINFormats

Webinaires PIN - 4 novembre 2020

2010's



Les travaux du Programme Vitam sur la préservation

Identifier les besoins en terme de préservation

- les fonctionnalités à intégrer au logiciel Vitam : Identification, Conversion de format...
 - https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/VITAM_Preservation.pdf
- les proportions des principaux formats (type de format, nombre de fichier, volume des fichiers) et ceux spécifiques à la France non identifiés par le référentiel PRONOM (<https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)

Un groupe de travail réuni pendant 2 ans

- Les membres du Programme Vitam
- Des interventions d'institutions extérieures en pointe sur le sujet : Bibliothèque nationale de France (BNF), Institut national de l'audiovisuel (INA), Centre Informatique National de l'Enseignement Supérieur. (CINES)...

Participation à des événements internationaux : IPRES, contact DPC, OPF...

- iPRES : Conférence internationale sur la préservation numérique (Boston 2018, Amsterdam 2019)
- DPC : Digital preservation Coalition, <https://www.dpconline.org/>
- OPF : Open preservation Foundation, <https://openpreservation.org/>

Les livrables du chantier préservation

Identification des formats de fichiers :

https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/20200131_NP_Vitam_preservation-identification-format-v2.0.pdf

Validation des formats de fichiers :

https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/20200131_NP_Vitam_preservation-validation-format-v2.0.pdf

Extraction de métadonnées :

https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/20200131_NP_Vitam_preservation-extraction-MD-v2.0.pdf

L'étude du SIAF sur la préservation

Identifier les ressources à allouer au sujet de la préservation par les services d'archives

- Définir le sujet : périmètre, connaissances, concepts et normes en vigueur
- Modéliser les processus et des scénarios à mettre en œuvre : les ressources nécessaires...
- Les livrables sont consultables sur le site FranceArchives : <https://francearchives.fr/fr/article/162808225>



Un sujet de niche ?

Un besoin impérieux de mutualisation : partager les connaissances, les savoir-faire

Un besoin de créer une communauté : uniquement quelques « sachants » à l'échelon national

- Un besoin de participer aux événements internationaux pour enrichir les réflexions et connaissances
 - Pose la problématique de la langue

Comment poursuivre les efforts ?

Partager la charge pour assurer une meilleure progression et prise en compte du sujet

- Ne pas repartir de 0
- Gagner du temps en partageant les travaux effectués

Créer une « cellule » sans complexité administrative

- Trouver une institution « neutre » pour le portage du sujet, afin de fédérer les établissements et institutions concernées

LA solution ? s'agréger au Groupe PIN déjà existant et dédié à la préservation en y apportant des forces vives et de nouvelles problématiques

Le Groupe PIN



Un lieu d'échange qui fête ses 20 ans, de réflexion et d'action pour faire émerger des solutions dans les domaines suivants :

- Les informations à pérenniser
- Les questions juridiques
- Les systèmes d'archivage
- Les problèmes d'organisation sous-jacents

2019



Cellule nationale de veille sur les formats

➤ Lancement en 2019 d'une cellule de veille nationale sur les formats de fichiers

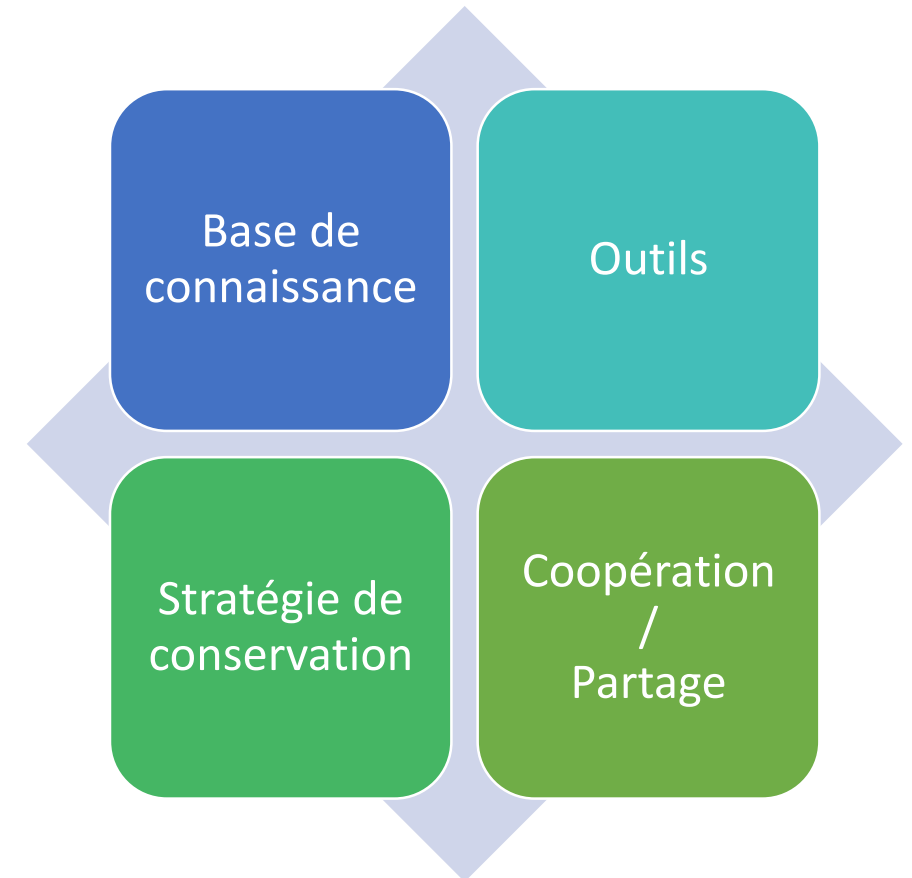
- Sous l'égide de l'association Aristote et de son groupe de travail PIN (Pérennisation de l'information numérique)
- Une dizaine de partenaires
- Participation ouverte, soumise à adhésion à l'association Aristote



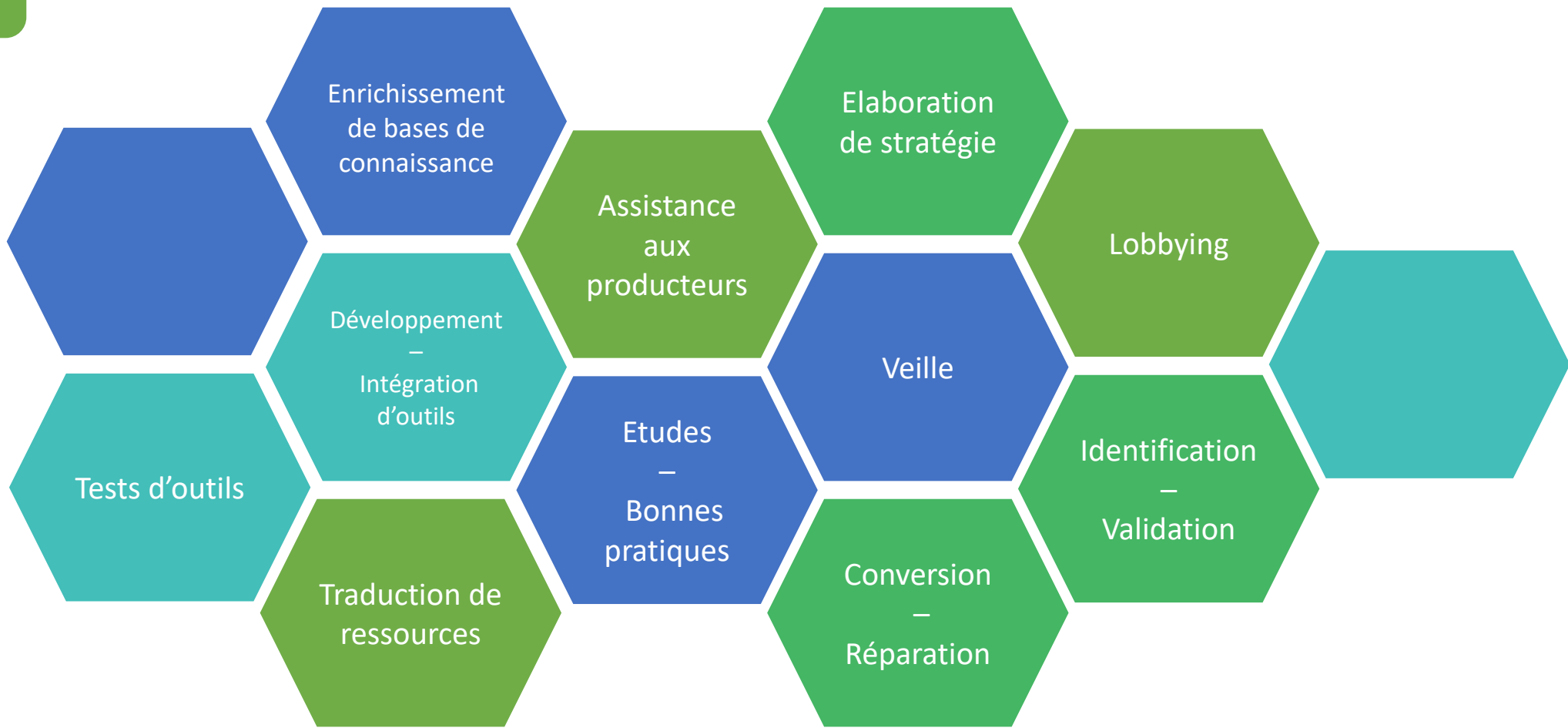
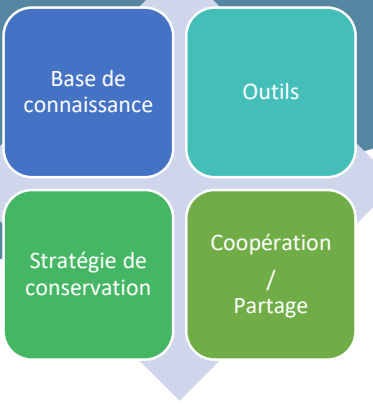
Cellule nationale de veille sur les formats

➤ Les objectifs

- Mutualiser les activités de veille
- Sensibiliser les professionnels sur le sujet
- Contribuer et influencer sur les outils
- Etre l'interlocuteur francophone dans les travaux internationaux

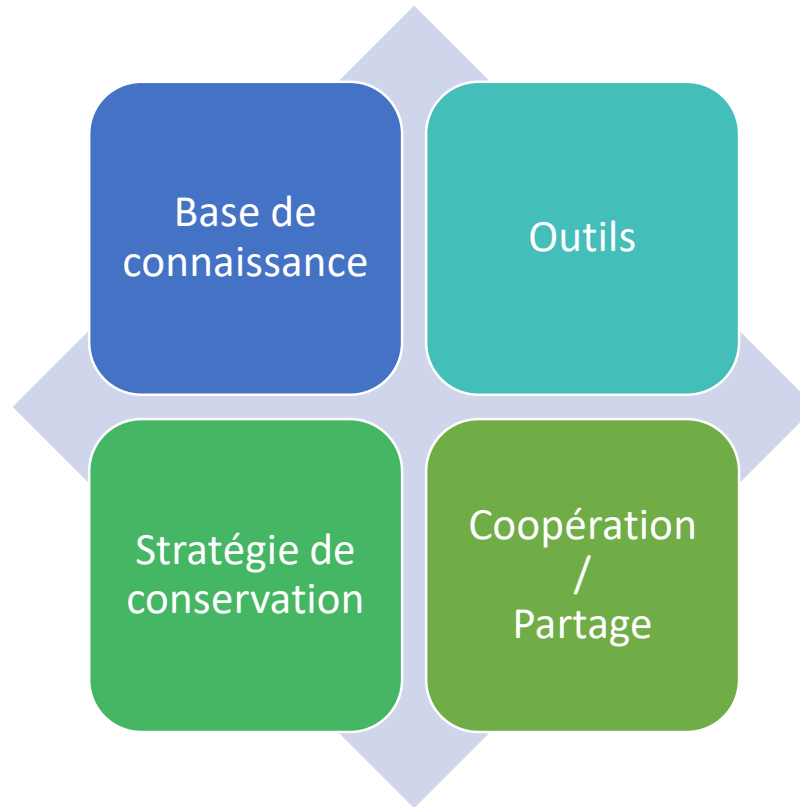


La pérennité des formats : panorama des activités



Axes de travail

- Identifier les expertises existantes sur les différents types de formats / Etablir une liste de sources génériques de connaissance sur les formats
- Etablir une cartographie des formats utilisés en France, notamment les formats spécifiques
- Classifier les formats en fonction de leur pérennité
- Mettre en place un dépôt national de conservation des spécifications des formats de fichiers
- Identifier les champs d'études nouveaux : par exemple, travailler sur les formats 2D, 3D et les formats géographiques
- Recenser les jeux de fichiers de tests existants / Mettre en place un corpus de fichiers de tests



- Partage d'expérience suite à tests d'outils : VeraPDF par ex
- Recenser les outils/méthodologie de réparation des formats
- Evaluer les problèmes spécifiques posés par la langue française vis-à-vis de formats et outils, produits par des anglo-saxons
- Echanger sur les outils d'extraction de métadonnées et sur les formats de métadonnées techniques
- Identifier des outils pour l'analyse automatique du contenu des fichiers

- Identifier et traduire les sources anglo-saxonnes indispensables
- Mettre en place des actions de communication, ateliers pratiques, webinaires, etc.
- Elaborer une documentation pédagogique sur la notion de « fichier » et de « format de fichiers »

- Echanger sur les pratiques de validation : quels outils pour quels formats ?
- Echanger / Définir les critères d'obsolescence des formats
- Identifier des critères pour le choix d'une stratégie de conservation : Conversion vs Emulation
- Améliorer les pratiques de préservation du XML

Composition de la cellule et programme de travail 2019-2020

Sous-groupe « Identification des expertises »

- Constitution d'un annuaire des expertises en France et à l'étranger

Sous-groupe « Connaissance des formats et critères d'obsolescence et de pérennisation »

- Recensement des formats existants, notamment les formats spécifiques
- Définition de critères d'obsolescence et de pérennité pour les formats

Sous-groupe « Outils et Corpus »

- Alimentation de Wikidata (utilisation du modèle de cycle de vie du DCC et du registre COPTR (Community Owned digital Preservation Tool Registry : https://coptr.digipres.org/Main_Page)

Sous-groupe « Traduction »

- Traduction de la grille d'évaluation de la NDSA (National Digital Stewardship Alliance : <https://ndsa.org/publications/levels-of-digital-preservation>)
- Traduction du Rapid Assessment Model du DPC : <https://www.dpconline.org/digipres/dpc-ram>
- Traduction du Handbook sur la préservation numérique élaboré par le DPC : <https://www.dpconline.org/handbook>

et vous, qu'en pensez-vous ?



Rendez-vous sur www.menti.com

menti.com



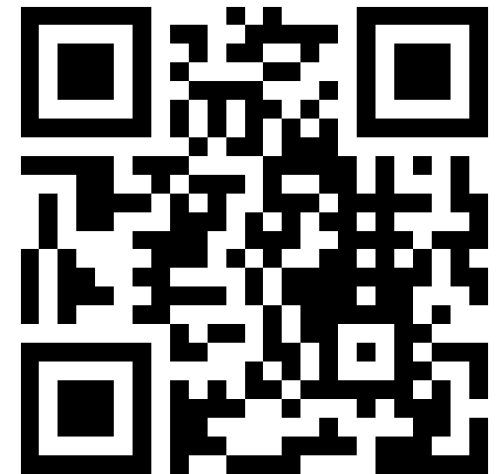
Please enter the code

12 34 56

Submit

The code is found on the screen in front of you

Sondage n°1 : Code 15 75 50 8



Rendez-vous sur www.menti.com

🔒 menti.com

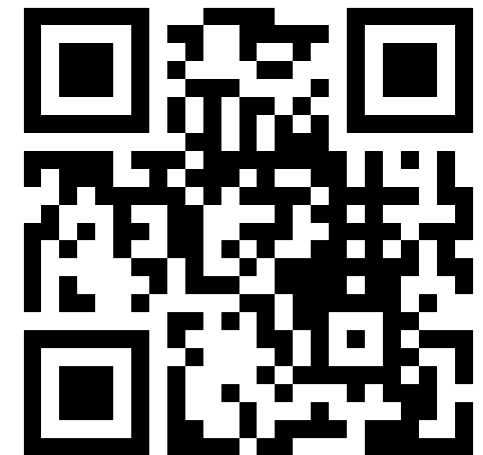


Please enter the code

Submit

The code is found on the screen in front of you

Sondage n°2 : Code 81 28 13 1



Qu'attendez-vous d'une cellule sur les formats ?

outils collaboratifs

recommandation et outils

Guides de bonnes pratiques

Des retours d'expériences

Connaître les formats pérennes à privilégier dans nos administrations pour chaque format que nous utilisons surtout lorsque l'on a pas de SAE.

Expérimentations sur les outils et recommandations

Recommandation MD exif et IPTC

Mutualisation des outils et méthodes d'analyse des formats de fichiers

Préconisations et aides à la gestion des formats de mon institution

Des questions ?



Merci de votre attention !

Contact : secretariat@association-aristote.fr

LES FORMATS, C'EST LONG,
C'EST TECHNIQUE, ÇA
N'INTÉRESSE PERSONNE...
MAIS ÇA PEUT SAUVER VOS
DOCUMENTS ET DONNÉES !

