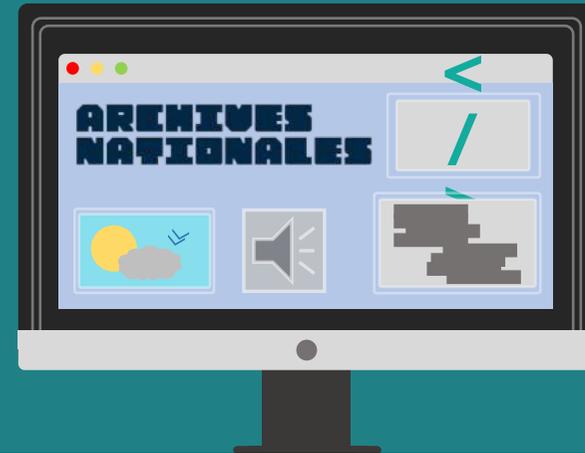


Stratégie de préservation

Retour d'expérience des Archives nationales

Webinaire PIN n°4
09/11/2021

#PINFormats

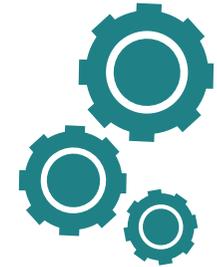


2021
2025



Programme de la séance

#PINFormats



01.

**Retour
d'expérience des
Archives nationales**

*Par Emeline
Levasseur, cheffe de
projet « reprise des
données numériques
et préservation »*

02.

**Actualité
internationale de la
préservation
numérique**

*iPres
World digital
Preservation Day*

03.

**Actualités de la
cellule**

*Derniers travaux
& objectifs 2022*



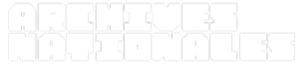
**ARCHIVES
NATIONALES**



Stratégie de préservation numérique – RetEx Archives nationales

Webinaires PIN, cellule de veille sur les formats - 9 novembre 2021 #PINFormats





- **Héritage et nécessaire renouvellement**
- **La préservation numérique : objectif de la stratégie 2021-2025 des Archives nationales**
- **Chantiers en cours : reprise des données, recommandations sur les nouvelles entrées, ...**
- **Les perspectives**

Héritage et nécessaire renouvellement

L'héritage de CONSTANCE

Première collecte d'archives numériques en 1983, suite à la mise en place de **CONSTANCE**

CONSTANCE (CONSeRvation et TraiteMent des Archives Nouvelles Constituées par l'Électronique) = politique, processus et méthode de **traitement et de conservation des données numériques et de leurs métadonnées**

Méthode adaptée à la collecte et à la préservation des données des enquêtes qui ont constitué **les premières bases informatiques** créées par des ministères et des institutions pionniers dans ce domaine (**l'INSEE et l'INED**, ou encore **le ministère de l'Agriculture** et l'ancien **ministère des Transports**).

Héritage et nécessaire renouvellement

L'héritage de CONSTANCE

Méthode pensée pour la collecte et la conservation des données issues des enquêtes.

Prévoit de :

- Pérenniser les fichiers de données sur des bandes LTO (stockage à froid), à plat, et nommés de manière homogène :
NoticeProducteur_NuméroVersement_NumeroArticle_NomFichier.extension
- Gérer les métadonnées qui décrivent les données dans des bases documentaires (Cindoc)
(« **informations de représentation** »)
- Conserver la documentation associée aux entrées pour comprendre le contexte de production et documenter les opérations d'archivage
- **L'OAIS avant la norme OAIS** qui définit le contenu d'information comme le plus petit élément archivable = objet-données + information de représentation

Héritage et nécessaire renouvellement

L'héritage de CONSTANCE

Les données :

- Représentent les réponses aux questionnaires
- Encodées en ASCII (American Standard Code pour Information Interchange, norme d'encodage des caractères) selon le principe 1 donnée = 1 à n caractères, et il y a autant d'octets que de caractères
- Archivées à plat dans des documents de type texte et stockées sur bande LTO
- Les fichiers de données sont gérés et décrits dans une base documentaire



Héritage et nécessaire renouvellement

L'héritage de CONSTANCE



Les **métadonnées** (informations de représentation) :

Elles sont réalisées grâce à une association entre le producteur et les Archives Nationales et sont indispensables pour accéder aux données et les comprendre.

Dans les bases de données CINDOC ont été conservées les fiches d'application (information sur l'application versée), dictionnaires de données, fiches de structure, dictionnaires de codes

La **documentation associée papier** (numérisée) :

Elle contextualise la production des données (instructions aux enquêteurs, questionnaires d'enquête vierges, bilan de l'enquête, publications de la recherche) et documente les échanges entre le producteur et les AN

Cet ensemble fournit la signification des données et décrit comment les données sont représentées dans les fichiers.

Un exemple

Enquête mobilité géographique et insertion sociale en 1992 (AN, 19990409), enquête de l'INSEE : une enquête sur la vie des immigrants et de leurs enfants en France.

Les données

```
FR110170212006971191135 0002 090 01 01131129365012 1 1
0011017020069720511OUI91135 0002
1100011533617000773207000110220203103
0112360123531 1 12 12
0111017022 00 1222 4 0 01 99226175056012202
11 0122 222229901201201201222 698788883309080100000004111 2 462
250 2502501411521198469011 312017225010172122
031101702 0100202
041101702010272KARIM 12509811
051101702011 320031 270852 1 11 2 2
061101702KABYLE 170FRANCAIS
11122321211155121221221223322
2 2 1 2623503222062511121 12202039005007732
071101702010370726828433921126
071101702020368706828433921126
071101702030357676725532121126
101101702040153571 20526740
```

Les métadonnées

SYMB	N-STRUCT	POS-DEPALONG	POS-ARRI	TYP-DONN
TYPE	Enregistrement 00	1	2	2
IDENT		3	7	9
NENQ		10	5	14
SEXE		15	1	15
JOUR		16	2	17
MOIS		18	2	19
CONF		20	3	22
D		23	2	24
C		25	3	27
CIL		28	3	30
IL		31	4	34
FIL		35	1	35
NLOT		36	3	38
NCHIF		39	4	42
R1		43	1	43
R2		44	2	45
R3		46	2	47
R4		48	2	49
R5		50	2	51
R6M		52	2	53
R6A		54	2	55

Extrait de la fiche de structure

Extrait du dictionnaire des données

NOM	A-DEB	A-FIN	SYMB
SEXE DE L'ENQUETE (CALCULE)	1992	1992	SEXE

Extrait des codes

SEXE

Sexe de l'enquêté (calculé)

1 : homme

2 : femme

Un exemple

Enquête mobilité géographique et insertion sociale en 1992 (AN, 19990409), enquête de l'INSEE : une enquête sur la vie des immigrants et de leurs enfants en France.

```
FR110170212006971191135 0002 090 01 01131129365012 1 1
00110170200697205110UI91135 0002
1100011533617000773207000110220203103
0112360123531 1 12 12
0111017022 00 12222 4 0 01 99226175056012202
11 0122 222229901201201201222 698788883309080100000004111 2 462
250 2502501411521198469011 312017225010172122
031101702 0100202
041101702010272KARIM 12509811
051101702011 320031 270852 1 11 2 2
061101702KABYLE 170FRANCAIS
11122321211155121221221223322
2 2 1 2623503222062511121 12202039005007732
071101702010370726828433921126
071101702020368706828433921126
071101702030357676725532121126
101101702040153571 20526740
```

Le sexe de l'enquêté se trouve en position 15 (15e caractère de l'enregistrement), et est de longueur 1.

1 = homme, 2 = femme

Dans cet enregistrement, la personne enquêtée est une femme.

Héritage et nécessaire renouvellement

Préserver des archives de plus en plus nombreuses et diverses

A partir des années 2000 :

- De plus en plus de services sont concernés par la production d'archives nativement numériques
- La collecte d'archives électroniques augmente de manière exponentielle
- Les typologies des archives collectées se diversifient :
 - Données structurées : données d'enquêtes statistiques, données extraites d'applications
 - Documents bureautiques
 - Documents audiovisuels et sonores
 - Photographies numériques natives
 - Messageries

La méthode s'adapte dans une certaine mesure, par ex : conversion des fichiers Word en PDF

Héritage et nécessaire renouvellement

Adamant : l'ouverture d'une nouvelle phase

2015 : lancement du programme **ADAMANT** (Administration des archives et de leurs métadonnées aux Archives nationales dans le temps). Vise à adapter les outils, les procédures et l'organisation de la chaîne archivistique aux enjeux du traitement des données nativement numériques

Novembre 2018 : mise en service de la plate-forme d'archivage électronique

- S'appuie sur la solution logicielle Vitam (*back office*)
- S'interface avec le SIA
- Prend en charge de grands volumes d'archives nativement numériques

Renouvellement nécessaire des méthodes de traitement et de conservation.

En premier lieu :

- Nouvelle infrastructure d'hébergement de données, localisées sur deux sites distants
- Service de stockage à froid

La préservation numérique : objectif de la stratégie 2021-2025 des Archives nationales

La stratégie des Archives nationales suit les grands principes du **cadre stratégique commun de modernisation des archives**.

Construction des « Archives nationales numériques » et des nouveaux usages du numérique : collecte, pérennisation, services à distance, présence sur le web.

Volonté de mettre en place une **mutualisation des savoirs sur la pérennisation pour construire une stratégie de préservation des données et des métadonnées** renouvelant l'approche du système et de la méthode Constance (objectif 22 de l'axe 4, « une mémoire collective pour les générations futures »)

Stratégie qui doit être adaptée aux spécificités des archives collectées par les Archives nationales => **il n'y a pas de solution « clé en main »**

La préservation numérique : objectif de la stratégie 2021-2025 des Archives nationales

Les actions en cours ou à engager :

- Reprise des données du patrimoine numérique des Archives nationales (1983-2018)
- Recommandations sur la collecte : première politique formats, pérennisation des messageries électroniques
- Prise en charge des archives audiovisuelles
- S'inscrire dans les réseaux professionnels, nationaux et internationaux
- Rédiger une stratégie de préservation
- Création et maintien d'une base de connaissance sur les formats
- Fixer un programme d'analyse des formats (actuels et nouveaux)
- Réflexion sur les fonctionnalités à développer liées à la préservation dans la plate-forme (audit, pilotage, actions de préservation)

Les chantiers en cours

La reprise du patrimoine numérique des Archives nationales (1983-2018)

Une histoire qui s'inscrit dans la continuité des 6 migrations précédentes (CONSTANCE)

Objectif : Garder les données et les métadonnées intègres, authentiques, lisibles et accessibles dans le temps

Le chantier couvre 73 téraoctets de données.

Comment ?

- Rassembler toutes les données et les métadonnées pour créer les moyens de les faire converger vers le SIA numérique
- Préparer les données et les métadonnées pour les rendre conformes au Standard d'Echange des Données pour l'Archivage (SEDA 2.1 compatible VITAM)
- Des données à plat à transformer en données dans une vue arborescente structurée par les métadonnées

Les chantiers en cours

La reprise du patrimoine numérique des Archives nationales (1983-2018)

Organisation du chantier :

- Définition et reprise de lots présentant des caractéristiques communes permettant d'industrialiser le plus possible les processus de reprise :
 - Enquêtes statistiques (gérées et décrites par la méthode Constance), déjà versés dans la plateforme ou en cours
 - Données (très souvent bureautiques) décrites par des instruments de recherche EAD
 - Reportages photos et vidéos présidentiels et Premier ministre
- Reprise de petits ensembles de versements (itératifs et ouverts), notamment de bureautiques ou de données structurées extraites d'applications métier.
- Réflexions à mener sur les données qui n'entrent pas dans les périmètres ci-dessus

Les chantiers en cours

Recommandations en entrée : une première politique formats

Mise à disposition de préconisations sur les formats, à côté des fiches-guides sur la constitution des paquets à verser (SIP), à destination des archivistes des missions dans les ministères.

Attention, ces choix sont valables dans le contexte Archives nationales, selon son offre de service et ses moyens pour assurer la lisibilité et l'exploitabilité des données sur le long-terme

Les Archives nationales indiquent ainsi :

- Des formats « acceptés » : les AN s'engagent à les préserver avec audit régulier
- Des formats « tolérés » : les AN les collectent sans garantie de préservation
- Des formats « refusés » : les AN ne les collectent pas pour des raisons techniques ou en l'absence de valeur juridique ou d'intérêt historique.

Elles précisent aussi les traitements à prévoir si cela est nécessaire

Cette liste est vouée à évoluer et à être complétée en fonction des formats de fichiers proposés dans les projets d'entrée (découverte de formats « exotiques »)

Types de données	Formats collectés	Acceptés par les AN, Tolérés, Refusés	Traitement à prévoir avant entrée aux AN
Données structurées, systèmes d'information, applications métier, bases de données	Contacter le DAEAA : daeea.archives-nationales@culture.gouv.fr		
Bureautique	ODF "Open Document Format" (ODT, ODP, ODS, etc.)	Acceptés	Sans traitement
	Suite Microsoft Office (DOC, DOCX, XLS, XLSX, PPT, PPTX, etc.)	Acceptés	Sans traitement
	Suite Microsoft Office avec Macro (DOCM, PPTM, etc.)	Tolérés	Sans traitement
	TXT (codage UTF-8)	Acceptés	Sans traitement
	PDF, PDF/A (ISO 19005), PDF/A-1b, PDF/A-2	Acceptés, avec les fichiers dans leur format d'origine s'il s'agit d'une conversion	Sans traitement

Compression bureautique	ZIP, TAR, TARBZ2, TARGZ2	Refusés	Décompression dans ReSIP Décompression en dehors de ReSIP
	7ZIP, RAR, etc.	Refusés	
	EML	Acceptés	Extraction des pièces jointes et du contenu textuel dans un fichier TXT
Messageries	Autres formats de message (MSG, etc.)	Tolérés	Conversion en EML et traitement (<i>cf. ci-dessus</i>)
	Conteneur PST	Acceptés	Extraction des messages au format EML (<i>cf. ci-dessus</i>)
	Autres formats de conteneur courriels (Mbox, etc.),	Tolérés	Extraction des messages au format EML (<i>cf. ci-dessus</i>)
	Carte de visite (.vcf)	Refusés	

Pages Web (html, css, javascript, etc.)	Contacter le DAEAA : daaaa.archives-nationales@culture.gouv.fr		
Fichiers système, fichiers temporaires, raccourcis	Thumbs.db, Desktop.ini, pagefile.sys, etc.	Refusés	

Images fixes	RAW, JFIF, "JPEG Image File Format" (JPG)	Acceptés	Sans traitement
	TIFF	Acceptés	Sans traitement
	PNG	Tolérés	Sans traitement
	GIF	Tolérés	Sans traitement
Images animées	<i>Conteneurs non compressés</i> (RVB 4.4.4 et YUV 4.2.2) AVI, MOV, MXF, MKV <i>Fichiers contenus</i> FFV1, DV	Acceptés	Sans traitement
	<i>Fichiers vidéo compressés (codec)</i> : mpeg1, 2, 4/H264, H265, cinepak, WMV	Tolérés	Sans traitement
	<i>Fichiers vidéo non compressés (codec)</i> DPX, JPEG 2000, Apple ProRes, AV1	Tolérés	Sans traitement, puis conservés en l'état ou convertis en mpeg4 par les AN avant archivage
Fichiers son	<i>Fichiers non compressés</i> : WAV(BWF), AIFF, PCM	Acceptés	Sans traitement
	<i>Fichiers compressés</i> : MP3, WMA, FLAC, AAC, OGG	Tolérés	Sans traitement

Les chantiers en cours

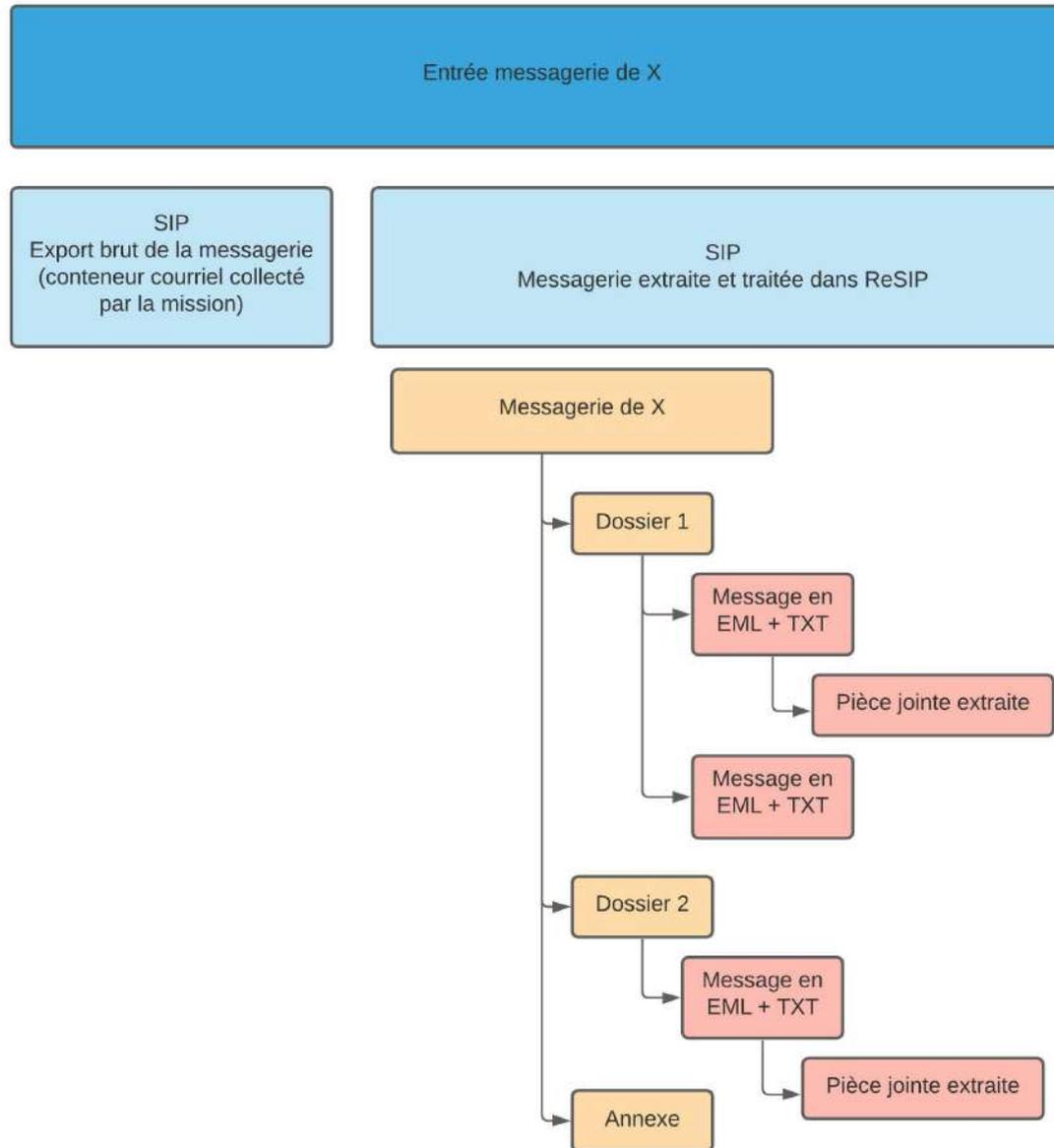
Préservation des messageries électroniques

Objectifs :

- Garantir la pérennisation des messages et de leurs pièces jointes
- Répondre si besoin à des demandes d'accès
- Garder des possibilités futures de représentation des messageries dans des environnements extérieurs au SIA

Préconisations => verser 2 SIP :

- Un SIP avec la messagerie extraite et traitée dans ReSIP, représentée dans une arborescence. Les messages sont archivés en EML et en TXT, les pièces jointes sont extraites et reliées aux mails.
- Un SIP avec l'export brut (PST ou Mbox), tel que collecté par la mission



Les chantiers en cours

Extension sur les métadonnées des archives audiovisuelles

Enjeu : préserver et décrire des archives audiovisuelles en utilisant le SEDA 2.1 dans le contexte Vitam

Constat :

- le SEDA 2.1 n'est pas adapté à nos besoins de description de ces archives (exemple : témoignages oraux, enregistrements de colloques, etc.) et nous avons besoin de pérenniser et d'accéder aux métadonnées techniques des fichiers
- Le copil SEDA a choisi de ne pas faire évoluer la prochaine version en intégrant les balises pour l'audiovisuel

Solution : création d'une extension dans la partie solution (SIA des AN) pour ajouter de nouvelles métadonnées qui seront acceptées dans le logiciel Vitam

Comment ? S'appuyer sur des schémas existants et notamment le MPG-7, norme de description de contenu multimédia, implémenté également à la BnF (pour les métadonnées techniques)

Les chantiers en cours

Extension sur les métadonnées des archives audiovisuelles

Lister les métadonnées dont nous avons besoin : durée, licence de réutilisation, codec, fréquence d'images, taille de l'image, profondeur de codage, fréquence d'échantillonnage, nombre de canaux, ...

A partir de celles-ci, définir :

- Noms des balises (en se rapprochant le plus possible de l'existant)
- Organisation
- Cardinalité
- Formats des valeurs (texte, nombre entier, date, ...)

Et après ? Réflexion sur l'automatisation de l'extraction de ces métadonnées (MediaInfo)

Les perspectives

Pour que le renouvellement de stratégie soit total, il faut que les Archives nationales :

- Se dotent et rédigent une politique de préservation, complètement adaptée à son contexte
- Travaillent de manière plus collaborative avec des experts en interne et à l'extérieur (national et international)
 - Cellule nationale de veille sur les formats
 - Dutch Digital Heritage Network (DDHN)
 - Archives nationales du Duché du Luxembourg
 - Bibliothèque et Archives nationales du Québec (BAnQ)
- Approfondissent leur connaissance des formats : bases de connaissance, programme de formats à analyser
- Travaillent sur les outils (à l'intérieur et en dehors du SIA) : audits, pilotage et actions de préservation

Merci de votre attention !
Des questions ?

Emeline Levasseur : emeline.levasseur@culture.gouv.fr

Actualités internationales

iPRES 2021 | World Digital Preservation Day

Par Yannick Grandcolas (BnF)

#PINFormats



iPRES 2021



General operation of iPRES2021

Onsite: National Science Library, CAS, Beijing



National Science Library, CAS



Exhibition Hallway



Academician's Hall



7D Meeting Room



Lecture Hall

iPRES 2021

Zoom Réunion

Vous voyez actuellement l'écran de Zoom#1-Speaker

Options d'affichage

Enregistrement



Wang Xudong
Director of the Palace Museum



2D Collection Digitization

数字故宫

Accelerating 2D collection photography since 2020

- 15 collection photography units, digitizing 80,000 artifacts, approx. 225,000 images/year
- entire collection expected to be digitized within 10 years

690,000

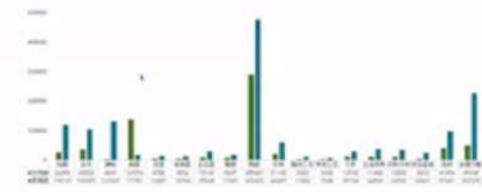
artifacts digitized

1.47 million

digital images

190 TB

data



Year	Artifacts Digitized	Digital Images	Data (TB)
2020	10,000	30,000	10
2021	20,000	60,000	20
2022	30,000	90,000	30
2023	40,000	120,000	40
2024	50,000	150,000	50
2025	60,000	180,000	60





iPRES 2021



General operation of iPRES2021

iPres202 Program, <https://ipres2021.ac.cn/>

Date	Beijing	Washington DC	Chicago	London	Canberra	Berlin	Zoom#1	Zoom#1	Zoom#2	Zoom#3	Zoom#4
	UTC+8	UTC-5	UTC-6	UTC+0	UTC+11	UTC+1	Lecture Hall	Lecture Hall	Academics Hall	TD Meeting Room	Small Meeting Room
Oct 19	18:00-19:30	06:00-07:30	04:00-05:30	11:00-12:30	21:00-22:30	12:00-13:30		Tutorial 1	Workshop 1	Tutorial 2	Hackathon
	20:00-21:30	08:00-09:30	07:00-08:30	13:00-14:30	23:00-00:30	14:00-15:30		Tutorial 1	Workshop 2	Workshop 3	Hackathon
Oct 20	16:00-17:30	04:00-05:30	03:00-04:30	09:00-10:30	19:00-20:30	10:00-11:30	Plenary 1- Opening Ceremony				
	18:00-19:30	06:00-07:30	05:00-06:30	11:00-12:30	21:00-22:30	12:00-13:30		Paper Session 1	Paper Session 2	Panel 1	
	20:00-21:30	08:00-09:30	07:00-08:30	13:00-14:30	23:00-00:30	14:00-15:30		Paper Session 3	Paper Session 4	Panel 2	
Oct 21	16:00-17:30	04:00-05:30	03:00-04:30	09:00-10:30	19:00-20:30	10:00-11:30	Plenary 2- Plenary Speech Session				
	18:00-19:30	06:00-07:30	05:00-06:30	11:00-12:30	21:00-22:30	12:00-13:30		Paper Session 5	Paper Session 6	Paper Session 7	
	20:00-21:30	08:00-09:30	07:00-08:30	13:00-14:30	23:00-00:30	14:00-15:30		Panel 3	Paper Session 8	Panel 4	
	22:00-23:30	10:00-11:30	09:00-10:30	15:00-16:30	01:00-02:30 (+1day)	16:00-17:30					STC Meeting
Oct 22	08:00-11:30	21:00-21:30 (-1day)	20:00-22:30 (-1day)	02:00-11:30	12:00-14:30	03:00-06:00	Professional Visit				
	16:00-17:30	04:00-05:30	03:00-04:30	09:00-10:30	19:00-20:30	10:00-11:30	Open Local Event				
	18:00-19:30	06:00-07:30	05:00-06:30	11:00-12:30	21:00-22:30	12:00-13:30		Session 9 (Lightning Talks 1)	Session 10 (Posters/Dissemination)	Panel 5	
	20:00-21:30	08:00-09:30	07:00-08:30	13:00-14:30	23:00-00:30	14:00-15:30		Session 11 (Lightning Talks 2, Hackathon Presentation)	Paper Session 12	Paper Session 13	
	22:00-23:30	10:00-10:30	09:00-10:30	15:00-16:30	01:00-02:30 (+1day)	16:00-17:30	Plenary 3- Closing Ceremony				

Paper Sessions	10
Panels	5
Poster Session	1
Lightning Talks + Hackathon	2
Hackathon Event	2
Workshops Session	3
Tutorials Session	3
Plenary Sessions	3
Open Local Event	1
Professional Visit	1

iPRES 2021

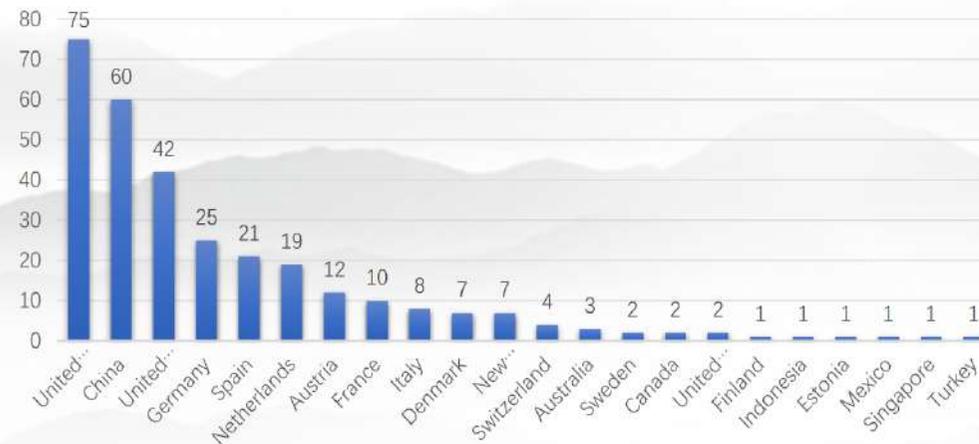


Basic Statistics of iPRES2021

Contributions

	Accepted
Shor Papers	23
long Papers	18
Panels	5
Posters	9
Demonstrations	1
Workshops	3
Tutorials	3
Hackathons	1
Lightning talks	25

305 authors from 22 Countries



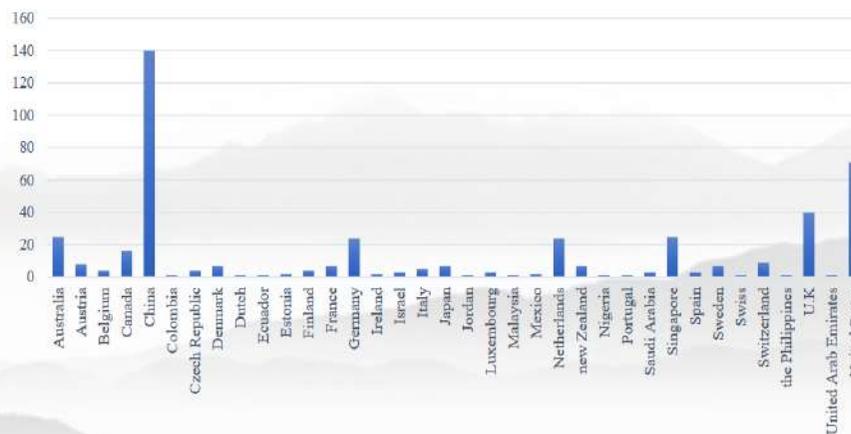
iPRES 2021



Basic Statistics of iPRES2021

Participation

462 Registrants from 36 Countries



Participation at iPRES2021

Oct.19: 205+45/1800-1930; 203+35/2000-2130

Oct.20: 177+140/1600-1730, **Opening Ceremony**

Oct.20: 263+55/1800-1930

Oct.20: 244+34/2000-2130

Oct.21: 152+50/1600-1730, **Plenary Keynote Speeches**

Oct.21: 201+45/1800-1930

Oct.21: 211+35/2000-2130

Oct.22: 87+35/1600-1730, **Open Local Event**

Oct.22: 208+44/1800-1930

Oct.22: 202+35/2000-2130

Oct.22: 154+5/2200-2330

iPRES 2021

HOME
ABOUT
REGISTRATION
SPONSORS
CALL FOR CONTRIBUTIONS
PROGRAM
BEIJING
CONTACT
Congress has ended

Program Structure

19 Oct. Program >

20 Oct. Program

21 Oct. Program

22 Oct. Program

Keynotes

Panels

Workshops/Tutorials

COP TR-EDIT-A-THON

iPRES2021 Awards

Professional Visit

UTC+0 Europe/Africa UK(London)
 UTC+11 Oceania-Australia(Canberra)
 UTC+1 Europe/Africa-German (Berlin), Danmark(Copenhagen), Netherlands(Amsterdam)

Beijing UTC+8	Washington DC UTC-5	Chicago UTC-6	London UTC+0	Canberra UTC+11	Berlin UTC+1	Zoom#1	Zoom#1	Zoom#2	Zoom#3	Zoom#4
						Lecture Hall	Lecture Hall	Academics Hall	7D Meeting Room	Small Meeting Room
1800-1930	0600-0730	0400-0530	1100-1230	2100-2230	1200-1330		Tutorial 1	Workshop1	Tutorial2	Hackathon
2000-2130	0800-0930	0700-0830	1300-1430	2300-0030	1400-1530		Tutorial 3	Workshop2	Workshop3	Hackathon

1800-1930
Lecture Hall

TUTORIAL 1
Zoom#1

UNDERSTANDING AND IMPLEMENTING METS - A TUTORIAL
Karin Bredenberg (Kommunalförbundet Sydskövle, Sweden), Juha Lehtonen (CSC – IT Center for Science, Finland), Tobias Steinke (Deutsche Nationalbibliothek, Germany)

1800-1930
Academician's Hall

WORKSHOP 1
Zoom#2

WORKSHOP ON TRUSTWORTHINESS AUDITING FOR SUSTAINABLE DIGITAL PRESERVATION SERVICES
XL Zhang (National Science Library, CAS, China), ZK Wu (National Science Library, CAS, China), YH Tong (Library of Peking University, China)

1800-1930
7D Meeting Room

TUTORIAL 2
Zoom#3

RIGHTS MANAGEMENT FOR RESEARCH DATA MANAGEMENT, SHARING, AND PRESERVATION
Zeyu Zhang (Computing and Networking Information Center, CAS, China), Lulu Jiang (Computing and Networking Information Center, CAS, China), Zhihong Shen (Computing and Networking Information Center, CAS, China)

2000-2130
Lecture Hall

TUTORIAL 3
Zoom#1

UNDERSTANDING AND IMPLEMENTING PREMIS - A TUTORIAL
Karin Bredenberg (Kommunalförbundet Sydskövle, Sweden), Eld Zierau (The Royal Danish Library), Michelle Lindlar (TIB Leibniz Information Centre for Science and Technology)

2000-2130
Academician's Hall

WORKSHOP 2
Zoom#2

GUIDELINES FOR THE SELECTION OF DIGITAL HERITAGE FOR LONG-TERM PRESERVATION. WHAT, WHY, WHEN
Dr. Ingrid Parent (University of British Columbia / IFLA, Canada), Dr. Anthea Seles (ICA), Jenna Murdock Smith (Library and Archives Canada), Claire McGuire (The International Federation of Library Associations and Institutions / IFLA, Netherlands)

2000-2130
7D Meeting Room

WORKSHOP 3
Zoom#3

FROM CAPTURE TO REPLAY: WEB ARCHIVING WITH WEBRECORDER TOOLS
Lorena Ramirez Lopez (webrecorder USA), Ilse Krueger (webrecorder USA), Emma Dickson (webrecorder

iPRES 2021

ENHANCING FRENCH DIGITAL PRESERVATION COOPERATION !

Formats knowledge ? Tools?
Lost in translation ?

Translation !
#NDSA Level 2.0
#Handbook of Preservation

Experts !

Tools !

#Aristote: BnF + SIAF
+ Archives Nationales...

<https://association-aristote.fr>



WDPD 2021



The screenshot shows the Open Preservation Foundation website. The header features the logo on the left, a navigation menu with items like 'About', 'Products', 'Collaboration', 'Resources', 'Blogs', 'Events', 'News', and 'Membership', and a 'Member area | Login' link on the right. A search icon is also present. Below the header, there is a 'Back to all blogs' link. The main content area displays a blog post with the title 'Tout transformer pour que rien ne change,' by Yannick GRANDCOLAS, dated 4 November 2021. Below this, there is another section titled 'La stratégie de préservation de la BnF face au PSD' with a logo for 'Journée mondiale de la préservation numérique' dated 4 novembre 2021. At the bottom, there is a paragraph of text and a link 'Dans les épisodes précédents...'. A vertical scrollbar is visible on the right side of the page.

Open Preservation Foundation

Member area | Login

About Products ▾ Collaboration Resources ▾ Blogs Events News Membership ▾

Back to all blogs

Tout transformer pour que rien ne change,

 Yannick GRANDCOLAS
4 November 2021

La stratégie de préservation de la BnF face au PSD

 Journée mondiale de la préservation numérique
4 novembre 2021

Par Alix Bruys, Bertrand Caron, Yannick Grandcolas, Thomas Ledoux, & Anne Paounov. Journée mondiale de la préservation numérique 2021.

Dans les épisodes précédents...

WDPD 2021

The screenshot shows the header of a website for the Open Preservation Foundation. The header includes the logo, navigation links (About, Products, Collaboration, Resources, Blogs, Events, News, Membership), and a search icon. Below the header is a paragraph of text and a table with 9 rows of file analysis results. Each row contains a file name, a status icon (green smiley, red frowny, or yellow neutral), a file type, and a detailed format description.

Member area | Login

About Products Collaboration Resources Blogs Events News Membership

La présence de fichiers PSD, très minoritaires dans les trois fonds, parfois même « cachés » derrière une extension .tif, a été révélée par un outil développé en interne appelé Frontin. Mis à la disposition des responsables de collections qui trient des fonds numériques, cet outil identifie et caractérise le format des fichiers analysés et émet un premier diagnostic sur l'acceptation des fichiers lors du versement.

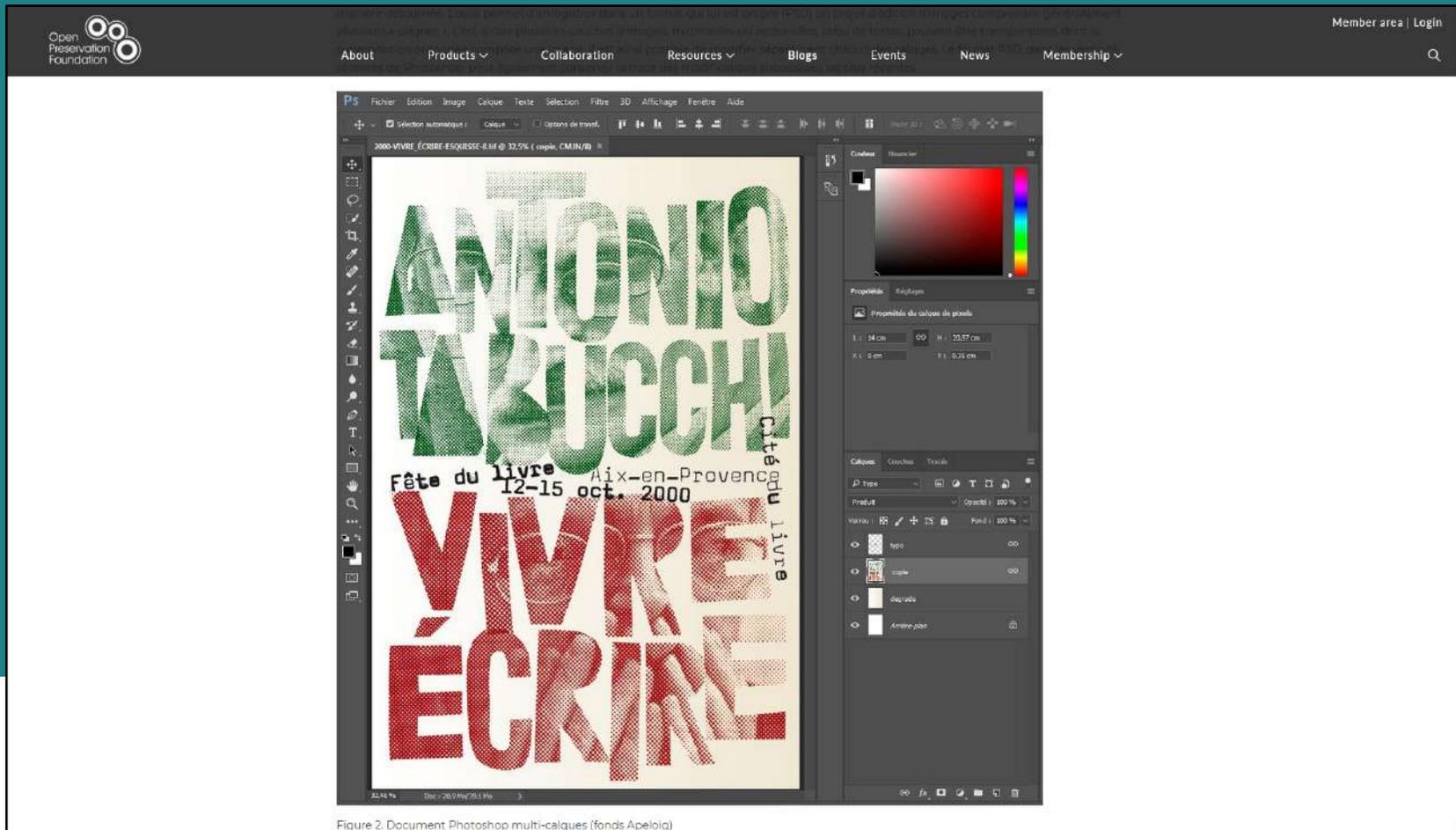
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-1.jpg	😊	image/jpeg	Format : JPEG, Dimension [2799x4136], Compression : JPEG, Résolution [300x300], Profondeur : 24
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-2.tif	😞	image/vnd.adobe.photoshop	Format : bytestream,
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-3.jpg	😊	image/jpeg	Format : JPEG, Dimension [2797x4128], Compression : JPEG, Résolution [300x300], Profondeur : 24
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-4.pdf	😊	application/pdf	Format : PDF 1.4, Pages : 0, Cryptage : non, Status : Well-Formed and valid
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-5.pdf	😊	application/pdf	Format : PDF 1.4, Pages : 0, Cryptage : non, Status : Well-Formed and valid
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-6.tif	😐	image/tiff	Format : TIFF 6.0, Dimension [1772x2817], Compression : uncompressed, Résolution [300x300], Profondeur : 32
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-7.tif	😐	image/tiff	Format : TIFF 6.0, Dimension [1701x2383], Compression : LZW, Résolution [150x150], Profondeur : 48
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-8.tif	😞	image/vnd.adobe.photoshop	Format : bytestream,
2000-VIVRE_ÉCRIRE-ANTONIO_TABUCCHI/2000-VIVRE_ÉCRIRE-ESQUISSE-9.tif	😞	image/vnd.adobe.photoshop	Format : bytestream,

Figure 1. Résultat d'analyse d'un lot de fichiers par l'outil Frontin

A partir de là, trois options ont rapidement été éliminées :

- demander une relivraison dans un format accepté par la BnF (cette option s'est révélée impossible) ;
- exclure les fichiers (cela n'avait pas de sens pour ces fonds destinés à la recherche) ;
- accepter les fichiers tels quels (ils n'étaient pas compatibles avec la politique de la BnF en matière de formats)

WDPD 2021



WDPD 2021

The screenshot shows the Digital Preservation Coalition (DPC) website. The header is dark green with the DPC logo and name on the left, and navigation links (ABOUT, NEWS, DIGITAL PRESERVATION, EVENTS, BLOG) and a Login link on the right. The main content area is white. On the left, there are links for 'Also in this section', 'Job Vacancies', and 'New Members of the DPC', along with a green 'DPC NEWS RSS FEED' button. The main article is titled 'Breaking Down Barriers: DPC shares translations of its popular Digital Preservation Handbook', dated 4 November 2021. It includes a CERN logo and text about the handbook's availability in French and Italian.

dpc Digital Preservation Coalition

ABOUT NEWS DIGITAL PRESERVATION EVENTS BLOG

Login

Also in this section

Job Vacancies

New Members of the DPC

DPC NEWS
RSS FEED

ENHANCED BY Google

Home > News > Breaking Down Barriers: DPC shares translations of its popular Digital Preservation Handbook

Breaking Down Barriers: DPC shares translations of its popular Digital Preservation Handbook

Added on 4 November 2021

[la version française suit / segue la versione italiana]

The Digital Preservation Coalition (DPC) is delighted to share French and Italian translations of its popular **Digital Preservation Handbook**, adding to a range of other translated resources.

Now available in the two new languages, the Digital Preservation Handbook identifies good practice in creating, managing and preserving digital materials. By providing a strategic overview of the key issues,

Actualités de la cellule

Derniers travaux

Par Yannick Grandcolas (BnF)

#PINFormats



P *r*éservation
I *n*formation
N *u*mérique



Traduction



Connaissance
des formats



Expertises



Outils & corpus

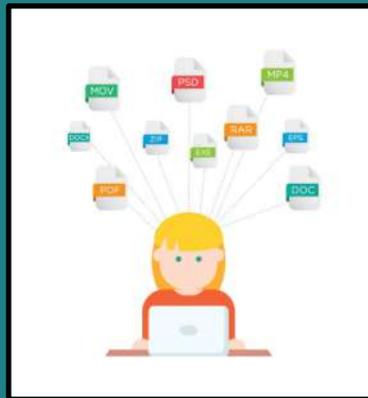
Merci de votre attention !

#PINFormats



Retrouvez nos précédents webinaires

Présentation et vidéos disponibles sur le site d'Aristote



Webinaires de présentation de
la cellule
Novembre 2020



Digital Preservation Handbook
Juin 2021