

BIBLIOTHÈQUE
DE GENÈVE
ICONOGRAPHIE



Le fonds Revillard

retour d'expérience d'un projet pilote

4 septembre 2023 – Bibliothèque nationale, Berne



Sommaire

1. Le projet Revillard (Eloi)
2. Extraction des métadonnées, avec Exiftool, Saxon, Python (Sylvain)
3. Evaluation de la masse: les différents outils (Cécile)

1. Le projet Revillard

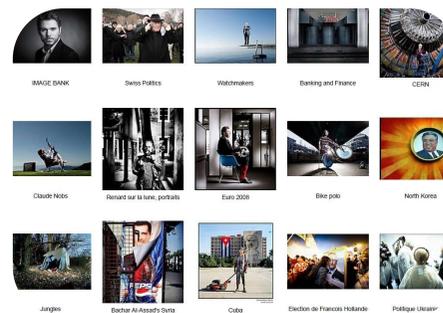
Le fonds Jean Revillard en quelques mots



Archives analogiques
(4.8 ml)



122 grands tirages
d'exposition



Site Rezo.ch (66'000
fichiers dont 20'000 de
Jean Revillard)



Disques durs (environ
41.6 To)





Objectif général

Remplir les conditions de conservation et de valorisation du fonds Jean Revillard



Objectifs spécifiques

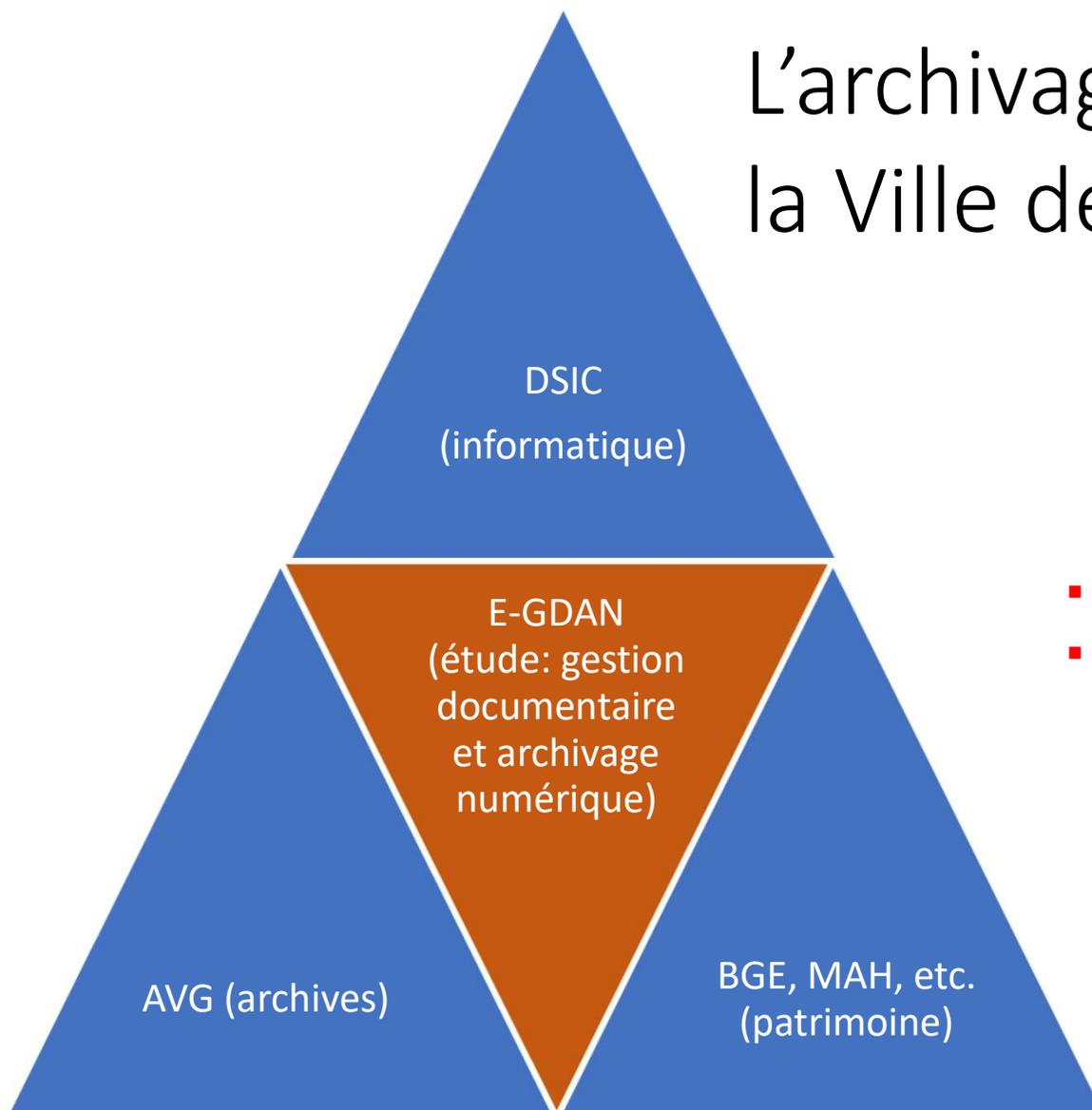
1. Sélectionner et importer dans Museum Plus les données et métadonnées du fonds
2. Sécuriser les données et métadonnées en attendant la faisabilité de son archivage pérenne
3. Acquérir les compétences pour le traitement de fonds numériques acquis ultérieurement



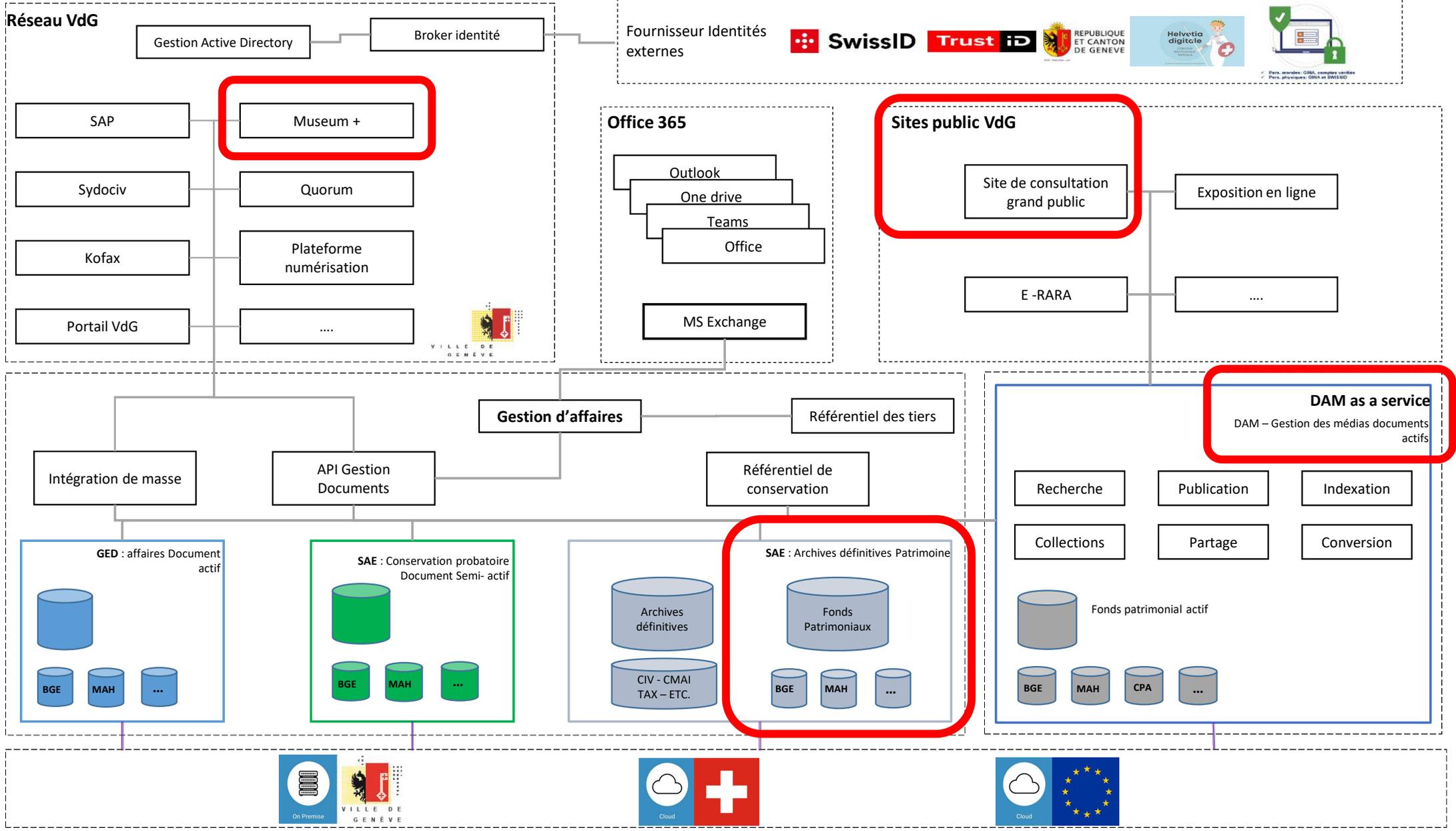
Hors périmètre

Archivage pérenne (cf. GDAN)

L'archivage numérique à la Ville de Genève



- **E-GDAN:** macro-analyse, 2021-2022
- **GDAN:** étude de marché dès 2023 (DAM et SAE) en vue d'une demande de crédit au législatif de la Ville de Genève



CAROUGE 110
Contenu
VES-SO - MOBILITEC
880
PROPERTY OF
ELEVATION STUDIO LTD

D12
2012
L-Ce 2012

Solar
2013
back up - cross-america

Solar 2016
+
CD Jean
2002 - 2009

Back up
ordi Jean
2016 - 2017

Bak up Solar
2016
back up S12 TDW

10ans
Rezo
back up

D14
2012
back up photo

Solar Impulse
2015-2016

Solar
2016
Solar TDW

(light room)
2017

MEAN Kevin
Solar JR
Conference
CRFS?

Base public
Photoshelter

2013
disques

Solar
back up
2015-2016

Solar Impulse
2016
Solar TDW 3 unite: 1

Expo 2016
+
Mandat 2017
2018
Light room
Expo book 2016

37

D11
JR ST
attache a gas

Solar Impulse
2013
Sia Crosse

2015 - 2016

2015
2017

Solar
2016
back up TDW 3

Back up ordi Jean
2011-2012 (on photo)
I take sauvegarde
photos Sandy

Catégories de documents

Images diffusées

Travaux de commandes
Travaux personnels
Solar impulse

Aussi
dans
Rezo.ch

Archives

Travaux de commandes
Travaux personnels
Solar impulse
Agence Rezo.ch

Nouveau



VILLE DE
GENÈVE

2. Extraction de métadonnées



Rezo.ch: extraction des métadonnées

1. Exiftool

- Production des fichiers XML: 1 fichier = métadonnées de 1 image
- Contrôle d'exhaustivité: 1 fichier XML par image

2. Saxon

- Extraction des métadonnées pertinentes selon indications CD
- Formatage et distribution des métadonnées dans des colonnes Excel selon indications CD
- Contrôle d'exhaustivité: 1 ligne Excel = 1 image

3. Python

- Sélection des métadonnées des images dont l'auteur est explicitement Revillard
- Sélection des métadonnées des images dont l'auteur n'est pas identifié et donc potentiellement de Revillard
- Repérage des images dont les noms de fichier sont identiques
- Copie des images sélectionnées dans un dossier pour traitement par SP
- Identification des images existant à la fois dans Listed et Unlisted





Rezo.ch: traitements Exiftool et Saxon

1. Traitement Exiftool

Commande Exiftool

```
C:\> Invite de commandes
Microsoft Windows [version 10.0.18363.2094]
(c) 2019 Microsoft Corporation. Tous droits réservés.

P:\> C:\>

C:\> perl C:\Windows\exiftool\exiftool.pl -@ U:\collections_bge\centre_iconographie\cig_3\w_03_traitement\revillard_fejoz\photoshelter_unlisted_metadata\exiftool_xml_prod\unlisted6_image_list.txt -X -w %d%f.xml
```

Chemins des images

```
unlisted6_image_list.txt - Bloc-notes
Fichier Edition Format Affichage Aide
U:\collections_bge\centre_iconographie\cig_3\w_03_traitement\revillard_fejoz\photoshelter_unlisted\unlisted6\LODH by Rezo\FW03061.jpg
U:\collections_bge\centre_iconographie\cig_3\w_03_traitement\revillard_fejoz\photoshelter_unlisted\unlisted6\LODH by Rezo\FW03066.jpg
U:\collections_bge\centre_iconographie\cig_3\w_03_traitement\revillard_fejoz\photoshelter_unlisted\unlisted6\LODH by Rezo\FW03067.jpg
U:\collections_bge\centre_iconographie
```

2. Traitement Saxon

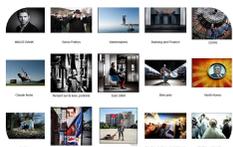
Appel de Saxon

```
def transform(input, output):
    #https://stackoverflow.com/questions/29443364/use-saxon-with-python
    subprocess.call(f"java -cp C:\saxon\SaxonHE10-6J\saxon-he-10.6.jar net.sf.saxon.Transform -t -s:{input} -xsl:{xslt} -o:{output}")
```

Fichier transfo XSLT

```
147 <!-- //XMP-dc:Creator -->
148 <xsl:variable name="XMP-dc:Creator">
149 <xsl:choose>
150 <xsl:when test="//XMP-dc:Creator">
151 <xsl:choose>
152 <xsl:when test="//XMP-dc:Creator != "">
153 <xsl:value-of select="//XMP-dc:Creator"/>
154 </xsl:when>
155 <xsl:otherwise>
156 <xsl:text>XMP-dc:Creator VIDE</xsl:text>
157 </xsl:otherwise>
```





Saxon: traitement machine le plus long

dossier	traitement	date	minutes	nombre de fichiers traités = lignes du fichier csv	check nombre de fichiers traités	error_log.csv : nb erreurs	commentaire error log	images revillard	dossier tra
photoshelter unlisted 1	revillard_saxon	08.02.2022	14	1016	1016	0		81	U:\collecti
photoshelter unlisted 2	revillard_saxon	08.02.2022	170	12290	12317	27	pas des images revillar	47	U:\collecti
photoshelter unlisted 3	revillard_saxon	02.02.2022	5	397	397	0		105	U:\collecti
photoshelter unlisted 4	revillard_saxon	09.02.2022	33	2422	2422	0		709	U:\collecti
photoshelter unlisted 5	revillard_saxon	08.02.2022	175	12970	12970	0		9196	U:\collecti
photoshelter unlisted 6	revillard_saxon	09.02.2022	42	3103	3103	0		2108	U:\collecti
photoshelter unlisted 7	revillard_saxon	09.02.2022	4	324	324	0		268	U:\collecti
	TOTAL DUREE :		443	32522	32549	27	TOTAL REVILLARD :	12514	

7 heures pour 32'522 fichiers
0.8 seconde par fichier

Avant analyse des
images «sans auteur»

- **Critères sélection Revillard** : « jr » ou « revillard » dans le nom du fichier (métadonnée <System:FileName>) ou « revillard » dans une de ces métadonnées : <IFDO:Artist>, <IFDO:Copyright>, <IPTC:By-line>, <IPTC:CopyrightNotice> et <XMP-dc:Creator>
- **Critères sélection Sans auteur** : « jr » ou « revillard » pas dans le nom du fichier (métadonnée <System:FileName>) et les métadonnées <IFDO:Artist>, <IFDO:Copyright>, <IPTC:By-line>, <IPTC:CopyrightNotice> et <XMP-dc:Creator> sont vides ou inexistantes





Disques durs: extraction des métadonnées

1. Exiftool

- Production des fichiers XML: 1 fichier = métadonnées de 1 image
- Contrôle d'exhaustivité: 1 fichier XML par image & autre type de fichier
- Réplique des arborescences (dossiers, sous-dossiers...) de chaque disque dur

2. Saxon

- Extraction des métadonnées pertinentes selon indications CD
- Formatage et distribution des métadonnées dans des colonnes Excel selon indications CD
- Contrôle d'exhaustivité: 1 ligne Excel = 1 image ou autre type de fichier

3. Python

- Tout ou partie des opérations faites pour Rezo.ch, notamment des opérations de contrôle des données
- Identifier les images ayant des métadonnées de contenu et absentes de Rezo.ch
- Dans les arborescences, ajout d'un fichier XML (ou txt?) par fichier non image
- Croisement avec les données obtenues par CD au moyen des outils qu'elle emploie afin de compléter les métadonnées des fichiers non images
- Import des données obtenues dans une base FileMaker afin de pouvoir faire des requêtes facilement





Exiftool: *Rezo.ch* VS *Disques durs*



1. Traitement Exiftool *Rezo.ch*

```
Invite de commandes Commande Exiftool
Microsoft Windows [version 10.0.18363.2094]
(c) 2019 Microsoft Corporation. Tous droits réservés.

P:\>C:

C:\>perl C:\Windows\exiftool\exiftool.pl -@ U:\collections_bge\centre_iconographie\cig_3\w_03_traitement\revillard_f_ejoz\photoshelter_unlisted_metadata\exiftool_xml_prod\unlisted6_image_list.txt -X -w %d%f.xml
```

-X -w %d%f.xml

2. Traitement Exiftool *Disques durs*

```
Invite de commandes Commande Exiftool
Microsoft Windows [version 10.0.19045.3324]
(c) Microsoft Corporation. Tous droits réservés.

P:\>c:

C:\>exiftool -@ M:\docs_sylvain\listes_fichiers\d09_image_list.txt -X -w M:\docs_sylvain\listes_fichiers\tous_disques_xml\%:1d%f.xml -efile1 M:\docs_sylvain\listes_fichiers\d09_errors.txt
```

-X -w new_path\ %:1d%f.xml



VILLE DE
GENÈVE



Disques durs: deux livrables particuliers

1. Arborescence complète: fichiers xml

svc (S:) > F_Logistique > F05_Informatique > Travaux > revillard > extraction_cisel > cisel_disques_xml > tous_disques_xml > 1049-1_Revillard D09 >

Nom	Modifié le	Type	Taille
CHAMMARTIN_Christophe dossier	25.07.2023 13:47	Dossier de fichiers	
ECRAN1	25.07.2023 13:47	Dossier de fichiers	
ECRAN2	25.07.2023 13:47	Dossier de fichiers	
ECRAN3	25.07.2023 13:47	Dossier de fichiers	
ECRAN4	25.07.2023 13:47	Dossier de fichiers	
ECRAN5	25.07.2023 13:48	Dossier de fichiers	
EXPO_FINAL	25.07.2023 13:48	Dossier de fichiers	
EXPO_ZRH	25.07.2023 13:48	Dossier de fichiers	
HUMEROSE_Alan dossier 2	25.07.2023 13:48	Dossier de fichiers	
interviewexpo	25.07.2023 13:48	Dossier de fichiers	
JEAN_VIDEOS_HOMMAGE	25.07.2023 13:48	Dossier de fichiers	
MERZ dossier	25.07.2023 13:49	Dossier de fichiers	
PHOTOS_INTERVIEW	25.07.2023 13:50	Dossier de fichiers	
REVILLARD dossier	25.07.2023 13:51	Dossier de fichiers	
righetti4 dossier	25.07.2023 13:51	Dossier de fichiers	
shortcut2	25.07.2023 13:51	Dossier de fichiers	
VIDEOS_MAKINGOF	25.07.2023 13:52	Dossier de fichiers	
WAVRE dossier	25.07.2023 13:52	Dossier de fichiers	
BEAMER_1.xml	25.07.2023 13:47	Document XML	13 Ko
BEAMER_2.xml	25.07.2023 13:47	Document XML	263 Ko
BEAMER_3.xml	25.07.2023 13:47	Document XML	13 Ko
CHAMMARTIN_CHRISTOPHE_H264.xml	25.07.2023 13:47	Document XML	7 Ko
HUMEROSE_ALAN_H262_3.xml	25.07.2023 13:47	Document XML	6 Ko
HUMEROSE_ALAN_H264_2.xml	25.07.2023 13:47	Document XML	6 Ko
MERZ_FRED_H264.xml	25.07.2023 13:47	Document XML	7 Ko
RIGHETTI_NICOLAS_H264.xml	25.07.2023 13:47	Document XML	7 Ko
WAVRE_Francois_H264.xml	25.07.2023 13:47	Document XML	7 Ko

Ce disque 009: 70 Mo vs Disque original: 123 Go

2. Base FileMaker pour faciliter les recherches

_DSC7302.JPG

Auteur	<input type="text" value="XMP-dc:Creator ABSENT"/>	Titre	<input type="text" value="XMP-dc:Description ABSENT"/>
Date de création	<input type="text" value="XMP-xmp:CreateDate ABSENT"/>		
Extension	<input type="text" value="jpg"/>		
Poids	<input type="text" value="11 MB"/>		
Largeur en pixels	<input type="text" value="4472 x 7952"/>	Résumé	<input type="text" value="IPTC:Caption-Abstract ABSENT"/>
Résolution en dpi	<input type="text" value="350"/>		
Image rating	<input type="text" value="XMP-xmp:Rating ABSENT"/>		
No disque cisel	<input type="text" value="D02"/>	Sujet	<input type="text" value="XMP-dc:Subject ABSENT"/>
Chemin complet dans cisel	<input type="text" value="M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D02\2018-07-07_DSC7302.JPG"/>		
enregistrement	<input type="text" value="saxon_ok"/>		

G E N È V E

3. Evaluation de la masse: les différents outils

Qu'y a-t-il sur les disques?

Comment est-ce classé?

Quels types de fichiers y a-t-il et quel volume occupent-ils

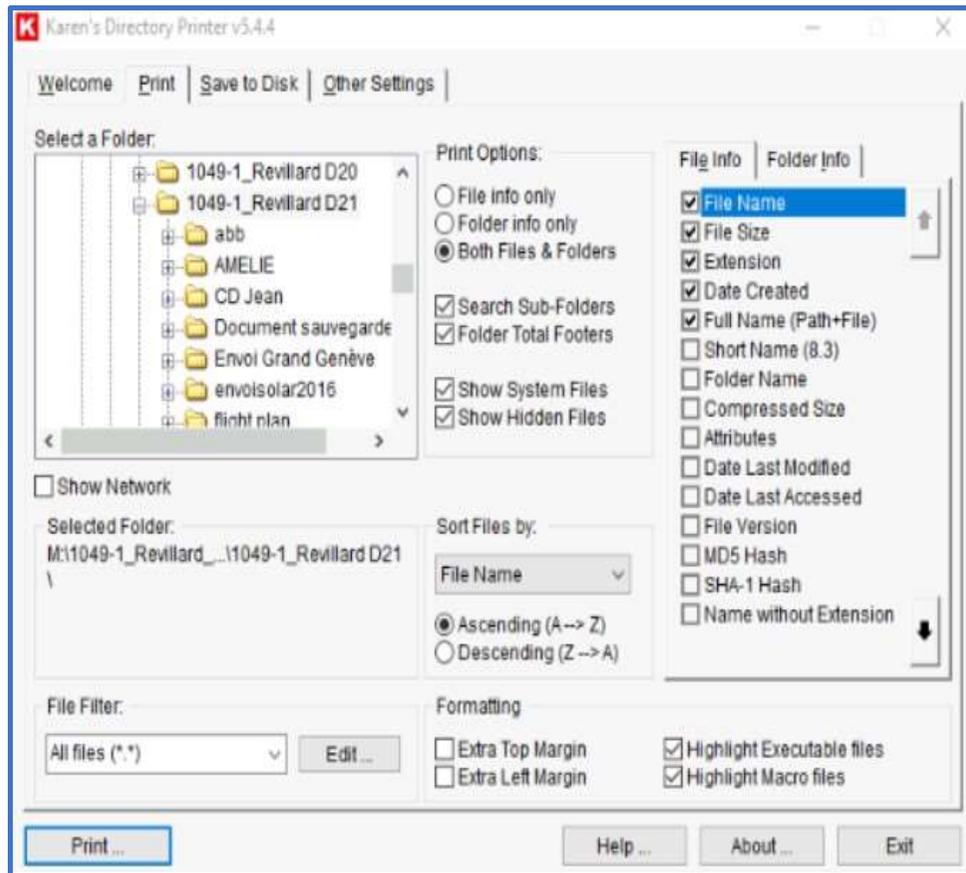
Y-a-t-il des fichiers ou dossiers à double?

Quels sujets apparaissent?

Quoi sélectionner et comment?

Qu'y a-t-il sur les disques?

Extraction d'une liste des dossiers/fichiers avec le logiciel Karen's Directory printer



Dossier/fichier	Nom du dossier/fichier	Taille en octets	Extension	Date de dernière modification	Nombre de dossiers	Nombre de fichiers	Chemin complet
Dossier	1049-1_Revillard D21	23 326 717	Dossier	05.07.20 21 13:45	16		M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\
Dossier	abb	78 864 977	Dossier	08.12.20 22 17:01	0	11	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\abb\
Fichier	01-supplier-information-REZO.CH2.pages	783 949	pages	15.05.20 16 18:02			M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\abb\01-supplier-information-REZO.CH2.pages
Fichier	1000.jpg	212 291	jpg	13.04.20 16 17:53			M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\abb\1000.jpg

Comment est-ce classé?

Extraction de l'arborescence avec Powershell tree

```
M:\1049-1_REVILLARD_TRAITEMENT_BES\1049-1_REVILLARD D09
1049-1_Revillard D09 - Shortcut.lnk
BEAMER_1.mov
BEAMER_2.mov
BEAMER_3.mov
CHAMMARTIN_CHRISTOPHE_H264.mov
HUMEROSE_ALAN_H262_3.mov
HUMEROSE_ALAN_H264_2.mov
MERZ_FRED_H264.mov
RIGHETTI_NICOLAS_H264.mov
WAVRE_Francois_H264.mov

CHAMMARTIN_Christophe dossier
  chamartin.aep
  chamartinCompte-rendu.txt
  CHAMMARTIN_SONMOTION.wav
  CHAMMARTIN_SONMOTION1.mov

  (Metrage)
    01-7hitr.jpg
    03 Sadir.m4a
    03-10tr.jpg
    0407-02.jpg
    06-07.jpg
    07.jpg
    08.jpg
    09.jpg
```

Chaque disque présente une arborescence de premier niveau différente

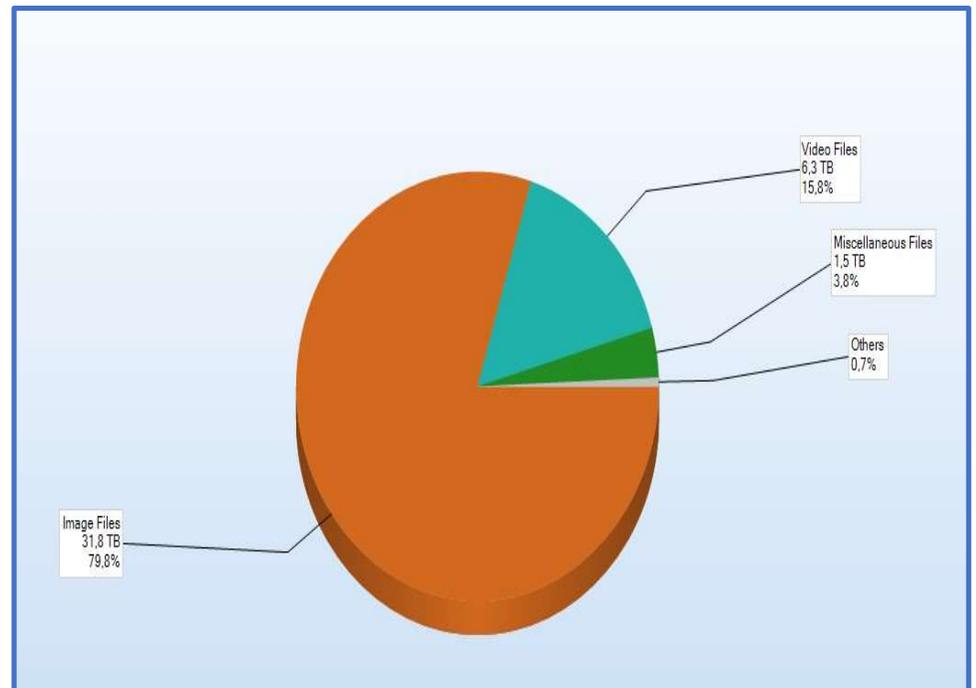
Les arborescences ne présentent pas de logique particulière

Chaque dossier a sa propre organisation

Quels types de fichiers y a-t-il et quel volume occupent-ils?

Extraction d'une liste des types et d'un graphique avec Tree Size Professionnal

Name	Size	Files	Type
Image Files	31.7 TB	1698560	Files containing pictures, images or mouse cursors
Video Files	6.3 TB	24011	Files containing videos or animations
Miscellaneous Files	1.5 TB	689520	Unknown file types
Audio Files	162.9 GB	15855	Files containing music, sounds or playlists
Container Files	76.0 GB	8878	Compressed Archives and disk images
Office Files and Documents	13.5 GB	7300	Documents and files of office programs and PDFs
Mail Files	10.2 GB	33729	Email messages and files of email clients
Program Files	2.7 GB	11800	Program Files, Libraries and other compiled resources
System Files	2.5 GB	2119	System Files
Database Files	2.1 GB	444	Files containing the data of client and server databases
Temporary and Backup Files	1.7 GB	82	Temporary files and backup copies containing previous version of current files
Software Development Files	906.5 MB	16783	Source and project files of software development projects
Internet Files	270.7 MB	36053	Files related to the WWW, like HTML files
Text Files	279.1 MB	12674	Plain text files, log files
MacOS Files	234.8 MB	1047	Files which are typically used on MacOS system
Help Files	74.0 MB	381	Files of the Windows help system
Configuration Files	10.3 MB	1303	Files containing configuration settings



95% des fichiers sont de type audiovisuel

Y-a-t-il des fichiers ou dossiers à double?

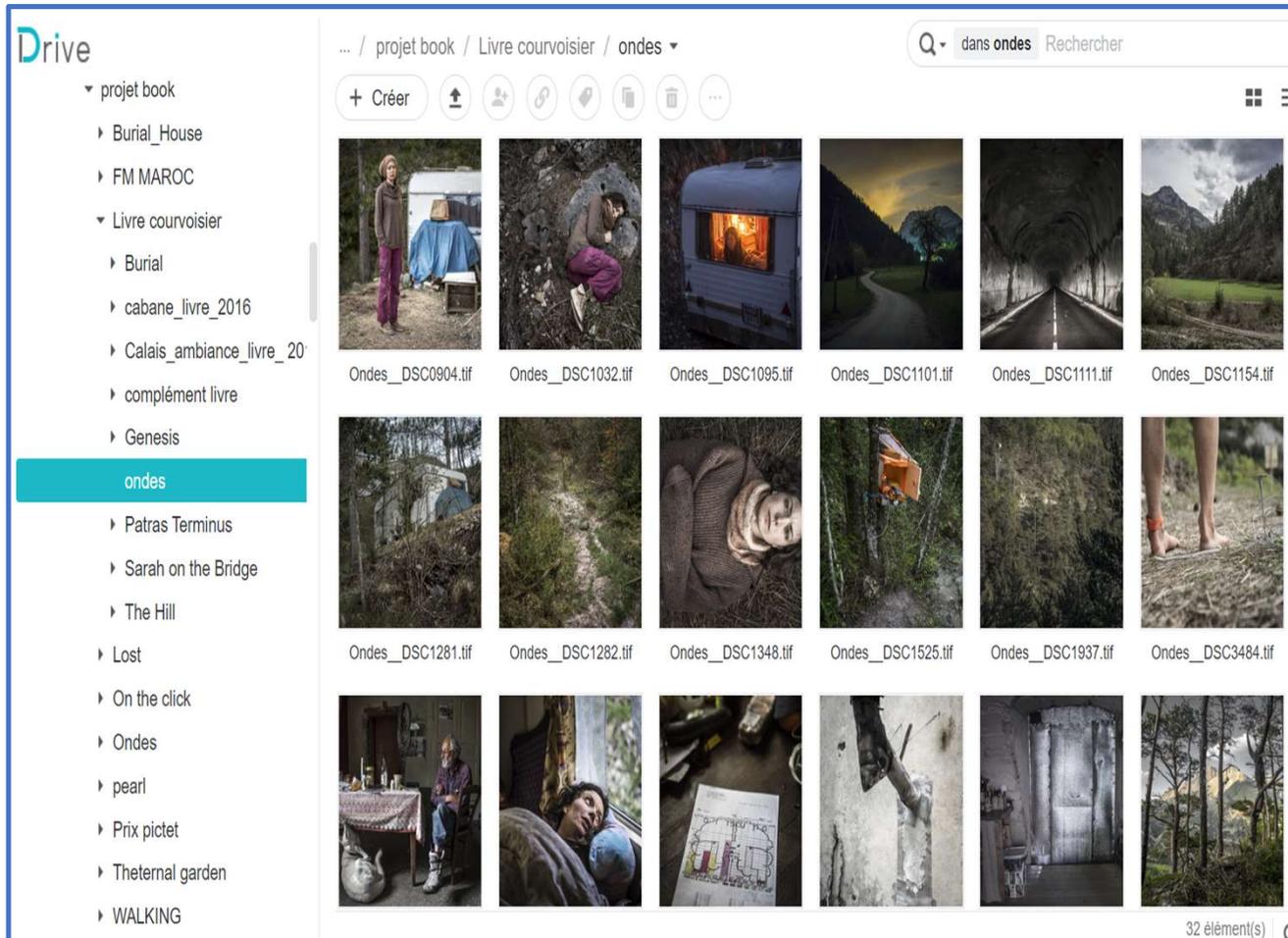
Extraction d'une liste des dossiers identiques avec TreeSize Professional

cabane_livre_2016	[multiple]
cabane_livre_2016	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
cabane_livre_2016	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
Calais_ambiance_livre_2016	[multiple]
Calais_ambiance_livre_2016	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
Calais_ambiance_livre_2016	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
Genesis	[multiple]
Genesis	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
Genesis	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
ondes	[multiple]
ondes	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
ondes	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
Patras Terminus	[multiple]
Patras Terminus	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
Patras Terminus	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
Sarah on the Bridge	[multiple]
Sarah on the Bridge	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
Sarah on the Bridge	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
The Hill	[multiple]
The Hill	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\Livre courvoisier\
The Hill	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D28\Livre courvoisier\
04_14_2016_SunCharge	[multiple]
04_14_2016_SunCharge	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D22\
04_14_2016_SunCharge	M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D30\

9846 dossiers en double

Quels sujets apparaissent?

Visionnage des dossiers sur le Cloud



Les noms de dossiers sont parlants ("envoisolar2016", "projet book", "perso TDM <<")

10 catégories de documents ont pu être identifiées : projets personnels, travaux de commande ...

Quoi sélectionner et comment?

Faire une recherche par mots-clés dans les noms de dossiers avec TreeSize Professional

The screenshot displays the TreeSize Professional File Search application. The interface is divided into several sections:

- Search Configuration:** Located on the left, it shows a list of search criteria. The criteria are: Name contains jungle, Name contains calais, Name contains cabane, Name contains patras, Name contains grece, Name contains onde, Name contains ehs, Name contains emma, Name contains sarah, Name contains bridge, Name contains hill, Name contains burial, Name contains manille, Name contains italie, and Name contains refugee.
- Search Configuration Panel:** Below the criteria, there are checkboxes for 'Name', 'Folder Path', and 'Path'. The 'Name' checkbox is checked.
- Search Configuration Table:** A table with three columns: Name, Folder Path, and Path. It lists various folders and their paths, such as 'Burial_House' at 'M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\projet book\' and 'WPP_burial' at 'M:\1049-1_Revillard_Traitement_Bes\1049-1_Revillard D21\Document sauvegarde\con...'. The 'Name' column is checked for all entries.
- Search Configuration Panel:** On the left side of the main window, there are buttons for 'Basic Search', 'Duplicate Search', 'Select Template', and 'recherche travaux personnels'.
- Search Configuration Panel:** At the top, there are tabs for 'Search Configuration' and 'Result list'. The 'Search Configuration' tab is active.
- Search Configuration Panel:** At the bottom, there are buttons for 'Start', 'Stop', 'New advanced search', 'Rename current search', 'Delete current search', 'Save current filters as template', 'Customize File Groups', 'Exclude Filter', 'Include files', and 'Include folders'.