

Conférence LibreABC - 2023.08.31

# Les logiciels libres pour l'aide à l'archivage numérique à la **Cinémathèque suisse**

Rebecca ROCHAT - Cheffe de programme  
Robin FRANÇOIS - Chef de projet, Archiviste numérique  
Unité Plateforme logicielle, Cinémathèque suisse

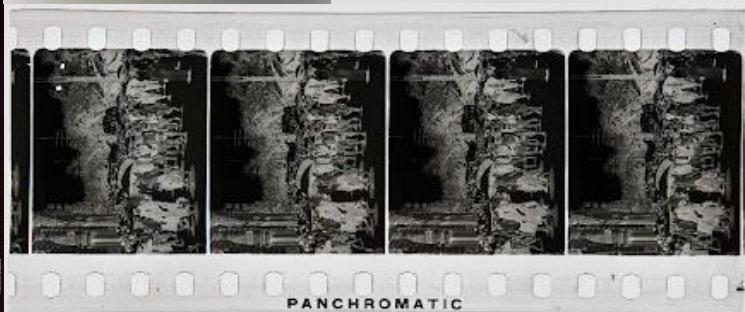
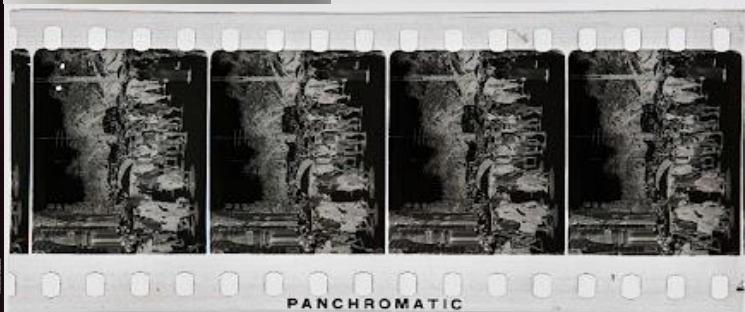
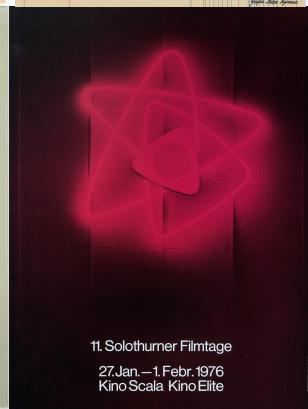
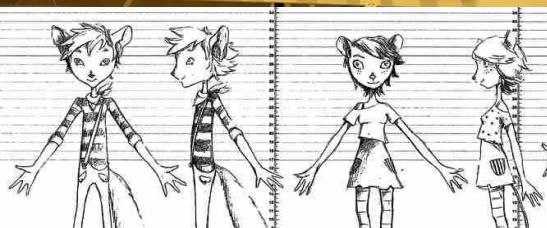


# Cinémathèque suisse en chiffres

<b>85'000</b>	titres de film
<b>700'000</b>	bobines
<b>300'000</b>	affiches
<b>2,8 mio</b>	photos (tirages, négatifs, positifs)
<b>200'000</b>	images numériques natives
<b>27'000</b>	livres
<b>10'000</b>	scénarios
<b>240'000</b>	dossiers documentaires
<b>3000</b>	appareils cinématographiques
<b>2000 mètres linéaires</b>	fonds d'archives (> 200 fonds)
<b>5 Po</b>	bibliothèque LTO

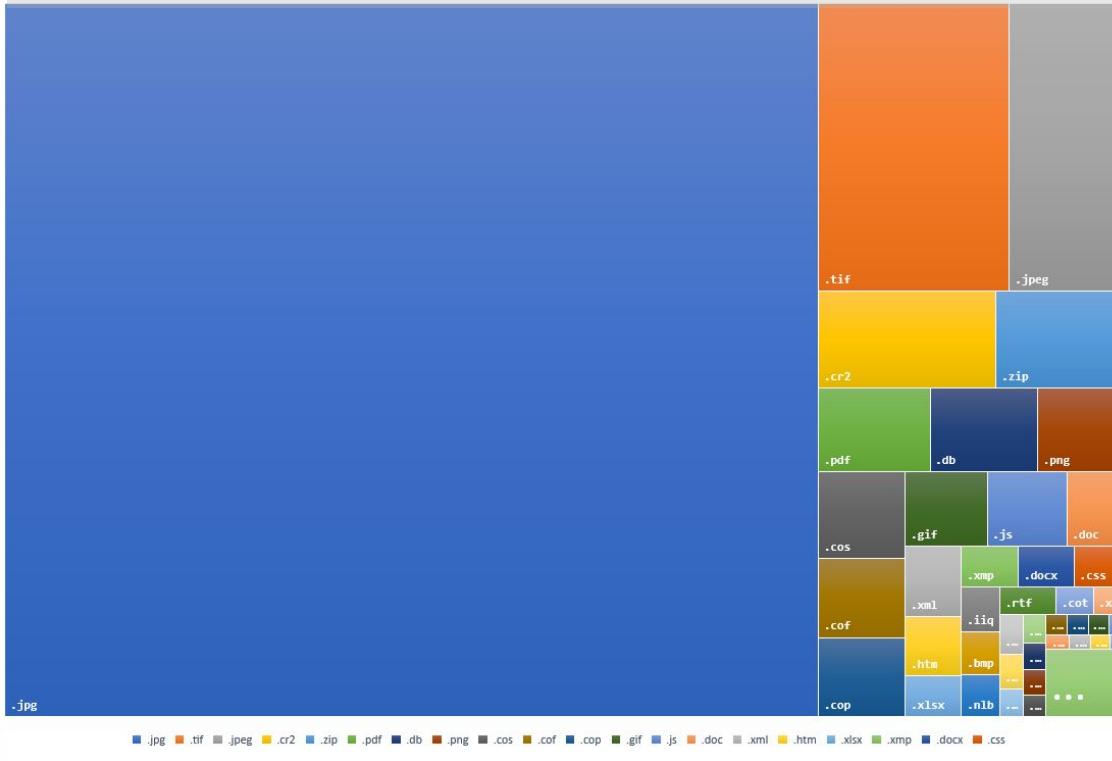
# Grande variété de matériels et typologies

## Des collections analogiques ...



# ... aux collections numériques

## Emplacement réseau téléchargement matériel promotionnel



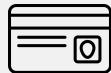
# Plateforme logicielle, quésako ?



**Optimiser et améliorer** les outils de travail actuels

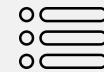
## État des lieux

- Obsolescence des outils
- Mises à jour compliquées
- Solutions **ne répondent pas aux exigences** des collections numériques



## Le principe

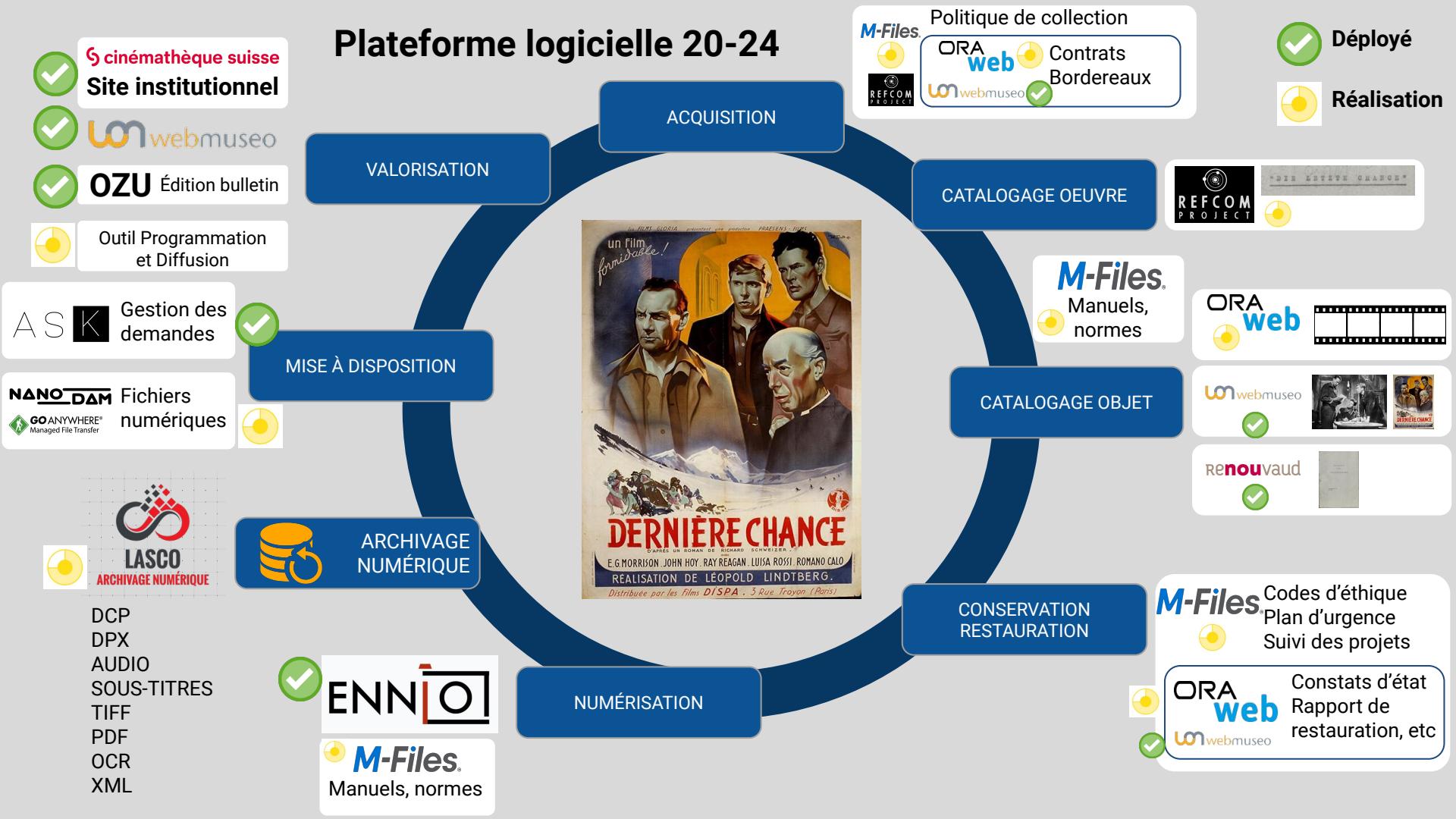
Repose sur plusieurs **briques logicielles** individuelles qui à terme communiqueront ensemble

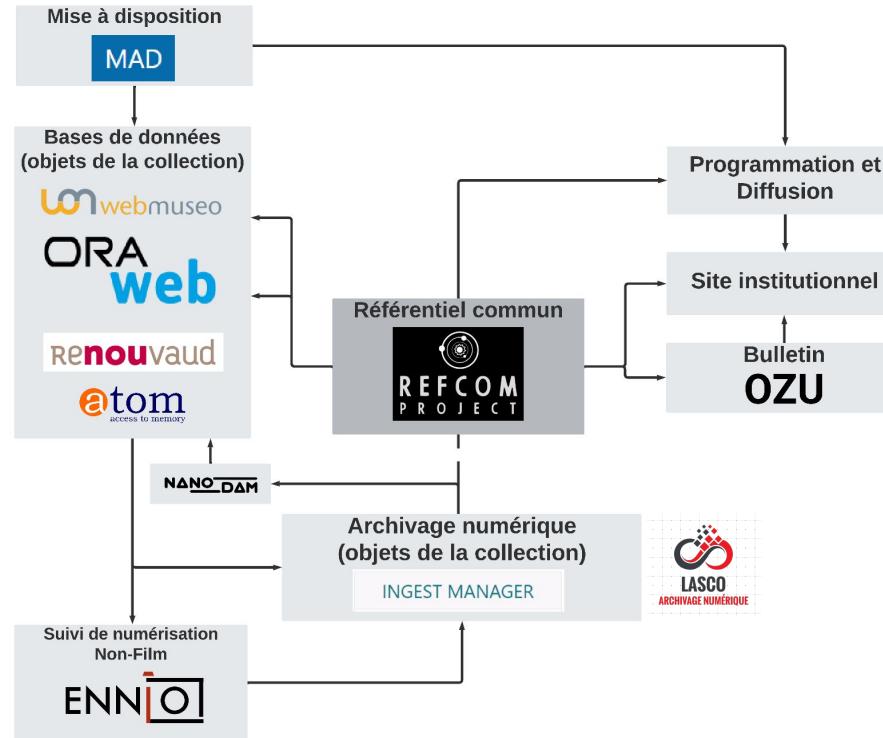


## Développement de nouveaux logiciels

- assurer les missions de la CS **à l'ère du numérique**
  - Documentation
  - Conservation
  - Archivage
  - Valorisation

# Plateforme logicielle 20-24





Gestion électronique des documents administratifs

**M-Files.**

# Dans le concret ...

66

**Context is everything** - John Sheridan, National Archives UK

# Paradoxe de l'archivage numérique des données

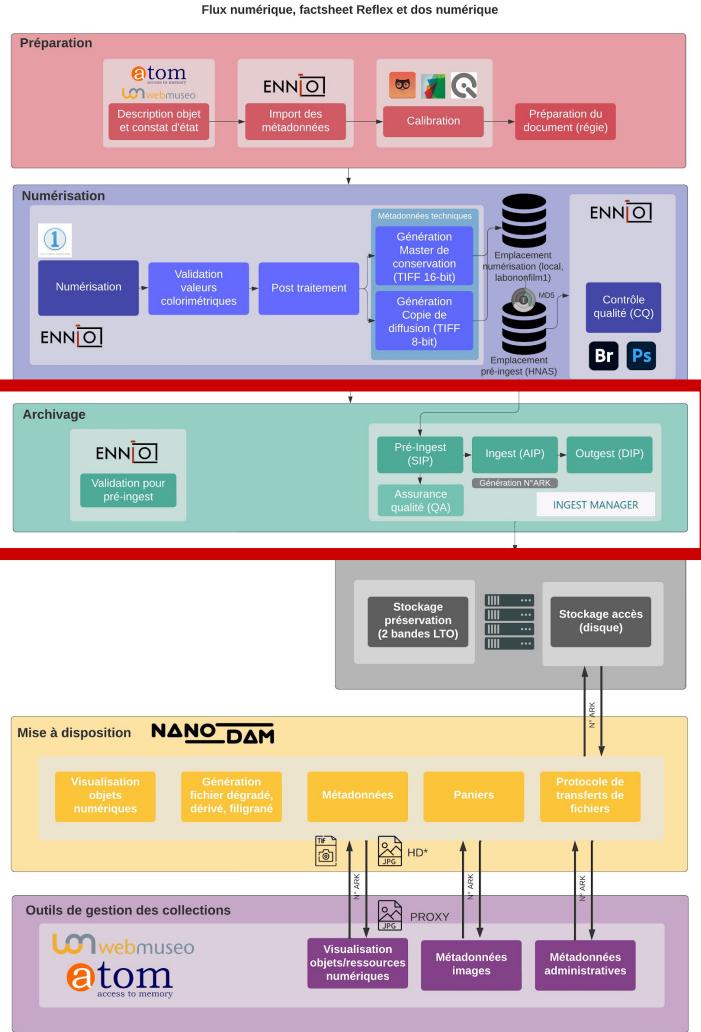
Assurer une **capsule temporelle** contextualisée

## Assurer l'**accès des données**

→ Sauvegarde numérique ≠ Archivage numérique

## **Exemple flux numérisation**

### **Tirages et affiches**



# LASCO

# Logiciel d'Archivage, de Sauvegarde et de Conservation des Œuvres



# ¿¿¿Lasco???:

## Objectifs :

- Solution informatique pour traiter les lots de données à archiver à long terme, conçue pour implémenter le modèle OAIS

## Dans les faits :

Une boîte à **outils** développée en interne en fonction de nos besoins pour automatiser au mieux les opérations/actions nécessaires pour archiver les fichiers numériques et permettre la mise en place des bonnes pratiques (ex: modèle OAIS).

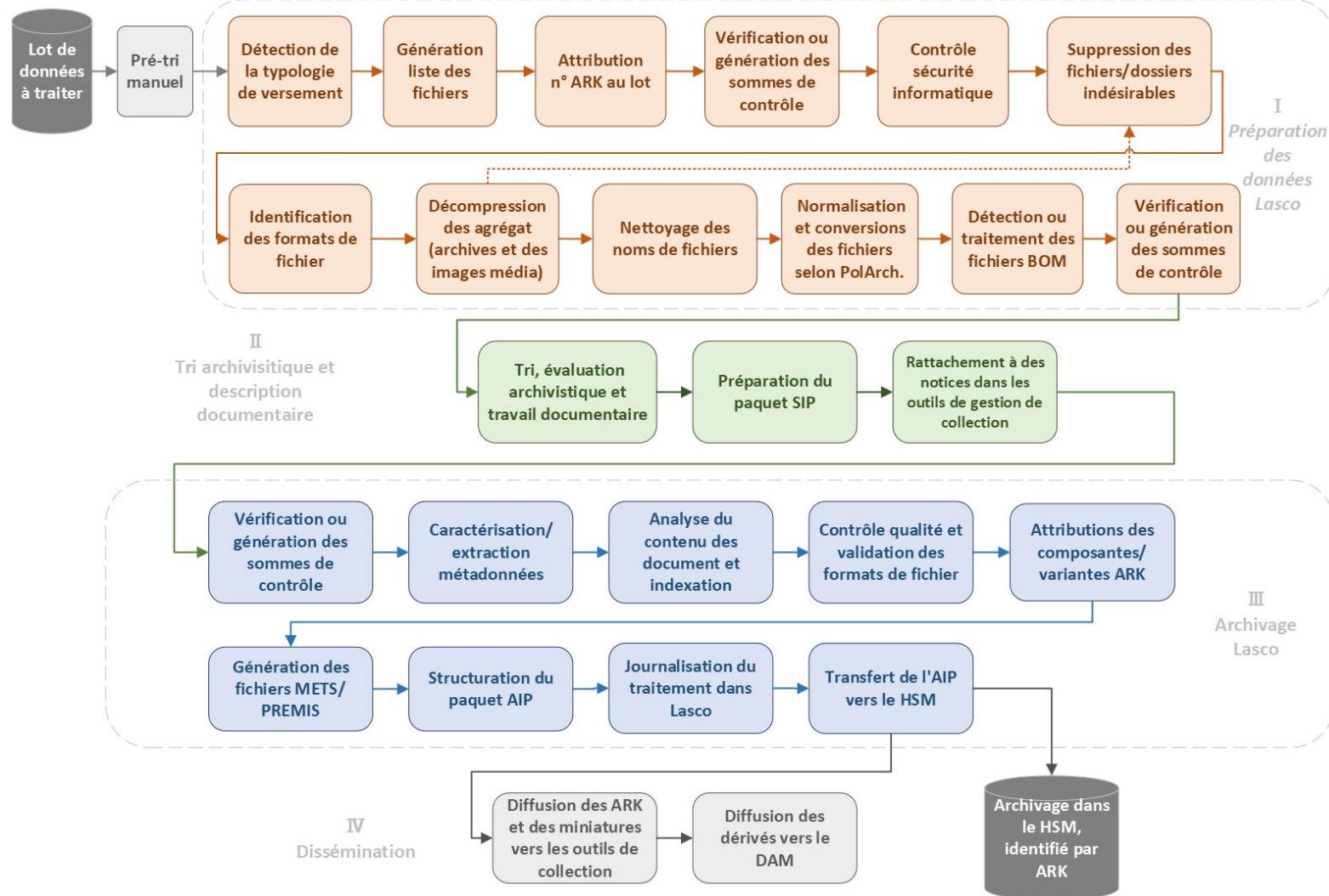
## Développement interne très fortement inspiré par **Archivematica**

Syndrome du “**Not Invented Here**”, en partie justifié

Un développeur à 80% sur le projet

# Historique

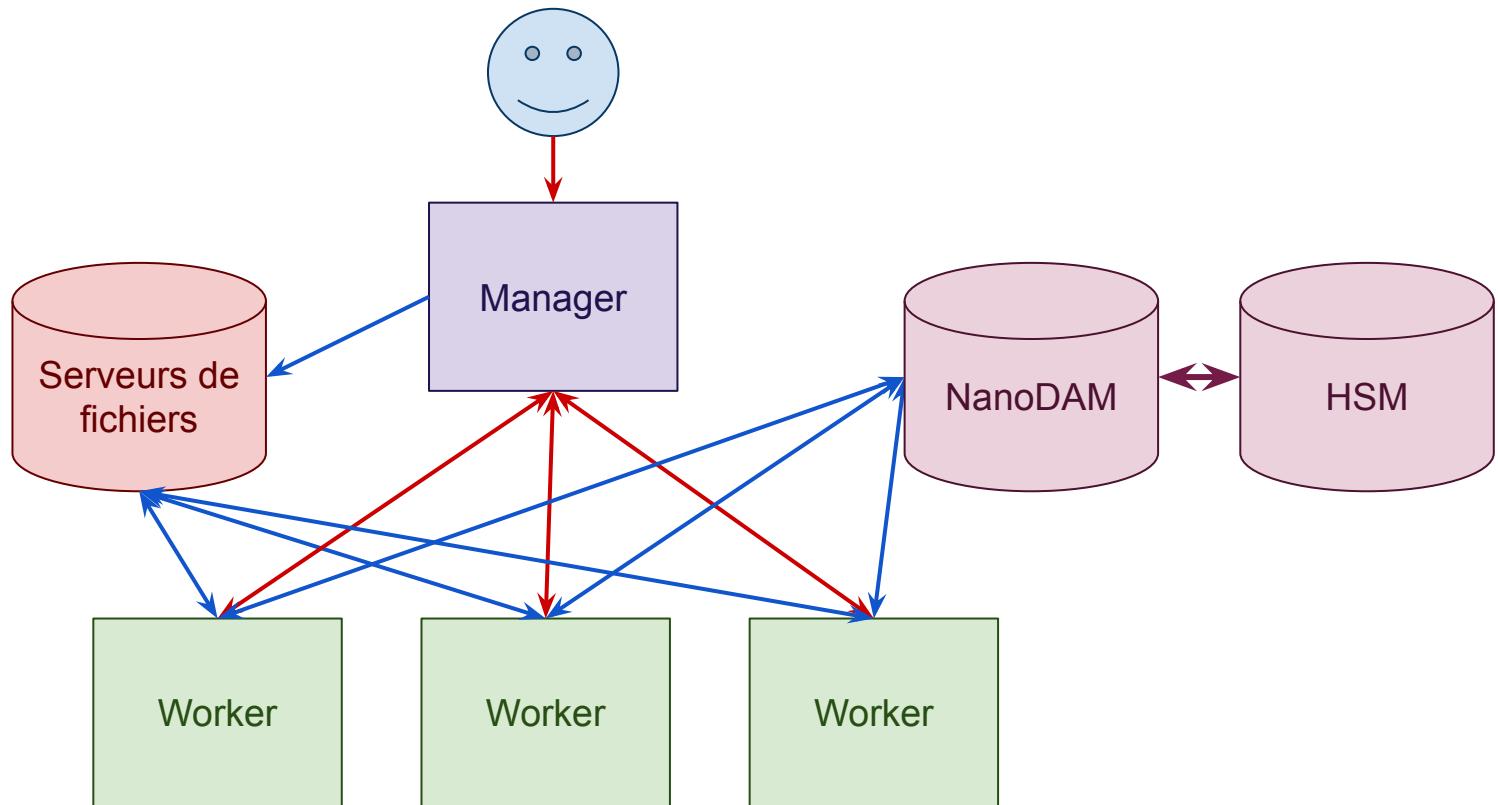
<b>Été 2020</b>	<b>État des lieux, traduire le modèle OAIS et à adapter à l'institution</b>
<b>Automne 2020</b>	Cahier des charges en vue d'une étude/évaluation des solutions existantes,
<b>Automne 2020</b>	<b>Tests avec Archivematica</b>
<b>Janvier 2021</b>	<b>Décision du département IT de partir sur un développement interne</b>
<b>Mars à juillet 2021</b>	Développement d'un prototype
<b>Été 2021</b>	Gel du développement en l'absence de spécifications précises du métier. Création d'un comité technique → "accorder les violons"
<b>Avril 2022</b>	<b>Départ du développeur principal. Projet mis en pause</b>
<b>Été/Automne 2022</b>	Rédaction de spécifications pour une v0.5
<b>Décembre 2022</b>	Reprise du développement avec un nouveau développeur
<b>Juillet 2023</b>	<b>Première mise en production (v0.5), surtout les besoins du dpt Film</b>



# Architecture simplifiée

CIFS/SMB

HTTP(S)



# Boîte d'outils très libre

## Identification du format :

Siegfried, file (Linux)

## Caractérisation (métadonnées)

Exiftool, MediaInfo ffprobe...

## Caractérisation (encodage des caractères)

dfeal, charset-normalizer

## Extraction

Image disque: Aaru, HxC Floppy Emulator...

Archive: 7zip, unar...

## Normalisation

FFmpeg, ImageMagick, Inkscape, Ghostscript, LibreOffice...

## Validation

ClairMeta, JHOVE

En rouge: les outils déjà intégrés à Lasco  
En noir: outils à venir dans Lasco

# ClairMeta

Outil open source en Python pour la validation et l'extraction des métadonnées des DCP

Plus de 130 points de contrôle qui permettent de valider la conformité par rapport aux différentes normes (SMPTE...)

Indispensable, car DCP non conforme potentiellement inutilisable en salle

Concurrent de EasyDCP (propriétaire),  
Cinémathèque suisse **contributeur** du projet

**DCP:** Digital Cinema Package, ensemble de fichiers pour la projection en salle

# ClairMeta

```
$ python3 -m clairmeta.cli check -type dcp /media/it_dev_lasco/IT_DEV_LASCO/01-DCP/01.05-DCP_Proxy_[0\]/  
[...]
```

2023-08-30 17:41:15,521 - Clairmeta - INFO - Check report:

**Status : Success**

Path : /media/it\_dev\_lasco/IT\_DEV\_LASCO/01-DCP/01.05-DCP\_Proxy\_[0]

Size : 1.57 GiB

Total check : 78

Total time : 19.06 sec

**Warning(s) :**

- + /media/it\_dev\_lasco/IT\_DEV\_LASCO/01-DCP/01.05-DCP\_Proxy\_[0]
  - . Foreign files detection (not listed in AssetMap).  
test.rtf

- + CPL\_Von-Fischen-und-Menschen\_TLR\_S\_GSW-DE\_51\_2K\_20201201\_TPX\_IOP\_OV\_1ec9858c-4cd7-43.xml  
(Von-Fischen-und-Menschen\_TLR\_S\_GSW-DE\_51\_2K\_20201201\_TPX\_IOP\_OV)
  - . Digital Cinema Naming Convention compliance (9.3).

- ContentTitle should have 12 parts to be fully compliant with ISDCF naming convention version 9.6, 10 part(s) found

ContentTitle Film Name does not respect naming convention rules : Von-Fischen-und-Menschen

Field TerritoryRating not found in ContentTitle

Field Studio not found in ContentTitle

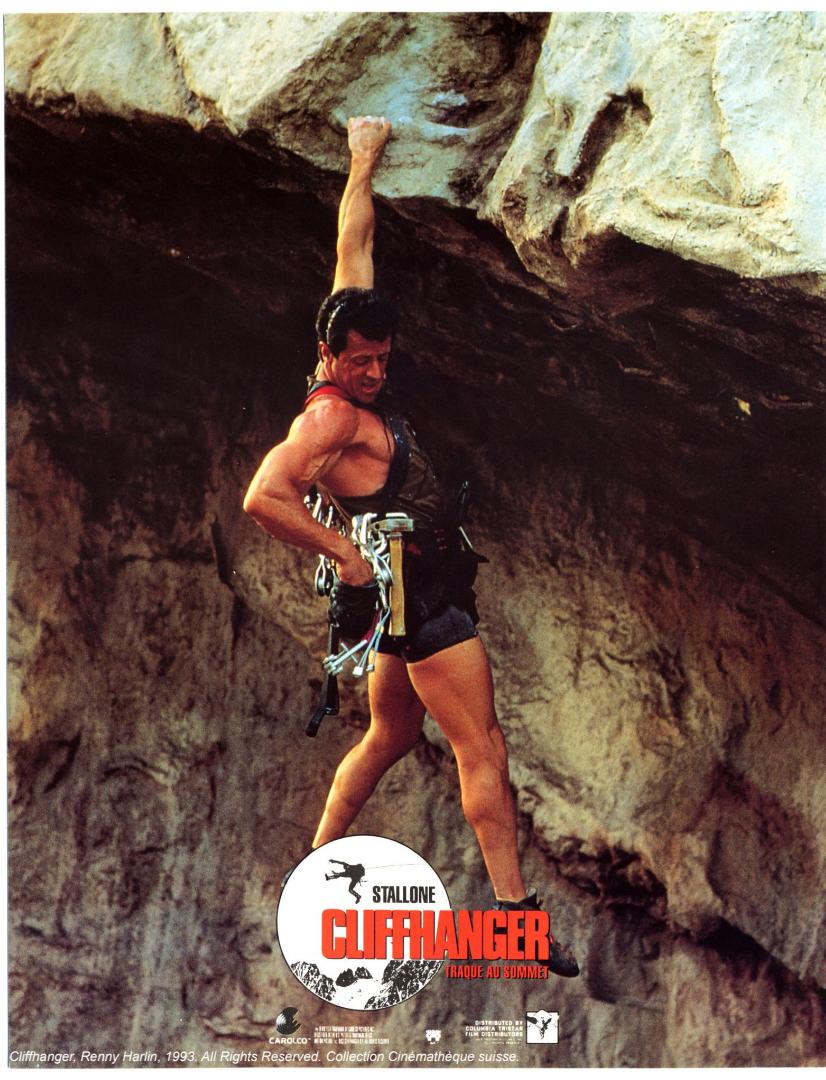
. Subtitle (presence) from CPL and ContentTitleText shall match.

ContentTitle suggest Subtitle but CPL contains No Subtitle

**DCP - /media/it\_dev\_lasco/IT\_DEV\_LASCO/01-DCP/01.05-DCP\_Proxy\_[0]/ - Check succeeded**



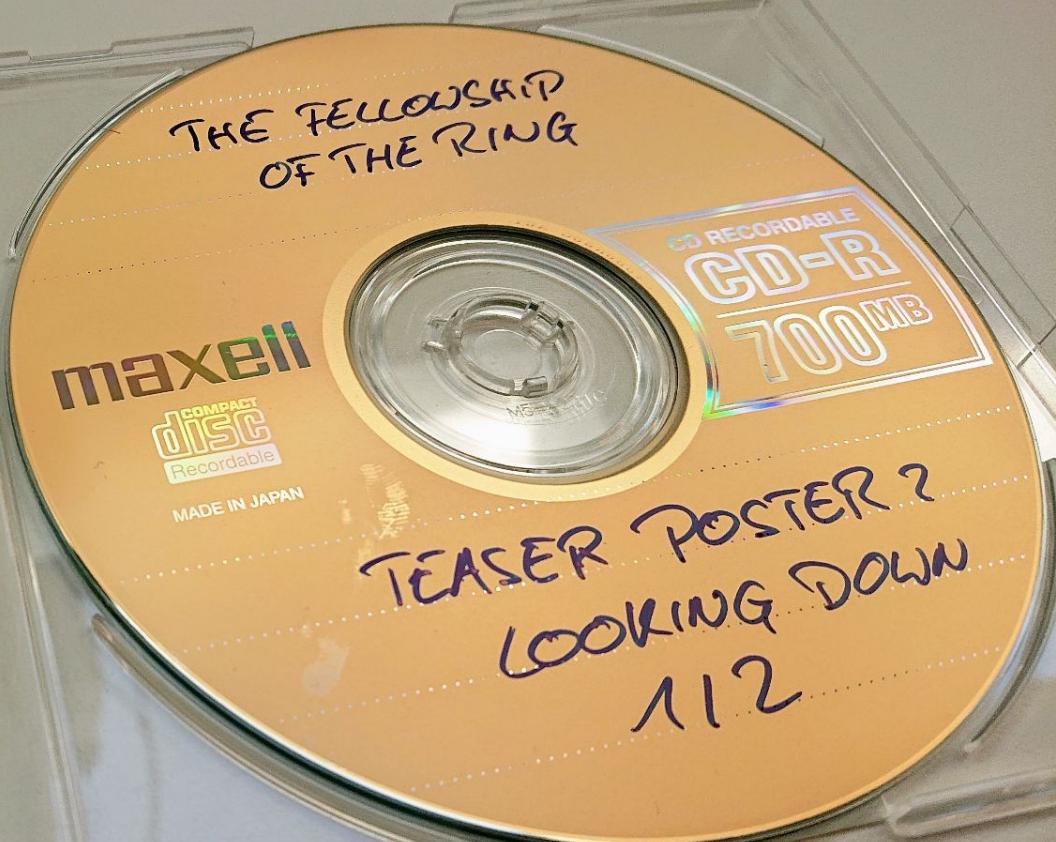
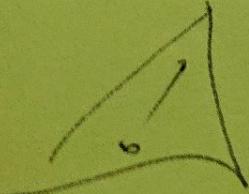
# Demo





# Les logiciels libres pour la préservation et extraction des données des supports informatiques

Un  
readable





# Brocki BOX

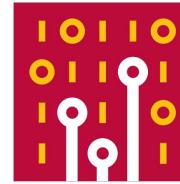
Pour le  
déménagement :

- Cuisine
- Salon
- Salle de bain
- Chambre à coucher
- Galetas

Photos

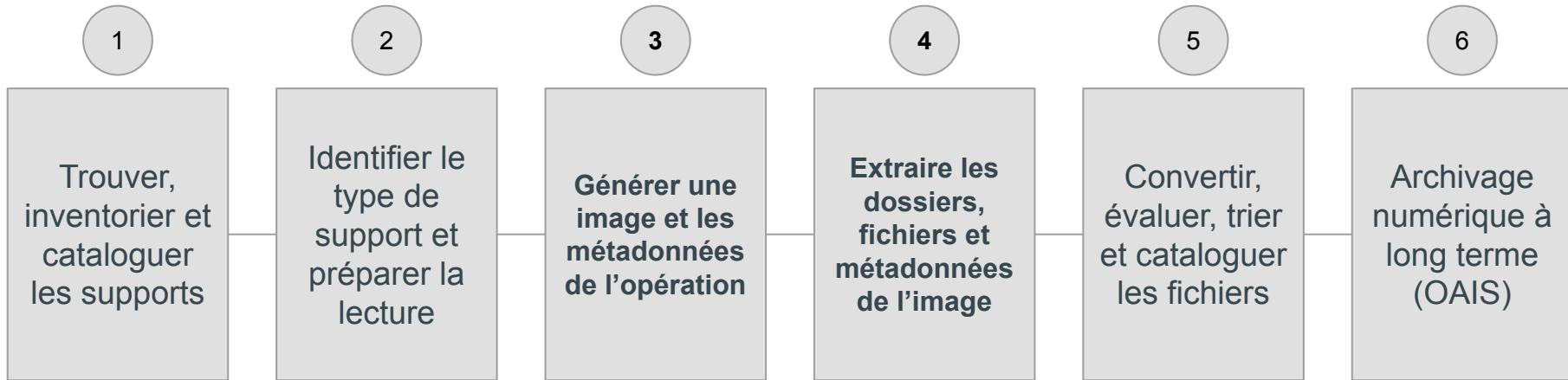
ne lange  
(TV - festivals - cartes postales)

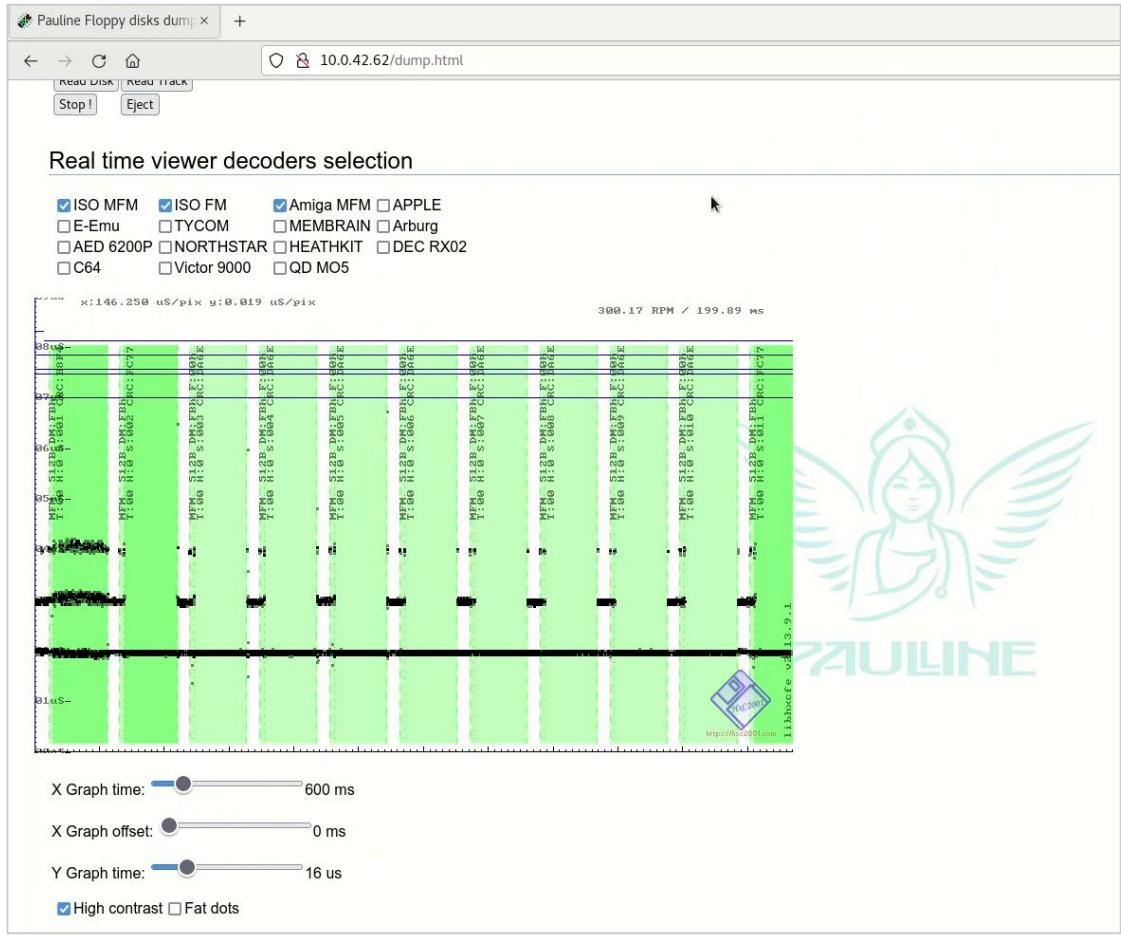
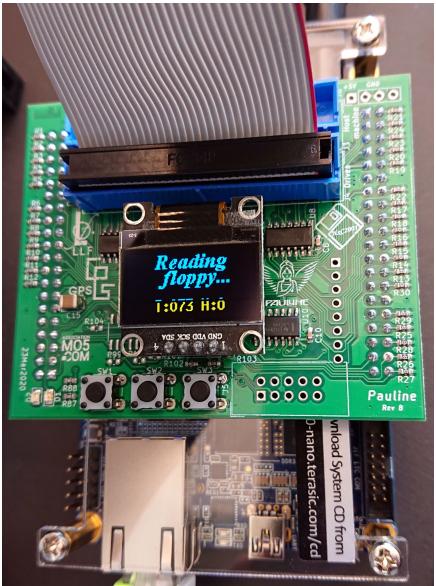
A TRIER



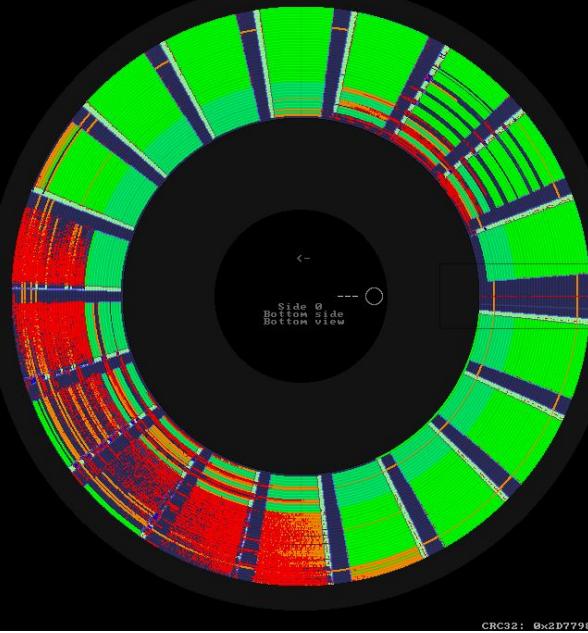
Computer  
History  
Museum



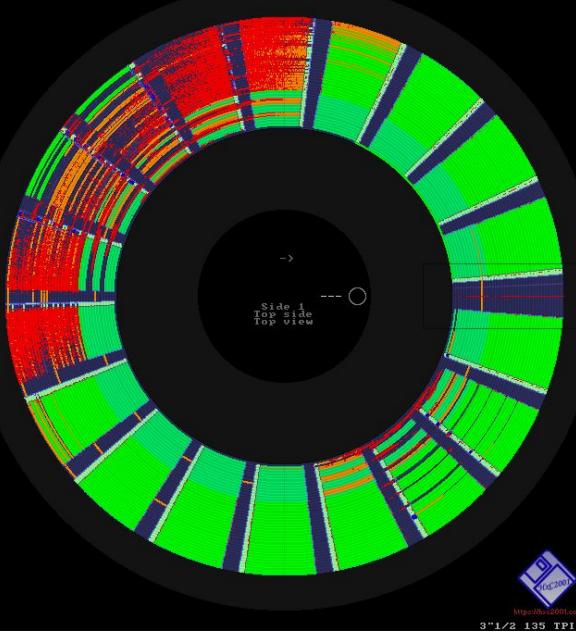




Visual Floppy disk  
Side 0: 41 tracks  
1324 Sectors, 366 bad  
651264 Bytes  
ISO MFM



track11.0.hexstream  
Side 1: 41 tracks  
1328 Sectors, 358 bad  
652544 Bytes  
ISO MFM





## Fonctionnalités majeures

Faire une image d'un support

Analyse d'image  
Conversion d'image  
Comparaison d'images

Identification des systèmes de fichiers

Extraction des dossiers/fichiers

```
# aaru i info Fellowship_of_the_ring_unreadable.aaru
aaru 6.0.0-alpha8+b1ac16c5 built by pcw-pz-extra01
in Release
Copyright © 2011-2022 Natalia Portillo
```

Image format identified by Aaru Format  
(49360069-1784-4a2f-b723-0c844d610b0a).

#### Image information:

**Format:** Aaru version 1.0

Was created with Aaru version 5.3

Image without headers is 169828977 bytes long

Contains a media of 69988 sectors with a maximum sector size of 2048 bytes (if all sectors are of the same size this would be 143335424 bytes)

Created on 21.01.2022 08:42:35

Last modified on 21.01.2022 08:51:38

Contains a media of type CDR and XML type

OpticalDisc

Has partitions

Has sessions

Contains 2 readable media tags:

CD\_FullTOC CD\_ATIP

Contains 9 readable sector tags:

CdSectorSync CdSectorHeader CdSectorSubHeader  
CdSectorEdc CdSectorEccP CdSectorEccQ CdSectorEcc  
CdSectorSubchannel CdTrackFlags

#### CompactDisc Table of Contents contained in image:

First complete session number: 1

Last complete session number: 1

Session 1

First track number: 1 (Data track, recorded uninterrupted)

Disc type: 0

Last track number: 1 (Data track, recorded uninterrupted)

Lead-out start position: 15:35:13

Lead-out is data type

Data track 1 starts at: 00:02:00 (Data track, recorded uninterrupted)

**CompactDisc Absolute Time In Pregroove (ATIP) contained in image:**

Indicative Target Writing Power: 0x04

Disc is CD-R

Disc is medium type B, low beta category (B-) CD-R

Disc use is restricted

ATIP Start time of Lead-in: 97:25:24

ATIP Last possible start time of Lead-out: 79:59:74

Disc uses long strategy type dye (Cyanine, AZO, etc...)

Disc manufactured by: Hitachi Maxell, Ltd.

**Image sessions**

Session	First track	Last track	Start	End				
1	1	1	0	69987				

**Image tracks**

Track	Type	Bps	Raw bps	Subchannel	Pregap	Start	End			
1	CdMode1	2048	2352	Raw	150	0	69987			

**Image indexes**

Track	Index	Start	
1	0	-150	
1	1	0	

```
# aaru fs info Fellowship_of_the_ring_unreadable.aaru
aaru 6.0.0-alpha8+b1ac16c5 built by pcw-pz-extra01 in
Release
Copyright © 2011-2022 Natalia Portillo
```

Image format identified by Aaru Format.

2 partitions found.

**Partition 0:**

Name	Track 1
Type	CdMode1
Start	sector 0, byte 0
Length	69988 sectors, 0 bytes
Scheme	Optical disc track
Description	

Identified by ISO9660 Filesystem.

ISO9660 file system

Joliet extensions present.

**VOLUME DESCRIPTOR INFORMATION:**

System identifier: APPLE COMPUTER, INC., TYPE: 0002  
Volume identifier: IVY\_HILL\_GRAPHICS  
Volume set identifier:  
Publisher identifier:  
Data preparer identifier:  
Application identifier: TOAST ISO 9660 BUILDER COPYRIGHT  
(C) 1997 ADAPTEC, INC. - HAVE A NICE DAY

Volume creation date: 28.02.2001 07:16:28  
Volume modification date: 28.02.2001 07:16:39  
Volume does not expire.  
Volume has always been effective.  
Volume has 69836 blocks of 2048 bytes each

**JOLIET VOLUME DESCRIPTOR INFORMATION:**

System identifier: APPLE COMPUTER,  
Volume identifier: IVY HILL GRAPHIC  
Volume set identifier:  
Publisher identifier:  
Data preparer identifier:  
Application identifier: Toast Joliet Builder ©1997  
Adaptec, Inc.  
Volume creation date: 28.02.2001 07:16:28  
Volume modification date: 28.02.2001 07:16:39  
Volume does not expire.  
Volume has always been effective.

## Partition 1:

Name	Toast 3.5.7 PPC HFS Optimizer
Type	Apple_HFS
Start	sector 56, byte 114688
Length	69779 sectors, 142908416 bytes
Scheme	Apple Partition Map
Description	Partition flags: Partition is valid. Partition entry is allocated. Partition is readable.

Next allocation block: 1.

34888 volume allocation blocks.

4096 bytes per allocation block.

4096 bytes to allocate when extending a file.

4096 bytes to allocate when extending a Extents B-Tree.

4096 bytes to allocate when extending a Catalog B-Tree.

Sector of first allocation block: 12

Next unused CNID: 219

0 unused allocation blocks.

65536 bytes in the Extents B-Tree

0 bytes in the Catalog B-Tree

Volume name: IVY HILL GRAPHICS

Finder info:

CNID of bootable system's directory: 0

CNID of first-run application's directory: 0

CNID of previously opened directory: 0

CNID of bootable Mac OS 8 or 9 directory: 0

CNID of bootable Mac OS X directory: 0

0 blocks in volume cache

4096 blocks in volume bitmap cache

0 blocks in volume common cache

Volume is not bootable.

Identified by **Apple Hierarchical File System**.  
**Apple Hierarchical File System**

HFS uses 512 bytes/sector while device uses 2048 bytes/sector.

Master Directory Block:

Creation date: 28.02.2001 07:11:47

Last modification date: 28.02.2001 07:16:27

Volume has never been backed up

Volume was unmounted.

2 files on root directory

1 directories on root directory

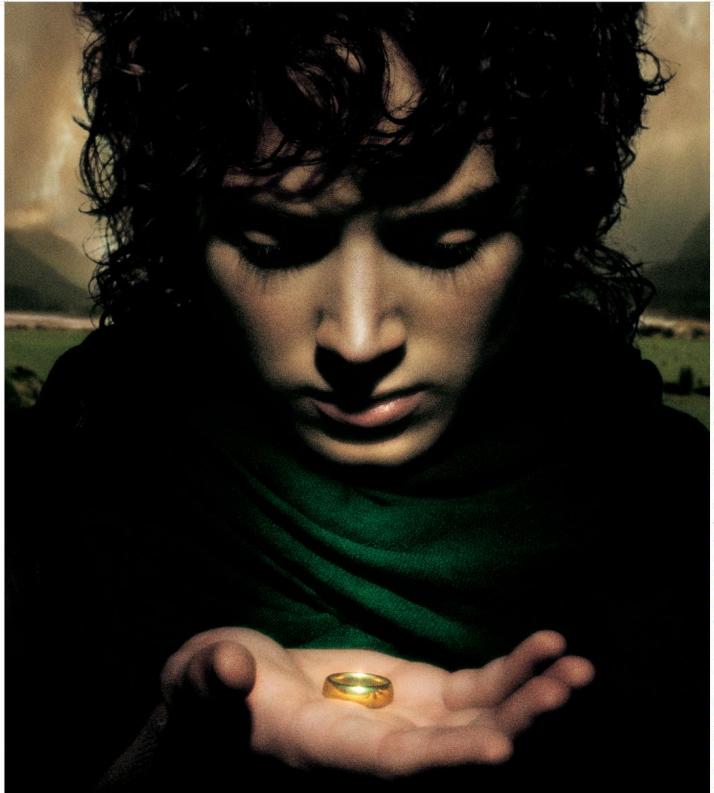
195 files on volume

8 directories on volume

Volume units count: 212

[Feb 28 2001] FELLOWSHIP OF THE RING 1SHT  
[Feb 24 2001] 8015556Q02r2.qxd  
[Feb 28 2001] FONTS  
[Feb 17 2001] 8015556FONTS  
[Feb 17 2001] 8-015556FONTS  
[Aug 30 1988] Clare  
[Aug 30 1988] ClareBol  
[...]  
[Jun 25 1998] WBRFonEle  
[Jun 25 1998] WBRFonTwe  
[Feb 28 2001] Icon  
[Feb 26 2001] PC FONTS  
[Jan 27 2000] AGENLEC\_.12  
[Jan 27 2000] AGENLEC\_.PFB  
[Mar 29 1994] FCEBO\_\_\_.PFB  
[...]  
[May 10 1994] KBW\_\_\_\_\_.PFB  
[Jan 25 2000] NSGOGBEC.PFB  
[Feb 26 2001] pfm  
[Jan 27 2000] AGENLEC\_.PFM  
[Dec 22 1994] FCEBO\_\_\_.PFM

[...]  
[Jan 25 2000] NSGOGBEC.PFM  
[Jun 7 2000] SAR\_\_\_\_\_.PFM  
[Dec 22 1994] UR\_\_\_\_\_.PFM  
[May 17 1994] SAR\_\_\_\_\_.PFB  
[Feb 16 2001] TT  
[Jan 2 2001] ABLEFONT.TTF  
[Aug 24 1996] ARIBLK.TTF  
[Jun 27 2000] EDDMB\_\_\_.12  
[Jun 27 2000] EDDMB\_\_\_.TTF  
[Mar 18 1994] UR\_\_\_\_\_.PFB  
[Feb 28 2001] IMAGE  
[Feb 24 2001] 8015556P02r3.tif  
[Feb 28 2001] Icon  
[Feb 28 2001] SUPPORT FILES  
[Aug 3 2000] 4-c SLUG.eps  
[Feb 24 2001] 8015552I05r1.eps  
[Feb 28 2001] 8015556I01r1.eps  
[Feb 28 2001] 8015556I02r2.eps  
[Feb 28 2001] 8015556I03r1.eps  
[Feb 28 2001] 8015556I05r1.eps  
[Feb 28 2001] 8015556I06r0.eps  
[Feb 28 2001] Icon



POWER CAN BE HELD  
IN THE SMALLEST OF THINGS

THE TRILOGY BEGINS DECEMBER 2001

THIS FILM IS NOT YET RATED

America's Oscar® Keyword: Lord of the Rings [www.lotrthelring.com](http://www.lotrthelring.com)

New Line Cinema

© 2001 New Line Cinema Productions, Inc. All Rights Reserved. Collection Cinémathèque suisse.

ivyhill



*The Lord of The Rings: The Fellowship of the Ring*. Peter Jackson, 2001. © 2001 New Line Cinema Productions, Inc. All Rights Reserved. Collection Cinémathèque suisse