
Rapport de stage du Master 1

Cybersécurité

Fournir un service numérique innovant à l'Institut
mémoires de l'édition contemporaine.

Matthias David

Stage effectué du 03/03/2025 au 04/07/2025

Tuteurs de stage : Emmanuel Giguet, Tanguy Gernot, Christophe Rosenberger

Enseignants référents : Gaétan Richard, Patrick Lacharme

Établissement d'enseignement : Université de Caen Normandie, Ensicaen

Laboratoire d'accueil du stage : Laboratoire CNRS UMR 6072, GREYC

Remerciements

Je tiens tout d'abord à exprimer ma gratitude envers **Emmanuel Giguet**, **Tanguy Gernot** et **Christophe Rosenberger**, mes tuteurs de stage au sein du laboratoire **GREYC**. Leurs conseils, la confiance qu'ils m'ont accordée, ainsi que la liberté d'action et l'autonomie qu'ils m'ont laissée tout au long de mes travaux ont été déterminants pour la réussite de ce stage.

Je remercie également **Gaétan Richard** et **Patrick Lacharme** pour la qualité de leurs enseignements, leur écoute attentive et leur aide précieuse dans la recherche de ce stage au cours de cette année universitaire.

Je souhaite exprimer ma reconnaissance à l'ensemble de l'équipe **SAFE** pour leur accueil, leur esprit collaboratif et leur soutien au quotidien, tout au long de mon stage. Je tiens aussi à remercier tout particulièrement les autres stagiaires ayant travaillé à mes côtés, notamment **Guillaume Hautot** et **Titouan Le Bret**, avec qui il a toujours été un plaisir de collaborer et d'échanger sur les difficultés rencontrées. Je remercie également **Louise** et **Juliette** de l'**IMEC**, qui ont su écouter nos longues présentations tout en gardant un intérêt sincère et passionné pour nos avancées.

Je n'oublie pas mes proches, qui m'ont constamment soutenu dans la construction de mon projet professionnel et m'ont encouragé à chaque étape de la rédaction de ce rapport.

Remerciements spéciaux à mes relecteurs et correcteurs qui ont contribué, grâce à leurs conseils et recommandations, à l'écriture et au bon déroulement de mon rapport de stage.

Abstract

Depuis plusieurs décennies, la croissance exponentielle des données numériques et l'obsolescence rapide des technologies ont rendu leur préservation et leur exploitation particulièrement complexe. Pour répondre à ces enjeux, le **ministère de la Culture** a lancé le projet **IANEC**, en partenariat avec l'équipe **SAFE** du laboratoire **GREYC** et l'**IMEC** (*Institut Mémoires de l'Édition Contemporaine*). Ce projet vise à développer des solutions innovantes permettant d'identifier, classifier et exploiter des archives numériques privées, souvent issues d'environnements informatiques anciens.

Dans ce cadre, mon stage a notamment porté sur deux axes complémentaires :

- L'analyse et l'émulation d'anciens systèmes d'exploitation (**Mac OS Classic** et **Mac OS X**), largement représentés dans les collections de l'**IMEC**, afin d'en comprendre le fonctionnement, en particulier au niveau des systèmes de fichiers.
- Le développement d'outils logiciels pour unifier l'exploitation et l'émulation de plusieurs systèmes, tout en assurant leur compatibilité avec les infrastructures existantes de l'**IMEC**.

Ces travaux combinent des techniques de rétro-ingénierie, d'émulation et de conception logicielle, dans une démarche collaborative visant à renforcer les capacités d'investigation numérique de l'**IMEC**. Les résultats obtenus ouvrent des perspectives pour la préservation à long terme des archives numériques et leur exploitation dans des environnements modernes.

Mots clés : **Préservation numérique, Forensique numérique, Rétro-ingénierie, émulation, Mac OS Classic/X, Ingénierie logicielle.**

Table des matières

1	Introduction	1
1.1	Contexte	1
1.2	Problématique	1
1.3	Méthodologie	2
1.4	Structure du rapport	2
2	Contexte et cadre du projet IANEC	3
2.1	La préservation numérique	3
2.1.1	L'essor des données numériques	3
2.1.2	État de l'art de la préservation numérique	4
2.2	Le projet IANEC : une réponse aux défis de la préservation	5
2.2.1	Origine et objectifs du projet	5
2.2.2	Les partenaires et leurs rôles	5
3	Travaux réalisés	7
3.1	Répartition des tâches et organisation du travail	7
3.1.1	Objectifs personnels	8
3.2	Analyse et émulation des systèmes Mac OS Classic et Mac OS X	8
3.2.1	Contexte des archives de l'IMEC	8
3.2.2	Méthodologie d'analyse	9
3.3	Développement d'outils logiciels	15
3.3.1	Difficultés et solutions apportées	18
4	Résultats et compétences acquises	20
4.1	Intégration et validation des résultats	20
4.1.1	Tests et validation	20
4.1.2	Documentation et pérennisation	20
4.2	Compétences techniques acquises	21
4.2.1	Technologies et outils	21
4.2.2	Rétro-ingénierie	21
4.2.3	Intégration et automatisation	22
4.3	Compétences transversales	22
4.3.1	Gestion de projet collaboratif	22
4.3.2	Documentation et vulgarisation technique	23
4.3.3	Capacité d'adaptation	23
4.4	Perspectives	23
5	Conclusion	25

Chapitre 1

Introduction

1.1 Contexte

La transformation numérique a profondément modifié notre rapport à l'information. Chaque année des volumes colossaux de données sont produits, et une partie d'entre elles possède une valeur culturelle, historique ou juridique majeure. Préserver ces archives est un enjeu stratégique pour assurer leur transmission et leur exploitation à long terme. Cependant cette mission se heurte à deux problèmes majeurs : l'obsolescence rapide des formats et des supports, et la complexité des environnements logiciels d'origine, souvent indispensable pour interpréter correctement les données.

1.2 Problématique

C'est dans ce contexte que le **Ministère de la Culture** a lancé le projet **IANEC**, en collaboration avec l'équipe **SAFE** du laboratoire **GREYC** et l'**IMEC** (*Institut Mémoires de l'Édition Contemporaine*). L'objectif est de concevoir et d'évaluer des méthodes innovantes permettant l'exploitation, l'identification et la classification d'archives numériques privées collectées par l'**IMEC**.

L'expertise de l'équipe **SAFE** en science de l'investigation (forensique), ainsi que ses compétences en conception logicielle, combinées aux ressources et au savoir-faire archivistique de l'**IMEC**, ont permis à six stagiaires — encadrés par **Emmanuel Giguet**, **Tanguy Gernot**, **Christophe Rosenberger** — de prendre part à ce projet.

Dans le cadre de ce stage, j'ai intégré le groupe dont la mission s'articulait autour de deux axes principaux :

- L'analyse et la compréhension des systèmes d'exploitation et logiciels anciens, en particulier sous **Mac OS Classic** et **Mac OS X**, très présents dans les archives de l'**IMEC** ;
- Le développement d'outils facilitant l'exploitation, le filtrage et la production de statistiques à partir des données extraites.

Ces objectifs visent à répondre à un enjeu central : **comment préserver et exploiter des archives numériques issues de systèmes obsolètes, tout en les intégrant dans des infrastructures modernes ?**

1.3 Méthodologie

Pour atteindre les objectifs fixés, plusieurs approches complémentaires ont été mises en œuvre. Mes travaux ont notamment porté sur :

- l'exploration des technologies d'émulation afin de reproduire fidèlement les environnements des fonds archivés ;
- des activités de rétro-ingénierie pour analyser notamment les systèmes de fichiers et certains logiciels ;
- le développement d'un logiciel modulaire avec une interface intuitive permettant d'unifier différentes technologies.

Bien que menés indépendamment, ces travaux ont bénéficié de l'aide de mes collègues, assurant une cohérence globale avec leurs propres travaux initiaux sur une base de données servant de support pour les statistiques et le filtrage. Ce travail collaboratif s'inscrit dans une vision commune visant à renforcer les capacités d'investigation numérique de l'IMEC.

1.4 Structure du rapport

Ce rapport est structuré de la manière suivante. Dans une première partie, nous présenterons le contexte de la préservation numérique, puis nous détaillerons le projet **IANEC**, en mettant en avant ses partenaires ainsi que les objectifs qu'ils se sont fixés

La deuxième partie sera consacrée aux travaux réalisés : nous y décrirons la répartition des tâches, les technologies et outils utilisés pour atteindre les objectifs définis.

La troisième partie exposera les résultats obtenus, en mettant en évidence les avancées effectuées, les difficultés rencontrées au cours du stage, ainsi que les compétences acquises sur le plan personnel et professionnel.

Enfin, le rapport se conclura par une synthèse du projet, accompagnée de pistes d'évolution et de perspectives pour de futurs travaux de recherche.

Chapitre 2

Contexte et cadre du projet IANEC

2.1 La préservation numérique

2.1.1 L'essor des données numériques

Depuis le début des années 1980, la démocratisation de l'informatique personnelle et l'explosion d'internet ont entraîné une croissance exponentielle de la production de données numériques. En 2025, on estime que plus de 180 zettaoctets de données sont générés à l'échelle mondiale ([Bar25]). Cette masse colossale de données se répartit sur une grande diversité de supports : d'abord magnétiques (disquettes, bandes, cassettes), puis optiques (CD, DVD, Blu-Ray), et plus récemment électroniques (clés USB, SSD, disques durs).

Cette évolution témoigne bien d'un développement technologique important. Cependant elle soulève également de nombreuses problématiques relatives à la conservation des données. L'obsolescence des supports et des formats constitue en effet une menace majeure : perte d'accès aux contenus, dégradation physique des supports, disparition des matériels de lecture, mais aussi incompatibilité des vieux formats avec les environnements et logiciels actuels.

Ces risques concernent particulièrement les institutions publiques, les administrations et les entreprises, qui se trouvent confrontés à la nécessité de préserver des fonds numériques variés et parfois uniques. Les enjeux sont multiples :

- Conservation du patrimoine culturel et scientifique, afin d'assurer une transmission efficace aux générations futures ;
- Garantir un accès long terme aux documents, au-delà des contraintes techniques et technologiques ;
- Respect du cadre juridique, qu'il s'agisse de la protection des données personnelles (**RGPD**), du droit d'auteur ou encore des obligations légales de conservation. A titre d'exemple les données de facturation doivent être conservées dix ans en application du Code de commerce, même si la personne concernée n'est plus cliente ([CNI20]).

La préservation numérique apparaît donc comme un défi actuel majeur, au croisement de la technique, du juridique et du sociétal.

2.1.2 État de l'art de la préservation numérique

Face à ces enjeux globaux, des solutions collaboratives et professionnelles ont été développées. Et ceux afin de garantir un accès à plus ou moins long terme aux données numériques.

Les méthodes couramment employées se regroupent principalement en trois catégories :

- **La duplication**, qui cherche à créer et maintenir plusieurs copies des données sur différents supports ou lieux géographiques. Son objectif est de protéger contre les modifications ou pertes accidentelles, les pannes matérielles ou les sinistres (incendies, inondations, cyberattaques). Cependant elle ne se suffit pas à elle seule, elle vise surtout à conserver les données sans prendre en compte leur exploitation futur. De plus cette technique engendre des coûts supplémentaires en stockage et en maintenance.
- **La migration**, qui consiste à transférer régulièrement les données vers de nouveaux supports ou formats afin d'éviter l'obsolescence matérielle ou logicielle. Cette approche est largement utilisée dans les administrations et les entreprises, mais elle requiert une vigilance constante ainsi qu'une connaissance approfondie de la documentation pour éviter toute perte d'information ou altération de l'intégrité des fichiers ;
- **L'émulation / La virtualisation**, qui vise à reconstituer l'environnement matériel ou logiciel d'origine afin de rendre accessibles des données ou applications anciennes sur des systèmes modernes. Cette méthode est de plus en plus exploitée dans les bibliothèques numériques et les laboratoires d'archives. De nombreux projets collaboratifs visent à augmenter les possibilités de cette technologie. Or, la mise en place de cette technologie requiert une documentation approfondie sur les systèmes émulés et implique une forte complexité technique, en particulier pour les environnements les plus anciens.

Ces méthodes sont aujourd'hui soutenues par de nombreux acteurs institutionnels et scientifiques. Parmi eux, on peut citer le **NIST** (*National Institute of Standards and Technology*), qui maintient des bases de référence dédiées aux formats, logiciels et systèmes d'exploitation. Les **Archives Nationales**, qui développent des plans de préservation à long terme. Ainsi que plusieurs projets européens, tels que **PREFORMA**, qui visent à définir et tester des standards ouverts pour améliorer la durabilité des formats.

Cependant, malgré ces avancées, des limites importantes subsistent. Les coûts élevés liés à la migration régulière ou au maintien d'infrastructures d'émulation/virtualisation peuvent freiner les institutions. De plus, la complexité technique de certaines solutions limite leur accessibilité aux non-spécialistes, et la question de la longévité de ces outils reste ouverte. Enfin, l'absence d'un cadre universel et normalisé complique la standardisation des pratiques entre les différents acteurs.

Ainsi, l'état de l'art montre l'urgence de préserver les données numériques face aux risques de perte ou d'obsolescence. Et d'autre part, la complexité de concevoir des solutions à la fois robustes, pérennes et économique. C'est précisément dans ce contexte que s'inscrit le projet **IANEC**, en cherchant à développer et évaluer des approches innovantes adaptées aux besoins spécifiques de l'**IMEC**.

2.2 Le projet IANEC : une réponse aux défis de la préservation

2.2.1 Origine et objectifs du projet

Le projet **IANEC** (*Investigation et Analyse des Archives Numériques de l'Édition Contemporaine*) est né d'une initiative du **ministère de la Culture**, dans le cadre de sa politique de préservation et de valorisation du patrimoine numérique.

Pour répondre aux enjeux vu par la préservation numérique, un partenariat a été mis en place entre le laboratoire **GREYC**, spécialisé notamment en forensique numérique et en ingénierie logicielle, et l'**IMEC** (*Institut Mémoires de l'Édition Contemporaine*), qui conserve et valorise un vaste ensemble d'archives privées d'écrivains, d'éditeurs et d'acteurs de la vie intellectuelle et culturelle.

Suite aux échanges entre ces différents partenaires, plusieurs objectifs principaux ont été fixés :

- **Classification automatiques** des archives numériques privées collectées par l'**IMEC**, en tenant compte de la grande variété de formats ;
- **Développement d'outils et logiciels** permettant l'analyse et l'exploitation des données archivées, afin de faciliter leur consultation et leur valorisation ;
- **Intégration des résultats** avec les méthodes déjà existantes de l'**IMEC**, en veillant à la compatibilité avec les pratiques archivistiques actuelles ;
- **Production de documentation** destinée à assurer la pérennité des travaux de recherches effectués et la réutilisabilité des outils développés.

Le projet **IANEC** se positionne donc comme un projet de recherche cohérent avec limites identifiées dans l'état de l'art, en cherchant à concilier rigueur scientifique, faisabilité technique et besoins concrets des professionnels de l'archivage.

2.2.2 Les partenaires et leurs rôles

Le projet **IANEC** repose sur une collaboration étroite entre plusieurs acteurs institutionnels et scientifiques, chacun apportant des compétences spécifiques et complémentaires.

Le GREYC et l'équipe SAFE

Le laboratoire **GREYC** (*Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen*), affilié au **CNRS** et à l'**Université de Caen**, joue un rôle central dans le projet. Au sein de ce laboratoire, l'équipe **SAFE** (*Sécurité, Analyse Forensique et protection de l'Information*) a mis à profit son expertise en investigation numérique (forensique), ingénierie logicielle et analyse de données. Expertise développée et mise en avant à travers la plateforme **G'DIP**.

Cette équipe a assuré :

- **l'encadrement technique et scientifique des stagiaires**, en leur fournissant un cadre méthodologique ;

- **la coordination des travaux de recherche**, afin de garantir la cohérence des travaux effectués par les stagiaires ;
- **le développement et l'évaluation des outils numériques**, en veillant à leur pertinence scientifique et technique.

Grâce à cette expertise, le **GREYC** constitue la base technique du projet, en apportant des solutions adaptées aux défis liés à l'obsolescence des environnements numériques.

L'IMEC

L'IMEC (*L'Institut Mémoires de l'Édition Contemporaine*) est une institution dédiée à la conservation et à la valorisation du patrimoine littéraire, éditorial et intellectuel contemporain. Son rôle dans le projet est double :

- **Fournir les archives numériques** issues des fonds privés d'auteurs, d'éditeurs et de chercheurs, qui présentent une grande diversité de formats et de systèmes d'exploitation, en particulier **Mac OS Classic et Mac OS X** ;
- **Exprimer ses besoins**, afin que les solutions développées soient compatibles compatibles avec les outils existant et adaptés aux travaux archivistiques.

Ces exigences permettent de cadrer la conception des outils et garantissent leur utilisations à long terme au sein de l'IMEC.

Une collaboration basée sur la complémentarité

Le succès du projet repose sur de nombreux échanges et la complémentarité des compétences entre les partenaires.

- **Les archivistes**, garant des normes de conservation et de la valeur culturelle des documents ;
- **Les chercheurs et ingénieurs en informatique**, responsables de concevoir des solutions techniques durables et d'assurer la coordination entre les archivistes et les stagiaires.
- **Les stagiaires**, qui contribuent à la mise en oeuvre des méthodes et au développement des outils.

Cette collaboration permet une approche pluridisciplinaire, essentielle pour relever les défis de la préservation numérique.

Chapitre 3

Travaux réalisés

3.1 Répartition des tâches et organisation du travail

Le stage s'est déroulé au sein de l'équipe **SAFE** du **laboratoire GREYC**, en collaboration avec cinq autres stagiaires et les encadrants du projet **IANEC**. Une répartition en deux groupes de travail a été définie dès le début afin de répondre aux différents objectifs du projet.

Le premier groupe a adopté une approche dite plus « *haut niveau* », c'est-à-dire l'utilisation d'outils simplifiant le développement afin de favoriser l'expérimentation. Leurs travaux portaient notamment sur le développement de solutions de classification automatique des documents en fonction de plusieurs critères : le type (image, texte, photo...), le contenu (privé, professionnel, médical...), la communicabilité, etc. Pour cela, plusieurs techniques d'intelligence artificielle ont été employées afin d'atteindre ces objectifs.

Le second groupe, dit « *bas niveau* », était composé de deux autres stagiaires et de moi-même. Nos objectifs se concentraient principalement sur la mise en oeuvre d'outils et de méthodes facilitant l'exploitation des fonds pour les archivistes de l'**IMEC**, ainsi que sur la réalisation d'outils statistiques et de filtres permettant une analyse rapide d'un fonds. Nos travaux ont notamment dû s'adapter, d'une part, à l'environnement (**Ubuntu 22**, fonds en grande partie sous **Mac OS**, etc.) et, d'autre part, aux outils déjà utilisés par l'**IMEC** (**Autopsy** pour analyse forensique, **SheepShaver** pour l'émulation Mac OS Classic, etc.).

Les encadrants, quant à eux, assuraient la coordination globale, la validation des choix techniques et méthodologiques, ainsi que le lien entre les attentes de l'**IMEC** et les objectifs scientifiques.

Pour maintenir une cohérence dans l'avancement, nous avons mis en place plusieurs méthodes de travail collaboratif :

- **Réunions hebdomadaires** avec l'ensemble des stagiaires et les encadrants, permettant de partager les avancées, d'identifier les difficultés et de réfléchir aux prochaines étapes ;
- Utilisation d'**outils de communication**, notamment un groupe **Discord** afin de faciliter les échanges entre stagiaires et archivistes de l'**IMEC** ;
- Mise en place d'un **dépôt Git partagé** pour versionner les codes développés et simplifier la collaboration ;
- Organisation de **revues techniques** (via **Unicloud**) régulières entre stagiaires afin de formaliser les avancées et éviter les doublons.

3.1.1 Objectifs personnels

Mes objectifs spécifiques dans le cadre du projet **IANEC** s'articulaient autour des axes suivants :

- Réaliser une **analyse approfondie des systèmes Mac OS Classic et Mac OS X**, en se concentrant notamment sur leurs systèmes de fichiers, leurs mécanismes de stockage et leurs logiciels ;
- Expérimenter et développer des **solutions d'émulation**, afin de reproduire les environnements d'origine et d'accéder aux contenus des archives sans altérer les données originales ;
- Concevoir et développer un **logiciel graphique modulable**, intégrant les différentes technologies d'émulation, permettant une exploitation simplifiée des données, tout en assurant sa compatibilité avec l'environnement présent ;
- Documenter les choix techniques, les procédures expérimentales et les résultats afin de garantir la **pérennité et la réutilisabilité** du travail.

Ces objectifs ont été atteints grâce à un esprit collaboratif au sein du groupe et avec les encadrants, ainsi qu'avec l'aide des travaux complémentaires des autres stagiaires.

3.2 Analyse et émulation des systèmes Mac OS Classic et Mac OS X

3.2.1 Contexte des archives de l'IMEC

Une part importante des archives numériques conservées par l'**IMEC** provient d'auteurs, éditeurs et institutions ayant travaillé sur des systèmes **Macintosh** au cours des années 1990 à 2010. À cette époque, les ordinateurs Apple étaient très répandus dans les milieux littéraires, artistiques et éditoriaux en raison de leur interface graphique intuitive et de leurs outils performants pour la création graphique et la mise en page. Une grande partie des supports collectés (disques durs, CD, disquettes) contiennent donc des données issues des systèmes **Mac OS Classic** (versions 6 à 9) et **Mac OS X** (versions 10.0 à 10.5).

Ces environnements ne sont plus supportés par les matériels actuels. De plus, certains formats propriétaires (documents, bases de données, logiciels) ne disposent plus d'outils natifs pour leur ouverture, ce qui complique le travail des archivistes.

Défis liés à l'obsolescence et à la compatibilité

Les principaux défis rencontrés peuvent être regroupés en trois axes :

- **Obsolescence matérielle** : Les systèmes **Mac OS Classic** fonctionnaient sur des architectures *PowerPC* ou *M68k*, aujourd'hui totalement abandonnées. Trouver des machines physiques en état de marche est devenu rare, cher et leur maintenance pose des problèmes logistiques.
- **Obsolescence logicielle** : Les applications et formats de fichiers (par exemple, **ClarisWorks**, **MacWrite**, **QuarkXPress** des années 90) ne sont pas directement compatibles avec les systèmes modernes. Certains formats sont peu documentés ou

entièrement propriétaires, ce qui nécessite des techniques de rétro-ingénierie pour comprendre leur fonctionnement interne.

- **Contraintes d'interopérabilité** : Même lorsqu'un fichier peut être lu, il est nécessaire de garantir sa lisibilité dans des environnements récents sans perdre d'information ni modification des métadonnées. Cette exigence est nécessaire pour garantir la préservation à long terme et se conformé aux standards archivistiques.

Face à ces problématiques, l'émulation et la virtualisation s'impose comme une solution clé. Elle consiste à reproduire un environnement logiciel ou un système d'exploitation ancien sur une machine moderne, permettant d'exécuter les systèmes et applications d'époque sans altérer les données originales. Ce choix technique a guidé une partie essentielle de mes travaux au cours de ce stage, en particulier pour les systèmes **Mac OS Classic** et les premières versions de **Mac OS X**.

3.2.2 Méthodologie d'analyse

Virtualisation et émulation

SheepShaver

Après une rapide analyse des fonds disponibles, cherchant à identifier les systèmes d'exploitation et leurs versions, nous avons choisi de concentrer nos efforts sur les environnements **Mac OS Classic**, en particulier les versions comprises entre 7 et 9.

En nous appuyant sur les travaux réalisés par un stagiaire lors de l'année précédente ([OMA24]), le choix s'est porté sur l'utilisation de **SheepShaver**, un émulateur open source capable de reproduire l'architecture *PowerPC*. Cet outil permet d'exécuter des systèmes **Mac OS** allant de la version **7.5.2** à **9.0.2**, assurant ainsi une compatibilité avec les archives analysées.

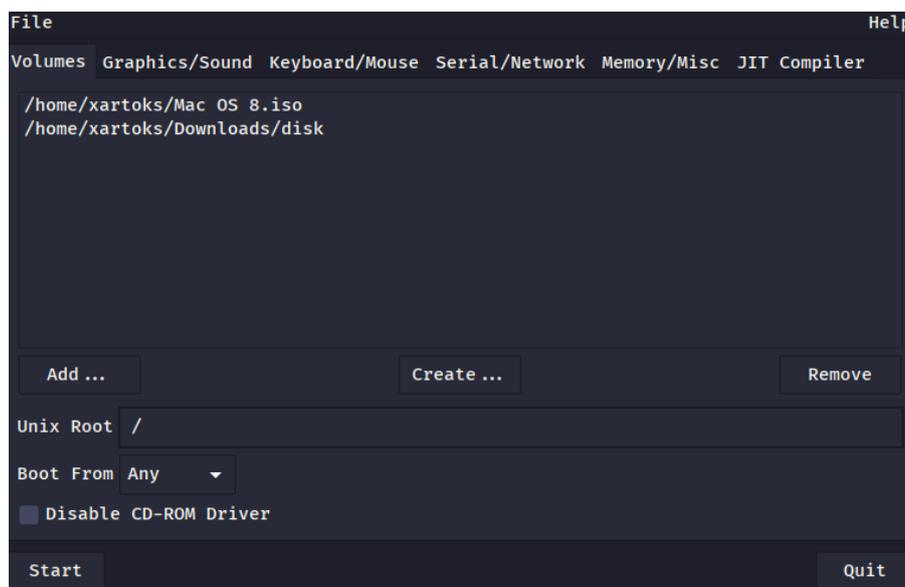


FIGURE 3.1 – Interface graphique permettant la configuration de **SheepShaver**

La procédure d'utilisation de **SheepShaver** est la suivante :

1. **Téléchargement et compilation** (ou installation) de SheepShaver ;

2. **Configuration des paramètres** (mémoire vive, résolution, périphériques) ;
3. **Création d'un disque virtuel** au format **HFS** ;
4. **Installation du système** via une image ISO d'un CD original.

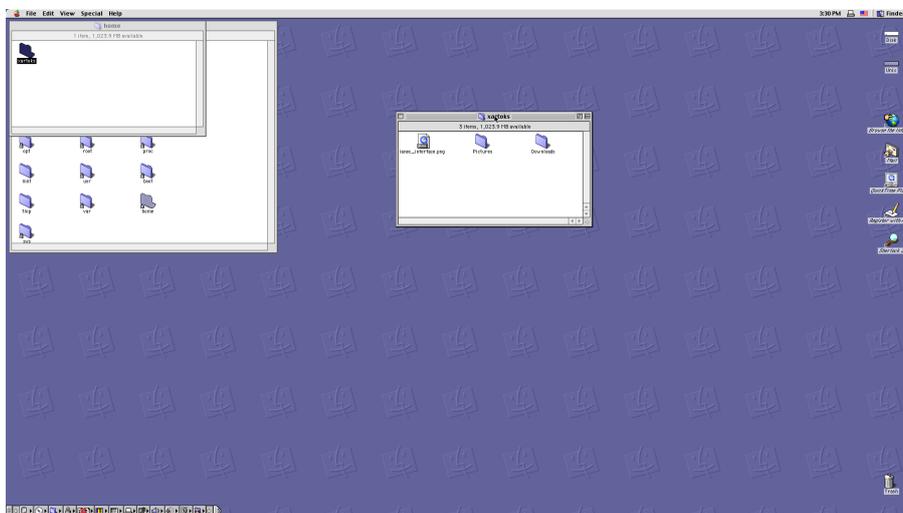


FIGURE 3.2 – Mac OS 9 émulé avec SheepShaver

L'utilisation de **SheepShaver** présente plusieurs atouts majeurs pour les travaux des archivistes :

- **Performances satisfaisantes**, offrant un environnement fluide sans bugs majeurs ;
- **Échanges simplifiés de fichiers**, en temps-réel, entre la machine hôte et l'environnement émulé, tout en conservant les caractéristiques propres au système **HFS** ;
- **Disponibilité d'une documentation** et d'une communauté active, facilitant la résolution de problèmes techniques.

Cependant, l'émulateur présente également certaines limitations :

- **Absence de support** pour les architectures *PowerPC G3/G4*, limitant l'émulation à **Mac OS 9.0.2** et excluant les premières versions de **Mac OS X** ;
- **Gestion limitée des périphériques**, pas de support USB, impression non fonctionnelle ;
- **Paramétrage complexe pour certaines fonctionnalités**, par exemple, export PDF.
- **Installation de logiciels** non standardisée, avec plusieurs versions disponibles en ligne, chacune nécessitant une procédure d'installation spécifique.

Pour fonctionner correctement, **SheepShaver** nécessite plusieurs éléments :

- **Un fichier ROM Apple**, indispensable pour reproduire fidèlement l'architecture *PowerPC*, trouvable sur **Internet Archive** ;
- **Une image disque d'installation de Mac OS** CD original ou ISO, trouvable sur **WinWorldPC** ou sur **Macintosh Repository** ;
- **Un disque virtuel HFS**, utilisé comme support pour le système et les applications, pouvant provenir des archives de l'**IMEC** ou être créé vierge si nécessaire.

QEMU PPC

Suite à la prise en main de **SheepShaver** et à l'analyse des premières archives, il est apparu nécessaire d'explorer une solution complémentaire capable de prendre en charge des environnements plus récents, notamment les premières versions de **Mac OS X** (10.0 à 10.5).

Pour répondre à ce besoin, notre choix s'est porté sur **QEMU** (*Quick EMUlator*), un logiciel open source largement utilisé dans le domaine de l'émulation et de la virtualisation. **QEMU** se distingue par sa liberté dans les configurations possible et sa capacité à émuler une large gamme d'architectures processeur, dont *PowerPC*. Dans le cadre de ce projet, nous avons exploité la version **PPC** de **QEMU** pour recréer un environnement **PowerMac G4**, permettant ainsi l'installation et l'exécution des premières versions de **Mac OS X**.

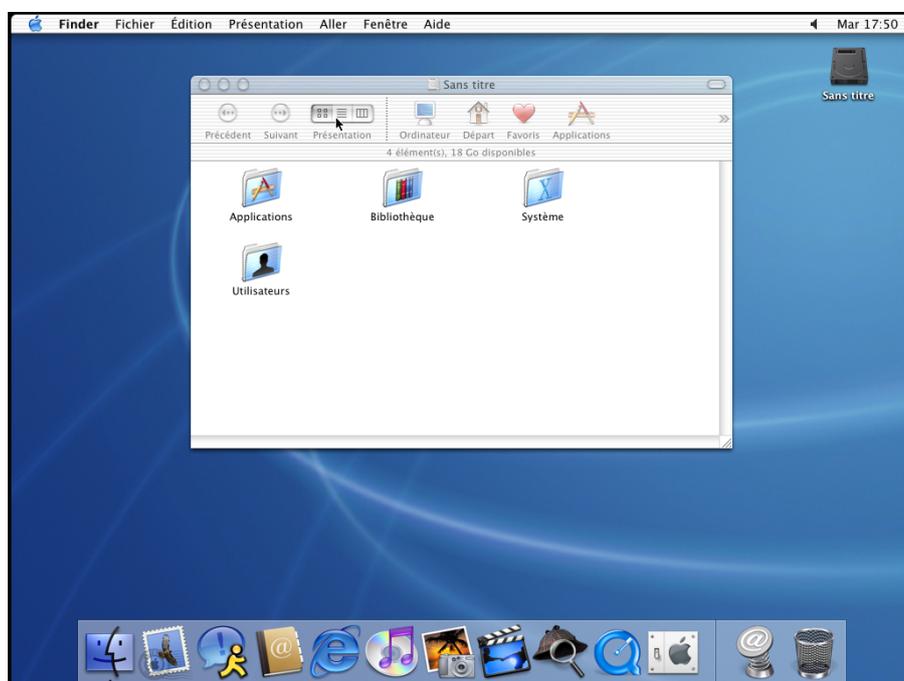


FIGURE 3.3 – Mac OS X 10.2 émulé sous QEMU PPC

QEMU PPC a été retenu pour plusieurs raisons :

- **Compatibilité étendue** avec les processeurs *PowerPC G3/G4*, nécessaires pour émuler **Mac OS X** ;
- **Support et documentation**, facilitant la configuration et le dépannage ;
- **Configurable**, permettant de simuler différents modèles de machines et de périphériques.

Contrairement à **SheepShaver**, **QEMU** prend également en charge la virtualisation des périphériques réseau et le montage de disques virtuels au format **QCOW2** (croissance dynamique).

L'utilisation de **QEMU** offre une meilleure couverture des systèmes visés, notamment pour **Mac OS X** (aussi possible d'émuler les dernières versions de **Mac OS Classic 9**,

mais présente aussi certaines contraintes :

- **Performances limitées**, l'émulation complète d'un G4 étant très gourmande en ressources ;
- **Complexité de configuration**, requiert la connaissance de ligne de commande complexe ;
- **Échanges de fichiers**, nécessitant différentes procédures suivant le sens de l'échange. Les échanges en temps réel ne sont pas possibles sans une configuration complexe, dépendant de versions spécifiques.

Grâce à **QEMU**, nous avons pu accéder à des environnements hybrides entre **Mac OS 9** et les premières versions de **Mac OS X**, permettant d'explorer des documents et logiciels absents des systèmes émulés par **SheepShaver**. L'utilisation combinée de ces deux solutions nous a permis de couvrir un large spectre des systèmes présents dans les fonds de l'IMEC.

TABLE 3.1 – Comparaison entre **SheepShaver** et **QEMU PPC**

Critères	SheepShaver	QEMU PPC
Compatibilité	Mac OS 7.5.2 à 9.0.2	Mac OS 9.2 à Mac OS X 10.5
Architecture	PowerPC (603/604)	PowerPC G3 / G4
Performances	Excellentes (émulation rapide et fluide)	Moyennes (émulation coûteuse en ressources et quelques bugs de souris)
Utilisation	Simple (interface graphique disponible)	Complexe (requiert des lignes de commande)
Installation de logiciels	Complexe (procédures différentes suivant la source)	Simple (copier-coller de l'installateur)
Échange de fichiers	Oui (montage HFS intégré)	Limité (nécessite procédure spécifique)
Support périphériques	Limité (pas d'USB, impression non supportée)	Avancé (réseau, périphériques virtuels)
Format disque	Fichiers HFS	Formats RAW, HFS+ et QCOW2 (<i>copy-on-write</i>)
Documentation et communauté	Active et accessible	Très active, mais ressources éparpillées

Documentation

En complément de la mise en place des environnements virtuels, une documentation technique détaillée a été rédigée pour chaque technologie d'émulation utilisée (**SheepShaver** et **QEMU PPC**). Cette documentation couvre :

- **L'installation** : prérequis, téléchargement des dépendances et configuration initiale ;
- **La configuration** : réglages mémoire, périphériques, montage des volumes, paramètres réseaux ;

- **L'utilisation** : procédures pour lancer les systèmes, échanger des fichiers avec l'hôte, installation de logiciels.

En parallèle, une base de connaissances a été constituée concernant les applications présentes dans les fonds analysés. Pour chaque application, un tableau récapitulatif a été produit contenant :

- **Nom de l'application** ;
- **Version** et langue ;
- **Source d'obtention** (Liens vers le site source et vers une copie sur Unicloud) ;
- **Compatibilité système** (Version de Mac OS Classic/Mac OS X) ;
- **Procédure d'installation** détaillée.

Cette documentation vise à pérenniser les travaux réalisés, à faciliter la réutilisation des environnements, et à garantir une exploitation fiable des applications présentes dans les archives.

Rétro-ingénierie

Après avoir mis en place les environnements émulés avec **SheepShaver** et **QEMU**, il a été possible de lancer des tests les proches de celles des systèmes d'origine. Cependant, l'émulation ne suffit pas à comprendre en profondeur la structure des applications et des fichiers, ni pour identifier les mécanismes propres à Mac OS Classic/X. Pour cette raison, une phase de rétro-ingénierie a été menée afin d'analyser les systèmes de fichiers (**HFS**, **HFS+**), les formats binaires et l'organisation des logicielles.

Cette analyse s'appuie sur des outils adaptés à ce contexte historique : **hfsutil** pour extraire et inspecter les disques et partitions, **hexdump** pour explorer les fichiers en hexadécimal, et **ResEdit** (exécuté dans Mac OS émulé) pour examiner et modifier les ressources internes des applications. Des ressources documentaires issues de la communauté ont également été utilisées notamment des manuels consacrés à la programmation sous Mac OS. notamment des manuels de programmation Mac OS disponible sur le site **Vintage Apple**.

Système de fichiers HFS/HFS+

Les supports analysés utilisaient principalement **HFS** (*Hierarchical File System*) et, dans certains cas, **HFS+**, des systèmes de fichiers caractéristiques de **Mac OS Classic** et des premières versions de **Mac OS X**. **HFS** a introduit une organisation hiérarchique des données avec dossiers et sous-dossiers, en remplacement de l'ancien **MFS** (*Macintosh File System*), qui ne proposait qu'un répertoire unique.

Une particularité essentielle de ces systèmes est la division de chaque fichier en deux forks :

- **Data Fork** (obligatoire) : contient les données principales du fichier (par exemple, le texte d'un document ou une image brute) ;
- **Resource Fork** (optionnelle) : stocke des informations supplémentaires structurées, telles que des icônes, des menus, des polices ou des paramètres liés à l'application.

Ce mécanisme permet d'associer à un même fichier des données binaires et des ressources graphiques ou textuelles sans les mélanger. Or, cette architecture étant spécifique à **Mac OS**, lorsqu'un disque **HFS/HFS+** est monté sur un système moderne, seul le **Data Fork** est généralement accessible, la **Resource Fork** étant ignorée. Cependant, cette dernière peut contenir des informations précieuses, notamment pour l'analyse archivistique.

Pour pallier cette limitation, **SheepShaver** implémente une solution consistant à créer, dans chaque répertoire, un sous-dossier nommé *.rsrc*. Ce dernier contient, pour chaque fichier du répertoire principal (fork de données), un fichier du même nom représentant son **Resource Fork**.

Afin de reproduire ce comportement lors du montage d'un disque **HFS** sur un système moderne, un outil spécifique a été développé par **Guillaume Hautot**, basé sur **hfsutil**, permettant d'émuler la gestion des forks de manière similaire à **SheepShaver**.

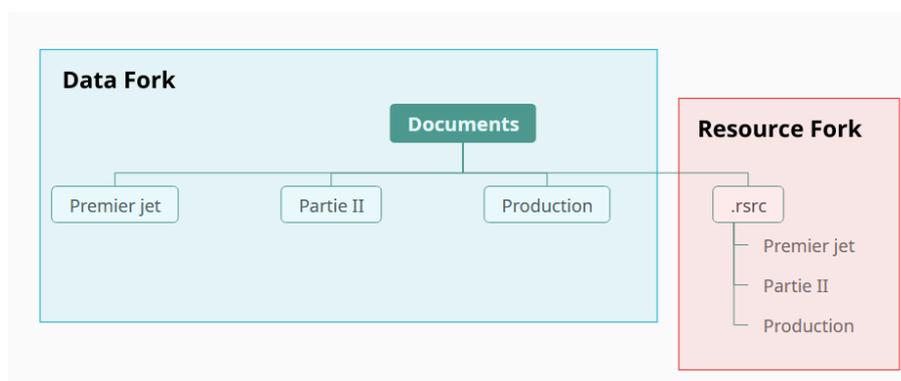


FIGURE 3.4 – Reproduction du système **HFS** par **SheepShaver**

Analyse des fichiers et formats propriétaires

Suivant les travaux précédents menée par **Thorsten Ries** -([Rie22]) sur «*la reconstruction des processus d'écriture numériques de Derrida grâce à la computer forensics*», les recherches se sont concentrées sur des documents issus de logiciels d'édition rencontrés dans les fonds et populaires tels que **ClarisWorks** ou **Microsoft Word**. Ces formats posent encore aujourd'hui des défis :

- Certains logiciels peuvent encore être installés dans les mêmes versions sous **Windows**, mais le contenu la mise en forme n'est pas identique (problème d'encodage, différences d'alignement, de polices ou de styles).
- **LibreOffice**, via des bibliothèques comme **libmwaw**, offre une compatibilité relativement bonne, mais des écarts subsistent (marges, polices propriétaires non reconnues).

Grâce à la rétro-ingénierie, il a été possible de comprendre comment ces paramètres de mise en forme sont stockés dans les fichiers d'origine :

- Inspection des entêtes et des blocs dédiés à la mise en page.
- Identification des zones contenant les informations sur les polices, les styles et les marges.

- Vérification de la présence éventuelle de dépendances avec des ressources externes (modèles, icônes).

L'utilisation de hexdump a permis d'examiner les structures binaires, souvent en big-endian, et d'identifier les segments liés aux métadonnées de formatage. Ces observations expliquent pourquoi certaines pertes ou modifications interviennent lors de la migration vers des formats modernes (ODT, DOCX).

Les métadonnées `type/creator`

Un autre aspect révélé par cette analyse est le système `type/creator`, utilisé par **Mac OS** pour identifier les fichiers :

- **Type** : nature du contenu (ex. TEXT pour un document texte, PICT pour une image).
- **Creator** : application d'origine (ex. MSWD pour Microsoft Word, BOBO pour ClarisWorks).

Ces informations, stockées dans le catalogue **HFS**, sont essentielles pour ouvrir un fichier avec la bonne application, même sans extension. Dans un contexte archivistique, elles constituent un élément clé pour la préservation, car elles permettent d'associer un fichier à son environnement logiciel d'origine. Cela donne également la possibilité d'analyser et de produire des statistiques sur les pratiques et comportements des utilisateurs.

Une liste non exhaustive de `TYPE/CREATOR` peut être trouvée sur **MacDisk**.

Intérêt pour la préservation numérique

Cette phase de rétro-ingénierie a permis :

- D'identifier les structures responsables de la mise en page dans les formats propriétaires.
- De comprendre les causes des écarts de rendu après migration.
- De documenter le rôle des forks et des métadonnées `type/creator` dans les systèmes logiques Mac OS Classic.

Ces connaissances sont essentielles pour définir des stratégies de conservation (choix des formats, traitement des forks, documentation des dépendances) et pour améliorer la fidélité des environnements émulés.

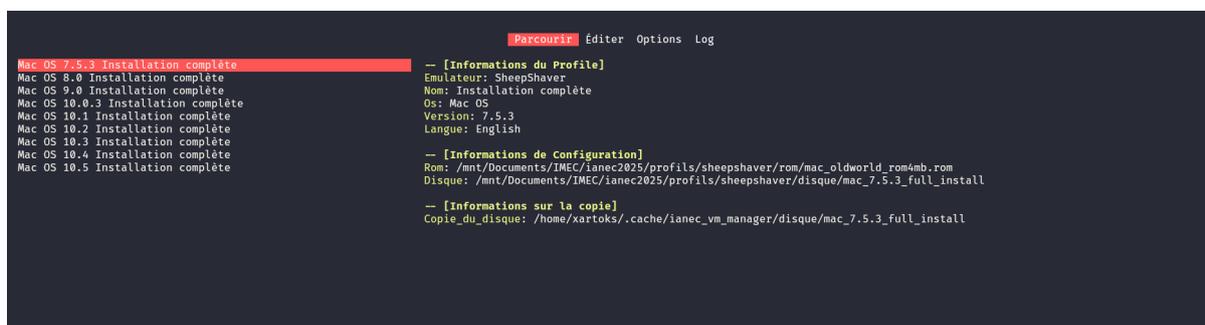
3.3 Développement d'outils logiciels

Dans le cadre du projet, un outil logiciel a été développé pour faciliter l'accès et l'exploitation des environnements **Mac OS Classic** et **Mac OS X** au sein des archives de l'**IMEC**. Cet outil répond à plusieurs objectifs liés à la conservation numérique, à l'ergonomie pour les archivistes et à la compatibilité avec les infrastructures existantes.

Objectifs du logiciel développé

Le développement a visé plusieurs objectifs majeurs :

- **Unification des technologies d'émulation** : l'outil regroupe dans une interface graphique commune différentes solutions d'émulation (SheepShaver pour **Mac OS Classic** et **QEMU** pour **Mac OS X**), afin d'éviter aux utilisateurs de devoir connaître et manipuler en détails plusieurs logiciels.
- **Interface intuitive** : l'application fournit une interface textuelle ergonomique (basée sur **ratatui**) permettant la manipulation et le lancement de plusieurs environnements sans recourir à des lignes de commandes complexes. L'objectif est de rendre l'utilisation accessible aux équipes non techniques de l'IMEC.
- **Une intégration fluide avec les autres outils** : l'architecture du logiciel est conçue de manière à permettre l'ajout facile de nouvelles technologies d'émulation, tout en garantissant la possibilité pour des logiciels externes d'interagir avec ses données si nécessaire.



```
Parcourir Éditer Options Log
Mac OS 7.5.3 Installation complète
Mac OS 8.0 Installation complète
Mac OS 9.0 Installation complète
Mac OS 10.0.3 Installation complète
Mac OS 10.1 Installation complète
Mac OS 10.2 Installation complète
Mac OS 10.3 Installation complète
Mac OS 10.4 Installation complète
Mac OS 10.5 Installation complète

-- [Informations du Profil]
Emulateur: SheepShaver
Nom: Installation complète
Os: Mac OS
Version: 7.5.3
Langue: English

-- [Informations de Configuration]
Rom: /mnt/Documents/IMEC/ianec2025/profils/sheepshaver/rom/mac_oldworld_rom4mb.rom
Disque: /mnt/Documents/IMEC/ianec2025/profils/sheepshaver/disque/mac_7.5.3_full_install

-- [Informations sur la copie]
Copie_du_disque: /home/xartoks/.cache/ianec_vm_manager/disque/mac_7.5.3_full_install
```

FIGURE 3.5 – Interface du logiciel

Architecture et technologies

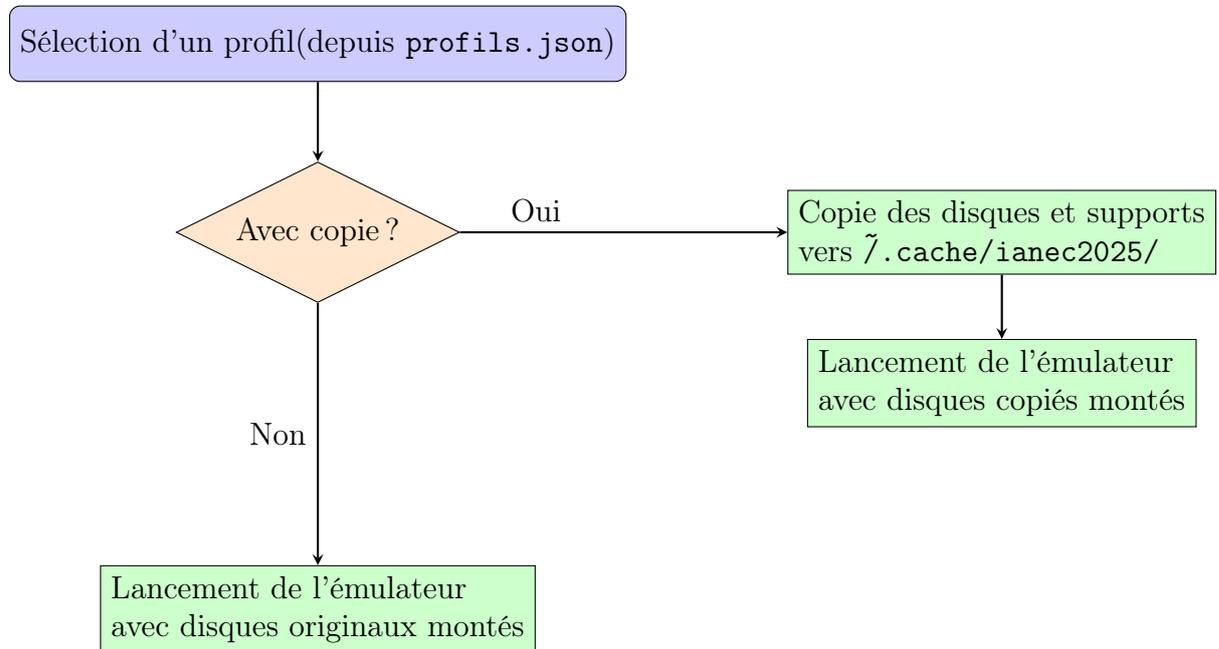
L'architecture du logiciel repose sur des principes de modularité et de compatibilité. Le langage **Rust** a été choisi pour sa sécurité mémoire, sa fiabilité et ses performances élevées. Sa bibliothèque standard riche, et l'écosystème étendu des crates, ont permis d'accélérer le développement tout en assurant une base robuste.

L'interface en mode texte est construite avec la bibliothèque **ratatui**, offrant une navigation fluide au clavier et à la souris, directement dans le terminal, ce qui la rend particulièrement adaptée aux environnements Linux.

L'architecture de l'application a été conçue pour être facilement extensible : il est possible d'ajouter de nouveaux émulateurs ou des versions supplémentaires de Mac OS (ou d'autres systèmes) sans réécrire le cœur du programme. Le seul prérequis est que chaque émulateur soit associé à une commande générique permettant le lancement du système choisi.

Des disques préconfigurés ont été générés pour chaque version majeure, de **Mac OS 9** à **Mac OS 10.5**, afin d'éviter la configuration manuelle fastidieuse (création du disque et installation du système).

L'ensemble des profils configurés et des informations associées par l'utilisateur (ou le développeur) est structuré dans un fichier JSON, tandis que les disques correspondants sont stockés dans le cache de l'application, garantissant un accès rapide et la possibilité pour des outils tiers d'interagir facilement avec ces profils.



Fonctionnalités clés

Le logiciel intègre plusieurs fonctionnalités essentielles pour le travail archivistique :

- **Disques préconfigurés pour chaque version majeure** de Mac OS Classic et Mac OS X (de 7.5.2 à 10.5) permettant un usage rapide et standardisé des environnements.
- **Support simple des périphériques et images externes** : l'utilisateur peut attacher des disques ou des supports à l'émulateur via l'interface.
- **Travail par défaut sur des copies** : toutes les opérations se font sur des duplicatas afin de préserver l'intégrité des originaux.
- **Création et duplication d'environnements** : possibilité de copier un environnement existant ou d'en créer un nouveau basé sur une configuration existante.
- **Compatibilité avec les infrastructures IMEC** : fonctionnement optimisé sous Linux, avec support WSL pour Windows, afin d'éviter les contraintes matérielles.

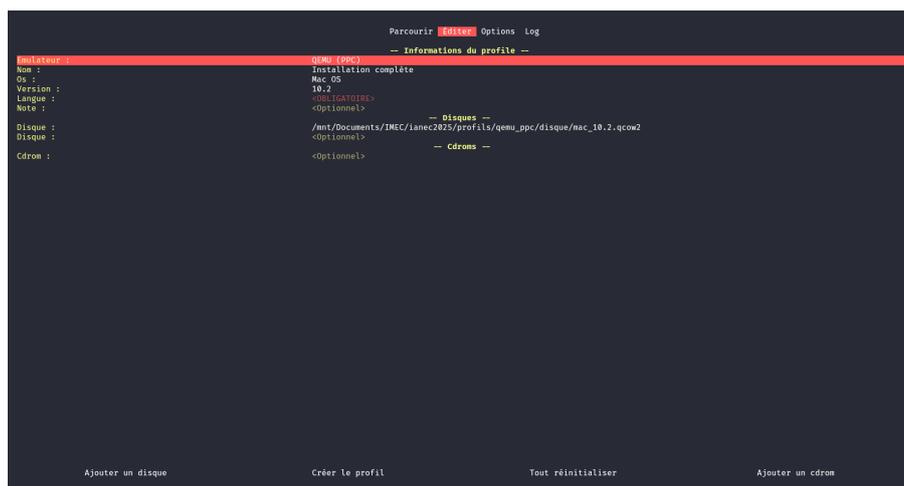


FIGURE 3.6 – Interface de création/modification de profils

Intérêt pour la préservation numérique

Cet outil répond à un besoin critique : **normaliser l'accès aux systèmes historiques tout en sécurisant les données archivées.**

Il permet de :

- **Simplifier la manipulation des environnements** pour les archivistes, même sans expertise technique ;
- **Réduire les risques d'erreurs** en automatisant la préparation des disques, l'attachement de support et la gestion des copies ;
- **Assurer la pérennité** en facilitant l'ajout de nouveaux émulateurs ou de nouvelles versions de systèmes sans modifications de l'architecture existante.

3.3.1 Difficultés et solutions apportées

Défis techniques

Le travail sur des systèmes historiques et des formats propriétaires a soulevé plusieurs obstacles techniques :

Problèmes liés à la mise en forme des fichiers

Lors de l'analyse des fichiers provenant de logiciels d'édition de texte comme **Claris-Works** ou **Microsoft Word**, nous avons constaté que certains paramètres de mise en page (notamment la taille physique d'une page) dépendaient directement de la configuration de l'imprimante utilisée sur la machine d'origine. Ces informations n'étant stockées ni dans le fichier ni dans les métadonnées associées, il est **impossible de garantir une restitution parfaitement fidèle** à la mise en forme voulue par l'auteur.

Solution : Nous avons documenté cette limitation et identifié les éléments critiques qui ne peuvent être conservés. La bibliothèque **libmwaw de LibreOffice** proposant une mise en forme satisfaisante sans altérer le contenu, cette solution a été préférée par l'IMEC.

Impossibilité de démarrer depuis les disques originaux

Malgré nos efforts, il n'a pas été possible de démarrer directement sur les disques issus des archives dans les environnements émulés.

Solution : La solution la plus simple consiste à monter ces disques en tant que volumes secondaires, afin de permettre l'accès à leur contenu.

Incompatibilités avec certains systèmes de fichiers HFS+

Certaines archives identifiées comme HFS+ ne pouvaient pas être montées directement par les outils d'émulation. Toutefois, il était possible d'en extraire le contenu en ligne de commande grâce à **hfsutils**.

Solution : Nous avons documenté la procédure d'extraction sur de telles volumes.

Documentation incomplète des outils d'émulation

La connaissance de la configuration des outils comme **QEMU** et **SheepShaver** était documentée de manière parcellaire, notamment en ce qui concerne les options avancées (gestion du BIOS, configuration des périphériques virtuels, mapping des disques). Cela a nécessité de nombreux tests empiriques pour déterminer les commandes adaptées.

Solution : Après une phase d'expérimentation, nous avons mis au point une commande générique, intégrée dans le logiciel développé, permettant de lancer les émulateurs avec les paramètres adéquats de manière reproductible. Ces difficultés techniques ont été surmontées par un travail collaboratif, combinant *recherche documentaire* (notamment sur les forums spécialisés et archives de documentation Apple) et *tests empiriques* dans différents environnements.

Défis organisationnels

Le projet offrait une grande liberté d'exploration, ce qui constituait un avantage pour la recherche, mais également un défi en termes de *gestion du temps* et de *priorisation des tâches*. Les axes d'investigation étaient nombreux (émulation, rétro-ingénierie, conversion de formats, développement d'outils, établissement d'une base de hash, plugin Autopsy), et il n'était pas possible de traiter toutes les pistes dans le temps imparti. **Solutions mises en œuvre :**

- Mise en place de **réunions régulières avec les tuteurs** et l'IMEC afin de recentrer les objectifs et valider les priorités.
- **Collaboration étroite avec les autres stagiaires** pour mutualiser les compétences et éviter les redondances.
- Documentation continue des travaux non réalisés, afin de faciliter leur reprise par de futurs projets.

Cette organisation souple, mais suivie, a permis d'aboutir à des résultats solides malgré la complexité et la diversité des tâches initialement envisagées.

Chapitre 4

Résultats et compétences acquises

4.1 Intégration et validation des résultats

Le développement des outils n'était pas vu comme une finalité : une étape essentielle a consisté à valider leur bon fonctionnement, à vérifier leur compatibilité avec les infrastructures de l'IMEC, et à préparer leur pérennisation par la documentation.

4.1.1 Tests et validation

La validation des outils a reposé sur une combinaison de méthodes techniques et organisationnelles :

- **Collaboration entre stagiaires** : l'ensemble des outils développés a été testé collectivement, chaque stagiaire utilisant les solutions des autres. Cette approche croisée a permis de détecter rapidement les anomalies et de bénéficier d'un retour d'expérience ;
- **Suivis régulier des encadrants** : des points de suivi hebdomadaires avec les tuteurs ont permis de contrôler l'avancement, d'identifier les priorités et d'ajuster les développements ;
- **Échanges avec l'IMEC** : des réunions régulières ont été organisées avec l'équipe archivistique afin de présenter les avancées, de recueillir leurs besoins et d'intégrer leurs retours dans le développement ;
- **Vérification multi-systèmes** : les outils et émulateurs ont été installés et testés sur plusieurs systèmes d'exploitation, via des machines virtuelles, afin de garantir leur portabilité.

Une réunion spécifique a également été organisée avec l'IMEC afin d'installer et tester les outils directement sur leurs propres machines. Cette étape a permis d'identifier les dernières lacunes, notamment au niveau de la documentation, et de consolider les procédures.

4.1.2 Documentation et pérennisation

L'intégration à long terme des outils nécessitait un travail de documentation et de transmission :

- **Guides techniques** : des documents détaillant l'installation et la configuration des émulateurs, ainsi que l'utilisation des outils développés, ont été rédigés. Ces guides incluent également les scripts d'installation permettant d'automatiser la mise en place sur différents systèmes ;
- **Documentation continue** : la production de documentation a accompagné le développement tout au long du stage, évitant les pertes d'information ;

- **Formation et appropriation** : une partie des réunions avec l'IMEC a été consacrée à la présentation pratique des outils, à travers des démonstrations, afin que les archivistes puissent se les approprier.

Cette démarche d'intégration et de validation a permis d'assurer que les outils développés sont fonctionnels dans un environnement de recherche, mais aussi véritablement utilisables et maintenables à long terme par l'IMEC. Elle a également instauré un esprit collaboratif entre les stagiaires, encadrants et archivistes, garantissant la pertinence des résultats par rapport aux besoins concrets de la conservation numérique.

4.2 Compétences techniques acquises

Au cours de ce stage, j'ai pu développer et consolider un de nombreuses compétences techniques couvrant différents domaines, allant de la rétro-ingénierie à l'automatisation, en passant par la maîtrise des environnements d'émulation et des systèmes historiques.

4.2.1 Technologies et outils

Maîtrise des émulateurs et systèmes historiques

J'ai acquis une expertise dans l'installation, la configuration et l'utilisation d'émulateurs tels que **QEMU** et **SheepShaver**, permettant de reproduire des environnements Mac OS Classic (versions 7.5.2 à 9.x) et Mac OS X (jusqu'à la version 10.5). Ces compétences incluent la compréhension des options avancées (gestion des périphériques virtuels, mapping des disques, configuration des BIOS) et leur adaptation pour assurer une compatibilité maximale avec les archives de l'IMEC.

Gestion des systèmes de fichiers HFS/HFS+

J'ai travaillé sur des systèmes de fichiers historiques comme **HFS** (Hierarchical File System) et **HFS+**, en apprenant à *monter, explorer et extraire* des données à partir d'images disque. Une attention particulière a été portée à la gestion des **forks** (*Data Fork* et *Resource Fork*), spécifiques à Mac OS, afin de garantir une analyse complète des fichiers et de leurs métadonnées.

Développement logiciel avec Rust et Ratatui

Le projet m'a permis de développer un outil en **Rust**, un langage moderne reconnu pour ses performances et sa sécurité mémoire. L'interface textuelle (TUI) a été réalisée avec la bibliothèque **Ratatui**, offrant une ergonomie adaptée aux besoins des archivistes et une modularité permettant d'ajouter facilement de nouvelles fonctionnalités (gestion d'émulateurs, manipulation de disques, automatisation des processus).

4.2.2 Rétro-ingénierie

Analyse binaire et manipulation des ressources

J'ai approfondi mes compétences en rétro-ingénierie grâce à l'utilisation d'outils comme **hexdump** pour l'exploration hexadécimale des fichiers binaires et **ResEdit** (au

sein des environnements Mac OS émulés) pour examiner et modifier les ressources internes des applications (icônes, menus, polices, etc.).

Compréhension des formats propriétaires et logique Type/Creator

L'analyse des fichiers historiques m'a permis de comprendre le fonctionnement des formats propriétaires des applications Mac (notamment ClarisWorks et Microsoft Word), ainsi que la logique des champs **Type** et **Creator** utilisée par Mac OS pour identifier la nature et l'application associée à un fichier. Ces connaissances sont essentielles pour garantir une compatibilité optimale lors de la conversion et la préservation des documents.

4.2.3 Intégration et automatisation

Déploiement multi-OS et scripts d'installation

Pour assurer la pérennité des outils, j'ai travaillé sur un **déploiement compatible avec plusieurs systèmes** : Linux (en production) et Windows via WSL. Des **scripts d'installation automatisés** ont été rédigés et testés sur différentes configurations afin de simplifier la mise en place des émulateurs et du logiciel développé.

Rédaction de guides techniques

Enfin, la documentation technique produite (guides utilisateurs, procédures d'installation, description des environnements émulés) a été pensée pour être claire, exhaustive et directement exploitable par les archivistes et futurs développeurs de l'IMEC.

4.3 Compétences transversales

En parallèle des compétences techniques, ce stage m'a permis d'acquérir des aptitudes transversales essentielles dans un contexte professionnel, notamment en gestion de projet collaboratif, communication et adaptation à un environnement complexe.

4.3.1 Gestion de projet collaboratif

Travail en équipe et intégration des outils

L'un des points clés de ce stage a été la collaboration étroite avec les autres stagiaires, chacun travaillant sur des aspects complémentaires (extraction de données, génération de bases de hachage, outils d'émulation). Nous avons mis en place un système de **retours croisés réguliers**, permettant d'intégrer rapidement les nouvelles fonctionnalités et de corriger les incohérences.

Coordination avec les tuteurs et échanges avec l'IMEC

Des réunions hebdomadaires ont été organisées avec les tuteurs afin de présenter l'avancement des travaux et de valider les choix techniques. Par ailleurs, nous avons eu des échanges réguliers avec les archivistes de l'IMEC pour ajuster les outils aux besoins réels de l'institution, garantissant l'adéquation de la solution développée avec son usage final.

4.3.2 Documentation et vulgarisation technique

Rédaction de guides accessibles

J'ai participé à la rédaction de **guides techniques et procédures d'installation** destinés à des utilisateurs non informaticiens, notamment les archivistes de l'IMEC. L'objectif était de produire une documentation claire, illustrée et utilisable sans connaissances techniques approfondies.

Communication des enjeux techniques

J'ai appris à **vulgariser des concepts complexes** (émulation, systèmes de fichiers HFS/HFS+, rétro-ingénierie des forks) afin d'expliquer les contraintes techniques et les choix effectués.

4.3.3 Capacité d'adaptation

Résolution de problèmes dans un environnement obsolète

Le travail sur des systèmes et logiciels obsolètes (Mac OS Classic, formats propriétaires) a nécessité une forte capacité d'adaptation. Face à des problèmes imprévus (fichiers corrompus, incompatibilités entre versions, documentation parcellaire), il a fallu mener des expérimentations et ajuster nos solutions en continu.

Recherche et exploitation de ressources historiques

Pour surmonter ces difficultés, j'ai développé une compétence en **recherche d'informations spécialisées**, en m'appuyant sur des ressources historiques comme la bibliothèque en ligne **Vintage Apple** ou des forums d'archivage numérique. Cette démarche a permis de trouver des solutions à des problèmes non documentés dans la littérature récente.

4.4 Perspectives

Les travaux menés durant ce stage ont permis de poser les bases facilitant exploitation des archives numériques de l'IMEC, en particulier celles issues de Mac OS Classic et des premières versions de Mac OS X. Néanmoins, plusieurs pistes d'évolution peuvent être envisagées afin de consolider et d'élargir ces résultats.

1. Amélioration et extension des outils développés

Le logiciel conçu constitue un premier prototype fonctionnel. Il pourrait être enrichi par l'ajout :

- d'une interface graphique plus avancée, facilitant encore l'usage par des non-informaticiens ;
- d'un module d'automatisation permettant l'extraction et la conversion des données en formats actuels ;
- une documentation plus détaillée sur le code (diagramme de composant) .

2. Ouverture à d'autres systèmes d'exploitation et formats

Si le stage s'est concentré sur les environnements Macintosh, d'autres fonds archivés à l'IMEC proviennent également de systèmes Windows ou Linux anciens. L'extension des solutions d'émulation et d'analyse à ces environnements permettrait de couvrir d'avantage des besoins archivistiques.

3. Standardisation et mutualisation

La documentation et les procédures établies pourraient être diffusées auprès d'autres institutions confrontées aux mêmes problématiques. La mutualisation des environnements émulés (images disques préconfigurées, guides d'installation) pourrait favoriser la création d'une base de connaissances commune.

4. Contribution aux projets existants

Le travail de rétro-ingénierie sur les formats des anciens logiciels d'édition a mis en évidence que certains paramètres restent encore mal interprétés par LibreOffice. Un approfondissement de cette analyse permettrait d'apporter une contribution utile à la bibliothèque `libmwaw`.

Chapitre 5

Conclusion

Le stage réalisé dans le cadre du projet **IANEC** s'inscrit dans une problématique actuelle : celle de la préservation et de l'exploitation des archives numériques issues d'environnements informatiques obsolètes. La croissance exponentielle des données et l'obsolescence rapide des technologies, impose de concevoir des solutions innovantes garantissant durablement l'accès aux documents à forte valeur historique, culturelle ou scientifique. Ce travail a donc répondu à un double enjeu : préserver l'intégrité des données tout en rendant leur exploitation possible dans un contexte technique profondément différent de celui d'origine.

Les travaux menés au cours de ce stage se sont articulés autour de deux axes majeurs :

- **L'analyse et l'émulation des environnements Mac OS Classic et Mac OS X**, très présents dans les fonds de l'**IMEC**, afin d'en comprendre les particularités et de développer des solutions documentaires et techniques visant à faciliter leur utilisation.
- **Le développement d'un logiciel unifiant les technologies d'émulation**, visant à simplifier leur usage et à automatiser les opérations dangereuses, telles que la gestion des disques et la configurations des environnements.

La mise en place de solutions d'émulation (via **SheepShaver** et **QEMU**) a constitué une réponse solide face à l'impossibilité d'utiliser les matériels d'origine. Ces outils ont permis de reproduire fidèlement des environnements **Mac OS Classic (7.5.2 à 9.x)** et **Mac OS X (jusqu'à la version 10.5)**, offrant ainsi aux archivistes un moyen d'explorer les archives sans altérer les supports originaux. L'étude des systèmes de fichiers **HFS** et **HFS+** a permis de comprendre les spécificités de tels systèmes, comme la séparation entre **Data Fork** et **Resource Fork**, ou encore la logique **Type/Creator**, éléments essentiels pour restituer la structure et la logique des documents dans leur contexte d'origine. Ces connaissances ont non seulement servi à mieux comprendre les archives, mais également à concevoir des solutions permettant leur exploitation dans un environnement moderne.

Le développement logiciel réalisé durant ce stage constitue une avancée significative pour l'**IMEC**. Grâce à une architecture modulaire et à l'utilisation d'un langage moderne comme **Rust**, il a été possible de concevoir un outil performant et sécurisé, intégrant plusieurs émulateurs et offrant une interface intuitive adaptée aux besoins des archivistes. Ce choix technologique, associé à des principes d'extensibilité, permettra à terme d'intégrer de nouveaux émulateurs ou systèmes, garantissant la pérennité et l'évolution de cette solution. Les fonctionnalités développées – notamment la sélection de profils, la gestion automatisée des copies de disques et la compatibilité avec les infrastructures existantes – répondent à une exigence essentielle : sécuriser les données tout en facilitant leur manipulation par des utilisateurs non spécialisés.

Plus que des résultats techniques, ce stage a été une expérience enrichissante sur le plan professionnel et humain. Il m'a permis d'acquérir et de consolider des compétences clés dans plusieurs domaines : rétro-ingénierie (analyse de formats propriétaires, exploration des structures binaires), ingénierie logicielle (développement en **Rust**, conception d'interfaces textuelles ergonomiques), et virtualisation/émulation (maîtrise des outils **SheepShaver** et **QEMU**). De plus, la conduite du projet dans un cadre collaboratif m'a permis de développer des compétences transversales essentielles, telles que la gestion de projet en équipe, la communication avec des acteurs aux profils variés (informaticiens, archivistes), la rédaction de documentation technique claire et la vulgarisation de concepts complexes.

Les difficultés rencontrées, qu'elles soient techniques (problèmes d'incompatibilité, documentation difficile d'accès) ou organisationnelles (priorisation des tâches dans un contexte d'exploration), ont été autant d'opportunités d'apprentissage. Elles ont renforcé ma capacité d'adaptation et ma rigueur méthodologique, deux qualités indispensables dans le domaine de la cybersécurité et de la forensique numérique.

Enfin, ce travail ouvre la voie à de nombreuses perspectives. À court terme, l'amélioration du logiciel développé, par l'ajout de modules d'automatisation pour la conversion des formats. À moyen terme, l'extension des solutions d'émulation à d'autres environnements (Windows, Linux historiques) élargirait le champ des archives exploitables. À long terme, la mutualisation des connaissances et des outils produits avec d'autres institutions confrontées aux mêmes défis offrirait un cadre propice à la normalisation et à la standardisation des pratiques dans le domaine de la préservation numérique.

Ce stage a représenté plus qu'une expérience académique : il a constitué une immersion dans un domaine avec un enjeu sociétal, au croisement de la technologie, de la culture et de la science. Il m'a permis de mesurer l'importance d'une approche pluridisciplinaire pour traiter des problématiques complexes et en constante évolution. Les acquis techniques et méthodologiques obtenus durant cette période constituent des atouts majeurs pour mon projet professionnel et renforcent mon ambition de poursuivre dans le domaine de la recherche en cybersécurité. Ce travail, en posant les bases d'outils pérennes et évolutifs, contribue à la mission de préservation et de valorisation du patrimoine numérique menée par l'**IMEC**, et s'inscrit dans un objectif de transmission des savoirs aux générations futures.

Bibliographie

- [Rie22] Thorsten RIES. “Pour la reconstruction des processus d’écriture numériques de Derrida grâce à la computer forensics : reconstruction des données et matérialité numérique historique”. working paper or preprint. Juin 2022. URL : <https://hal.science/hal-03692433>.
- [OMA24] Hamid BEN OMAR. *Installation de Systèmes Mac OS*. 2024. URL : https://forensics.greyc.fr/interns/Rapport_de_Stage-Hamid-BEN-HOMAR.pdf.
- [Bar25] Kevin BARTLEY. *Big data statistics : How much data is there in the world?* 28 mai 2025. URL : <https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/>.
- [CNI20] CNIL. *Les durées de conservation des données*. 28 juillet 2020. URL : <https://www.cnil.fr/fr/passer-laction/les-durees-de-conservation-des-donnees>.