

L'archivage du web à l'Institut National de l'Audiovisuel : entre déploiement technologique, enjeux et valorisation

Mémoire de Master II
Valorisation numérique du patrimoine : Archives et Images
Sous la direction de Stefan Du Château

Remerciements

Je remercie tout d'abord mon directeur de mémoire Monsieur Stefan Du Château d'avoir guidé mes réflexions, de m'avoir encadré, aidé et conseillé.

Je souhaite aussi remercier Mme Sophie Gillery de l'INA Marseille qui m'a accueilli, aidé et orienté dans mes recherches.

Merci à mes parents et à ma famille pour leur soutien et encouragements pendant l'année.

Sommaire

Partie I : Rapport de stage..... p. 7

Les archives de l'INA : le fonds de « La Chaîne Marseille »

Introduction..... p. 8

Chapitre 1 : Présentation de la structure d'accueil..... p. 10

A- L'INA et ses missions..... p. 10

B- La délégation Méditerranéenne de l'INA..... p. 20

Chapitre 2 : Cadre et missions du stage.....p. 25

A- L'équipe de la délégation méditerranéenne..... p. 25

B- Le fonds de la chaîne LCM..... p. 29

C- Le traitement du fonds..... p. 41

Chapitre 3 : Apports et difficultés du stage.....p. 63

A- Autonomie et polyvalence.....p. 63

B- Difficultés matérielles et taille du fonds..... p. 65

Conclusion..... p. 67

Bibliographie..... p. 69

Sitographie..... p. 72

Tables des illustrations..... p. 74

Annexes..... p. 76

Partie II : Mémoire théorique.....p. 83

L'archivage du web à l'Institut National de l'Audiovisuel :

entre déploiement technologique, enjeux et regards historiques

Introduction..... p. 84

Chapitre 1 : Une évolution de l'archivage web des années 1990 à

aujourd'hui..... p. 90

Partie 1 : L'évolution numérique de l'INA..... p. 91

A- Structure actuelle de l'INA.....p. 91

B- Le virage numérique de l'INA.....	p. 96
C- L'INA et le web français.....	p. 99
Partie 2 : De l'archivage web au dépôt légal du web, une lente évolution ?	p. 104
A- Le dépôt légal comme une photographie du foisonnement culturel français.....	p. 104
B- Genèse de l'archivage web : des approches diverses pour un but commun.....	p. 106
C- Contexte juridique et acteurs.....	p. 109
Partie 3 : L'archive du web comme miroir de l'internet français.....	p. 112
A- L'archiva web : patrimoine fragile.....	p. 112
B- Limites de l'archivage web.....	p. 116
C- Enjeux de gouvernance.....	p. 118
Chapitre 2 : Mise en pratique de l'archivage du web.....	p. 121
Partie 1 : Principe et fonctionnement de l'archivage web.....	p. 122
A- Définition du périmètre de l'archivage.....	p. 122
B- Collecter et indexer.....	p. 124
C- Stocker et pérenniser.....	p. 128
Partie 2 : Techniques, outils et mise en place de collectes spécifiques.....	p. 131
A- Utilisation des métadonnées et du format DAFF.....	p. 131
B- La collecte des UGC.....	p. 134
C- L'archivage de Twitter.....	p. 136
Partie 3 : Les aspects documentaires de l'archivage web : de la veille à la création d'une notice.....	p. 140
A- Mise en place d'une veille.....	p. 140
B- Permettre la consultation d'une archive web.....	p. 142
C- Mise en place d'une notice.....	p. 145
Chapitre 3 : Diffusion et enjeux des archives du web.....	p. 147
Partie 1 : Mise en consultation des archives du web.....	p. 148
A- Pour qui et où consulter.....	p. 148
B- Critères de recherche.....	p. 150

C- Multiples outils pour la recherche détaillée sur Twitter.....	p. 154
Partie 2 : Valorisation des archives du web.....	p. 157
A- Des valorisations en ruptures avec les archives papiers et audiovisuelles.....	p. 157
B- Exemples de valorisation des archives du web.....	p. 159
Partie 3 : Enjeux historiques.....	p. 162
A- Construction d'un regard d'historien sur les archives du web.....	p. 162
B- Le nouveau rapport à l'archive (reborn digital heritage).....	p. 164
<i>Conclusion.....</i>	p. 166
Partie III : Projet professionnel.....	p. 169
Bibliographie.....	p. 171
Sitographie.....	p. 175
Tables des illustrations.....	p. 178
Annexes.....	p. 179

Partie I :

Rapport de stage

Les archives de l'INA : le fonds de « *La Chaîne Marseille* »



Établissement d'accueil : INA Méditerranée à Marseille
Image provenant du site web de l'INA Méditerranée (annexe 1)

Introduction

Dans le cadre du Master 2 Valorisation Numérique du Patrimoine : Archives et Images, j'ai eu l'opportunité d'effectuer un stage à l'antenne méditerranéenne de l'Institut National de l'Audiovisuel (INA). Ce stage s'est déroulé entre le 28 mars et le 1 juillet 2022 dans les locaux de l'INA Méditerranée. Lors de ce stage, ma tutrice a été Madame Sophie GILLERY, documentaliste au sein de la structure. L'INA Méditerranée est une des 6 structures régionales de l'INA et a pour vocation de créer des partenariats régionaux, établir une relation avec des institutions ou collectivités locales ou encore promouvoir les recherches en lien avec l'audiovisuel à l'échelle de la région. Diplômé d'un Master 2 en Histoire monde moderne et contemporain à l'Université d'Aix-Marseille, j'ai déjà eu l'occasion durant les deux années de recherches en Master 1 et Master 2 d'effectuer des recherches à la délégation marseillaise de l'INA. Ces visites m'ont permis de définir un corpus de recherche et par la suite de visionner les sources pertinentes pour mon sujet d'étude. Ainsi je connaissais, avant mon stage, les locaux de l'INA et la recherche historique grâce aux ressources de l'INA. C'est donc naturellement que je me suis tourné vers la délégation méditerranéenne de l'INA pour solliciter et obtenir un stage afin de mettre en pratique mes compétences acquises lors de l'année du master 2 VALARI.

Ainsi dans le cadre de ce stage, pour la première fois, j'ai pu observer, analyser et comprendre comment les archives sont mises à la disposition des chercheurs. J'ai pu avoir accès à l'interface professionnelle des logiciels que l'INA utilise et enrichir mes connaissances sur les archives audiovisuelles.

Ce stage de 14 semaines m'a permis de me confronter au traitement documentaire des archives de l'INA. Plusieurs compétences et connaissances ont été mises à profit durant le stage, notamment la connaissance des techniques de mise en ligne d'un fonds, la connaissance dans le traitement documentaire des archives audiovisuelles, les connaissances en conservation des archives audiovisuelles ainsi que des compétences dans l'enrichissement de base de données. J'ai pu mettre à profit des savoirs et compétences appris lors de l'année de Master 2 VALARI pour mener à bien les tâches qui m'ont été confiées. Ma mission principale lors du stage a été le catalogage et l'indexation d'un fonds audiovisuel régional. Il

s'agissait du fonds de la chaîne de télévision « La Chaîne Marseille » (LCM) qui a émis de 2005 à 2016. La mission consistait à visionner les émissions de cette chaîne de télévision et à produire des notices contenant diverses informations sur les émissions pour enrichir la base de données de l'INA. Par la suite, j'ai été amené à construire la grille de programme de cette chaîne et à enrichir la base de données de l'INA avec les différents journalistes et acteurs de la chaîne. J'ai notamment traité les archives de l'année 2010 de cette chaîne de télévision. J'ai eu la possibilité d'utiliser plusieurs logiciels internes à l'INA notamment « Totem »¹ lors de la création de notice ou encore « Sumoco »² lors du support de l'archives. J'ai par ailleurs eu l'opportunité de participer à des actions de valorisation notamment lors de deux conférences s'inscrivant dans un cycle de conférences intitulé le « Le temps des archives » en partenariat avec le MUCEM³ et par la suite de participer aux réflexions sur la communication de ce cycle de conférences sur les réseaux sociaux. Ainsi lors de ce stage, j'ai participé à la réalisation de plusieurs actions qui ont enrichi mes compétences et mon approche des sources audiovisuelles.

¹ Logiciel propriétaire INA

² Logiciel propriétaire INA

³ Musée des civilisations de l'Europe et de la Méditerranée

Chapitre 1 : Présentation de la structure d'accueil

A- L'INA et ses missions

L'INA est un établissement public à caractère industriel et commercial (EPIC), principalement chargé d'archiver et de conserver les productions audiovisuelles françaises. L'INA exploite également ses collections d'archives en produisant des contenus originaux et met à disposition de divers publics ses archives. L'établissement a la charge du dépôt légal de la télévision, de la radiodiffusion et du web. De plus, la structure contribue à la recherche dans le domaine de la production et de la communication audiovisuelle. Enfin, par le biais de ses formations, l'INA contribue à former différents publics en communication, production, réalisation et documentation audiovisuelle. C'est donc une structure qui a la charge d'un grand nombre de missions sur lesquelles nous reviendrons plus tard. Tout d'abord, l'INA est une structure gérée dans le cadre juridique d'un EPIC, au même titre que les grandes entreprises publiques. Un EPIC est en premier lieu un établissement public avec le statut d'une personne morale de droit public, c'est-à-dire une personne morale soumise au droit public, qui détient la gestion d'une activité de service public de nature industrielle et commerciale. Un EPIC est forcément créé par une personne publique en l'occurrence l'État, qui a créé l'établissement public l'INA en 1974.

L'INA est donc né de la loi du 4 août 1974⁴, à la suite de l'écclatement et du démantèlement de l'ORTF. Sept sociétés sont créées⁵ :

- **Radio France** qui englobe les 4 chaînes de radios nationales de l'ORTF
- **Télévision Française 1 (TF1)** qui succède à la Première Chaîne
- **Antenne 2 (A2)** qui succède à la Seconde Chaîne
- **France Régions 3 (FR3)** qui succède à la Troisième Chaîne et englobe les bureaux régionaux de la radio et de la télévision de l'ORTF

⁴ Loi du 4 août 1974 : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000333539> > (consulté le 05/03/22).

⁵ HOOG Emmanuel, *L'INA*, Paris, PUF, 2006, p. 7.

- **Télédiffusion de France (TDF)** qui a la charge de la gestion des émetteurs de télédiffusion
- **Société française de production (SFD)** qui a la charge des activités de production.
- **Institut national de l'Audiovisuel (INA)** qui a la charge de la gestion des archives de la radio et de la télévision.

L'INA est la seule société à n'avoir pas changé de statut, et est resté un établissement public depuis 1974. L'institut est sous la tutelle du ministère de la Culture et de la Communication. Si l'INA existe depuis 1974, la structure conserve des fonds bien antérieurs, remontant jusqu'à 1905. Cependant la plupart des premiers fonds datent des années 1940. Aujourd'hui, l'INA conserve près de 15 millions d'heures de diffusion qui augmentent d'environ 1 million d'heures par an, par la captation en continu de 168 chaînes de télévision. En parallèle sont également conservées les archives de la radio et depuis 2006 les archives du Web.

Lors de la création de la structure, la télévision française ne possédait que des chaînes publiques. Et durant les 17 années qui suivirent, malgré l'avènement des chaînes de télévisions privées, la mission de collecte de l'INA concernait uniquement les archives des chaînes publiques.

Après la mise en place de la télévision dans les foyers français, et à la suite de la disparition de l'ORTF, l'INA hérite des archives de l'ORTF et donc en même temps, de celles de la Radiodiffusion française entre 1945 et 1949 (RDF) et de la Radiodiffusion-télévision française entre 1949 et 1964 (RTF). L'institut s'est alors trouvé dépositaire des archives de la Première Chaîne puis, au moment de leur création, de celles de la Seconde Chaîne en 1964 et de la Troisième Chaîne en 1972. Cependant, face à la privatisation de TF1 en 1986 et à l'émergence d'une multitude d'autres chaînes de télévision, privées et publiques, la collecte des archives des trois premières chaînes ne permettait plus à l'INA de posséder des archives reflétant le patrimoine de la télévision française dans son intégralité. C'est pour pallier cette lacune dans les fonds de l'INA qu'a été élargie en 1992 la loi sur le dépôt légal, qui existait déjà depuis le XVIème siècle pour l'écrit et l'imprimé.

Le dépôt légal est défini ainsi dans le code du Patrimoine à l'article L131-2⁶ :

« Les documents imprimés, graphiques, photographiques, sonores, audiovisuels, multimédias, quel que soit leur procédé technique de production, d'édition ou de diffusion, font l'objet d'un dépôt obligatoire, dénommé dépôt légal, dès lors qu'ils sont mis à la disposition d'un public. »

La loi du 21 juin 1943⁷ stipule que le dépôt légal inclut désormais le cinéma : une copie de chaque film français doit être déposée à la BNF. C'est le 20 juin 1992 que la loi sur le dépôt légal est élargie pour prendre en compte toutes les œuvres audiovisuelles : cinéma, radio, télévision, documents sonores, informatique (les logiciels et les bases de données). Trois organismes sont définis : la BNF, le CNC pour le cinéma et la photo et l'INA pour la télé et la radio. Un conseil scientifique du dépôt légal est créé dans le but d'unifier les méthodes de collecte et de conservation. Le dépôt légal est élargi en 2002 aux chaînes du câble et du satellite, puis en 2005 à la télévision numérique terrestre (TNT). Le dépôt légal est à nouveau modifié le 1er août 2006 par la loi DADVSI⁸ qui étend le dépôt légal au Web.

Pour lutter contre la disparition des archives audiovisuelles notamment à cause des dégradations chimiques ou de l'obsolescence des supports, l'INA instaure en 1999 un plan de sauvegarde et de numérisation (PSN). Dans un premier temps, la structure a agi sur les archives les plus en danger : il s'agit des cassettes des années 1970-80 qui ne supportent que peu de lecture. Ces fonds en danger n'étaient pas les plus intéressants d'un point de vue historique. Ainsi dans un second temps, l'INA a réalisé une cartographie des fonds avec un fort intérêt historique et en mauvais état. En 2019, c'est 335 000 heures de télévision et près de 500 000 heures de radio⁹ qui ont été numérisées grâce au PSN. Grâce à l'expérience du PSN, l'INA a développé un savoir-faire autour de la sauvegarde et de la mise en valeur de fonds

⁶ Code du Patrimoine : < https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000020905828/ > (consulté le 05/03/22).

⁷ Loi du 21 juin 1943 : < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006071152/> > (consulté le 05/03/22).

⁸ Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).

⁹ CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).

audiovisuels. Ainsi cela permet à l'INA de proposer à d'autres détenteurs d'images de bénéficier de cette expertise, en prenant notamment en charge leurs collections à travers des mandats commerciaux. L'INA sauvegarde et exploite alors sous forme de mandats les archives de l'AFP, de TF1 pour ses programmes d'actualité et de sports ou encore de Radio Nova, ainsi que celles de nombreux producteurs audiovisuels privés¹⁰. De plus, l'INA accueille des collections audiovisuelles à valeur culturelle et historique. L'institut assure alors la conservation et la mise en consultation. Parmi ces collections, celle des théâtres nationaux (Théâtres de l'Odéon, de la Colline, de Chaillot, Opéra de Paris...) ou encore celle des Rencontres d'Arles¹¹.

Pour donner un accès à ces archives numérisées à tous, l'INA crée en 2006 le site « INA.fr »¹². Mais cet accès aux œuvres en accès libre est limité puisque seulement 50 000 heures de programmes sous forme d'extraits sont disponibles et le reversement des droits de consultation de ces œuvres est fait par la publicité sur le site. Sur INAMediaPro¹³ c'est 2 millions d'heures supplémentaires qui sont accessibles mais uniquement réservées aux professionnels de l'audiovisuel. Ainsi l'unique façon de consulter les quelques 22 millions d'heures d'archives¹⁴ est de se rendre à la Bibliothèque nationale de France (BNF) ou à l'INA thèque qui est un service de l'INA au sein de la BNF. Il existe également des postes de consultation multimédia (PCM) en dehors des centres de l'INA. Ces postes permettent de consulter les notices des archives conservées par la structure et d'avoir accès à quelques archives. Ils sont souvent installés dans les bibliothèques municipales ou dans les bibliothèques universitaires. Ainsi le public de l'INA est assez varié à la BNF, dans les centres régionaux ou sur les postes PCM. Toutefois le public principal de l'organisme est composé d'universitaires, de chercheurs et d'étudiants. Il y a également un public de professionnels de l'audiovisuel, des producteurs et des réalisateurs. Enfin un des publics de l'INA est constitué par les institutions culturelles.

¹⁰ CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).

¹¹ CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).

¹² Site web de l'INA : < <https://www.ina.fr> > (consulté le 05/03/22).

¹³ Site web de INAMédiaPro : < <http://www.inamediapro.com> > (consulté le 05/03/22).

¹⁴ Site web de l'INA : < <https://www.ina.fr> > (consulté le 05/03/22).

La structure de l'INA n'a donc connu aucune évolution de son cadre juridique. Sa structuration a peu évolué durant plusieurs décennies. De même, les missions de l'INA ont peu évolué et sa mission principale reste l'archivage. Comme dans un service d'archives papier, l'INA assure, en plus de la collecte responsable de la sauvegarde des archives, leur valorisation auprès des publics. L'INA définit elle-même ses missions en ces termes :

« Les missions de service public de l'INA, définies par l'article 49 de la loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication¹⁵ modifiée, sont aujourd'hui les suivantes :

- L'INA assure la conservation des archives audiovisuelles des entreprises du secteur audiovisuel public ainsi que de toute entreprise, désireuse de lui confier la conservation et l'exploitation de ses archives audiovisuelles.
- Il exploite ses collections d'archives en éditant et en produisant des contenus originaux pour tous les supports ainsi qu'en mettant ces collections à disposition des chercheurs, des créateurs, des producteurs, des autres médias, des institutions culturelles, etc.
- Il assure le dépôt légal de la télé et de la radiodiffusion et, avec la Bibliothèque nationale de France, du web.
- Il contribue à l'innovation et à la recherche dans le domaine de la production et de la communication audiovisuelle.
- Il contribue à la formation continue et initiale et à toutes les formes d'enseignement dans les métiers de la communication audiovisuelle. »¹⁶

Ainsi ces missions sont fixées dans un cahier des charges fixé par décret et validé en amont par le Conseil Supérieur de l'Audiovisuel (CSA), nouvellement l'Autorité de régulation de la communication audiovisuelle et numérique (ARCOM).

¹⁵ Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068930/> > (consulté le 05/03/22).

¹⁶ Site web de l'INA : <https://www.INA.fr/institut-national-audiovisuel/organisation/statuts-et-missions> > (consulté le 05/03/22).

L'INA, comme nous l'avons indiqué plus haut, de la même manière qu'un service d'archives, va collecter les archives de l'audiovisuel français et du web, classer les archives, les conserver, puis les communiquer. Le rôle des archivistes de l'INA est de conseiller les différents publics. Ainsi les missions sont identiques à celles des centres d'archives. Du point de vue de la collecte des archives, l'INA enregistre, 24h/24h et 7 jours sur 7, plus de 180 chaînes de radio et de télévision au titre du dépôt légal. L'INA organise également la collecte du web, comme indiqué plus haut et cela depuis 2006. Cette collecte s'organise en relation avec la BNF, l'INA a pour mission de collecter les services de communication audiovisuelle. En décembre 2020, la structure avait archivé environ 16.000 sites web¹⁷ dont environ 12.000 actifs¹⁸ qui sont captés en continu représentant 104 milliards¹⁹ de versions d'URL. L'INA estime que le nombre d'URL augmente de 13 milliards par an²⁰. Cette collecte de web s'effectue également sur les réseaux sociaux ainsi que les « hashtag » les plus populaires. L'institut a donc collecté plus de 15.000²¹ comptes de réseaux sociaux et environ 1.500²² « hashtag » se rapportant aux émissions et événements audiovisuels. Certaines vidéos publiées sur internet se rapportant à la communication audiovisuelle sont également collectées. Ainsi il s'agit de plus de 15.000²³ comptes sur 18 plateformes de publications, cela représente plus de 28 millions²⁴ de vidéos pour un total de plus de 2.860.000 heures²⁵. Les podcasts se retrouvent collectés par la structure et représentent 1,2 millions²⁶ d'émissions et 402.000 heures²⁷ de

¹⁷ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

¹⁸ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

¹⁹ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁰ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²¹ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²² Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²³ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁴ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁵ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁶ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁷ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

programme. En plus des podcasts de Radio France, l'INA collecte depuis peu des podcasts des médias traditionnels et de nouveaux acteurs indépendants qui sont regroupés en 31²⁸ web radios.

En plus de ces missions qui incombent à tous les centres d'archives, l'institut a une politique de valorisation très ambitieuse. En effet, l'INA produit des contenus originaux grâce à ses collections. Ainsi le rapport d'activité de 2020 de la structure établit un total de 1,132 milliard²⁹ de vidéos vues par le grand public sur 5 canaux différents :

- 622 millions de vidéos vues sur Facebook
- 362 millions de vidéos vues sur YouTube
- 23,1 millions de vidéos vues sur INA.fr
- 33 millions de vidéos vues sur Twitter
- 89 millions de vidéos vues sur Instagram

L'INA avec plus d'un milliard de vidéos vues compte 401.000 heures de programmes en ligne avec un accès direct pour le grand public. Ce qui est encore loin des plus de 22 millions³⁰ d'heures d'archives de programme de télévision et de radio consultables par les chercheurs et étudiants à l'INA thèque ou dans les centres régionaux. Cette valorisation des collections, que l'on peut qualifier de valorisation à grande échelle, s'accompagne d'une grande fidélisation du public grâce aux réseaux sociaux. L'INA compte 2,5 millions³¹ de fans sur Facebook, 2,5 millions³² d'abonnés répartis sur les 41 chaînes YouTube de la structure, 220.000³³ followers sur Twitter ainsi que 710.000³⁴ abonnés sur Instagram. Cette politique

²⁸ Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

²⁹ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁰ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³¹ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³² Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³³ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁴ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

ambitieuse de valorisation s'accompagne de la production de films³⁵, de la diffusion de programmes inédits³⁶, de la création de sites web patrimoniaux³⁷. La valorisation se poursuit également sur papier avec 7 ouvrages édités³⁸ uniquement pour l'année 2020. La valorisation des collections de l'INA a connu un point d'orgue par le lancement de l'offre de SVOD « Madelen ». Il s'agit d'un abonnement à 2,99 euros³⁹ par mois qui ouvre l'accès à 13.000⁴⁰ archives en intégralité. Cette plateforme basée sur le modèle des géants du secteur comme Netflix ou Disney + a pour vocation à présenter des programmes en lien avec « les tendances et les événements de la société »⁴¹. Ainsi la plateforme met en avant des archives en relation avec l'actualité pour apporter un éclairage nouveau ou plus simplement rendre hommage à une personne. On peut notamment citer des archives de l'émission Thalassa pour rendre hommage au présentateur Georges Pernoud ou encore la série Arsène Lupin datant de 1980, en parallèle de la diffusion de la nouvelle série Lupin sur Netflix. « Madelen » est accessible sur un ordinateur et également sur smartphone grâce à une application. En décembre 2020, cette offre a compté environ 40.000⁴² abonnés. L'INA propose également des émissions originales chaque semaine. On peut notamment citer l'émission « INAttendu », un magazine de 26 minutes qui décrypte l'actualité à partir d'images d'archives. De plus, en coproduction avec la chaîne LCP, l'émission « Rembob'INA » « revisite l'histoire contemporaine »⁴³ par le prisme d'archives audiovisuelles. Ainsi cette émission, par le biais d'experts et de témoins de l'époque, éclaire les problématiques du présent. En conclusion, cette mission de valorisation est fortement traitée et mise en valeur par l'institut. L'ensemble de cette valorisation s'inscrit

³⁵ 38 films produit en 2020 par l'INA, rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁶ 148 heures de programmes inédits diffusés en 2020, rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁷ 73 sites web, rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁸ Rapport d'activité INA, p.4 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

³⁹ Site web Madelen : < <https://madelen.INA.fr> > (consulté le 05/03/22).

⁴⁰ Rapport d'activité INA, p.6 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴¹ Rapport d'activité INA, p.6 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴² Rapport d'activité INA, p.6 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴³ Rapport d'activité INA, p.6 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

dans la volonté d'éclairer et de contextualiser l'actualité du temps présent par l'utilisation des archives audiovisuelles.

L'une des missions de l'INA réside également dans l'enseignement, en formation continue et initiale, dans le domaine de la communication audiovisuelle. Ainsi l'INA dirige « INA sup », une école supérieure de l'audiovisuel et des médias numériques. L'institut propose 15 cursus différents de la Licence 1 au Master 2. Le site web de la structure définit l'école en ces termes :

« À destination des jeunes talents et des jeunes professionnels, ces cursus préparent et forment à l'ensemble des métiers technico-artistiques, de production ou encore de gestion des contenus sur tous supports. Ces formations initiales ou professionnalisantes sont reconnues par l'État (ministères de l'Éducation nationale, de la Culture, de l'Enseignement supérieur, de la Recherche et de l'Innovation) et par la CNC. »⁴⁴

Cette école vise à donner une formation sur l'ensemble de la chaîne de traitement de l'image et du son. Sur les 15 formations, 6 diplômes⁴⁵ sont délivrés par l'INA et les 9⁴⁶ autres sont des diplômes en partenariat avec d'autres établissements supérieurs comme l'École Normale Supérieure ou encore Paris 1 Panthéon-Sorbonne. Ces cursus forment 400 étudiants par an.

Enfin la dernière mission de l'INA, en lien avec l'enseignement, est la recherche. L'institut a développé de nombreux partenariats scientifiques, l'objectif de la structure étant de faire de l'INA un « laboratoire des médias »⁴⁷. La recherche scientifique de l'institut recoupe plusieurs thématiques comme « la restauration et la numérisation, l'analyse des images et des sons, la reconnaissance des formes, la fouille de données, le traitement

⁴⁴ Site web de INA sup : < <https://www.INA-expert.com/enseignement-superieur> > (consulté le 05/03/22).

⁴⁵ Rapport d'activité INA, p.26 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴⁶ Rapport d'activité INA, p.26 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴⁷ Rapport d'activité INA, p.14 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

sémantique des contenus et l'analyse des médias. »⁴⁸. L'INA lance également des projets scientifiques, nous pouvons notamment citer le projet Gender Equality Monitor (GEM)⁴⁹ qui rassemble des partenaires académiques de l'INA pour mener des recherches sur le genre dans les médias. Ces recherches transdisciplinaires portées par l'INA et l'entreprise Deezer avec le concours de 2 laboratoires de recherche en sciences de l'information et de la communication spécialisés dans l'extraction d'information automatique à partir du texte et de la parole et 3 laboratoires de recherche en sciences humaines et sociales spécialisés dans l'étude du genre et des médias. Le projet est qualifié en ces termes par l'agence nationale de la recherche :

« Le projet GEM vise à décrire de manière automatique les différences de représentation et de traitement existant entre les femmes et les hommes dans les médias en langue française que sont la TV, la radio, la presse écrite et les collections musicales.

L'ambition du projet est d'analyser plusieurs millions de documents échantillonnés sur une période de plus de 80 ans, pour réaliser la plus vaste étude portant sur la place des hommes et des femmes dans les médias. »⁵⁰

En parallèle de ses partenariats dans la recherche et l'innovation scientifique, l'INA récompense chaque année, depuis 1997, la recherche sur les médias audiovisuels en décernant les Prix de l'INA thèque. De plus, depuis 2017 la structure a mis en place deux dispositifs de soutien à la recherche et à la valorisation scientifique de ses collections par un statut de chercheur associé à l'INA ainsi que l'attribution de bourses de recherche. Les lauréats du prix de l'INA thèque ou des bourses de recherche doivent participer aux événements organisés par la structure et partager leurs résultats scientifiques sous différentes formes, à la fois des publications mais aussi des articles dans *La revue des médias*⁵¹. Il s'agit d'une revue en ligne traitant des mutations des médias ainsi que de leurs enjeux. La revue analyse 3 grands points : l'offre éditoriale des médias, l'écosystème des médias et les usages des médias. Ainsi

⁴⁸ Rapport d'activité INA, p.14 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁴⁹ Projet Gender Equality Monitor (GEM) : < <https://anr.fr/Project-ANR-19-CE38-0012> > (consulté le 05/03/22).

⁵⁰ Projet Gender Equality Monitor (GEM) : < <https://anr.fr/Project-ANR-19-CE38-0012> > (consulté le 05/03/22).

⁵¹ Rapport d'activité INA, p.4 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

nous observons que l'INA, par ses actes dans la recherche, souhaite jouer un rôle de médiateur entre ses collections et les chercheurs. L'institut organise des séminaires sous la forme de cartes blanches dans lesquels des intervenants spécialisés proposent une lecture des archives des collections de la structure. Ainsi l'axe de la recherche permet à l'INA de faire vivre ses collections ; nous comptabilisons 8 projets⁵² de recherche menés en France et à l'international sur les collections de l'INA pendant l'année 2020 et 17 publications⁵³ scientifiques sur l'année 2020.

En conclusion, les missions de l'INA sont assez nombreuses et dépassent le cadre des missions d'un centre d'archives. L'INA, en plus des missions qui incombent à un service d'archives, propose un grand nombre de valorisation qui font le succès et la renommée de la structure. La valorisation de ces collections se décline sur différents supports, par le biais d'émissions télévisées, de contenu web, de contenu sur les réseaux sociaux, par des expositions ou encore par des partenariats avec des entreprises ou des services publics comme récemment l'organisation d'une fresque sur le site web de la SNCF pour commémorer les 80 ans de l'entreprise publique. La recherche et la formation sont également des points centraux dans les missions de l'INA qui permettent à la fois de maintenir une haute exigence des recherches sur les collections de l'institut, et également de former des futurs spécialistes de l'audiovisuel. L'INA par son histoire et sa grande diversité de missions définies par l'ARCOM s'impose en France comme la structure principale dans le traitement des archives audiovisuelles. Au-delà des locaux de l'INA en région parisienne, les différents centres de l'INA mettent à profit les missions déclinées plus haut, au service d'une mémoire et d'archives locales.

B- La délégation méditerranéenne de l'INA

C'est à Bry sur Marne que se situe la maison mère de l'institut. L'INA est également composé de 6 centres relais en région répartis dans toute la France métropolitaine. Ainsi il y

⁵² Rapport d'activité INA, p.4 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁵³ Rapport d'activité INA, p.4 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

a un centre INA à Marseille, Rennes, Strasbourg, Lyon, Toulouse et Lille. Le stage s'est déroulé dans le centre régional de Marseille, à l'INA Méditerranée. Les délégations régionales ont le même rôle et les mêmes missions que l'INA. L'ancrage régional implique que les délégations ont des partenariats locaux et s'intéressent plus précisément aux fonds et à la culture locale. Ainsi l'INA définit les délégations régionales en ces termes :

« Les délégations régionales initient de nombreuses actions éducatives et culturelles valorisant la mémoire régionale, en partenariat avec les acteurs économiques et culturels locaux. »⁵⁴

L'INA Méditerranée relaie ainsi les missions de l'INA dans l'ensemble de la région Sud et en Corse. La délégation est installée à Marseille depuis 1982. Au-delà de la délégation de l'INA à Marseille, l'INA dispose de 4 centres de consultation dans le région Sud. Il y a le centre de consultation d'Aix-en-Provence situé à la Maison Méditerranéenne de Sciences de l'Homme⁵⁵ (MMSH) et disposant de 2 postes PCM, le centre de Nice installé dans la bibliothèque Louis Nucéra⁵⁶ disposant également de 2 postes PCM, le centre dans la bibliothèque Ceccano⁵⁷ d'Avignon disposant de 2 postes PCM et enfin le centre de consultation dans la cinémathèque Casa di Lume⁵⁸ à Porto-Vecchio.

L'INA Marseille est au cœur de l'écosystème audiovisuel et numérique de la région Sud. La délégation conserve et numérise des fonds audiovisuels issus des chaînes de télévision et de radio régionales. Ainsi la délégation détient des fonds allant de la RTF jusqu'à France 3 région Sud. La délégation estime que les fonds s'enrichissent chaque année de plusieurs milliers d'heures⁵⁹. L'INA Marseille collecte également des fonds de diffuseurs privés et d'institutions régionales comme notamment Canal Marseille, LCM ou encore M6 Marseille. La délégation, au même titre que l'INA, affiche une politique de valorisation ambitieuse. En effet,

⁵⁴ Rapport d'activité INA, p.28 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

⁵⁵ Site web de la MMSH : < <https://www.mmsh.fr> > (consulté le 05/03/22).

⁵⁶ Site web bibliothèque Louis Nucéra : < <https://bmvr.nice.fr> > (consulté le 05/03/22).

⁵⁷ Site web bibliothèque Ceccano : < <https://bibliotheques.avignon.fr/in/faces/details.xhtml?id=816093b1-a9ef-42d4-8430-3c370fe36fd3> > (consulté le 05/03/22).

⁵⁸ Site web cinémathèque Casa di Lume : < <https://www.casadilume.corsica> > (consulté le 05/03/22).

⁵⁹ Rapport d'activité INA, p.28 : < <https://www.ina.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).

elle a pour ambition de valoriser des fonds régionaux auprès de tous les publics. Ainsi l'INA Marseille développe des offres digitales très variées. On peut notamment citer la création du site web « Sudorama », qui regroupe plus de 500 archives régionales de l'INA contextualisées par des universitaires. Les archives sont classées dans 8 catégories : Arts et Culture, Économie et Société, Espace et Paysage, Périodes, Relations internationales, Sciences et Techniques, Sports et enfin Vie Politique⁶⁰. Ces 8 parcours thématiques s'intéressent à l'histoire de la région depuis 80 ans. L'ensemble des archives est classé également par chronologie ou encore par géographie. De plus, le site a une ambition éducative avec une entrée par niveau scolaire et par discipline. Sudorama est un exemple fort de la valorisation des fonds de l'INA Méditerranée permettant d'ouvrir ses collections aux scolaires et à l'ensemble des populations voulant s'informer sur la région. Un autre exemple intéressant de valorisation réalisée par la délégation, réside dans la création d'un site web⁶¹ traitant de l'histoire de la ville de Cannes. Ce site web regroupe 6 thèmes (Patrimoines, Sports, Célébrités, Tourisme, Environnement, Évènements culturels) comprenant chacun des archives audiovisuelles éclairant les thèmes. Sur le même principe, l'INA a construit un site web⁶² traitant des médias et des mobilisations en Corse. Ce site a été élaboré en partenariat avec l'université de Corse Pascale Paoli afin de retracer les grandes mobilisations sociales, politiques et environnementales de 1945 à nos jours en Corse. L'ensemble des archives a été sélectionné par les chercheurs de l'Université. Ainsi la valorisation sur le web vise à faire découvrir le patrimoine culturel et les différents territoires de la région à différents publics.

La valorisation s'accompagne également de création de films documentaires ou encore d'organisation de rencontres et d'évènements. La délégation organise par exemple à la bibliothèque l'Alcazar à Marseille un projet nommé les « Les escales de l'INA »⁶³. A la suite de la projection d'un film, l'INA propose un débat-discussion animé par différents historiens du cinéma. De la même manière, la délégation propose également un projet intitulé « Instant

⁶⁰ Site web Sudorama : < <https://fresques.ina.fr/sudorama> > (consulté le 05/03/22).

⁶¹ Site web Histoire de la ville de Cannes en vidéo : < <https://sites.ina.fr/ville-de-cannes/focus/chapitre/6> > (consulté le 05/03/22).

⁶² Site web Médias et mobilisations en Corse de 1945 à nos jours : < <https://sites.ina.fr/mobilisations-en-corse/tempo/> > (consulté le 05/03/22).

⁶³ Site web délégation INA Méditerranée : < <https://www.ina.fr/institut-national-audiovisuel/en-regions/INA-mediterranee> > (consulté le 05/03/22).

plaisir INA »⁶⁴. Il s'agit d'une immersion dans les archives de la télévision au sein du café culturel Aglaé et Sidonie. La délégation a également des partenariats avec des musées et notamment avec le Mucem (Musée des civilisations de l'Europe et de la Méditerranée). L'institut organise des rencontres intitulées « Le Temps des Archives »⁶⁵ dans lesquelles un journaliste interroge des chercheurs ou des témoins d'un événement historique, et le tout est enrichi par un corpus d'archives de l'INA. Le partenariat avec le Mucem s'appuie sur l'organisation d'un espace appelé « MédINathèque » dédié aux archives méditerranéennes animé par la délégation. Le partenariat avec les institutions muséales ne s'arrête pas ici. En effet, l'INA fournit des images pour l'écran géant en courbe de 240° pour le musée national du sport de Nice. Ainsi la délégation INA Méditerranée mène de nombreuses actions éducatives et culturelles mettant en scène la mémoire régionale avec des acteurs économiques et culturels locaux notamment des entreprises, des musées, des établissements scolaires ou des associations.

Les missions de l'INA sont également la formation et la recherche ; dans ce domaine, la délégation, en partenariat avec le CMCA⁶⁶ (Centre Méditerranéen de la communication audiovisuelle) propose des sessions de formation continue consacrées à l'écriture de reportage avec des archives de l'INA pour les journalistes des télévisions méditerranéennes. De plus l'INA co-organise avec le CMCA le prix Primed⁶⁷ (Prix du documentaire méditerranéen). La formation initiale n'est pas uniquement l'apanage de INA sup, et la délégation marseillaise propose des ateliers d'éducation aux médias (EMI) pour les lycéens du territoire. Enfin, en partenariat avec l'université de Côte d'Azur, l'INA propose des formations dans le cadre des « Summer Schools de Cannes », il s'agit de cours à destination des étudiants et également des chercheurs et des professionnels dans le domaine de la création audiovisuelle. En conclusion, les missions de la délégation marseillaise sont calquées sur les missions de l'INA : la collecte, le traitement ainsi que la conservation de plusieurs fonds régionaux, couplées à une valorisation ambitieuse de ses archives avec de nombreux

⁶⁴ Site web délégation INA Méditerranée : < <https://www.INA.fr/institut-national-audiovisuel/en-regions/INA-mediterranee> > (consulté le 05/03/22).

⁶⁵ Site web délégation INA Méditerranée : < <https://www.INA.fr/institut-national-audiovisuel/en-regions/INA-mediterranee> > (consulté le 05/03/22).

⁶⁶ Site web CMCA : < <https://cmca-med.org/formation-INA-cmca/> > (consulté le 05/03/22).

⁶⁷ Site web Primed : < <https://primed.tv> > (consulté le 05/03/22).

partenaires régionaux. Enfin la délégation propose des formations professionnelles et des partenariats avec des universités pour la recherche.

La délégation est donc un acteur majeur de la valorisation d'archives régionales, et à ce titre, elle est le moteur de plusieurs événements culturels locaux comme nous l'avons énoncé plus haut. Il convient à présent de présenter plus en détail le fonctionnement de la délégation ainsi que les missions que j'ai pu mener à bien durant l'ensemble du stage.

Chapitre 2 : Cadre et missions de stage

A- L'équipe de la délégation méditerranéenne

Avant d'analyser mes missions de stage, il convient d'expliciter le fonctionnement de la délégation. Les délégations en région, à l'instar de la délégation marseillaise, comptent uniquement 4 employés. Tout d'abord, la structure s'appuie sur une déléguée régionale en la personne de Madame Mireille Maurice.

Le poste de déléguée régionale est un poste de direction de la délégation⁶⁸ et l'une des tâches principales est d'assurer et de développer le rayonnement en région des activités de l'INA. Plus précisément Madame Mireille Maurice a à sa charge la représentation et le développement de la présence de l'INA auprès des partenaires potentiels que sont les pouvoirs publics, les collectivités locales, les institutions régionales et locales, les diffuseurs régionaux et locaux, les entreprises, les professionnels de l'audiovisuel et enfin le monde universitaire et de la recherche. C'est à la déléguée régionale de définir la stratégie de la délégation pour mener à bien les missions définies par l'INA et de veiller à la cohérence des actions menées. La déléguée se doit également de concevoir, proposer et piloter les projets de développement à caractère commercial, culturel ou éducatif et pour ce faire la déléguée assure le suivi du portefeuille des différents clients et partenaires de l'INA et développe les marchés de la délégation. Ce poste de direction a également pour mission de piloter des projets de valorisation des fonds d'archives régionaux et plus particulièrement des collections de l'INA et également de valorisation des actions et des outils de communication en cohérence avec la stratégie de communication de l'entreprise. Étant à la tête de la délégation, la déléguée régionale encadre et anime l'équipe en poste à la délégation. Ainsi la déléguée doit garantir la réalisation des actions de collecte, de conservation, de traitement des fonds, de communication et de valorisation. Elle planifie les activités et affecte les ressources humaines et matérielles. Enfin, elle coordonne les actions au sein de la délégation, définit les priorités et fixe les objectifs de la structure. Pour terminer, la déléguée a aussi la charge de la gestion

⁶⁸ Fiche de poste déléguée régionale, à consulter en annexe page 78

administrative et de la communication des données de la délégation à l'INA basé à Bry-sur-Marne. C'est donc un poste de direction à responsabilité et comme pour l'ensemble des postes à la délégation, qui nécessitent une grande autonomie et une polyvalence accrue par rapport aux postes situés au siège de l'INA.

La déléguée régionale est assistée dans ces tâches par une assistante de direction⁶⁹. Dans la structure marseillaise, c'est Madame Valérie Perey qui occupe ce poste. Dans les délégations, les assistantes de directions ont des activités plus larges et une polyvalence plus importante que les assistantes de direction du siège de l'INA. Dans un premier temps l'assistante de direction se doit d'assister la déléguée régionale dans la gestion des activités de la délégation. Elle réalise alors les tâches classiques du secrétariat, comme l'accueil téléphonique et physique des visiteurs, la gestion du courrier, la prise de rendez-vous, la gestion des agendas ainsi que l'organisation des réunions et des missions. Elle a également la charge de la gestion et du suivi des budgets par l'élaboration de tableaux de bord, de la transmission des éléments de facturation, de la gestion administrative comme les contrats ou encore de la gestion de l'intendance. Les missions de communication sont à la charge de l'assistante de direction en région, ce n'est pas le cas pour les assistantes de direction à Bry-sur-Marne. Les missions de communication consistent à élaborer des outils de communication de la délégation en lien avec la direction de la communication mais aussi à contribuer à l'animation et la mise en place de nouveaux vecteurs de communication comme des films promotionnels à destination des réseaux sociaux. L'ensemble des outils peuvent être des affiches, des kakémonos ou des publications sur les réseaux sociaux. Enfin l'assistante de direction doit coopérer à l'organisation des événements et en assurer le suivi. Pour terminer Madame Valérie Perey est chargée du traitement des dossiers à caractère commercial, administratif et culturel dans un objectif de suivi de la clientèle. L'assistante de direction centralise pour la délégation l'ensemble des propositions commerciales et les dossiers en cours de traitement. Elle tient à jour les tableaux de synthèse et de suivi d'activités, enfin elle finalise les offres commerciales comme les devis d'achat de droits d'exploitation d'archives. Toutes ces missions de traitement des dossiers commerciaux incombent uniquement à

⁶⁹ Fiche de poste assistante de direction en région, à consulter en annexe page 77

l'assistante de direction dans les délégations régionales. C'est en collaboration étroite avec la responsable documentaire que le traitement des offres commerciales est mené à bien.

Le troisième poste à la délégation est celui de responsable documentaire⁷⁰, il est occupé par Madame Sophie Gillery qui est ma tutrice de stage, et assure de nombreuses missions. Tout d'abord, la responsable documentaire a la charge de la gestion des données et des techniques documentaires, cela signifie que la collecte, la description, l'indexation, la valorisation, la structuration, l'éditorialisation ainsi que la communication des fonds d'archives multimédia sur tout type de supports sont réalisées par la responsable documentaire en accord avec la déléguée régionale. Il s'agit d'un travail assez long que la responsable documentaire ne peut effectuer sans l'aide des autres membres de la structure. Le poste demande également de coopérer lors d'une expertise des fonds, ce fut notamment le cas lors de mon stage. En effet, l'entreprise Pernod Ricard a demandé une expertise de leurs fonds, qui a été réalisée par la responsable documentaire et la déléguée régionale. La responsable documentaire a la charge des recherches de sons et images à proposer et/ou à commercialiser aux clients en tenant compte des contraintes juridiques et commerciales notamment grâce à l'apport du service juridique de l'INA basé à Bry-sur-Marne. Ainsi des structures ou des particuliers demandent à la délégation des archives sur un sujet, le travail de recherche est donc fait par la responsable documentaire. Durant mon stage, le musée du Mucem à Marseille souhaitait faire une exposition sur les animaux en ville. La responsable documentaire a donc proposé une sélection d'archives sonores et vidéos pour enrichir la future exposition. Le poste demande de participer au développement régional de l'INA par des actions de communication et de valorisation des fonds. C'est donc la responsable documentaire en lien avec la déléguée régionale qui met en place des projets de valorisation souvent en lien avec les acteurs culturels régionaux. La responsable documentaire a également la charge de l'accueil et de l'accompagnement des usagers lors de la consultation des archives de la collection de l'INA. Ainsi elle accueille les étudiants, chercheurs ou professionnels de l'audiovisuel, et propose une présentation des outils et des ressources de la structure. De plus, elle apporte une assistance aux usagers en matière de consultation des archives et d'utilisation des outils documentaires. Lors de l'accueil d'étudiants, public principal

⁷⁰ Fiche de poste responsable documentaire en région, à consulter en annexe page 79

de la délégation marseillaise, la responsable documentaire explique le fonctionnement des archives à l'INA et énumère les fonds disponibles à la consultation. Puis elle propose une formation sur le logiciel HyperBase qui permet de naviguer dans les fonds de l'INA. Enfin au besoin, elle présente les logiciels documentaires annexes de l'INA comme MédiaCorpus permettant de créer un corpus et d'organiser un grand nombre d'archives. Le rôle d'un responsable documentaire est aussi de prendre en charge des interventions pédagogiques. L'INA avec ses partenariats dans le monde universitaire propose des formations, c'est donc la responsable documentaire qui dispense des cours de différents niveaux, à la fois des cours pour des étudiants en sciences humaines, mais aussi en journalisme ou encore à des professionnels de l'audiovisuel. Lors de mon stage, elle a assuré un cours dans les locaux de l'INA portant sur la colonisation française dans le cadre d'un master de l'Université d'Aix-Marseille intitulé « Métiers des archives, des bibliothèques. Médiation de l'Histoire et Humanités numériques »⁷¹. Enfin le poste demande d'effectuer une veille documentaire pour se tenir informé et être en quête permanente d'informations pertinentes et utiles. Ce poste que j'ai le plus observé, durant mon stage, apparaît comme très complet, avec des missions polyvalentes et une grande charge de travail.

Le dernier poste à la délégation est celui de technicien sons, images et multimédia. Le poste est pourvu par Monsieur François Landriot. Il s'agit d'un poste de cadre technique. A ce titre Monsieur Landriot conçoit, assure et garantit les opérations techniques audiovisuelles à la fois pour une prise de vues, de sons, de montage, d'étalonnage ou encore de restauration numérique. Dans les délégations, le métier demande une grande polyvalence, ce n'est pas le cas pour les techniciens au siège de l'INA qui ne travaillent que sur leur spécialité. A la délégation, le cadre technique travaille aussi bien sur les images que sur le son, sur des archives audiovisuelles et sur des projets de valorisation ou de restauration. Le technicien doit aussi assurer des missions de maintenance des appareils utilisés à l'INA. De plus, le technicien a la charge de l'envoi des archives au siège de l'INA pour une numérisation avant une mise en ligne sur la base de données de l'entreprise. Le technicien encadre également l'ensemble des prises de vues de l'INA comme notamment la retransmission sur les réseaux sociaux de l'INA

⁷¹ Page web du master : « Métiers des archives, des bibliothèques. Médiation de l'Histoire et Humanités numériques », consultable en ligne : < <https://formations.univ-amu.fr/ME5HHC-PRHHC5AC.html> >, (consulté le 04/07/22).

des activités de valorisation. Lors de mon stage ce fut par exemple le cas pour la captation de l'émission « Le Temps des Archives »⁷². Ainsi le technicien sélectionne, organise et configure les moyens techniques en vue d'une production, puis exploite les images. Il se doit d'assurer une veille technologique dans ses domaines d'expertise. Enfin à l'instar de la responsable documentaire et de l'assistante de direction, il peut proposer une sélection d'archives lors des demandes commerciales. A titre d'exemple, lors de mon stage, le technicien a proposé des images de reportages télévisés portant sur la guerre opposant la Russie et l'Ukraine à la suite de la demande d'un client. Ainsi du fait de l'équipe assez réduite à la délégation, l'ensemble des membres a une grande polyvalence et traite des tâches assez variées.

B- Le fonds de la chaîne LCM

Les missions de stages ont été définies par ma tutrice, la responsable documentaire, Madame Sophie Gillery. Avant de commencer le stage, après plusieurs échanges avec ma tutrice, elle a proposé que mes missions se déclinent dans le traitement documentaire d'un fonds. Il s'agissait pour moi tout d'abord de visionner le fonds, puis d'effectuer un travail d'indexation et de description documentaire. Il m'a été également demandé de réaliser un travail de catalogage du fonds. Par la suite, je devais avoir la charge de la création de notices documentaires pour les archives ainsi que la création de collection pour cataloguer les archives. L'une de mes missions serait aussi de préparer des lots d'archives pour les envoyer en numérisation, cette tâche renvoie à la politique du plan de sauvegarde et de numérisation mise en place par l'INA en 1999. Avant le stage, ma tutrice m'a informé que le choix du fonds serait affiné au démarrage du stage en fonction des besoins de la délégation. Elle m'a tout de même informé que je travaillerais sur un des trois fonds suivants :

Le fonds de la chaîne LCM : il s'agit d'une chaîne de télévision aussi appelée La chaîne de Marseille (LCM), une chaîne de télévision locale généraliste et privée diffusée dans la région marseillaise et aixoise par le réseau TNT entre 2005 et 2016. La chaîne propose des émissions d'actualités, comme le « 19/30 » qui traite de l'actualité locale, l'émission « Talk » qui présente un débat autour d'une thématique d'actualité ou encore l'émission « Tryptik »

⁷² Émission disponible sur la chaîne YouTube du Mucem, consultable en ligne : <https://www.youtube.com/watch?v=T16MqUb8LA&t=6s>, (consulté le 05/07/22)

consacrée à l'actualité urbaine. La chaîne a diffusé plusieurs émissions de divertissement et de sport.

Le second fonds est celui de M6 Marseille : il s'agit d'une chaîne régionale du groupe M6 qui diffuse des programmes similaires à la chaîne M6, aujourd'hui la chaîne ne diffuse plus.

Enfin le dernier fonds à traiter est celui de l'association SOS Méditerranée⁷³ : C'est une association de défense des réfugiés. La mission principale de l'association est la sauvegarde des réfugiés, venant en grande majorité d'Afrique et voulant traverser la mer Méditerranée sur des embarcations précaires. A la suite des sauvetages, l'association protège les réfugiés d'abord en proposant des soins gratuits, de la nourriture, de l'eau et des produits de première nécessité, ensuite en proposant un toit aux réfugiés. SOS Méditerranée s'est donné comme mission de témoigner sur les difficultés et la violence dont sont victimes les réfugiés. Ainsi l'association organise des événements notamment des concerts, des conférences ou des ateliers pédagogiques pour sensibiliser un maximum de personnes et mobiliser l'opinion publique sur ces questions. Le fonds de SOS Méditerranée⁷⁴ contient des archives audiovisuelles assez diverses. Il est constitué à la fois par des vidéos de sauvetage en pleine mer, des moments de vie filmés sur le bateau de SOS Méditerranée, et également des conférences et interviews de personnalités de l'association. On y trouve également des reportages sur la vie à bord des bateaux, les expositions et événements réalisées par l'association ou encore des témoignages de réfugiés. Il s'agit donc d'un fonds assez éparpillé traitant de sujets d'actualité. Le producteur de ce fonds est l'association elle-même qui embarque à bord de ses bateaux des caméras pour répertorier et expliquer ses actions.

Lors de mon arrivée à la délégation, ma tutrice m'a annoncé que la tâche la plus urgente pour la délégation était l'envoi en numérisation des cassettes d'archives de la chaîne LCM. Le traitement de ce fonds d'archives fut donc ma première mission au sein de l'INA. Dans un premier temps, j'ai effectué des recherches complémentaires sur la chaîne de télévision LCM pour définir le type d'archives audiovisuelles et pour fournir des informations portant sur les

⁷³ Site web SOS Méditerranée, consultable en ligne : < <https://www.sosmediterranee.fr> > (consulté le 05/03/22).

⁷⁴ Lien vers quelques vidéos du fonds de SOS Méditerranée présentées sur le site web de l'association, consultable en ligne : < <https://www.sosmediterranee.fr/journal-de-bord?type=video> > (consulté le 05/03/22).

origines de la chaîne, son évolution et quelques données significatives sur LCM comme l'audience cumulée. Ces recherches se sont basées sur environ une dizaine de documents sous format Word ou Excel appartenant à la chaîne LCM mis à la disposition de l'INA. Effectivement, le projet d'archivage de cette chaîne est impulsé par Madame Maurice la déléguée régionale à l'INA Marseille. Après avoir récupéré ce fonds lors de la fin de la chaîne en 2016, elle a pris contact avec Madame Anne-Sophie Maxime aujourd'hui rédactrice en cheffe à France 3 PACA et anciennement rédactrice en cheffe de la chaîne LCM. Madame Anne-Sophie Maxime en s'appuyant sur ses contacts à la chaîne LCM, nous a fait parvenir quelques documents de travail en format Word et Excel trouvés sur des anciennes clefs USB appartenant à des membres de la chaîne. Ces documents sont dans leur grande majorité des conducteurs d'émissions qui n'ont que peu d'intérêt pour l'INA, mais il y a des brouillons de grilles de programmes ainsi qu'une présentation des ambitions de la chaîne. En m'appuyant sur ces documents et mes recherches, j'ai pu rédiger une note de synthèse pour comprendre l'histoire de la chaîne, enrichir les futures notices et poser les bases de la provenance du fonds car aucune autre recherche n'a été faite par l'INA au sujet de cette chaîne.

Fig. 1 : Exemple d'un document fourni par Anne Sophie Maxime, brouillon d'une liste d'émissions classées par genre

<u>Emissions</u>	<u>Info</u>	<u>Magazines</u>	<u>Sport</u>	<u>Mini-série/séries</u>	<u>Documentaires</u>	<u>Politique</u>	<u>Divertissements</u>
LCM +	Bonjour Marseille	Les snippets de l'info	X-Trem	La chaîne à moi	Les grandes évasions	Portraits politiques	C'est du Jazz
La Grande émission	LCM Matinale	Tranche de Vie	C Foot	Funky style	Collard « Coupable hier »	1 candidat/ 1 jour	Concerts
Emission Spéciale	J.T de la rédaction	Vous qui partez en voyage	Vélo	Dédé le Fachomako	Collard « Coupable aujourd'hui »	Elections : (régionales/nationales/ Législatives...)	Live
	22/Minuit	Grand Tourisme	Planet.pétanque	Jackpot 13	Documentaires non identifiés	Face à la presse	Cinéma
	J.T de la semaine	C' Nouveau	C Sport				Théâtre
	C'est l'Info Présentateur : Guillaume Bladorne Réalisation : Christophe Liotard	Handiversité (1 et 2)	Bonne mer				Musique
	J.T 19h	C' Aix					

Fig. 2 : Exemple d'un document fourni par Anne Sophie Maxime, brouillon d'une grille des programmes de 2009 à 2010

GRILLE SAISON 2009-2010									
	LUNDI	MARDI	MERCREDI	JEUDI	VENDREDI	SAMEDI	DIMANCHE		
7:00	C LA SANTÉ	JT REDIF	JT REDIF	JT REDIF	JT REDIF	JT REDIF	JT SEMAIN	7:00	ACTU
7:15								7:15	ACTU EXT
7:30	METEO	METEO	METEO	METEO	METEO	METEO	METEO	7:30	ECO
7:45	BIN' O BIN	BIN' O BIN	BIN' O BIN	BIN' O BIN	BIN' O BIN	BIN' O BIN	JT SEMAIN	7:45	OM
8:00	METEO	METEO	METEO	METEO	METEO	METEO	METEO	8:00	SPORT
8:15	TOTAL LO	JT REDIF	JT REDIF	JT REDIF	JT REDIF	JT REDIF	JT SEMAIN	8:15	LOISIRS
8:30	VOYAGE							8:30	HUMOUR
8:45	METEO	METEO	METEO	METEO	METEO	METEO	METEO	8:45	MAGAZINE
9:00	C NOUVEA	C NOUVEA	EUROSHO	C NOUVEA	C NOUVEA	EUROSHO	TOP DE L'	9:00	TENDANCE
9:15								9:15	COMMERCIAL
9:30								9:30	FILMS
9:45								9:45	INSTITUTIONNELS
10:00								10:00	SANTÉ
10:15								10:15	SERIE
10:30	C L'ECO	C L'OM			TOP DE L'	C LA SANTÉ		10:30	CULTURE
10:45					TOTAL LOOK			10:45	
11:00	MON MAR	C LA SANTÉ	C SPORT	CINEMA	DOCU	C L'OM	HANDIVER	11:00	
11:15			TOP DE L'	CELESTE				11:15	
11:30	C L'IMMO	CELESTE	CINEMA	C L'IMMO		C SPORT	LCM CAFE	11:30	

Comme nous l'avons évoqué plus tôt, la chaîne LCM est une chaîne de télévision généraliste et dans cette optique, elle diffuse des journaux télévisés, des reportages, des émissions de débat, de divertissement ou des fictions. Ce sont donc des archives de plusieurs genres différents ayant des statuts juridiques également assez différents. En effet, l'INA ne peut pas traiter de la même manière un direct d'émission et une fiction. LCM est une chaîne locale, elle a pour but d'informer et de divertir la population locale, celle de l'agglomération marseillaise et aixoise, à ce titre la chaîne émet sur la TNT uniquement en local. Enfin il est à noter que nous avons à faire à une chaîne privée dont l'actionnariat repose sur les acteurs économiques de la région marseillaise. Il s'agit ici d'un élément important car la grille des programmes se retrouve influencée par la mise en avant de certains acteurs économiques investissant sur la chaîne.

Le projet de la chaîne LCM est impulsé par le CSA, nouvellement l'ARCOM, à la suite d'un appel à projet datant du 18 mars 2003⁷⁵. Cet appel à projet s'inscrit dans le développement des télévisions locales en France impulsé par les pouvoirs publics et aboutira à un rapport⁷⁶ du Sénat produit le 22 juin 2005, écrit par le rapporteur et sénateur Claude Belot⁷⁷ au nom de la commission des finances. C'est en octobre 2005 que le projet « TV7 Marseille » est sélectionné par le CSA pour émettre dans la région PACA, « TV7 Marseille » va alors changer de nom et devenir LCM (La Chaîne Marseille). Ce projet est financé par la Caisse d'épargne d'Alpes-Corse, la Caisse nationale d'épargne et le quotidien La Provence⁷⁸. La nouvelle chaîne de télévision s'inscrit alors dans le cadre juridique obligatoire qui définit une chaîne de télévision locale. Effectivement, le cadre général d'une télévision locale est fixé par la loi 86-1067 du 30 septembre 1986 relative à la liberté de communication⁷⁹. L'une des prérogatives d'une télévision locale est fixée par l'alinéa 5 de l'article 41-3 de la loi numéro 86-1067 considérant qu'une télévision est considérée comme locale lorsqu'elle dessert une population de moins de 10 millions d'habitants. Le sénateur Claude Belot ajoute dans le rapport datant du 22 juin 2005 que :

« De façon plus pertinente, une télévision locale se définit concrètement par l'identification à un territoire et concourt, de ce fait, au pluralisme des médias : les dimensions de démocratie et de proximité sont étroitement imbriquées. Ces caractéristiques ne s'avèrent pas universelles : aux Etats-Unis l'élément décisif est le sentiment d'appartenance à une communauté.

L'identification à un territoire se traduit par une programmation spécifique des télévisions locales, lesquelles répondent ainsi à une demande des téléspectateurs que ne suffisent pas à satisfaire les décrochages régionaux de France 3 et M6. »⁸⁰

⁷⁵ Communiqué du CSA datant du 31 mars 2004 énumérant les candidats à la création d'une chaîne de télévision locale, consultable en ligne : < https://www.csa.fr/var/ezflow_site/storage/csa/rapport2004/donnees/communiques/555.htm >, (consulté le 05/07/22)

⁷⁶ Rapport du Sénat écrit par Claude Belot daté du 22 juin 2005, consultable en ligne : < https://www.senat.fr/rap/r04-417/r04-417_mono.html >, (consulté le 05/07/22)

⁷⁷ Claude Belot est sénateur du groupe Union pour un Mouvement Populaire entre 1989 et 2014

⁷⁸ Rapport du Sénat écrit par Claude Belot daté du 22 juin 2005, consultable en ligne : < https://www.senat.fr/rap/r04-417/r04-417_mono.html >, (consulté le 05/07/22)

⁷⁹ Loi numéro 86-1067 du 30 septembre 1986 relative à la liberté de communication, consultable en ligne : < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068930/> >, (consulté le 05/07/22)

⁸⁰ Introduction du rapport du Sénat écrit par Claude Belot daté du 22 juin 2005, consultable en ligne : < https://www.senat.fr/rap/r04-417/r04-417_mono.html >, (consulté le 05/07/22)

Cette mise en avant des chaînes de télévision locale est renforcée par les propositions du Sénat pour garantir la pérennité des chaînes. L'ensemble des propositions mises en avant dans le rapport de Claude Belot vise à favoriser la création de nouvelles chaînes locales, à accroître les revenus publicitaires des chaînes et permettre un soutien public et privé accru. Les propositions sont les suivantes :

« Assurer la présence des télévisions locales parmi les chaînes gratuites de la TNT, en préemptant au besoin une partie des futures fréquences disponibles grâce au progrès des normes de diffusion.

Dans le cadre du réaménagement des fréquences lié au passage à la TNT, fixer un second plafond des dépenses restant à la charge des télévisions locales. Exprimer ce plafond en euros selon un barème qui tienne compte du chiffre d'affaires des chaînes.

Permettre la création de télévisions locales en Ile-de-France, soit par une nouvelle définition des télévisions locales qui desserviraient des zones comptant jusqu'à douze millions d'habitants, soit par le lancement d'appels à candidatures par le CSA en Ile-de-France.

Encourager une procédure de labellisation des télévisions locales prenant en compte le contenu de la programmation.

Envisager la création d'un fonds d'amorçage des télévisions locales, transitoire, autofinancé grâce à l'amélioration du recouvrement des taxes sur le hors-média et la publicité télévisée. Ce fonds pourrait constituer une nouvelle section de l'actuel fonds de soutien à l'expression radiophonique.

Affecter une part du produit de la redevance audiovisuelle au financement des télévisions locales à caractère public ou associatif.

Favoriser la reconnaissance des versements aux associations de téléspectateurs comme des dons ouvrant droit à réduction d'impôt.

Généraliser le principe d'une redevance des câblo-opérateurs aux télévisions locales du câble.

Permettre aux collectivités territoriales d'obliger les câblo-opérateurs à offrir du temps d'antenne et du matériel aux télévisions locales associatives.

Envisager de réserver aux seules télévisions locales, dans un premier temps, la levée de l'interdiction de publicité télévisée pour la grande distribution dans le cadre des campagnes de promotion.

Afin d'élargir le champ de la publicité institutionnelle pour les télévisions locales, les utiliser comme supports des campagnes d'information gouvernementale. »⁸¹

Toutes ces mesures favorisent l'implantation d'une télévision locale et c'est dans ce contexte que la chaîne LCM a pu émettre le 7 octobre 2005 sous la direction de Pierre Boucaud le directeur général et Philippe Bes directeur de l'information et des programmes. La chaîne diffuse 24h sur 24H en priorité de l'information locale, elle souhaite devenir un média participatif et de proximité. Pour ce faire, elle ouvre ses micros aux mondes associatifs et aux acteurs de la vie locale. La chaîne est confrontée à des difficultés économiques à cause d'un désengagement de ces actionnaires du fait d'une perte de confiance à l'échelle nationale portant sur le concept de télévision locale. Le président du conseil d'administration, Jean-Pierre Foucault, en place depuis 2005 quitte son poste. La chaîne est reprise par l'entrepreneur Patrick Siri⁸² jusqu'en 2011 où la chaîne rejoindra le groupe Médias du Sud⁸³ déjà propriétaire des télévisions locales TV Sud Montpellier et TV Sud Camargues-Cévennes. La chaîne profite de ce rachat pour changer d'identité visuelle, de logo et même pour passer en format d'image 16 : 9. La grille est harmonisée avec les autres chaînes du groupe, ainsi la grille des programmes fait la part belle aux directs, aux événements sportifs et culturels. Pour marquer son affiliation avec le groupe Médias du Sud, LCM change de nom le 7 mai 2015, pour TV Sud Provence⁸⁴. Malheureusement, la chaîne a des grosses difficultés économiques, les conditions de travail sortent du cadre légal, les salariés sont payés avec de nombreux mois de retard. Tout cela pousse les salariés à faire grève et à interrompre les programmes. Le 3 mai

⁸¹ Rapport du Sénat écrit par Claude Belot daté du 22 juin 2005 consultable en ligne < https://www.senat.fr/rap/r04-417/r04-417_mono.html >, (consulté le 05/07/22)

⁸² Article de journal La Provence consultable en ligne : < <https://www.laprovence.com/actu/region-en-direct/151794/la-reprise-de-lcm-par-bertrand-bigay-et-patrick-siri-officialisee.html> >, (consulté le 05/07/22)

⁸³ Article du magazine Stratégue consultable en ligne : < <https://www.strategies.fr/actualites/medias/186545W/les-chaines-lcm-et-tv-sud-harmonisent-leurs-programmes.html> >, (consulté le 05/07/22)

⁸⁴ Article de journal La Provence consultable en ligne : < <https://www.laprovence.com/article/actualites/3431602/apres-lcm-tv-sud-provence-joue-la-difference.html> > (consulté le 05/07/22)

2016, le chaîne cessera sa diffusion⁸⁵ à la suite d'une liquidation judiciaire. L'ensemble de ces difficultés apparaissent lors d'un contexte assez précis de la télévision en France, notamment avec l'avènement de la télévision par TNT (télévision numérique terrestre). Cette technologie a permis d'obtenir une meilleure qualité d'image et de réduire les coûts d'exploitation et de transmission. Dans ce contexte, la télévision française devient de plus en plus concurrentielle avec la création le 31 mars 2005 de huit nouvelles chaînes de télévision : Direct 8, W9, TMC, NT1, NRJ12, La Chaîne parlementaire-Publique Sénat, France 4, BFM, I-Télé, Europe 2 TV, Gulli, France Ô. Ces nouvelles chaînes concurrencent les chaînes locales malgré une différenciation par la grille de lecture à l'échelle locale de LCM. L'arrivée en 2011 de 6 nouvelles chaînes en HD continue va rendre difficile le développement des chaînes locales. Ces six nouvelles chaînes sont : HD1, L'équipe 21, 6ter, Numéro 23, RMC découvertes et Chérie 25. Ainsi le contexte d'augmentation de l'offre audiovisuelle n'a pas permis à LCM de pouvoir tirer parti de sa spécificité locale. Cependant LCM a eu des ambitions assez importantes. Cela est notamment mis en avant dans un document interne à la chaîne que nous avons pu consulter grâce aux concours de Madame Sophie Maxime. Ce document est une présentation de la saison 2009 aux actionnaires de LCM. Nous pouvons y voir que LCM avait pour objectif en 2009 d'être la première chaîne locale de France en 2010. Ainsi la chaîne explicite son objectif en ces termes :

« S'imposer, d'ici fin 2010, comme la 1^{ère} chaîne locale de France, en poursuivant le développement de l'audience cumulée quotidienne, et en capitalisant sur une forte « mobilisation identitaire »⁸⁶

Pour y parvenir, LCM compte s'appuyer sur 4 facteurs : le premier, son référencement dans le bouquet TV de la télévision d'Orange, le second est d'accroître le trafic du site web de la chaîne où les émissions sont retransmises, le troisième facteur est le renforcement de la durée d'écoute et la fidélisation des téléspectateurs autour d'une émission quotidienne forte, enfin le dernier facteur est l'augmentation de la qualité des partenariats et de l'attractivité des programmes. L'ensemble de ces points est mis en avant dans le document sous cette forme :

⁸⁵ Article du site d'information Marsactu consultable en ligne : < <https://marsactu.fr/bref/une-chaine-bfm-marseille-remplacera-bientot-azur-provence-tv/> >, (consulté le 05/07/22)

⁸⁶ Document interne à LCM, daté de 2009.

Fig. 3 : Les objectifs 2010 de LCM⁸⁷

LCM Objectif 2010



Les objectifs :

- ➔ S'imposer, d'ici fin 2010, comme la **1ère chaîne locale de France**, en poursuivant le développement de l'audience cumulée quotidienne, et en capitalisant sur une forte « mobilisation identitaire ».

4 leviers identifiés :

1. Accroître l'initialisation de la chaîne, via son référencement imminent dans le **bouquet TV d'Orange**
2. Accélérer le **trafic sur le site internet www.lachainemarseille.fr**
3. Renforcer la durée d'écoute moyenne quotidienne, en fidélisant les téléspectateurs tout au long d'un **magazine de 52'**, en direct, du lundi au vendredi, entre 17h30 et 18h30,
4. Favoriser la visibilité de la chaîne, via la **qualité des partenariats**, et l'**attractivité des programmes**.

C'est effectivement une période au cours de laquelle LCM peut avoir de l'ambition, la chaîne réalise des bons résultats calculés grâce au système Médiamétrie. Ainsi la chaîne affirme être la troisième chaîne locale en France derrière TV7 Bordeaux et TLM (Télé Lyon Métropole) comme le montre le graphique ci-dessous. Ce graphique met en évidence l'audience cumulée quotidienne des trois grandes chaînes de télévision locales (TV7 Bordeaux, TLM, LCM). L'audience cumulée est définie comme « le nombre ou la proportion de téléspectateurs ayant été en contact au moins une fois avec le média lors d'une période donnée et quelle que soit la durée du contact⁸⁸ ». La croissance de l'audience est assez bonne pour LCM notamment en septembre 2007 où les chaînes de télévision locale diffusent à la fois en analogique et en numérique sur décision de la Direction générale des médias et des industries culturelles⁸⁹. Il

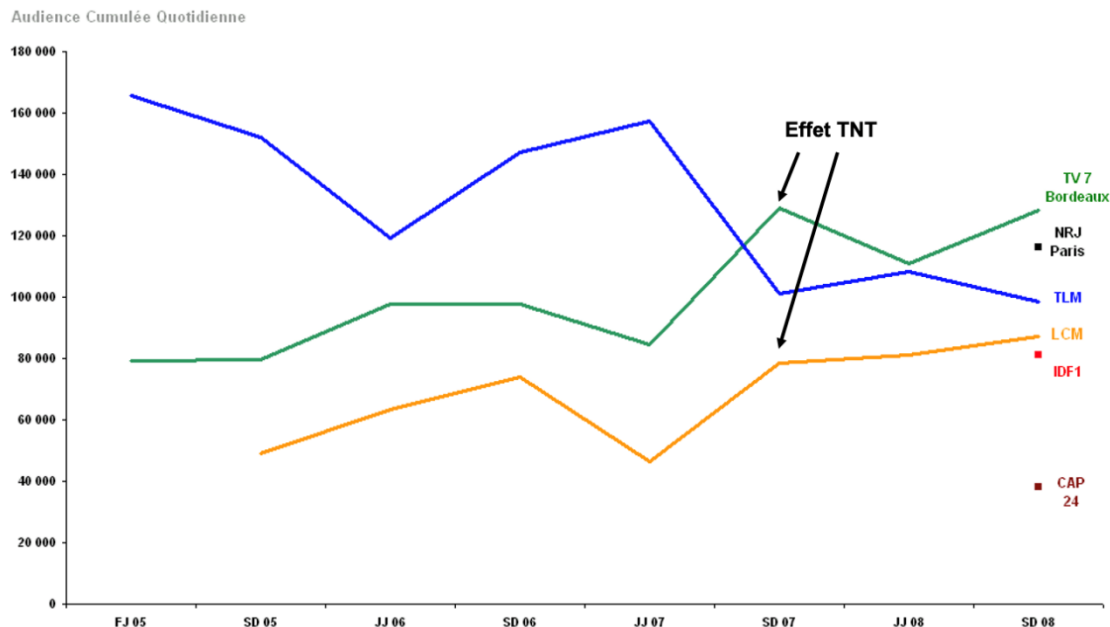
⁸⁷ Document interne à LCM, daté de 2009.

⁸⁸ Définition donnée par la revue spécialisée dans l'étude des médias intitulée « Effeillage » consultable en ligne : < <https://effeuillage-la-revue.fr/portfolio-item/laudience-cumulee/> >, (consulté le 05/07/22)

⁸⁹ Communiqué de la Direction générale des médias et des industries culturelles, datant du 12/05/2009, consultable en ligne : <

est à noter que la durée d'écoute du téléspectateur n'est pas comptabilisée ce qui peut faire perdre de la crédibilité à ces chiffres.

Fig. 4 : Croissance de l'audience entre 2005 et 2008⁹⁰



Le tableau ci-dessous est également un tableau interne à LCM qui m'a permis de cerner l'influence de la chaîne et de retracer son histoire. Nous y observons l'audience cumulée mesurée en individus sur la base d'individu de plus de 15 ans équipés d'une télévision sur la période 2008. Cette audience atteint 87.300 pour la chaîne LCM. Ainsi la chaîne se place comme la troisième télévision locale derrière TV7 Bordeaux et TLM. Il paraît pertinent de signaler que la chaîne NRJ Paris n'est pas comptabilisée comme étant une chaîne locale du fait qu'elle est diffusée dans une région de plus dix millions d'habitants.

https://archive.wikiwix.com/cache/index2.php?url=http%3A%2F%2Farchive.dgmic.culture.gouv.fr%2Farticle.php3%3Fid_article%3D1313#federation=archive.wikiwix.com >, (consulté le 05/07/22)

⁹⁰ Document interne à LCM, daté de 2009.

Fig. 5 : Tableau de l'audience cumulée des chaînes locales⁹¹

**Palmarès en audience cumulée en individus
sur une base totale 15+ équipés TV**

TV7 BORDEAUX	128 400
NRJ PARIS	116 100
TLM	98 500
LCM	87 300
IDF 1	81 300
TV8 MONT BLANC	74 300
CAP 24	38 000
DEMAIN TV IDF	33 400
TLT	31 800
TV TOURS	24 600
CLERMONT PREMIERE	24 500
NANTES 7	23 600
TELE GRENOBLE	23 500
CANAL 32	14 300
ANGERS 7	14 000
TV VENDEE	11 200
7L MONTPELLIER	10 700
ORLEANS TV	8 700
TELE MIROIR	5 600
CITYZEN TV	2 500
TELENANTES	2 300

Enfin le dernier tableau ci-dessous vise à démontrer la croissance rapide de la chaîne entre septembre 2007 et décembre 2008. On y observe notamment une audience cumulée quotidienne en augmentation de 11% entre la période allant de septembre à décembre 2007 et la période de septembre à décembre 2008. L'ensemble de ces chiffres relativement bons pour la chaîne LCM sont à mettre en perspective avec l'audience cumulée des chaînes de télévision plus généralistes. Effectivement, si en comparaison avec d'autres télévisions locales, LCM peut donner l'impression de profiter d'une audience importante, ces chiffres sont sans aucune mesure avec les chiffres de chaînes comme TF1 ou France 2 qui détiennent respectivement 26,1 et 17,5⁹² de la part de l'audience en 2008 avec notamment des pics d'audience cumulée quotidienne à environ 10 millions de téléspectateurs bien loin des 87.300 téléspectateurs quotidiens pour la période de septembre à décembre 2008 de la chaîne LCM.

⁹¹ Document interne à LCM, daté de 2009.

⁹² Article de la revue « Les Échos », publié le 30 décembre 2008 consultable en ligne : <
<https://www.lesechos.fr/2008/12/tf1-destabilisee-par-la-montee-en-puissance-des-chaines-de-la-tnt-506075>
>, (consulté le 08/07/22)

Fig. 6 : Tableau de croissance de l'audience de la chaîne LCM entre septembre 2007 et décembre 2008⁹³

LCM	sept-déc 2008	janv-juin 2008	sept-déc 2007	Evol. sd08/sd07
15 ans et + Equipés TV	1 241 200	1 195 800	1 060 500	17,0%
Notoriété	926 100	825 000	732 700	26,4%
soit	74,6%	69,0%	69,1%	
Initialisation	742 900	585 200	503 300	47,6%
soit	59,8%	48,9%	47,5%	
Habitudes d'Ecoute	542 600	438 900	365 800	48,3%
Audience 7 derniers jours	335 200	252 700	234 100	43,2%
Audience cumulée quotidienne	87 300	81 200	78 500	11,2%

Résultats 00h/24h en Lundi - Dimanche

En conclusion, ces informations m'ont permis d'écrire une histoire du fonds, de retracer les origines de la chaîne et de mettre en perspective quelques données pour fournir les éléments en vue d'une recherche plus approfondie de la part d'étudiants. Ces documents d'analyse du fonds ont été délivrés à ma tutrice de stage. Ainsi j'ai pu travailler sur ces éléments pour prendre en main le fonds et le traiter de façon plus pertinente. Comprendre la provenance du fonds et avoir réalisé un historique succinct de cette chaîne permet de mieux appréhender le visionnage et la rédaction des notices.

⁹³ Document interne à LCM, daté de 2009.

C- Le traitement du fonds

Le cœur de ma mission a été de visionner le fonds de la chaîne LCM et de fournir une notice pour chaque émission en vue d'une numérisation au siège de l'INA. Après avoir effectué une recherche sur le fonds de LCM, ma tutrice m'a demandé de commencer le travail de visionnage du fonds. Elle m'a demandé de regarder le maximum de cassettes en commençant par les cassettes datées à partir de janvier 2010, date à laquelle nous disposons de plus d'informations sur les grilles des programmes. Le travail consiste alors à visionner une journée de direct enregistrée sur une cassette puis à créer une notice par émission. Nous reviendrons plus en détail par la suite sur la définition de la notice et les étapes de la création. Les cassettes se trouvent dans les magasins de la délégation se situant dans les étages. J'ai obtenu l'autorisation de prendre des photos pendant le stage et j'ai pu capturer un magasin d'archives où sont stockées les cassettes LCM.

Fig. 7 : Un magasin d'archives à la délégation marseillaise de l'INA⁹⁴



⁹⁴ Image personnelle

A mon arrivée, j'ai appris que la climatisation des magasins d'archives était défectueuse, malheureusement cette panne a duré plusieurs mois. La climatisation n'a pas été remplacée durant cette période ce qui pourrait induire sur le long terme une dégradation des archives notamment sur les supports fragiles comme les bobines de projecteur de cinéma en 35mm.

Le fonds de la chaîne LCM se présente sur des cassettes de type DVCAM⁹⁵. Il s'agit d'un format propriétaire de la société japonaise Sony. Ce format réservé aux professionnels de l'audiovisuel permet de stocker des formats en Haute Définition⁹⁶ allant jusqu'à 1440 x 1080 pixels et cela pour un coût de production assez faible. Ces cassettes sont à l'INA dans un format compact pouvant contenir jusqu'à 3 heures de programme par cassette. En effet, il existe deux tailles⁹⁷ de cassettes DVCAM, la première taille est nommée « S » permet d'enregistrer jusqu'à 44 minutes tandis que la taille « L » utilisée par LCM permet d'obtenir jusqu'à 184 minutes d'enregistrement. La bande du format DVCAM mesure 0,6 cm⁹⁸, il s'agit d'une bande assez étroite qui ne permet pas d'enregistrer dans des conditions difficiles comme en extérieur ou de répondre à une sollicitation intensive en montage. Ainsi LCM a choisi d'enregistrer ces émissions en direct sur les cassettes du fabricant Sony qui convient à une utilisation pour de la télévision. Le visionnage se fait alors uniquement sur des magnétoscopes pouvant lire le format DVCAM. Il est à noter que la délégation ne dispose que de deux magnétoscopes pour cassettes DVCAM et que le siège de l'INA dispose d'un stock limité. De plus, le fabricant Sony a arrêté de produire ces machines. Ainsi le matériel utilisé pour lire les DVCAM est ancien, rare, et coûteux du fait de la rareté. Pour moi, il s'agit d'une chance de pouvoir visionner ces cassettes grâce au magnétoscope DSR-2000p. Avec l'accord de la délégation, j'ai pu prendre en photo le matériel utilisé lors du visionnage du fonds ainsi que les cassettes.

⁹⁵ Cassette en format propriétaire Sony

⁹⁶ La haute définition est une classification d'équipements de télédiffusion et de vidéo ayant une définition d'au moins 1280 x 720 pixels puis 1920 x 1080 pixels.

⁹⁷ Spécificités techniques inerrant aux deux tailles des cassettes en format DVCAM, consultable en ligne : < https://pro.sony.fr_FR/products/tape-media/dvcam >, (consulté le 08/07/22)

⁹⁸ Spécificités techniques des cassettes en format DVCAM, consultable en ligne : < https://pro.sony.fr_FR/products/tape-media/dvcam >, (consulté le 08/07/22)

Fig. 8 : Magnétoscope DSR 2000p servant à la lecture de DVCAM et cassette DVCAM⁹⁹



Les cassettes du fonds de la chaîne LCM contiennent les émissions en direct du jour. Un jour de direct est enregistré sur une cassette, hormis de rares cas où un jour de direct est enregistré sur deux ou trois cassettes pour des raisons que l'on ignore. Sur une cassette, nous trouvons en moyenne entre 4 et 6 émissions en direct pour une durée allant d'une heure à une heure et demie. Ces émissions sont donc les directs des émissions du jour, on y trouve des journaux télévisés, des débats, des reportages ou encore des émissions au format hybride

⁹⁹ Image personnelle

mêlant divertissement, reportage et information. A la demande de ma tutrice, j'ai donc visionné la première cassette de l'année 2010 datant du 4 janvier 2010. Sur la cassette étaient enregistrées 6 émissions de direct différentes. La première est l'émission « LCM Plus » d'une durée de 14 minutes 19 secondes, la seconde se nomme « C l'OM » d'une durée de 26 minutes 09 secondes, « LCM les titres » d'une durée de 2 minutes 35 secondes, « Vox Marseille » d'une durée de 23 minutes 17 secondes, l'émission « C Nouveau » d'une durée de 13 minutes 53 secondes et enfin « LCM le Journal » d'une durée de 14 minutes. Ainsi ces 6 émissions ont une durée globale de 94min 27 secondes. Lors du visionnage, il est important de noter les sujets abordés, les personnes interviewées ainsi que le nom des journalistes ou des intervenants. Toutes ces informations sont primordiales pour la création de la future notice.

Une notice de l'INA est une fiche détaillée de l'archive, à savoir l'émission enregistrée sur la cassette. La notice est présentée à l'INA comme la pièce d'identité de l'archive, on y inscrit des informations permettant de définir l'origine du fonds, le producteur de l'archive, la date de création ainsi que la date de diffusion avec les heures de diffusion, il faut également énumérer les acteurs de l'archive comme les journalistes ou les techniciens derrière la caméra si nous disposons de ces informations et noter le thème de l'archive. La notice permet de renseigner le thème de l'archive tout comme le genre audiovisuel que ce soit un journal télévisé, un magazine ou un reportage. La notice permet aussi de décrire l'archive pour éviter aux chercheurs de visionner l'archive dans son intégralité, s'ils recherchent uniquement un élément en particulier. Chaque notice a un identifiant propre. De plus, une notice correspondant à une archive doit être englobé dans une collection. La collection permet de regrouper l'ensemble des archives appartenant à un même groupe, par exemple l'ensemble des archives de l'émission « C Nouveau » sera englobé dans une collection intitulé « C Nouveau ». Ce système de rangement permet de naviguer plus facilement pour les chercheurs et professionnels utilisant le logiciel propriétaire de l'INA permettant la consultation de la base de données : HyperBase. Ainsi la création des notices après le visionnage des archives est devenue le centre des activités que j'ai réalisées durant mon stage. En effet, avant d'envisager une numérisation des cassettes, il est important que toutes les cassettes soient vues et les notices créées pour chaque émission. De plus les cassettes doivent être nommées et reliées avec les différentes notices que chaque cassette contient.

Pour créer ces notices, naviguer dans les bases de données et faire des recherches, les documentalistes utilisent l'application « Totem ». Il s'agit d'une application de recherche documentaire pouvant rechercher à la fois dans les fonds vidéo, radiophonique ou photographie. Cette application est utilisée quotidiennement par l'ensemble des membres de la délégation.

Fig. 9 : Capture d'écran de l'application « Totem »¹⁰⁰

Identifiant de la notice :	RAM22000941
Titre propre :	La Grande Émission : émission du 21 avril 2010 seconde partie
Titre collection (Aff.) :	La Grande Émission
Diffusion (aff.) :	21/04/2010 - type date: Diffusé -heure:18:30:00 - canal:La Chaîne Marseille (LCM)
Durée :	00:40:53
Producteurs (Aff.) :	Société de production - La Chaîne Marseille (LCM) - Marseille - 2010
Genre :	Journal télévisé :
Thématique :	Information :
Générique (Aff. Lia.) :	PRE Sahli, Igor : PAR Keramidas, Elane : PAR Aragones, Isidore : PAR Barlatier, Alain :
Résumé :	En plateau itw de Alain BARLATIER secrétaire départemental adjoint syndicat FSU (Fédération Syndicale Unitaire). Il évoque l'absentéisme à l'école et les éventuelles sanctions financières pour les familles. Présentation du Plan Local pour l'Insertion et l'Emploi, itw de François RANISE chef d'entreprise et parain d'Hassane LAHCENE chercheur d'emploi, itw de Hassane LAHCENE. Présentation de la "Femme de Noé", Une femme qui met en avant le développement durable, itw de Kiara volontaire originaire de Sardaigne, itw de Anita PRUNIER fondatrice de la ferme. En plateau itw de Lisa LE PANN-GOURITEN membre de la "Terre de Noé" et Alain MASSE agriculteur bio à Mallemond. Ils présentent l'agriculture biologique et ses bienfaits. En plateau itw de Gérard FILIPPI entomologiste. Il explique l'importance de la biodiversité dans le mode de vie des humains. Explication de l'affaire de la petite Élise. Il s'agit d'une affaire de kidnapping d'une petite fille par sa mère après un divorce difficile, itw de Jean-Michel ANDRÉ père d'Élise, itw de maître Isidore ARAGONES avocat spécialisé en droit des personnes. En plateau itw de maître ELIANE KERAMIDAS avocate au barreau de Marseille. Elle expose les différents types de contrats de divorce.
Identifiant Matériels (info.) :	MGRAM-MA0002747-AX_01
Matériels (Détail) :	[MSV] - MGRAM-MA0002747-AX_01/01 - Format: DVCAM - Définition: 625 - Format img.: 4/3 - Son: MONO - Coul.: COULEUR - Filière: MG - Type Mat.: MASTER - Durée: 03:00:00 - Stat.Vers.: Versé (16/06/2022) - Localisation: RAMASMAG (16/06/2022) - N° stock.: LCM568 - N° bande: -MA0002747-AX - Info. techniques document: Muet: Non - Vi: Non -
Origine du fonds (lib.) :	La Chaîne Marseille
Matériels dispo (Détail) :	[MSV] - MGRAM-MA0002747-AX_01/01 - Format: DVCAM - Définition: 625 - Format img.: 4/3 - Son: MONO - Coul.: COULEUR - Filière: MG - Type Mat.: MASTER - Durée: 03:00:00 - Stat.Vers.: Versé (16/06/2022) - Localisation: RAMASMAG (16/06/2022) - N° stock.: LCM568 - N° bande: -MA0002747-AX - Info. techniques document: Muet: Non - Vi: Non -
N° de stock :	LCM568 :
Type de date :	Diffusé

Cette capture d'écran de l'application montre une notice d'une émission, en l'occurrence la notice de « La Grande émission » datant du 21 avril 2010. Nous avons renseigné les informations suivantes :

- Le titre propre de l'archive (il s'agit du nom de la collection et de la date de diffusion)
- Le titre de la collection
- La date et l'heure de diffusion
- La durée
- Le nom du producteur
- Le genre
- Le thème
- Le générique (les intervenants de l'émission et les techniciens de l'émission)
- Le résumé

¹⁰⁰ Capture d'écran de l'application « Totem » datant du 24/06/22

- L'identifiant matériel
- Les données matérielles
- L'origine du fonds

Ces informations sont à renseigner par moi-même, hormis les données matérielles et l'identifiant matériel qui se renseignent automatiquement grâce à l'application « Sumoco » que nous aborderons plus tard. Ainsi après le visionnage, l'application « Totem » me permet d'éditer les notices, de les classer et hiérarchiser grâce aux collections. C'est donc dans une finalité d'enrichissement de la base de données que j'ai effectué ce travail. Ainsi lors de mon stage, j'ai visionné l'ensemble des cassettes de la chaîne LCM de janvier à mai 2010. Ce long travail m'a permis, au fur et à mesure des visionnages, de créer une grille de programmes pour la période, d'autant plus utile pour des futures recherches car les brouillons des grilles de programme de la chaîne LCM ne sont pas à jour. Avec ces visionnages, j'ai ainsi pu analyser 15 émissions différentes et rédiger plus de 400 notices.

Le champ demandant le plus de travail de recherche, de compréhension et d'analyse est celui du résumé. Après avoir échangé sur la méthode de résumé avec ma tutrice, nous avons convenu qu'il fallait être précis sans ajouter trop de détails pour permettre aux utilisateurs de la base de données de comprendre le sujet en peu de mots. Cela a représenté pour moi un travail considérable, il s'agissait de résumer de façon concise un nombre important d'émissions. J'ai réalisé ce travail en autonomie complète. Ma tutrice a corrigé mon écrit uniquement pour les premières notices, ensuite j'ai pu avancer rapidement sans correction. Pour cette tâche, il est important de décrire de façon détaillée les événements d'actualité ou les interviews susceptibles d'être utiles pour une recherche ou pour une commande d'archive. Ainsi les événements marquants de janvier à mai 2010 mis en avant par la chaîne LCM, comme par exemple la victoire en coupe de la ligue de l'Olympique Marseillais¹⁰¹, le séisme en Haïti¹⁰² ou encore les élections régionales¹⁰³, sont à décrire en

¹⁰¹ Victoire en coupe de la ligue de l'Olympique Marseillais obtenu le 27 mars 2010

¹⁰² Séisme en Haïti de magnitude de 7 à 7,3 ayant eu lieu le 12 janvier 2010 faisant plus de 229.000 morts et plus de 300.000 blessés, consultable en ligne : < <https://archive.wikiwix.com/cache/?url=http%3A%2F%2Fwww.insu.cnrs.fr%2Fa3348%2Cseisme-haiti-12-janvier-2010.html> >, (consulté le 09/07/22)

¹⁰³ Élections régionales en région PACA, Michel Vauzelle membre du parti socialiste est réélu à la tête de la région le 26 mars 2010 avec 44,11% des voix, consultable en ligne : < <https://www.lemonde.fr/elections->

priorité en précisant bien les noms des intervenants et des journalistes. Ces résumés se doivent de fournir assez d'informations. A titre d'exemple, le résumé ci-dessous est celui de l'archive de l'émission « LCM le journal » du 4 janvier 2010 :

« Attaque d'un fourgon blindé de la société Sazias à Gémenos, itw d'un témoin du braquage, itw David-Olivier REVERDY syndicat de police « Alliance », itw Jacques DALLEST procureur de la République de Marseille.

Grève de la police contre le projet de réforme de la police d'agglomération.

Vol d'une toile de Degas « Les Choristes » au musée Cantini. En plateau itw Daniel HERMANN adjoint au Maire de Marseille chargé de la culture. Il évoque le succès de l'exposition « De la scène au tableau », itw habitantes de Carry-le-Rouet, itw Eric DIARD maire UMP de Sausset-les-Pins.

Brève : retard des TER de la SNCF dans la Vaucluse (grève, neige et acte de vandalisme). Marseille capitale européenne de la culture en 2013 et le vol de la toile « les Choristes ».

Décès de la chanteuse Lhasa de Sela.

Mercato d'hiver de l'OM, départ supposé de Mathieu Valbuena.

Bain du nouvel an au cercle des nageurs de Marseille, itw Paul LECCIA président du cercle des nageurs de Marseille, itw d'adhérents au cercle des nageurs.

Micro-trottoir : les résolutions du nouvel an des Marseillais. »¹⁰⁴

Le résumé contient neuf informations traitées de manière différente lors du journal télévisé. Le point central du résumé est de mettre en avant l'information principale en ajoutant le nom des différents acteurs. Il peut s'agir d'interview en reportage dans ce cas nous utilisons l'abréviation « itw ». Dans le cas d'un interview en plateau, il est d'usage de le signaler. Il est à noter que lors de la description de l'archive, il est préférable de noter le style journalistique adopté comme la brève ou le micro-trottoir. Ainsi ce type de résumé a été réalisé pour plus de 400 archives provenant d'environ 15 émissions différentes ayant des thèmes et des structures inégales, nous reviendrons sur l'ensemble des émissions plus tard.

regionales/article/2010/02/25/michel-vauzelle-confirme-son-statut-de-favori-en-paca_1311005_1293905.html
>, (consulté le 09/07/22)

¹⁰⁴ Extrait de la notice de l'émission « LCM le journal » datant du 4 janvier 2010, à consulter en annexe page 80

Lors de la création d'une notice, en plus du résumé qui reste le point central, il y a plusieurs champs à décrire. Tout d'abord dans l'onglet « Diffusion », il faut renseigner la date de diffusion, l'heure de diffusion ainsi que le canal de diffusion comme montré lors de l'exemple de l'émission « La Grande émission » diffusé le 21 avril 2010.

Fig. 10 : Onglet « Diffusion » lors de la création d'une notice¹⁰⁵

Les champs suivants que je devais compléter se situent dans l'onglet « Production ». Il s'agit d'un onglet donnant des informations sur le producteur de l'archive. En l'occurrence, je renseigne le nom du producteur à savoir « La Chaîne Marseille » et également le sigle de la chaîne (LCM). L'application demande de fournir la date de production en année ainsi que le lieu de production. Enfin, dans une liste déroulante, je dois choisir le libellé du rôle du producteur, dans notre cas, il s'agit d'une société de production. Ce libellé est automatiquement attaché à un code rôle qui est « PRD ». Dans cette capture d'écran de la même émission du 21 avril, l'onglet « Production » est mis en lumière.

¹⁰⁵ Capture d'écran de l'application « Totem » datant du 24/06/22

Fig. 11 : Onglet « Production » lors de la création d’une notice¹⁰⁶

Code rôle	Libellé code rôle	Sigle	Libellé producteur
PRD	Société de production	LCM	La Chaîne Marseille

L’onglet « Typologie » permet de fournir les éléments de classification de l’archive. Ainsi dans une liste déroulante, nous pouvons définir la thématique et le genre de l’archive. Il s’agit de deux listes d’autorités fermées qu’il faut obligatoirement renseigner. Dans notre cas de l’émission du 21 avril, il y est renseigné « journal télévisé » pour le genre et « information » pour la thématique.

Fig. 12 : Onglet « Typologie » lors de la création d’une notice¹⁰⁷

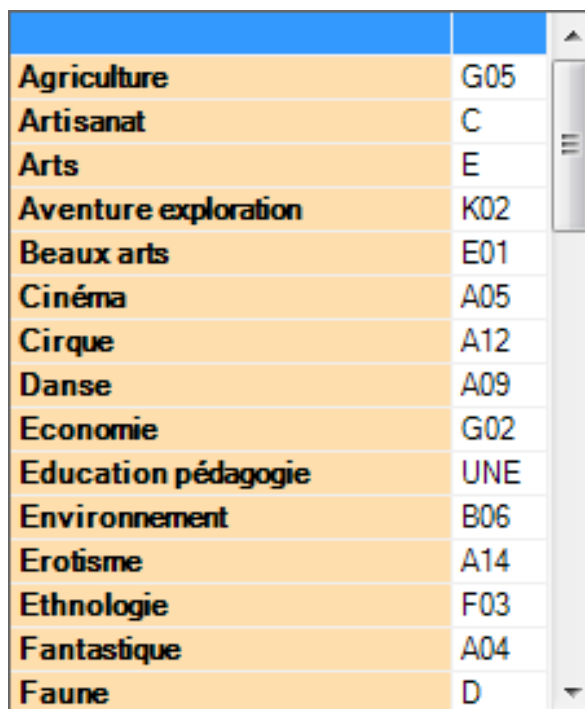
Il est possible de renseigner plusieurs genres ou plusieurs thématiques pour une seule archive. Cela peut être pertinent notamment dans le cas d’archive où se croisent des thématiques comme pour l’émission « C Nouveau ». La liste des thématiques étant précise, renseigner

¹⁰⁶ Capture d’écran de l’application « Totem » datant du 24/06/22

¹⁰⁷ Capture d’écran de l’application « Totem » datant du 24/06/22

plusieurs thématiques peut être approprié. Cette capture d'écran présente un aperçu des thématiques renseignées par l'INA.

Fig. 13 : Capture d'écran de la liste d'autorité des thématiques¹⁰⁸



Agriculture	G05
Artisanat	C
Arts	E
Aventure exploration	K02
Beaux arts	E01
Cinéma	A05
Cirque	A12
Danse	A09
Economie	G02
Education pédagogie	UNE
Environnement	B06
Erotisme	A14
Ethnologie	F03
Fantastique	A04
Faune	D

Enfin, le dernier onglet à renseigner est celui du « Générique ». L'ensemble des personnes prenant la parole à l'écran ou étant cité au générique de fin de l'émission doit figurer au générique. Il existe deux listes d'autorités pour documenter le générique. La première est un thésaurus fermé contenant le nom de personnes souvent citées dans les archives audiovisuelles de l'INA. La seconde liste est un lexique ouvert. Le lexique contient plus de nom que le thésaurus et est ouvert au documentaliste, c'est-à-dire qu'il est possible d'ajouter des personnes au lexique. Lors de l'ajout d'une personne, il est impératif de fournir le métier de la personne pour éviter autant que possible les doublons avec une personne homonyme déjà renseignée dans le lexique. Il est à noter que dans plusieurs émissions, le nom des présentateurs ou des journalistes n'est pas prononcé ou indiqué, il faut alors fournir un travail de recherche pour préciser le nom de la personne. Ce travail peut être assez long mais permettra d'enrichir la base de données et fournir des informations supplémentaires aux utilisateurs. J'ai pu ainsi dresser une liste des intervenants réguliers de la chaîne qui sera utile

¹⁰⁸ Capture d'écran de l'application « Totem » datant du 24/06/22

pour les personnes travaillant sur ce fonds. Cette liste est consultable en annexe 6¹⁰⁹. La capture d'écran suivante présente un exemple de générique complété. On y observe à la fois les participants c'est-à-dire les personnes interviewées et également l'ensemble du personnel technique et de rédaction crédité à la fin de l'émission. Est inscrit en face de chaque nom le libellé du rôle de l'individu ainsi que le code de ce libellé.

Fig. 14 : Onglet « Générique » lors de la création d'une notice¹¹⁰

Génériques			
	Code rôle	Libellé code rôle	Patronyme
☐	PRE	Présentateur	Floutier, Rosalie
☐	SON	Ingénieur du son	Gay, Guillaume
☐	RED	Rédacteur en chef	Maxime, Anne Sophie
☐	SCR	Scripte	Cordeau, Anne
☐	PAR	Participant	Lingelser, Lionel
☐	PAR	Participant	Chakri, Lotfi
☐	PAR	Participant	Sylvain, Elodie
☐	PAR	Participant	Teissier, Guy
☐	CED	Responsable d'édition	Dieuaide, Yann
☐	REA	Réalisateur	Liotard, Christophe
☐	CED	Responsable d'édition	Ricci, Alain

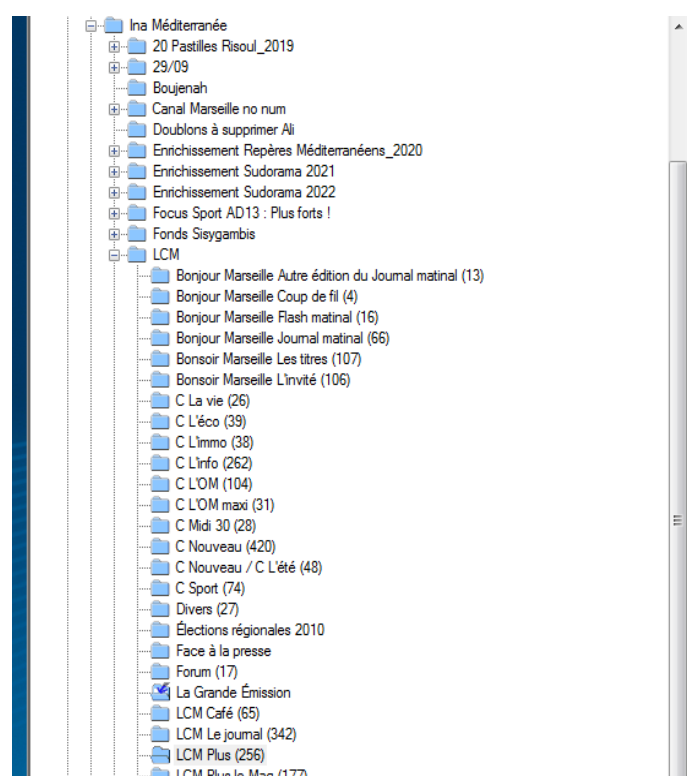
21/06/2022 - 10:57 La notice EMISSIONREGION 'RAM22000964' existe déjà dans le

L'ensemble de ces champs à remplir permet de créer une notice opérationnelle pour une archive se trouvant sur une cassette. Après la création de la notice, pour pouvoir la retrouver facilement, il est indispensable d'organiser l'information. Pour cela, la notice nouvellement créée est mise dans un dossier que l'on nomme folio. Ces folios sont des dossiers de travail partagés sur l'application « Totem ». L'ensemble des membres de l'INA peut intervenir sur les dossiers folios.

¹⁰⁹ Liste des intervenants réguliers de la chaîne LCM, à consulter en annexe page 82

¹¹⁰ Capture d'écran de l'application « Totem » datant du 21/06/22

Fig. 15 : Classement des folios de travail partagés¹¹¹



Lorsque l'on estime avoir suffisamment d'occurrences d'une émission, il est convenu de créer une notice de collection. Cette notice sera accessible depuis le dossier folio de l'émission. Il s'agit d'une notice basée sur le même principe qu'une notice d'archive expliquant le sujet de l'émission, le genre ainsi que les multiples spécificités. J'ai pu créer plusieurs notices de collection, nous prendrons ici l'exemple de la notice de collection de l'émission « La Grande émission ». Ainsi dans cette notice de collection, de la même façon qu'une notice d'archive, il faut renseigner le canal de diffusion, l'horaire de l'émission, la durée moyenne des émissions, la périodicité, la nature de production et le producteur. Puis il est demandé de fournir le genre et la thématique de l'émission. Avec l'accord de ma tutrice, nous avons défini le genre de magazine et choisi d'indiquer plusieurs thématiques pour avoir le plus d'ouvertures possible. Nous avons donc renseigné arts, cinéma, théâtre, spectacle, loisirs, information, politique, musique et tourisme. Ce choix a été motivé par l'idée de donner une idée générale des thèmes abordés dans l'ensemble des émissions. Contrairement aux notices d'archives où nous devons fournir un résumé de l'archive, dans le cas d'une notice de collection, il est demandé de fournir une description du contenu des archives appartenant à cette collection. Ainsi nous avons

¹¹¹ Capture d'écran de l'application « Totem » datant du 21/06/22

choisi d'apporter le plus de détail possible à cette partie pour permettre une recherche rapide et simplifiée des contenus des émissions. Nous avons pour cela fait le choix d'énumérer l'ensemble des chroniques présentes dans les émissions. La rubrique « Contenu » de la notice de collection contient les informations suivantes :

« Magazine quotidien culturel, politique et d'information en plateau avec des chroniqueurs et un présentateur. Alternance de plateaux et de reportages.

Diffusé à 18h00 du lundi au vendredi, d'une durée d'environ 1h30 pour les deux parties.

Cette émission arrive sur la grille à partir du 09/03/2010 et remplace « C Nouveau » qui avait lieu précédemment à 18h30 du lundi au vendredi de 2008 au 05/03/2010. Les différentes collections « C L'info », « C L'éco », « C L'OM », et « C l'immo » sont absorbées dans « La Grande émission » et n'apparaissent plus non plus à partir du 06/03/2010.

La première partie de l'émission (40') est consacrée aux thématiques culturelles, sur le même schéma que l'émission de la grille précédente « C Nouveau ». Cette partie est constituée d'une succession de chroniques et d'interviewes d'invités en plateau.

La deuxième partie (40') est axée sur l'information et la politique. Les lundi et vendredi, la fin de cette deuxième partie d'émission est consacrée au sport et présentée par Thierry TRESOR.

Première partie de l'émission :

Chronique « Le Bon Plan » diffusée du lundi au vendredi. Présentation d'un magasin ou d'un restaurant en partenariat avec l'émission. Des places ou des bons de réduction sont à gagner dans l'émission.

Chronique « MPM 2013 » : diffusée le lundi. Présentation d'un quartier de Marseille en partenariat avec MPM 2013.

Chronique « Rubrique quartier » : diffusée du lundi au vendredi. Remplace la chronique « MPM2013 » et est généralisée tous les jours de la semaine toujours en partenariat avec « MPM2013 »

Chronique « Mer » : diffusée le mardi. Présentation d'un navire de Marseille, bateau de plaisance d'un particulier, navette, vieux gréement, ferry, bateau de croisière.

Chronique « Innovation » : diffusée le lundi. Présentation d'un concept innovant ou d'un objet technologique.

Chronique « Tendances » : diffusée le mardi et le jeudi. Présentation de boutiques de vêtements ou de bijoux.

Chronique « Resto » diffusée le mardi et le vendredi. Présentation d'un restaurant et de sa recette emblématique

Chronique « Livre » diffusée le mercredi. Présentation d'une librairie et d'un ou plusieurs ouvrages coups de cœur par le libraire.

Chronique « People ciné » diffusée le mercredi. Présentation de la sortie cinéma ou du concert de la semaine.

Chronique « Culture » : diffusée le jeudi. Présentation d'une exposition muséale ou d'une galerie d'art.

Chronique « La dinette d'Annette » : diffusée le jeudi. Une recette est exécutée en plateau

Deuxième partie de l'émission

Chronique « Le coup de pouce » : diffusée du lundi au vendredi. Mise en avant et présentation d'une association caritative, sociale ou culturelle de quartier.

Chronique « L'OM » : diffusée le lundi et le vendredi. Discussion et débat autour de l'OM, avec 2 à 4 chroniqueurs en plus de Thierry TRESOR à la médiation.

Chronique « Vox Marseille » diffusée le mardi. Ancienne émission de la grille de 2008 au 05/03/2010 diffusée à 18h. Alternance de micro-trottoir et de discussions avec deux chroniqueurs sur des sujets variés (ex sous-titre de l'émission « Les Marseillais ont la parole »).

Chronique « Le droit et vous » diffusée le mercredi une seule fois par mois. Conseils juridiques et décryptage de dispositifs légaux délivré par Elianne KERAMIDAS. »¹¹²

Toutes ces informations, comme les jours de diffusion ou le déroulement de l'émission permettent aux utilisateurs de la base de données de trouver facilement l'archive recherchée et de permettre aux utilisateurs de comprendre les formes de l'émission sans la regarder. La

¹¹² Rebrique « Contenu » de la fiche de collection de la grande émission, datant du 21/06/2022

dernière rubrique importante d'une fiche de collection est la rubrique « Générique », il s'agit de la rubrique énumérant les individus récurrents des émissions ainsi que leur rôle.

Après avoir visionné une cassette, créé les notices pour chaque émission, organisé et classé les émissions dans un folios et créé une notice de collections, il convient de relier la notice nouvellement créée au matériel, à savoir la cassette. Il s'agit d'une étape primordiale pour retrouver la cassette dans les rayons du magasin d'archives et permettre son départ dans différent lieu comme le siège à l'INA pour une numérisation. Relier une cassette à des notices va permettre de fournir un numéro unique d'identification de la cassette. Pour ce faire, une application propriétaire est utilisée par l'INA. Cette application nommée « Sumoco » va permettre de faire les liens entre les notices et les cassettes ainsi que de fournir et de permettre l'impression des numéros d'identifiants. Pour cela, après avoir renseigné le numéro de la DVCAM, l'application demande les spécificités techniques de l'archive, dans notre cas, il sera toujours question d'enregistrement numérique, en standard couleur type « PAL » ou encore d'un format d'image en 4/3. Il faut par la suite sélectionner les notices que l'on souhaite attacher à la cassette comme nous le montre la figure suivante. Puis dans un dernier temps l'application fournit une côte de rangement. Il est à noter que l'application « Totem » et « Sumoco » fonctionnent de concert. Les notices créées sur l'application « Totem » sont retrouvables dans l'application « Sumoco » en renseignant la date de l'archive, la société de production, le canal de diffusion et l'origine du fonds. De la même façon, lorsque l'on relie un matériel avec des notices grâce à la l'application « Sumoco » cela se répercute sur les notices dans l'application « Totem ». En effet, sans rien renseigner, les notices se mettent à jour avec les éléments matériels comme la côte de rangement du matériel, le format de la cassette, le format de l'image, l'encodage du son, le profil de couleur ou encore la durée. L'interconnexion entre les deux applications permet de gagner du temps et d'éviter les erreurs de côte et d'identifiant du matériel et de la notice qui pourraient poser des problèmes aux utilisateurs ainsi qu'aux équipes de l'INA. Dans l'exemple ci-dessous, la cassette du 23 avril 2010 contient quatre émissions différentes pour cinq notices car l'émission « La Grande Émission » est en deux parties. Les cinq notices sont attachées au même matériel.

Fig. 16 : Application « Sumoco »¹¹³

Recherche docs sur la journée de diffusion

Critères de recherche
Date de Diffusion: Vendredi 23/04/2010 Thèque: RA Société Prod.: M N° Document: Titre d'ensemble commence par : Chaîne diff.: CM Ori. fonds: LB

☒ Toutes Diffusions ☐ Diffusés uniquement ☐ Conserver la sélection

N° de Matériel: MGRAM-MA0002749--AX

Select	N° Doc	Thq	Soc	Titre d'Ensemble	Titre du Document	Type	Date Diffusion	Heure Diffusion JS	Ch	Durée
☐	RAM22000964	RA	LCM	La Grande Émission	La Grande Émission : émission du	D	23/04/2010	18:00:00	CM	00:42:02
☐	RAM22000980	RA	LCM	La Grande Émission	La Grande Émission : émission du	D	23/04/2010	18:30:00	CM	00:41:55
☐	RAM22000981	RA	LCM	LCM Le journal	LCM Le journal : émission du 23 av	D	23/04/2010	18:45:00	CM	00:13:19
☐	RAM22000963	RA	LCM	LCM Café	LCM Café : émission du 23 avril 20	D	23/04/2010	19:25:00	CM	00:26:11

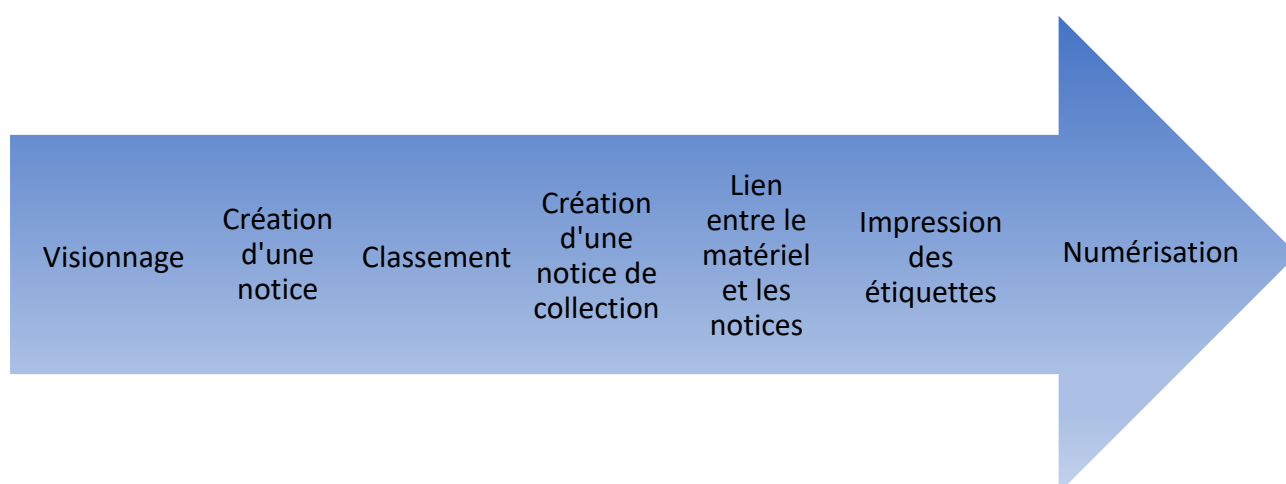
Nombre de documents sélectionnés : 0

Elmt	Document	Titre d'ensemble	Titre Propre	Date Diff	Heure Diff	JS	Durée doc
1			NOUVEL ELEMENT				
	RAM22000962	LCM Plus	LCM Plus : émission du 23 avril 2010	23/04/2010	12:15:00		
	RAM22000964	La Grande Émission	La Grande Émission : émission du 23	23/04/2010	18:00:00		
	RAM22000980	La Grande Émission	La Grande Émission : émission du 23	23/04/2010	18:30:00		
	RAM22000981	LCM Le journal	LCM Le journal : émission du 23 avril	23/04/2010	18:45:00		
	RAM22000963	LCM Café	LCM Café : émission du 23 avril 2010	23/04/2010	19:25:00		

Après avoir relié le matériel avec les notices correspondantes, il convient d'imprimer les étiquettes à coller sur les cassettes. Cette impression se fait sur une imprimante spécialisée dans l'impression d'étiquette. Ainsi sur ces étiquettes, nous avons plusieurs informations primordiales comme la côte de rangement, l'identifiant de la cassette et les informations techniques. Une fois les étiquettes sur la cassette sur la boîte de rangement, l'ensemble du processus de traitement de l'archive est terminé. Les boîtes partiront par la suite au siège de l'INA pour une numérisation. Le schéma ci-dessous reprend les grandes lignes d'un traitement documentaire d'un fonds à l'INA. Nous y observons l'ensemble de la chaîne de traitement que nous avons mis en place durant le stage.

¹¹³ Capture d'écran de l'application « Sumoco » datant du 21/06/22

Fig. 17 : Chaîne de traitement d'une archive mise en place lors du stage



Durant l'ensemble du visionnage, j'ai pu fournir à ma tutrice une liste des émissions sur la période étudiée allant de janvier à mai 2010. Les différentes émissions sont :

- LCM Plus : journal télévisé de la mi-journée, se terminant par un entretien en plateau
- LCM le Journal : journal télévisé de la fin de journée
- C l'OM : magazine sportif portant sur l'OM
- Les titres : programme court présentant les sujets d'actualités du jour
- Vox Marseille : débats et discussion entre deux journalistes de rédaction régionale sur des sujets d'actualité et de société entrecoupés de micro-trottoir
- C l'immo : magazine portant sur l'immobilier
- C l'info : débat sur un sujet d'actualité ou de société avec les acteurs du sujet
- C Nouveau : émission de divertissement présentant des activités culturelles
- C l'éco : magazine portant sur l'économie
- LCM Café : discussion entre quatre invités récurrents sur des sujets d'actualité et de société
- La grande émission : émission de divertissement présentant des activités culturelles, des recettes de cuisine, des reportages sur des quartiers de la ville de Marseille
- Les élections régionales : émission spéciale pendant les élections régionales présentant les candidats
- C sport : magazine portant sur le sport
- C la vie : magazine portant sur la santé
- Face à la presse : Entretien en plateau avec un politique

L'ensemble de ces émissions montre la diversité des programmes de la chaîne et l'importance pour LCM de s'ancrer dans le local pour être au plus proche des populations. Il est à noter que lors du mois de mars 2010, la chaîne LCM a modifié sa grille de programme. J'ai pu alors sur demande de ma tutrice fournir deux grilles des programmes de la chaîne. La première pour la période de janvier à février 2010 et la seconde pour la période de mars à mai 2010. Ce changement de grille de programme effectué au milieu de la saison interroge la documentaliste, il semble que cela est dû à la recherche d'un taux d'audience supérieur dans une période concurrentielle de l'offre télévisuelle. La première grille des programmes a été créée après avoir visionné un mois des directs de la chaîne. Ce fut un travail de recherche assez long du fait que les horaires ne sont pas communiqués dans les documents complémentaires fournis par LCM et que les présentateurs ne prononcent que rarement l'heure des émissions. Après un recoupage de plusieurs émissions et une mise en perspective avec les documents de la chaîne, j'ai pu organiser une grille. Cette grille compte dix émissions à laquelle viennent s'ajouter des programmes ponctuels comme un programme sur les élections régionales ou l'émission « Face à la presse ».

Fig. 18 : Grille des programmes de la chaîne LCM de janvier à février 2010¹¹⁴

	Lundi	Mardi	Mercredi	Jeudi	Vendredi
12h15	LCM Plus	LCM Plus	LCM Plus	LCM Plus	LCM Plus
17h45	C l'OM	C l'immo/ C sport	C l'éco/ vide	Vide	C l'OM
18h	Les Titres	Les Titres	Les Titres	Les Titres	Vide
18h02	Vox Marseille	C l'info	C l'info	C l'info	LCM Café
18h30	C Nouveau	C Nouveau	C Nouveau	C Nouveau	C Nouveau
18h45	LCM le Journal	LCM le Journal	LCM le Journal	LCM le Journal	LCM le Journal

¹¹⁴ Grille des programmes de la chaîne LCM réalisée par Sami Dendani

La grille est organisée autour de temps forts dans la journée. Le premier temps fort pour LCM est le journal du midi qui est quotidien, hormis le week-end. Le second temps fort est l'émission de divertissement « C Nouveau » directement suivi par le journal du soir. Le reste de la journée, différentes émissions sont diffusées en direct. Le matin, en soirée ainsi qu'entre 12h45 et 17h45, la chaîne diffuse des enregistrements des émissions ou des programmes dont elle a acheté les droits. Cette grille des programmes reste fixe jusqu'à mars 2010. A partir de mars 2010, l'ensemble de la grille des programmes est soumis à des modifications. Une nouvelle émission apparaît, il s'agit de « La Grande émission ». Ce programme inspiré de la formule du « Grand Journal » diffusé sur Canal + comme nous l'indique un document interne à la LCM¹¹⁵ vise à offrir à la fois du divertissement et de l'information entrecoupé de reportages sur des sujets variés. En s'appuyant sur l'émission « C Nouveau », le nouveau programme garde en son sein les intervenants récurrents de cette ancienne émission. Les deux journaux quotidiens sont toujours à l'antenne ainsi que l'émission de débat « LCM Café ».

Fig. 19 : Grille des programmes de la chaîne LCM de mars à mai 2010¹¹⁶

	Lundi	Mardi	Mercredi	Jeudi	Vendredi
12h15	LCM Plus	LCM Plus	LCM Plus	LCM Plus	LCM Plus
17h	La grande émission (partie 1)	La grande émission (partie 1)	La grande émission (partie 1)	La grande émission (partie 1)	La grande émission (partie 1)
17h45	La grande émission (partie 2)	La grande émission (partie 2)	La grande émission (partie 2)	La grande émission (partie 2)	La grande émission (partie 2)
18h45	LCM le Journal	LCM le Journal	LCM le Journal	LCM le Journal	LCM le Journal
19h	Vide	Vide	Vide	Vide	LCM Café

¹¹⁵ Document internet à LCM datant de septembre 2009

¹¹⁶ Grille des programmes de la chaîne LCM réalisée par Sami Dendani

Cette nouvelle grille supprime un grand nombre d'émissions en direct comme énoncé plus haut dans la notice de collection de l'émission. Pour remplacer ces émissions, il existe un grand nombre de chroniques sur des sujets variés. Après concertation avec ma tutrice, j'ai pris le temps de lister et répertorier l'organisation de ces chroniques. Ce travail a pu servir de base à la création de la notice de collection.

Fig. 20 : Répartition des chroniques de « La Grande émission » (partie 1) entre mars en mai 2010¹¹⁷

Lundi	Mardi	Mercredi	Jeudi	Vendredi
Le Bon Plan	Le Bon Plan	Le Bon Plan	Le Bon Plan	Le Bon Plan
MPM 2013	Chronique Mer	Chronique Livre	Culture	Chronique resto
Rubrique quartier	Rubrique quartier	Rubrique quartier	Rubrique quartier	Rubrique quartier
Innovation	Tendance	People Ciné	Tendance	
	Chronique resto		Dinette d'Annette	

Fig. 21 : Répartition des chroniques de « La Grande émission » (partie 2) entre mars en mai 2010¹¹⁸

Lundi	Mardi	Mercredi	Jeudi	Vendredi
Le coup de Pouce	Le coup de Pouce	Le coup de Pouce	Le coup de Pouce	Le coup de Pouce
Discussion et débat sur l'OM	Vox Marseille	Le droit et vous (Une fois par mois)		Discussion et débat sur l'OM

Mes missions se sont donc focalisées sur le fonds de la télévision LCM que j'ai visionné pendant les 5 mois de la durée du stage. La charge de travail a été relativement importante.

¹¹⁷ Répartition des chroniques de « La Grande émission » (partie 1) entre mars en mai 2010, réalisée par Sami Dendani

¹¹⁸ Répartition des chroniques de « La Grande émission » (partie 2) entre mars en mai 2010, réalisée par Sami Dendani

Les cassettes de LCM s'étalent de l'année 2005 à l'année 2016, ainsi il reste encore à effectuer un travail conséquent de visionnage et de description du fonds.

Outre le fonds de la chaîne LCM, j'ai eu l'opportunité de participer en tant qu'observateur à deux projets de valorisation des archives de la délégation. J'ai assisté aux rencontres intitulées « Le Temps des Archives »¹¹⁹ dans lesquelles un journaliste interroge des chercheurs ou des témoins d'un évènement historique, et les débats sont entrecoupés par des images d'archives sélectionnées par la délégation.

Le premier « Temps des Archives » en partenariat avec le Mucem s'est déroulé au Camp des Milles à Aix-en-Provence, sous la médiation du journaliste Hervé Brusini avec Maryse Burgot, journaliste et correspondante de guerre à France Télévisions ainsi que Jean-Christophe Klotz, journaliste et réalisateur spécialiste du génocide au Rwanda. Le sujet de cette rencontre était « L'information : dénoncer ou rendre compte ? ». Ainsi les deux journalistes ont pu développer leur point de vue sur des questions comme « peut-on rendre compte sans s'engager ? » ou « Débattre sans s'affronter ». Le débat est enrichi par des images d'archives, notamment des images de la guerre en Ukraine.

Le seconde rencontre du « Temps des archives » a eu lieu au musée du Mucem, toujours sous la médiation du journaliste Hervé Brusini avec Eric Scherer, directeur de la prospective et de la stratégie numérique de France Télévisions et Sylvain Louvet, directeur éditorial de Brut, et lauréat du Prix Albert Londres¹²⁰ en 2020 pour son documentaire « Tous surveillés : 7 milliards de suspects »¹²¹. Les intervenants questionnent la communication sur les réseaux sociaux et s'efforcent de répondre à la question suivante : « La technologie : communiquer ou influencer ? ». Ils évoquent le travail des influenceurs, le concept du nudge¹²², l'identité numérique ou encore l'utilisation des données personnelles.

Après avoir participé à ces deux rencontres, j'ai pu faire un retour, et donner mon avis personnel et mon ressenti lors d'une réunion de bilan avec les membres de la délégation.

¹¹⁹ Site web délégation INA Méditerranée : < <https://www.ina.fr/institut-national-audiovisuel/en-regions/ina-mediterranee> > (consulté le 05/03/22).

¹²⁰ Le prix Albert Londres est décerné aux grands reporters français, décliné en trois catégories : prix de la presse écrite, prix de l'audiovisuel, prix du livre

¹²¹ Documentaire diffusé sur « Arte » consultable en ligne : < https://www.youtube.com/watch?v=OWM3_6rSWyA >, (consulté le 11/07/22)

¹²² Le nudge est un concept de communication. C'est une incitation douce donnée à un individu pour modifier son comportement

Nous avons pu dessiner des axes de progression des rencontres notamment par un ancrage renforcé avec les archives de l'INA, des invitations auprès de personnalités pas uniquement issues du monde journalistique afin d'ouvrir le panel des participants et enfin la réalisation de communication des événements à destination des universités, des écoles de journalisme. Dans le même optique, j'ai participé à la création de vidéos à destination des réseaux sociaux de l'INA et du Mucem pour inciter les individus à visionner « les Rencontres » en rediffusion sur « YouTube » et à participer aux prochaines rencontres au Mucem.

En conclusion, mes missions de stage ont été assez diverses. La mission principale a été la description documentaire du fonds LCM, et j'ai démarré la mission en rédigeant un historique de la chaîne de télévision, puis j'ai commencé à visionner un grand nombre d'archives et me suis formé afin de maîtriser les logiciels propriétaires de l'INA pour mener à bien mes missions de description. Par la suite, j'ai eu l'opportunité de participer à des activités de valorisation à partir des rencontres du « Temps des archives » et à mettre en place une communication sur les réseaux sociaux.

Chapitre 3 : Apports et difficultés du stage

A- Autonomie et polyvalence

Le travail de description documentaire peut s'avérer relativement long et fastidieux. Cependant dans mon cas, la grande pluralité des émissions, couplée à l'utilisation de plusieurs logiciels m'ont permis d'avancer sans ressentir une quelconque monotonie. En effet, le nombre important de cassettes à visionner à l'INA montre que le travail de description de ce fonds en est encore à ses débuts. J'ai pu mettre en place des fiches portant sur l'histoire de la chaîne et sur les différentes collections afin de varier de la mission de visionnage. Ce travail réalisé servira aux utilisateurs mais aussi aux futurs stagiaires qui vont poursuivre la description de ce fonds. Les missions de ce stage à l'Ina m'ont permis de découvrir et travailler à partir de différents logiciels¹²³ et de me former à l'utilisation d'une base de données et à la réalisation des recherches sur cette même base. En effet, j'ai pu apprendre à rechercher dans une base de données notamment avec les différents opérateurs de recherche. La recherche sur cette base de données est très fine, pouvant aller jusqu'à une quarantaine de champs différents à remplir pour affiner au maximum la recherche. Ainsi lors des recherches, il m'a fallu un temps d'adaptation avant de maîtriser la recherche. De plus, il est possible de rechercher sur des fonds différents, aussi bien un contenu audiovisuel, qu'un contenu provenant de la radio ou de la photographie. Par la suite, j'ai enrichi ma technique de prise de notes lors des visionnages et affiné la méthodologie de synthèse, afin de rédiger un propos en quelques phrases. Par ma formation initiale en Histoire, au début il était un peu délicat pour moi de condenser les informations, mais avec les conseils et l'aide de ma tutrice, j'ai pu rapidement me focaliser sur les informations essentielles à mettre en avant dans les notices. Les informations essentielles sont le sujet, le contexte et les acteurs de l'information comme les personnes concernées ou les personnes interviewées. La mission du stage m'a permis de passer d'un travail relativement technique lors de l'utilisation des logiciels à un travail d'analyse et de rédaction pour la création des notices et cette diversité a été un atout dans ma formation lors du stage. Cette polyvalence m'a permis de plus de fournir des notices claires

¹²³ Logiciel « Totem » et « Sumoco »

et détaillées, d'effectuer des recherches sur les intervenants de la chaîne, notamment lorsque le nom des personnes n'est pas mentionné. Enfin, l'élaboration des deux grilles des programmes a été un travail qui s'est étendu sur l'ensemble du stage et qui s'est affiné au fur et à mesure des visionnages. Ce travail de synthèse a permis d'enrichir les connaissances sur ce fonds qui étaient relativement faibles et de fournir des informations aux prochains stagiaires. Lors de ce stage, j'ai travaillé en autonomie, tout en réalisant des points réguliers avec ma tutrice pour définir les éventuelles difficultés et faire un point sur l'avancement du travail. Ma tutrice s'est montrée très disponible durant l'ensemble du stage en répondant rapidement à mes questions. Durant le stage, j'ai également observé et j'ai participé à des missions de valorisation. Ces missions ont permis de comprendre le processus de création d'une mission de valorisation, tout comme les questions de logistique pour la réservation des hôtels et des restaurants pour les invités pris en charge par la directrice de la délégation, mais aussi le processus de recherche d'images d'illustration réalisé par ma tutrice, la responsable documentaire et le cadre technique. J'ai également observé la relation étroite de l'INA avec le musée national du Mucem, un des plus grands musées en Méditerranée. Cette relation a donné naissance aux rencontres du « Temps des archives ». Dans ce partenariat, le Mucem et la délégation se partagent les coûts de ces événements. Ces missions de valorisation m'ont permis de découvrir un autre versant du travail sur les archives. J'ai ainsi pu me rendre compte du travail à réaliser pour l'organisation d'un tel événement et observer les différentes problématiques. Malheureusement la qualité des rencontres n'est pas assez mise en valeur par le nombre de spectateurs relativement réduit. En effet, malgré une communication sur les réseaux sociaux le « Temps des archives » a regroupé entre 20 à 50 spectateurs sur place, 340¹²⁴ vues pour la rencontre portant sur l'information et 195¹²⁵ vues pour la rencontre sur la plateforme YouTube. Ce n'est donc pas une valorisation qui touche un grand public malgré la renommée des intervenants. J'ai pu notamment proposer des axes d'amélioration de la communication de l'évènement. Ces rencontres ont été l'occasion d'élargir mes missions au sein de la délégation et de compléter mes compétences lors de ce stage. Enfin, ce stage m'a permis de travailler en autonomie dans l'établissement, notamment du fait de la grande polyvalence et de la charge de travail importante des quatre membres de la délégation. J'ai

¹²⁴ Nombre de vues sur YouTube consultable en ligne : < <https://www.youtube.com/watch?v=yT16MqUb8LA&t=9s> >, (consulté le 11/08/22)

¹²⁵ Nombre de vues sur YouTube, consultable en ligne : < <https://www.youtube.com/watch?v=8XRuyXdouFY&t=6s> >, (consulté le 11/08/22)

ainsi travaillé dans une grande autonomie, sur un rythme soutenu pour visionner un maximum d'archives avant la fin de stage. En conclusion, ce stage m'a permis de compléter mes compétences informatiques tout en développant la prise de note et en confortant mes compétences en synthèse. J'ai pu comprendre et analyser le traitement d'une archive à l'INA, de la création de notices à la valorisation de l'archive dans le cadre d'un événement. Cependant j'ai également rencontré quelques difficultés.

B- Difficultés matérielles et taille du fonds

Durant le stage, j'ai tout de même rencontré des difficultés lors de la description documentaire. Tout d'abord, je n'ai pas abordé l'onglet « Descripteurs »¹²⁶ lors de la création de notice dans l'application « Totem ». C'était une demande de ma tutrice qui a souhaité que je puisse visionner un maximum d'archives sans pour autant me concentrer sur la mise en place de descripteurs qui aurait ralenti le visionnage mais un regret pour moi car je souhaitais me saisir de l'ensemble de la chaîne de l'indexation des archives du fonds de la chaîne LCM. De plus compte tenu de la taille du fonds, en 14 semaines, je n'ai pas pu traiter autant d'archives que j'aurais souhaité. En effet, le fonds s'étend de 2005 à 2016 et je n'ai traité que 5 mois de l'année 2010 en 14 semaines. Le traitement de l'ensemble du fonds demandera encore plusieurs mois voire années avant que l'ensemble du fonds soit disponible aux utilisateurs. Ce fonds constituera un terrain de recherche important pour la recherche historique locale ou l'étude des télévisions locales. Le traitement du fonds sera assuré par les prochains stagiaires de la délégation qui devront grâce aux documents internes à la chaîne LCM et grâce aux documents que j'ai produit, étendre les connaissances et le traitement du fonds. Par ailleurs, j'ai été confronté à de nombreux problèmes d'ordre matériel, liés à l'état ancien du matériel. En effet, tout d'abord l'imprimante reliée à l'application « Sumoco » a cessé de fonctionner dès le début du stage. Ainsi avec le cadre technique, nous avons démonté la machine et essayé de réparer sans réussite. Malgré l'échec de cette tâche, j'ai pu observer le travail du technicien et le seconder. J'ai également eu de nombreux problèmes avec le magnétoscope DSR 2000p servant à visionner les DVCAM. Le magnétoscope n'est pas récent et a des problèmes de moteur à l'éjection de la cassette. Une fois de plus, j'ai pu démonter le

¹²⁶ Onglet du logiciel « Totem » utilisé pour ajouter des descripteurs à une archive. Ces descripteurs permettent de définir le contenu d'un document pour le retrouver dans une base de données.

magnétoscope avec le technicien qui m’a alors expliqué en détail le fonctionnement de la machine. Ce problème était assez récurrent et j’ai pu par la suite en toute autonomie démonter la machine afin de régler le problème. Ainsi malgré ces problèmes de matériel, j’ai réussi à travailler avec une polyvalence certaine compte tenu des tâches qui m’incombaient. Enfin, un seul problème matériel fut handicapant dans mon travail. Il s’agit d’un problème de connexion au serveur de l’INA sur mon poste de travail. L’application « Totem » est installée sur mon poste de travail cependant l’application « Sumoco » est quant à elle sur un serveur sur lequel je devais me connecter. Son accès a été indisponible depuis mon poste pendant deux semaines. J’ai dû échanger par téléphone avec l’assistance informatique de l’INA sans pouvoir trouver de solution. Après deux semaines de recherche du problème par les équipes de l’assistance informatique, mon accès a pu être rétabli et j’ai pu reprendre la liaison entre le matériel et les notices. En conclusion, malgré des problèmes techniques et un fonds particulièrement important, ce stage a été très formateur et m’a permis de mieux comprendre les missions de l’INA et la portée du travail de responsable documentaire.

Conclusion

Ce stage d'une durée de 14 semaines dans le cadre du Master 2 Valorisation Numérique du Patrimoine : archives et images, m'a permis de m'immerger dans les missions, les problématiques et les enjeux de l'INA. J'ai ainsi pu mieux comprendre le rôle d'une délégation régionale de l'INA et observer de près ces missions. Ma mission principale de description d'un fonds documentaire a été réellement enrichissante. J'ai appris à utiliser différentes applications et j'ai renforcé des compétences de synthèse. La rédaction des notices ainsi que des documents historiques ou la création d'une grille des programmes ont été des tâches variées et instructives d'abord par le défi de la recherche et par le fait que les informations ont été relativement difficiles à trouver. Cette immersion sur le terrain, directement en lien avec les archives, m'a permis de consolider mes connaissances et de développer des compétences. J'ai particulièrement apprécié le travail en autonomie tout en restant sous l'autorité de ma tutrice. Le visionnage du fonds s'est révélé passionnant avec des nombreux liens à faire avec l'actualité de 2022 notamment la grippe A ou encore les élections régionales. J'ai pu m'immerger pleinement dans le travail journalistique réalisé par la chaîne dans l'année 2010, et prendre les actualités et informations de la chaîne avec un certain recul. L'aspect technique de la création des notices et de l'utilisation des applications m'ont permis de me confronter à l'utilisation de l'informatique dans le traitement des archives.

Concernant l'objectif de départ, qui était le traitement et l'indexation d'un fonds d'archive, je n'ai malheureusement pas complété entièrement l'indexation sur demande de ma tutrice, pour visionner un maximum d'archives et les décrire par la suite. Cependant j'ai pu décrire le fonds, mettre en place une cotation, relier les matériels aux notices et produire des documents complémentaires permettant de mieux cerner et comprendre le fonds.

Ce stage m'a donc apporté une grande plus-value professionnelle, notamment par l'acquisition de compétences et connaissances dans le milieu du traitement des archives mais pas uniquement. J'ai ainsi développé des connaissances en matière de communication d'évènement par la participation à la création de vidéo de promotion de l'évènement. Enfin d'un point de vue technique, j'ai appris à utiliser des applications de l'INA et à démonter un

magnétoscope. Je considère qu'il s'agit de compétences professionnalisantes dans un environnement comme une délégation régionale.

En conclusion, le stage a été très stimulant et professionnalisant par la grande autonomie dont j'ai bénéficié et par la polyvalence des tâches confiées. D'un point de vue personnel, j'ai rencontré plusieurs personnes dont d'anciens membres de la délégation, j'ai ainsi bien appréhendé les attendus du métier. En parallèle du stage, j'ai eu l'opportunité de candidater et d'être retenu dans le cadre d'un contrat de recherche à l'Université d'Aix-Marseille, portant sur une étude de la valorisation des activités de l'Université. J'envisage de me rendre de nouveau à l'INA pour étudier les fonds concernant ce sujet. Cette recherche sera simplifiée par l'acquisition des connaissances sur le fonctionnement et l'utilisation des outils de la délégation.

Bibliographie du rapport de stage

I- Ouvrages généraux sur les médias

- BALLE Francis, *Les médias*, Paris, PUF, 2004 [2ème ed. 2006]
- BALLE Francis, *Médias et sociétés : édition, presse, cinéma, radio, télévision, internet*, Paris, Domat Politique, 1994
- BALNAVES Marc, DONALD James et HEMELRYK DONALD Stéphanie, *Atlas des médias dans le monde*, traduit de l'anglais par PERRON Sylvie, Paris Autrement, 2001
- BARBIER Dominique et BERTHO-LAVENIR Catherine, *Histoire de médias : De Diderot à Internet*, Paris, U Histoire, 2009 [3ème ed. 2009]
- BROCHAND Christian, *Histoire générale de la radio et de la télévision en France*, Paris, La Documentation Française, volume 2, Paris, La documentation Française, 1994
- BROCHAND Christian, *Histoire générale de la radio et de la télévision en France*, Paris, La Documentation française, volume 3, 2006
- CLAVERT Frédéric et NOIRET Serge, *L'histoire contemporaine à l'ère numérique*, Bruxelles, Peter Lang, 2013
- DELPORTE Christian et D'ALMEIDA Fabrice, *Histoire des médias en France de la Grande Guerre à nos jours*, Paris, Champs Histoire, 2003 [ed. 2010]
- FERRO Marc et PLANCHAIS Jean, *Les médias et l'histoire*, Paris, Seuil, 1997
- HOOG Emmanuel, *La télé, une histoire en direct*, Paris, Découvertes Gallimard, 2010
- JEANNENEY Jean-Noël, *Une histoire des médias des origines à nos jours*, Paris, Seuil, 1996
- JOST François (dir.), *50 fiches pour comprendre les médias*, Paris, Bréal, 2012
- JOST François, *Comprendre la télévision et ses programmes*, Paris, Armand Colin, 2005 [2ème ed. 2009]
- JOST François, *Introduction à l'analyse de la télévision*, Paris, Ellipses, 1999 [3ème ed. 2007]

II- Ouvrages généraux sur les archives

- BABELON Jean-Pierre, BARBAT Philippe, FORLIVESI Luc, GILLET Patricia, *Les Archives Mémoire De La France*, Paris, Découvertes Gallimard Culture Et Société, 2008.
- COEURE Sophie, DUCLERT Vincent, *Les Archives. Nouvelle édition Entièrement Refondue Et Mise à Jour ed. 2011*, Paris, Collection Repères Histoire, 2011
- FARGE Arlette, *Le goût de l'archive*, Paris, Seuil, 1989
- FAVIER Jean, *Les Archives, 7e édition Mise à Jour ed. 2001*, Paris, Que Sais-je ? 2001.
- GUIET-MONCHAL, Lydiane (dir.), *Abrégé d'archivistique Principes et pratiques du métier d'archiviste*, 4ème édition refondue et augmentée, Paris : Association des archivistes français, 2020.

III- Ouvrages spécialisés sur les archives audiovisuelles

- GUYOT Jacques et ROLLAND Thierry, *Les archives audiovisuelles, Histoire, culture, politique*, Paris, Armand Colin, 2011
- LABONNE Sophie, BRAEMER Christine, *Les archives audiovisuelles*, Paris : Association des archivistes français, 2013
- HOOG Emmanuel, *L'INA*, Paris, PUF, 2006
- STOCKINGER Peter, *Les Archives Audiovisuelles Description, Indexation Et Publication*, Paris, Hermès Science Lavoisier, 2011
- STOCKINGER Peter, *Nouveaux Usages Des Archives Audiovisuelles Numériques*, Paris, Hermès Science Lavoisier, 2011

IV- Ouvrages spécialisés sur les archives du web

- BLANDIN Claire, ROBINET François, SCHAFER Valérie (dir.), *Penser l'histoire des médias*, Paris, CNRS Editions, 2019
- GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Editions, 2021
- MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie, THIERRY Benjamin G., *Qu'est-ce qu'une archive du web ?* OpenEdition Press, 2019
- SCHAFER Valérie, *Archives : comment le Web devient patrimoine*, The Conversation, 24 avril 2017

V- Articles

- CLAVERT Frédéric, MULLER Caroline (dir.), « Le goût de l'archive à l'ère du numérique », *La Gazette des archives*, n°253, 2019
- FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016
- OUERFELLI Tarek, « Archives audiovisuelles et valorisation du patrimoine à l'ère du numérique », dans *Les Cahiers du numérique*, 2015

Sitographie

I- Juridique

- Code du Patrimoine : < https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000020905828/ > (consulté le 05/03/22).
- CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).
- Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).
- Loi du 21 juin 1943 : < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006071152/> > (consulté le 05/03/22).
- Loi du 4 août 1974 : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000333539> > (consulté le 05/03/22).
- Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068930/> > (consulté le 05/03/22).

II- INA

- Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).
- Site web de INA sup : < <https://www.INA-expert.com/enseignement-superieur> > (consulté le 05/03/22).
- Site web de INAMédiaPro : < <http://www.INAmediapro.com> > (consulté le 05/03/22).
- Site web de l'INA : < <https://www.INA.fr> > (consulté le 05/03/22).
- Site web délégation INA Méditerranée : < <https://www.INA.fr/institut-national-audiovisuel/en-regions/INA-mediterranee> > (consulté le 05/03/22).

- Site web Madelen : < <https://madelen.INA.fr> > (consulté le 05/03/22).

III- Lieux de consultation PCM

- Site web bibliothèque Ceccano : < <https://bibliotheques.avignon.fr/in/faces/details.xhtml?id=816093b1-a9ef-42d4-8430-3c370fe36fd3> > (consulté le 05/03/22).
- Site web bibliothèque Louis Nucéra : < <https://bmvr.nice.fr> > (consulté le 05/03/22).
- Site web cinémathèque Casa di Lume : < <https://www.casadilume.corsica> > (consulté le 05/03/22).
- Site web de la MMSH : < <https://www.mmsh.fr> > (consulté le 05/03/22).

IV- Projet de valorisation et partenaires

- Projet européen NEDLIB (préservation des publications numériques) : < <https://cordis.europa.eu/project/id/LB5648/fr> > (consulté le 05/03/22).
- Projet Gender Equality Monitor (GEM) : < <https://anr.fr/Project-ANR-19-CE38-0012> > (consulté le 05/03/22).
- Site web CMCA : < <https://cmca-med.org/formation-INA-cmca/> > (consulté le 05/03/22).
- Site web Histoire de la ville de Cannes en vidéo : < <https://sites.INA.fr/ville-de-cannes/focus/chapitre/6> > (consulté le 05/03/22).
- Site web IESA : < <https://www.iesa.fr> > (consulté le 05/03/22).
- Site web Médias et mobilisations en Corse de 1945 à nos jours : < <https://sites.INA.fr/mobilisations-en-corse/tempo/> > (consulté le 05/03/22).
- Site web Primed : < <https://primed.tv> > (consulté le 05/03/22).
- Site web SOS Méditerranée : < <https://www.sosmediterranee.fr> > (consulté le 05/03/22).
- Site web Sudorama : < <https://fresques.INA.fr/sudorama> > (consulté le 05/03/22).

Table des illustrations

Figure 1 : Exemple d'un document fourni par Anne Sophie Maxime, brouillon d'une liste d'émissions classées par genre.....	p. 31
Figure 2 : Exemple d'un document fourni par Anne Sophie Maxime, brouillon d'une grille des programmes de 2009 à 2010.....	p. 32
Figure 3 : Les objectifs 2010 de LCM.....	p. 37
Figure 4 : Croissance de l'audience entre 2005 et 2008.....	p. 38
Figure 5 : Tableau de l'audience cumulée des chaînes locales.....	p. 39
Figure 6 : Tableau de croissance de l'audience de la chaîne LCM entre septembre 2007 et décembre 2008.....	p. 40
Figure 7 : Un magasin d'archives à la délégation marseillaise de l'INA.....	p. 41
Figure 8 : Magnétoscope DSR 2000p servant à la lecture de DVCAM et cassette DVCAM.....	p. 43
Figure 9 : Capture d'écran de l'application « Totem ».....	p. 45
Figure 10 : Onglet « Diffusion » lors de la création d'une notice.....	p. 48
Figure 11 : Onglet « Production » lors de la création d'une notice.....	p. 49
Figure 12 : Onglet « Typologie » lors de la création d'une notice.....	p. 49
Figure 13 : Capture d'écran de la liste d'autorité des thématiques.....	p. 50
Figure 14 : Onglet « Générique » lors de la création d'une notice.....	p. 51
Figure 15 : Classement des folios de travail partagés.....	p. 52
Figure 16 : Application « Sumoco ».....	p. 56
Figure 17 : Chaîne de traitement d'une archive mise en place lors du stage.....	p. 57
Figure 18 : Grille des programmes de la chaîne LCM de janvier à février 2010.....	p. 58

Figure 19 : Grille des programmes de la chaîne LCM de mars à mai 2010.....	p. 59
Figure 20 : Répartition des chroniques de « La Grande émission » (partie 1) entre mars en mai 2010.....	p. 60
Figure 21 : Répartition des chroniques de « La Grande émission » (partie 2) entre mars en mai 2010.....	p. 60

Annexes

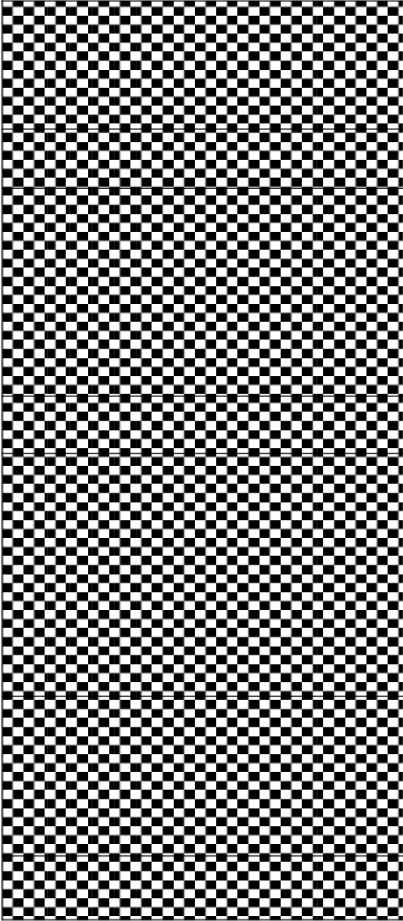
Annexe 1 : Image de la façade de l'INA au 23 rue Guibal, 13003, Marseille

Image d'illustration de la façade de la délégation de l'INA Méditerranée, provenant du site web de l'INA < <https://www.INA.fr/institut-national-audiovisuel/en-regions/centre-de-consultation-marseille> >, consulté le 04/07/22



Annexe 2 : Fiche des postes dans les délégations de l'INA (assistante de direction)

Fiche Fonction : Assistante de direction en région		Emploi (Assistant (e) de direction			
Famille : Administration					
Groupe : Groupe 5					
Finalité : Assurer l'assistantat d'une délégation					
Activités dominantes		Profil de compétences de la fonction			
N°	ACTIVITES DOMINANTES	CONNAISSANCES	SAVOIR FAIRE OPERATIONNELS	COMPORTEMENTS PROFESSIONNELS	METHODS ET OUTILS
		INTITULE	INTITULE	INTITULE	NATURE
1	<u>Assister le délégué dans la gestion des activités de la délégation</u> - Réaliser les tâches classiques de secrétariat pour sa direction : accueil téléphonique et physique des visiteurs, gestion des courriers, prise de rendez-vous, gestion des agendas, organisation de réunions, missions - Instruire et traiter les dossiers confiés : de gestion et suivi des Budgets notamment par l'élaboration de tableaux de bord, par la transmission des éléments de facturation, de gestion administrative (contrats, conventions..) de gestion de l'indemnité de la délégation	Culture Générale Langues étrangères	Savoir formaliser en se conformant à un modèle préalable	Adaptabilité Autonomie Capacité de synthèse	Bureautique Applications spécifiques IMA Procédures budgétaires
2	<u>Organiser et coordonner les informations internes et externes liées au fonctionnement de la direction</u> - Coordonner les activités des différents services de la délégation (reporting, tableaux de bord, suivi d'activités...) - Mettre à jour la base contacts de la hiérarchie	Métiers de l'entité Organisation de l'entité Organisation interne Produits /services Techniques de secrétariat		Capacité rédactionnelles Diplomatie Disponibilité Expression aisée et fluide Gestion des priorités Confidentialité Reactivité Sens du contact	Tableaux de bord TIC
3	<u>Prendre en charge des missions de communication</u> - Coopérer à l'organisation des événements et en assurer leur suivi - Elaborer des outils pour la délégation en lien avec la direction de la communication - Contribuer à l'animation des nouveaux vecteurs de communication				
4	<u>Réaliser le traitement de dossiers à caractère commercial, administratif et culturel dans un objectif de suivi de clientèle</u> - Centraiser pour la délégation l'ensemble des propositions commerciales et dossiers en cours de traitement - Tenir à jour les tableaux de synthèse et de suivi d'activités - Finaliser les offres commerciales (devis, contrats) - Traiter les dossiers d'archives sur références				



Annexe 3 : Fiche des postes dans les délégations de l'INA (déléguee régionale)

Fiche Fonction : Délégué(e) régional(e)		Emploi (e) Chef de service Niveau 2			
Famille :	Toutes familles				
Groupe :	Groupe 8				
Finalité :	Assurer la représentation de l'Ina en région et y déployer les missions et activités.				
Activités dominantes		Profil de compétences de la fonction			
N°	ACTIVITES DOMINANTES	CONNAISSANCES	SAVOIR FAIRE OPERATIONNELS	COMPORTEMENTS PROFESSIONNELS	METHODES ET OUTILS
		INTITULE	INTITULE	INTITULE	NATURE
1	<u>Assurer et développer le rayonnement en région des activités de l'Ina</u> - Représenter et développer la présence de l'Ina auprès des partenaires potentiels des régions : pouvoirs publics, collectivités locales, institutions régionales et locales, diffuseurs régionaux et locaux, entreprises, professionnels de l'audiovisuel, monde de l'université et de la recherche.	Culture audiovisuelle Environnement concurrentiel	Savoir convaincre et/ou négocier Savoir réguler en actionnant des leviers connus pour obtenir un résultat attendu	Autonomie Capacité à arbitrer et décider	Accords collectifs d'entreprise Bureautique
2	<u>Contribuer à la définition de la stratégie de l'Ina en région</u> - Décliner les missions définies par sa direction - Veiller à la cohérence des actions menées au sein des différentes régions de l'Ina	Langues étrangères Métiers de l'entité Organisation de l'entité Produits / Services		Capacité à encadrer une équipe Capacité d'analyse Capacité de synthèse Etre force de proposition	Applications spécifiques INA Bases documentaires Cahier des charges
3	<u>Encadrer et animer l'équipe</u> - Garantir la réalisation des actions de collecte, de conservation, de traitement des fonds, de communication et de valorisation	Stratégie de l'Ina		Sens du contact Sens du résultat Vision prospective Sens du service clients	Procédures liées à l'activité TTC
4	<u>Concevoir, proposer et piloter des projets de développement à caractère commercial, culturel et/ou éducatif</u> - Planifier l'activité, identifier et affecter les ressources humaines et matérielles - Définir et respecter les priorités, coordonner les actions au sein de son unité et avec les autres unités - Anticiper les changements, fixer les objectifs à moyen terme et planifier les actions	Techniques budgétaires Techniques de communication Techniques de gestion de projets			
5	<u>Mener des études et expertises sur tous les sujets relevant de l'activité en région</u>				
6	<u>Assurer la gestion administrative et le reporting à sa direction</u>				

Annexe 4 : Fiche des postes dans les délégations de l'INA (responsable documentaire)

Fiche Fonction : Responsable documentaire en région					Emploi (en région)	Responsable documentaire
Famille : Documentation						
Groupe : Groupe 6						
Finalité : Assurer l'ensemble de l'activité documentaire en région.						
Activités dominantes					Profil de compétences de la fonction	
N°	ACTIVITES DOMINANTES	CONNAISSANCES INTITULE	SAVOIR FAIRE OPERATIONNELS INTITULE	COMPORTEMENTS PROFESSIONNELS INTITULE	METHODES ET OUTILS NATURE	
1	<u>Assurer la gestion des données et techniques documentaires</u>	Culture audiovisuelle	Savoir contrôler ou effectuer un diagnostic avec une méthode préalable	Autonomie	Applications spécifiques INA	
	- Collecter, décrire, indexer, valoriser, structurer, éditorialiser et communiquer les fonds d'archives multimédia sur tout type de supports (Radio, Télé, Photo, Web, papier...)	Culture générale	Savoir mettre en œuvre des moyens appropriés pour atteindre des objectifs	Capacité d'analyse	Bases documentaires	
	- Coopérer, le cas échéant, à l'expertise des fonds	Métiers de l'entité		Capacité de synthèse	Bureautique	
2	<u>Définir et/ou coordonner des procédures, une méthodologie et/ou un suivi de productions et en garantir l'aboutissement.</u>	Organisation de l'entité		Capacités rédactionnelles	Procédures liées à l'activité	
	- Elaborer et suivre les méthodologies adaptées au domaine d'application	Produits / Services		Etre force de proposition	TIC	
3	<u>Rechercher et/ou proposer et/ou commercialiser des images et/ou sons aux clients internes et externes en tenant compte des contraintes juridiques et commerciales.</u>	Techniques de communication		Organisation, rigueur		
4	<u>Participer au développement régional de l'ina par des actions de communication et de valorisation des fonds d'archives multimédia auprès des clients et/ou partenaires.</u>	Techniques documentaires		Respect des délais		
5	<u>Utiliser des outils collaboratifs et/ou les médias réseaux</u>			Sens du contact		
5	<u>Accueillir et accompagner les usagers pour la consultation des collections Ina</u>			Sens du service clients		
	- Animer les présentations des outils et ressources					
	- Apporter une assistance aux usagers en matière de consultation des archives Ina et/ou d'utilisation des outils documentaires dédiés					
6	<u>Collaborer à des projets d'ingénierie documentaire et/ou patrimoniale</u>					
7	<u>Effectuer le reporting de son activité</u>					
8	<u>Prendre part aux réunions relevant de son domaine de compétences</u>					
	- Etre une personne ressource dans les réunions relevant de son périmètre de responsabilités					
9	<u>Effectuer une veille documentaire</u>					
10	<u>Prendre en charge, le cas échéant, des interventions pédagogiques</u>					

Annexe 5 : Exemple de deux notices rédigées lors du stage

Champs à modifier :

Titre propre : LCM Le journal : émission du 04 janvier 2010

Date de diffusion : 04/01/2010

Durée : 00:14:02

Genre : Journal télévisé

Thématique : Information

Générique :

Sahiri, Igor (pre)

Poncet Sylvain (jou)

Dallest Jacques (par)

Reverdy Olivier (par)

Hermann Daniel (par)

Diard Eric (par)

Leccia Paul (par)

Résumé :

Attaque d'un fourgon blindé de la société Sazias à Gémenos, itw d'un témoin du braquage, itw David-Olivier REVERDY syndicat de police « Alliance », itw Jacques DALLEST procureur de la République de Marseille.

Grève de la police contre le projet de réforme de la police d'agglomération

Vol d'une toile de Degas « Les Choristes » au musée Cantini. En plateau itw Daniel

HERMANN adjoint au Maire de Marseille chargé de la culture. Il évoque le succès

l'exposition « De la scène au tableau », itw habitantes de Carry-le-Rouet, itw Eric DIARD maire UMP de Sausset-les-Pins.

Brève : retard des TER de la SNCF dans le Vaucluse (grève, neige et acte de vandalisme).

Marseille capitale européenne de la culture en 2013 et le vol de la toile « les Choristes ».

Décès de la chanteuse Lhasa de Sela.

Mercato d'hiver de l'OM, départ supposé de Mathieu Valbuena.

Bain du nouvel an au cercle des nageurs de Marseille, itw Paul LECCIA président du cercle des nageurs de Marseille, itw d'adhérents au cercle des nageurs.

Micro-trottoir : les résolutions du nouvel an des Marseillais.

LCM Plus 04/01/2010

Durée 00:14:19

Titre propre : LCM Plus : émission du 04 janvier 2010

Date de diffusion : 04/01/2010

Genre : Journal télévisé

Thématique : Information

Générique :

Sahiri. Igor (pre)

Poncet. Sylvain (jou)

Dallest. Jacques (par)

Maxime. Anne-Sophie : Rédactrice en chef

Icard. Simon : Rédacteur en chef adjoint

Bonnard Florent : Chef d'édition

Yann Dieuaide : Chef d'édition

Christophe Liotard : Réalisation

Guillaume Gay Ingénieur du son

Cordeau Anne : Scripte

Présentation de la librairie "Maupeti", Geneviève GIMENO libraire fait découvrir le livre "RU" de Kim THUY, Gaëlle FARRE libraire présente l'ouvrage "Shéhérazade" de Béatrice FONTANEL.

Programme court fiction humoristique, intitulé "Dédé le Fachomako".

En plateau itw Mireille BENEDETTI présidente de l'association "La Ciotat, il était une fois".

Elle évoque le tissu associatif à La Ciotat et le festival.

Présentation d'un institut de massage ", itw Magali CAMPAN masseuse.

Annexe 6 : Listes des intervenants réguliers de la chaîne LCM

SAHIRI, Igor (présentateur)
DEPETRIS, Marina (journaliste)
PONCET, Sylvain (journaliste)
GALDON, Alexandra (présentatrice)
BLARDONE, Guillaume (présentateur)
BLISSON, Yves (présentateur et chroniqueur économique)
BONNARD, Florent (journaliste)
DIEUAIDE, Yann (journaliste)
TRESOR, Thierry (présentateur)
AUBERT, Céline (présentatrice)
CANUTI, Romain (journaliste)
GODARD, Hervé (journaliste)
FLOUTIER, Rosalie (présentatrice)
BLANC, Alexandra (journaliste)
OUDJANE, Karim (journaliste sportif)
FONTAN, Fanny (journaliste)
QUIOC, Margaïd (journaliste)
LIMBOUR, Anne (chroniqueuse culinaire)
FIORUCCI, Romain (rédacteur en chef)
LEGOUIT, Smahane (journaliste)
GAY, Guillaume (ingénieur du son)
MAXIME, Anne-Sophie (rédactrice en cheffe)
CORDEAU, Anne (scripte)
LIOTARD, Christophe (réalisateur)
RICCI, Alain (responsable d'édition)

Partie II :

Mémoire théorique

L'archivage du web à l'Institut National de l'Audiovisuel :
entre déploiement technologique, enjeux et
regards historiques



Logo de l'Institut National de l'Audiovisuel :

Image provenant du site web de l'INA (annexe 1)

Introduction

« En France comme ailleurs, le web est devenu un espace de mise en visibilité du passé à travers de multiples dispositifs mémoriels prenant des formes diverses et émanant d'acteur.rice.s varié.e.s, de l'internaute à l'institution muséale. »¹²⁷

Aujourd'hui le web et son utilisation encadrent nos pratiques culturelles, sociales et économiques. Son utilisation est omniprésente au quotidien. Il convient de s'interroger sur la mémoire du web, la diffusion de ses mémoires ainsi que la fabrique d'archives du web. Le web se modifie en permanence et en ce sens, un archivage régulier et complet du paysage du web doit être organisé pour permettre une mémoire collective et engendrer des recherches scientifiques, en histoire, sociologie, psychologie, économie ou tout autre domaine d'application du web. Le web régit nos pratiques, il doit donc être étudié comme une archive à part entière et peut à lui seul éclairer et expliquer des phénomènes du XXIème siècle. Pour pouvoir étudier le web, il faut avant toute chose un archivage cohérent et riche, englobant un maximum des facettes du web, allant des réseaux sociaux aux sites institutionnels en passant par les sites des géants du numérique que sont Apple, Amazon ou encore Microsoft. Cet archivage ne peut se faire avec les techniques conventionnelles de l'archiviste, il convient alors d'envisager un déploiement technique spécifique à cet archivage.

Tout d'abord, il paraît pertinent de définir l'archive et l'archivage. L'archive fait référence à la fois aux bâtiments d'archives, à un service administratif mais également aux documents. Ainsi le code du patrimoine modifié par la loi du 7 juillet 2016 relative à la liberté de la création, à l'architecture et au patrimoine produit la définition suivante pour une archive :

« Ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme ou leur support, produits ou reçus par toute

¹²⁷ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Editions, 2021, p. 9

personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice et leur activité »¹²⁸.

Par cette définition, le code du patrimoine englobe les archives du web dans le terme d'archive. De plus en plus d'archives sont numériques, peu importe le support, et sont traitées comme des archives papier. Les archives appartiennent toujours à un contexte de production. La date et l'époque de l'archive n'a pas d'importance car tout peut être considéré comme une archive. Les archives jouent un rôle juridique et administratif, et sont des outils de gouvernance et de gestion. Elles sont aussi des outils pour garantir les droits et la démocratie. Elles jouent un rôle de transparence administrative. Les archives assurent un rôle historique car grâce à elles, les historiens peuvent écrire et étudier les événements passés. Enfin l'aspect patrimonial des archives est à prendre en compte, elles fournissent l'identité du patrimoine voire l'identité d'une nation, c'est notamment le cas pour la France avec la « Déclaration des Droits de l'Homme et du Citoyen »¹²⁹ de 1789 qui est un symbole français marqueur de l'identité nationale. Les archives forment et documentent les objets de la civilisation. On peut les analyser comme la sauvegarde de la mémoire collective et personnelle. Il est à noter que le fonds d'archives est toujours un ensemble cohérent d'archives. On ne constitue pas de fonds artificiellement, nous verrons qu'il en va de même pour les archives web. Le fonds doit provenir d'un producteur. La notion de respect est énoncée par Natalis de Wailly chef de la Section administrative des Archives départementales au ministère de l'Intérieur en 1841 par une circulaire qui expliquait le respect du fonds en ces termes :

« Rassembler les documents par fonds, c'est-à-dire réunir tous les titres qui proviennent d'un corps, d'un établissement, d'une famille ou d'un individu, et disposer d'après un certain ordre les différents fonds. Les documents qui ont seulement rapport

¹²⁸ Code du patrimoine modifié par la loi du 7 juillet 2016 relative à la liberté de la création, à l'architecture et au patrimoine, article L211-1 consultable en ligne : <

https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032860025/ >, (consulté le 12/08/22)

¹²⁹ Déclaration des Droits de l'Homme et du Citoyen adopté en 1789. Il s'agit du texte fondateur de la Révolution française instaurant des droits aux individus et les conditions d'applications, consultable en ligne : < <https://www.legifrance.gouv.fr/contenu/menu/droit-national-en-vigueur/constitution/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789> >, (consulté le 12/08/22)

avec un établissement, un corps ou une famille ne doivent pas être confondus avec le fonds de cet établissement, de ce corps, de cette famille... »¹³⁰.

Ainsi cette notion du respect du fonds est primordiale lors de l'archivage. L'archivage se définit comme l'action de conserver et classer des archives. Elle s'est développée dès le 4^{ème} millénaire avec J-C. En effet en 1930 en Mésopotamie, on a découvert des archives dans un palais, dans une salle de conservation comme nous l'indiquent les auteurs Biga Maria-Giovanna, Charpin Dominique et Durant Jean-Marie dans la « *Revue d'assyriologie et d'archéologie orientale* »¹³¹. Ainsi dès qu'un pouvoir s'installe, il a besoin d'archives pas asseoir sa légitimité. Des découvertes similaires ont eu lieu en Égypte et à Rome. Ce n'est qu'avec Philippe Auguste que les archives du Royaume vont être centralisées et incorporées dans le trésor des Chartres. Les archives et l'archivage sont des éléments anciens et importants dans la création et la structuration d'un pouvoir politique, économique et social. Il en va de même avec les archives du web qui sont de plus en plus nombreuses dans notre monde interconnecté.

Tout d'abord, il est important de définir le web et plus globalement Internet. Internet est une invention des années 1960 utilisée dans un premier temps pour la communication de l'armée américaine. Pour les auteurs Dufour Arnaud et Ghernaouti-Hélie Solange¹³² dans l'ouvrage « *Qu'est-ce qu'Internet ?* », ils indiquent qu'internet est un réseau de télécommunication composé d'un grand nombre de réseaux interconnectés. Internet est alors qualifié de « réseau des réseaux » et offre une mise en relation et des transferts de données entre entités distantes. Internet est donc un réseau qui permet d'accéder à des ressources, de transférer des données et de réaliser des services. Le web (World Wide Web) est une invention du physicien au CERN (Conseil européen pour la recherche nucléaire) Tim Berners-Lee datant de 1989. Internet repose sur un protocole TCP/IP et un accès par un fournisseur

¹³⁰ DUCHEIN Michel, Le respect des fonds en archivistique : principes théoriques et problèmes pratiques (*La Gazette des Archives*, 97, 2e trimestre 1977, p. 71-96). Dans : *La Gazette des archives*, 1992. Études d'archivistique 1957-1992, p. 9-34.

¹³¹ BIGA Maria-Giovanna, CHARPIN Dominique, DURAND Jean-Marie, « Paolo Matthiae, une appréciation », *Revue d'assyriologie et d'archéologie orientale*, 2012/1 (Vol. 106), p. 1-4, consultable en ligne : < <https://www.cairn.info/revue-d-assyriologie-2012-1-page-1.htm> >, (consulté le 12/08/22)

¹³² DUFOUR Arnaud, GHERNAOUTI-HELIE Solange, « Qu'est-ce qu'Internet ? », dans : Arnaud Dufour éd., *Internet*. Paris cedex 14, Presses Universitaires de France, « Que sais-je ? », 2006, p. 5-17, consultable en ligne : < <https://www.cairn.info/--9782130554714-page-5.htm> >, (consulté le 12/08/22)

d'accès à Internet, le web lui repose sur un système hypertexte et nécessite un navigateur comme Chrome, Firefox ou encore Safari. Ainsi Internet est l'infrastructure de réseaux sur laquelle repose l'application web. Une archive web est alors une archive provenant de l'application web. Cependant une archive web ne peut en aucune façon être la copie d'un web passé notamment par les différences techniques des interfaces. Pour les auteurs de l'ouvrage « *Qu'est-ce qu'une archive du web ?* »¹³³, l'archive web est défini comme « un objet singulier, interactif, fluide et non figé. » Si des éléments rapprochent l'archive web d'une archive papier d'autres l'en éloignent. En effet, de la même façon qu'une archive papier, l'archive web se soumet à l'enjeu de l'exhaustivité et à l'authenticité des documents. La masse documentaire peut aussi s'approcher des archives papier contemporaines. De même l'aspect « éphémère » des archives du web, qui peut notamment poser des problèmes lors de la collecte, peut aussi s'apparenter à des archives papier comme les tracts politiques ou encore aux archives de cultures vernaculaires. Cependant l'archive web a des différences comme le montre l'auteur Bachimont Bruno qui voit l'archive web comme une collection plutôt qu'une archive, par le fait que l'archive permet de renseigner et de prouver les activités d'un producteur, alors que l'archive web, elle n'est pas forcément liée aux activités du producteur mais plutôt à la production d'idées. Il exprime cela en ces termes dans son ouvrage « *Patrimoine et numérique : technique et politique de la mémoire* » :

« Lorsque la constitution de l'ensemble documentaire obéit à une intentionnalité et un arbitraire lié non à la causalité de l'événement mais à la production des idées, on quitte le terrain de l'archive pour rejoindre celui de la bibliothèque »¹³⁴

L'archive web avec un statut hybride est accompagnée d'une chaîne de traitement différent des archives papiers. L'archive web est donc un objet qui évolue sous les différentes modalités de collecte, de la profondeur de collecte et des changements de techniques. L'archivage du web si jeune soit-il a déjà connu des évolutions fortes comme le souligne Merzeau Louise lors de l'assemblée générale de l'International Internet Preservation

¹³³ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 10

¹³⁴ BACHIMONT Bruno, *Patrimoine et numérique : technique et politique de la mémoire*, Bry-sur-Marne, INA-Éditions, 2017, p. 23

Consortium (IIPC)¹³⁵ où elle indique qu'après un archivage sur un modèle documentaire, l'archivage du web s'est transformé en un échantillonnage ciblé.

Deux entités en France organisent cet échantillonnage ciblé des archives du web, il s'agit de l'INA et de la BNF (Bibliothèque Nationale de France). Ces deux entités, grâce au dépôt légal du web, archivent le web français et permettent aux chercheurs des ressources conséquentes pour étudier l'histoire et l'évolution d'internet, ainsi que les phénomènes liés aux grands événements qui font l'actualité. Les deux institutions la BNF et l'INA collectent les archives du web français en choisissant les périmètres de collecte. Il s'agit donc d'une collecte orientée par des choix et il ne s'agit pas de l'ensemble du web français. Il y a alors par ces choix un impact sur la représentativité de certaines archives.

Ainsi durant l'ensemble du mémoire, nous nous efforcerons de nous questionner au sujet du traitement des archives du web, des modifications juridiques entraînant la mise en place d'un dépôt légal du web, mais aussi de mettre en perspective les avancées technologiques comme la mise en place de chaînes de traitement automatisé ou le format des archives. L'ensemble de ces points sera abordé. Après avoir traité l'archivage du web, il convient de nous interroger sur l'utilisation de ces archives, la mise en consultation pour des publics différents et également sur les projets de valorisation autour des archives du web. Ces projets de valorisation de ces archives portés souvent par des historiens peuvent nous questionner sur l'enjeu historique autour des archives du web.

Ainsi lors de cette étude, nous nous questionnerons sur la manière dont l'INA traite ces archives du web, de la collecte en passant par le dépôt légal du web, à la diffusion et la valorisation de ces archives ? Quelle est la chaîne de traitement des archives du web à l'INA ? Quels sont les enjeux historiques derrière l'utilisation de ces archives ? Et enfin nous nous demanderons quelles sont les évolutions de l'archivage du web de 1990 à aujourd'hui ?

Pour répondre à ces questions, nous analyserons dans un premier chapitre l'évolution de l'archivage web des années 1990 à aujourd'hui. Nous verrons alors l'évolution numérique de l'INA, tout d'abord en analysant la structure actuelle de l'INA, ainsi que le virage numérique

¹³⁵ International Internet Preservation Consortium, consultable en ligne : < <https://netpreserve.org> >, (consulté le 12/08/22)

de la structure et la place de l'INA avec le web français. Nous étudierons par la suite l'évolution de l'archivage du web, en passant par le dépôt légal, les premiers outils d'archivage et approches de l'archivage. Nous aborderons les évolutions juridiques et de son périmètre. Nous analyserons la fragilité d'une archive web, ses limites ainsi que les enjeux de gouvernance qui en découle. Notre deuxième chapitre mettra en lumière la mise en pratique de l'archivage du web français. Nous définirons les principes et fonctionnements de l'archivage du web en précisant le périmètre d'archivage automatisé et ciblé. Puis nous nous pencherons sur les outils pour collecter et indexer en terminant par les outils pour stocker et pérenniser les archives. Nous étudierons les protocoles technologiques mis en place pour archiver le web comme les robots de collectes, l'utilisation des métadonnées, le format « DAFF » propriétaire de l'INA, la collecte des UGC (User Generated Content) et de Twitter. Enfin nous nous pencherons sur l'aspect documentaire de l'archivage web avec la mise en place d'une veille, la consultation des archives et la mise en place de notice. Dans un troisième et dernier chapitre, nous analyserons la diffusion des archives du web, d'abord par une mise en consultation des archives en étudiant les publics et les lieux de consultation, ensuite en déchiffrant la recherche sur la base de données de l'INA et en mettant en perspective les multiples outils disponibles pour la recherche. Nous considérerons les valorisations des archives du web notamment en rupture avec une valorisation similaire aux archives papiers. Pour terminer, nous déchiffrerons les enjeux historiques derrière ces archives et notamment la construction d'un regard d'historien sur les archives du web et la mise en place d'un nouveau rapport à l'archive.

Chapitre 1 : Une évolution de l'archivage web

des années 1990 à aujourd'hui

L'archivage du web a commencé dès les années 1990, soit dès la création du web en 1989 par Tim Berners-Lee. Ce chapitre vise à étudier l'évolution de l'archivage web, en analysant l'évolution numérique de l'INA amorcée dès 1999. Ce virage explique aujourd'hui la structure de l'INA que nous sonderons en détail ainsi que la prise de conscience sur l'importance d'archiver le web français. Dans une seconde partie, nous étudierons l'évolution de l'archivage du web depuis la fondation de « Internet Archives » en 1996 par Brewster Khale, en abordant la « Wayback Machine »¹³⁶ créée de nouveau par Brewster Khale en 2001 jusqu'à la charte de l'Unesco¹³⁷ portant sur la conservation du numérique en 2003 et la mise en place d'un dépôt légal du web en 2006. Ainsi nous analyserons le contexte juridique de la loi DADVSI¹³⁸ mise en place en 2006, son périmètre et ses acteurs. Enfin dans une dernière partie, nous nous questionnerons sur le l'archivage du web comme le miroir de l'internet français. Ainsi nous verrons que l'archivage est en continuelle évolution, que les enjeux de gouvernance sont bien présents dans la manière d'archiver le web. Et pour finir ce premier chapitre, nous aborderons les limites de l'archivage comme la volatilité des contenus sur le web et la durée de vie moyenne des URL.

¹³⁶ Site web de la « Wayback Machine » : < <https://archive.org/web/> >, (consulté le 12/08/22)

¹³⁷ Charte portant sur la conservation du numérique, consultable en ligne : < http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html >, (consulté le 12/08/22)

¹³⁸ Loi du 1 août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information (DADVSI), consultable en ligne : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> >, (consulté le 12/08/22)

Partie 1 : L'évolution numérique de l'INA

A- Structure actuelle de l'INA

Si le cadre juridique de l'INA a peu évolué depuis sa création en 1974¹³⁹ à la suite du démantèlement de l'ORTF, l'institut s'est considérablement agrandi et structuré autour de différentes directions. Cette évolution de la structure de l'INA s'inscrit dans la visée de répondre toujours plus efficacement aux missions de l'entreprise. L'INA définit elle-même ses missions en ces termes :

« Les missions de service public de l'INA, définies par l'article 49 de la loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication¹⁴⁰ modifiée, sont aujourd'hui les suivantes :

- L'INA assure la conservation des archives audiovisuelles des entreprises du secteur audiovisuel public ainsi que de toute entreprise, désireuse de lui confier la conservation et l'exploitation de ses archives audiovisuelles.
- Il exploite ses collections d'archives en éditant et en produisant des contenus originaux pour tous les supports ainsi qu'en mettant ces collections à disposition des chercheurs, des créateurs, des producteurs, des autres médias, des institutions culturelles, etc.
- Il assure le dépôt légal de la télé et de la radiodiffusion et, avec la Bibliothèque nationale de France, du web.
- Il contribue à l'innovation et à la recherche dans le domaine de la production et de la communication audiovisuelle.

¹³⁹ Loi du 4 août 1974 : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000333539> > (consulté le 05/03/22).

¹⁴⁰ Loi du 30 septembre 1986 relative à la liberté de communication < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068930/> > (consulté le 05/03/22).

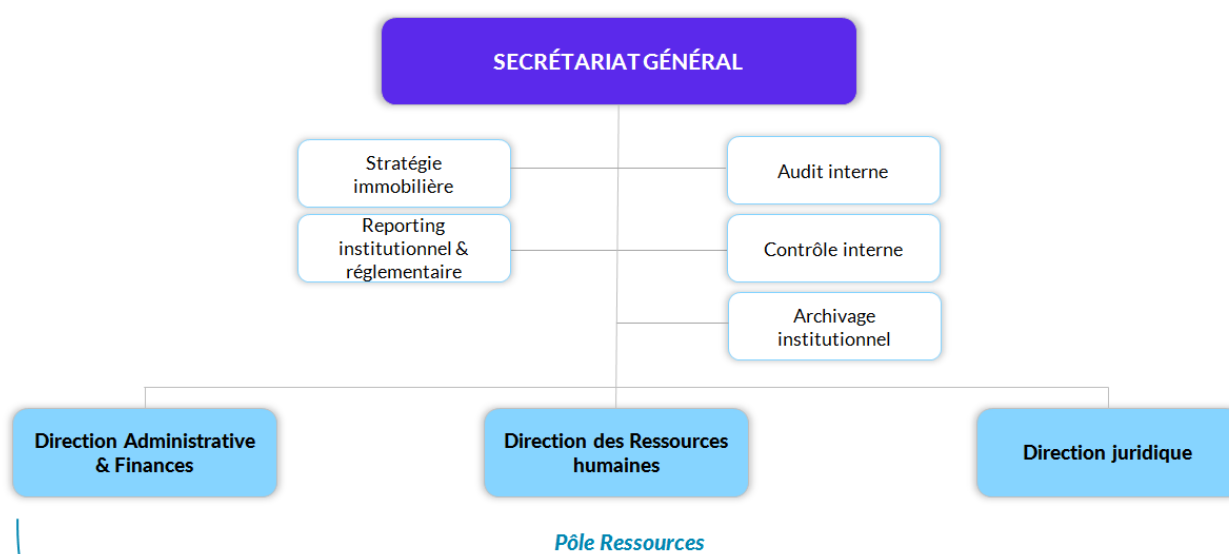
- Il contribue à la formation continue et initiale et à toutes les formes d'enseignement dans les métiers de la communication audiovisuelle. »¹⁴¹

Ainsi ces missions sont fixées dans un cahier des charges fixé par décret et validé en amont par le Conseil Supérieur de l'Audiovisuel (CSA) nouvellement nommé l'Autorité de régulation de la communication audiovisuelle et numérique (ARCOM). L'INA est structuré en trois grandes directions, elle-même sous la direction du président-directeur général Laurent Vallet nommé en conseil des ministres le 20 mai 2015 et renouvelé le 20 mai 2020¹⁴². Le président-directeur général est le garant des relations institutionnelles et extérieures, il définit la stratégie éditoriale et pédagogique ainsi que le développement international. Les trois grandes directions qui régissent l'INA sont le secrétariat général, la direction générale déléguée et la direction générale – campus. Le secrétariat général coordonne et anime un pôle « ressources » composé de trois directions supports : la direction des ressources humaines, la direction juridique et la direction administrative et financière. Ces directions accompagnent de façon quotidienne le pilotage économique, juridique et social de l'entreprise.

¹⁴¹ Site web de l'INA : <https://www.INA.fr/institut-national-audiovisuel/organisation/statuts-et-missions> > (consulté le 05/03/22).

¹⁴² Décret du 20 mai 2020 portant sur le renouvellement du mande de Laurent Vallet, consultable en ligne : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041898341> >, (consulté le 13/08/22)

Fig. 1 : Organigramme du secrétariat général de l'INA¹⁴³



La directrice générale déléguée a la charge du pilotage des priorités stratégiques de l'entreprise, de la poursuite de sa transformation dans le cadre du Contrat d'Objectifs et de Moyens (COM) ainsi que de son affirmation en tant que média patrimonial de service public. Les objectifs du COM ont été signés en mai 2021 et visent à transformer à la fois l'INA mais aussi l'ensemble de l'audiovisuel français. Le COM présente cinq priorités à mettre en avant durant la période 2020-2022 : la culture et la création, la jeunesse et l'éducation, l'information, la proximité et l'Europe et les actions extérieures. Pour mener à bien ses priorités, l'INA a défini des objectifs s'articulant autour de cinq axes :

« Garantir la sauvegarde pérenne, l'enrichissement et la modernisation de la gestion des collections audiovisuelles et web de l'INA au service au renforcement de ses offres et services.

Développer un pôle d'excellence dédié aux métiers de l'audiovisuel et du numérique favorisant l'émergence d'un outil de formation mutualisé au sein de l'audiovisuel public.

Restaurer dès que possible l'équilibre budgétaire et la capacité d'investissement de l'entreprise.

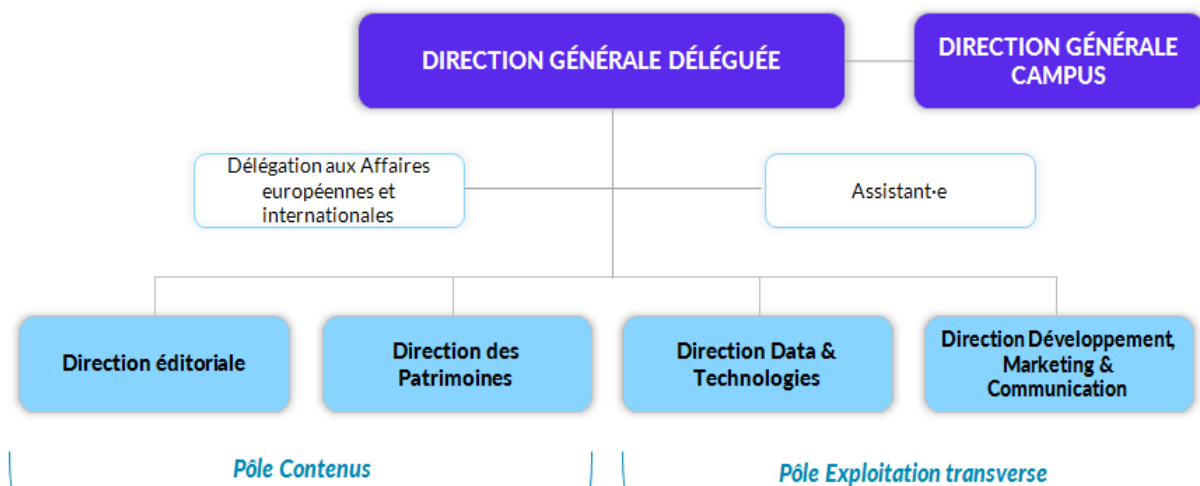
¹⁴³ Document interne à l'INA

Adapter l'organisation et les modes de travail de l'entreprise à la transformation de ses missions à l'ère numérique.

Poursuivre une politique immobilière et foncière ambitieuse et collaborative au service de la modernisation et de la mise aux normes environnementales des sites d'activité INA. »¹⁴⁴

Ainsi la direction générale déléguée doit se charger de mettre en place les stratégies pour mener à bien ses objectifs d'ici la fin de l'année 2022. Pour ce faire elle dispose de deux pôles. Le pôle « contenus » qui a la charge de faire travailler de concert la gestion des collections et la stratégie éditoriale. Ce pôle est divisé entre deux directions : la direction des patrimoines et la direction éditoriale. Le second pôle est celui de « l'exploitation transverse », il s'agit d'un pôle au service de l'ensemble des directions de métiers de l'INA. Ce pôle comprend deux directions : la direction « data et technologies » et la direction « développement marketing et communication ». Il est à noter que cette dernière est la direction régissant les délégations régionales, les rassemblant dans un service nommé « territoires et diversification ».

Fig. 2 : Organigramme de la direction générale déléguée¹⁴⁵

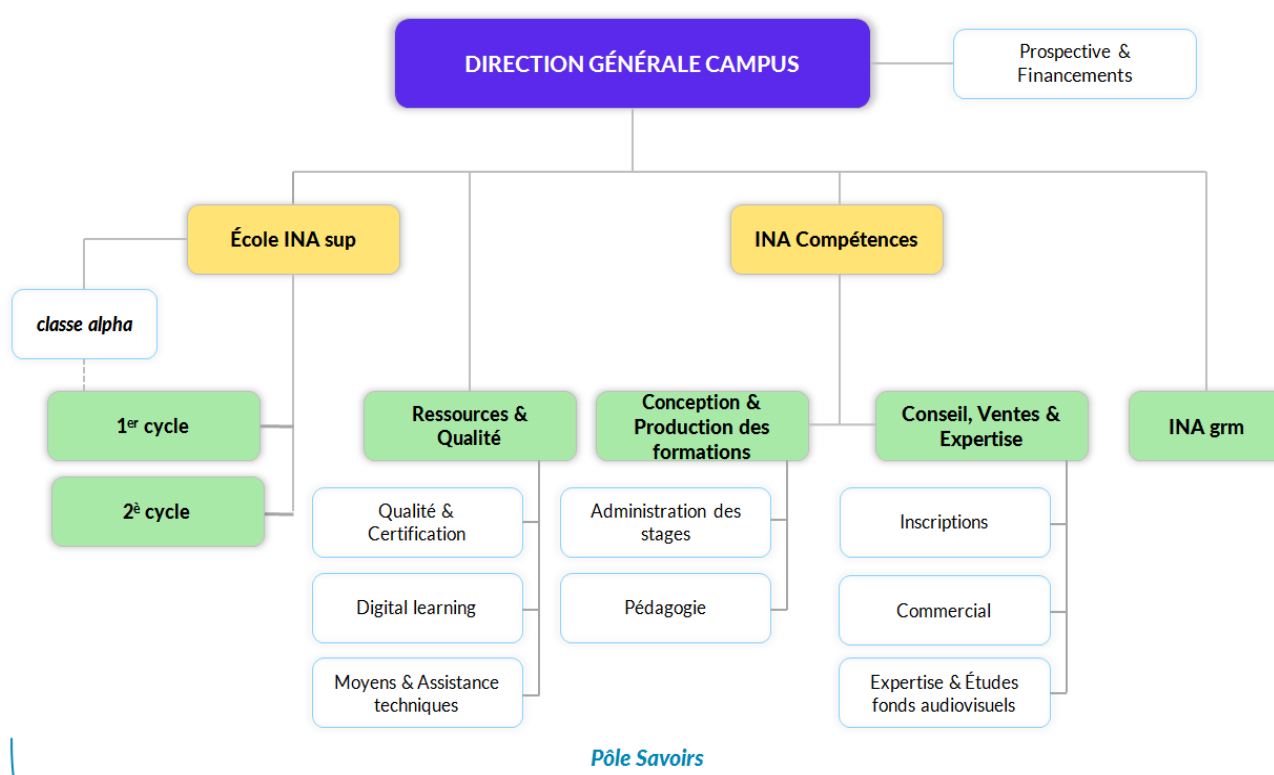


¹⁴⁴ Objectifs du COM5 présenté sur le site web de l'INA, consultable en ligne : < <https://www.ina.fr/institut-national-audiovisuel/organisation/statuts-et-missions> >, (consulté le 13/08/22)

¹⁴⁵ Document interne à l'INA

La dernière grande direction est celle de la direction générale Campus. Elle a pour mission de renforcer le positionnement de l'INA comme étant un établissement pédagogique de référence pour les métiers de l'audiovisuel et les médias numériques. Ainsi la direction regroupe à la fois les activités permettant d'accompagner l'apprentissage en formation initiale pour les étudiants mais aussi en formation continue pour les salariés. La direction regroupe dans le même temps les conseils et l'expertise que l'INA vend ainsi que l'« INA GRM » qui est un groupe de recherche musicale. C'est donc une direction avec des services variés qui sont inclus dans le pôle « savoirs ». Les deux départements les plus importants sont le département « INA SUP » une école rassemblant l'ensemble des volets pédagogiques en formation initiale, et le département « INA compétences », qui est le centre de formation professionnelle réunissant la formation, l'expertise des fonds tiers.

Fig. 3 : Organigramme de la direction générale campus¹⁴⁶



¹⁴⁶ Document interne à l'INA

En conclusion, l'INA a divisé ses activités en direction, département et service pour répondre au mieux aux missions fixées par l'ARCOM mais également à ses propres objectifs. Cette structuration en évolution constante, grâce au plan COM successifs, permet de répondre au mieux à la modification de l'audiovisuel français. Cette organisation pourra être remise en question si la suppression¹⁴⁷ de la redevance télé est effective dans les mois à venir. Enfin, il convient de présenter la direction, le département et les services en charge du dépôt légal du web. Le dépôt légal du web est organisé par la direction « data et technologie » qui est responsable de la mise en œuvre, de l'évolution, de la fiabilité et de la sécurité des systèmes technologiques permettant les missions de collecte, catalogage, numérisation, archivage numérique et valorisation. Le service « entrée et catalogage » conçoit les flux d'entrée et contrôle la cohérence des métadonnées¹⁴⁸ et médias entrants. Il assure alors l'ensemble des activités de collecte, de stockage et d'exploitation des objets Web. Le dépôt légal du web est également suivi par le département « Technologies » qui a, parmi ses missions, la tâche de sauvegarder et pérenniser sur supports numériques l'ensemble du patrimoine du web média. Pour terminer, l'INA est une structure importante avec de nombreux secteurs d'activités qui depuis le début des années 2000 entreprend un virage numérique.

B- Le virage numérique de l'INA

C'est à partir de 1999 que l'on peut considérer que l'INA amorce un virage numérique, qui sera primordial lors de l'établissement du dépôt légal du web. Comme nous l'avons abordé dans le rapport de stage, c'est dans un contexte de lutte contre la disparition des archives audiovisuelles en raison des dégradations des matériels et de l'obsolescence des supports que l'INA instaure en 1999 un plan de sauvegarde et de numérisation nommé PSN. En 2019, c'est 335 000 heures de télévision et près de 500 000 heures de radio¹⁴⁹ qui ont été numérisées grâce au PSN. Le PSN a permis à l'INA de proposer des images et vidéos dans un format

¹⁴⁷ Article du Monde portant sur la suppression de la redevance télé, consultable en ligne : < https://www.lemonde.fr/politique/article/2022/07/20/suppression-de-la-redevance-la-piste-de-la-tva-pour-financer-l-audiovisuel-public_6135472_823448.html >, (consulté le 13/08/22)

¹⁴⁸ Données servant à caractériser une autre donnée physique ou numérique.

¹⁴⁹ CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).

restauré, cette numérisation met en lumière la communication de ces archives. Avec l'expérience du PSN, l'INA a pu développer un savoir-faire qui lui permet désormais de proposer à des détenteurs d'images une expertise en prenant en charge les collections de tiers à travers des mandats commerciaux. Grâce à cette mise en place du PSN et à l'expertise acquise par l'entreprise, elle a pu sauvegarder et exploiter des fonds privés comme les archives de TF1, de l'AFP ou encore de Radio Nova¹⁵⁰.

Le virage numérique a continué à la suite de l'application du PSN, notamment grâce à la mise en place en 2004 du site web « INAMediaPro ». Ce site à destination des professionnels de l'audiovisuel, des journalistes ou encore des chercheurs permet de visionner 1,4 millions¹⁵¹ d'heures d'archives. Parmi ces archives, l'utilisateur trouvera les émissions nationales et régionales diffusées sur les chaînes françaises de télévision depuis 1949, ainsi que plusieurs fonds cinématographiques dont les "Actualités françaises" (1914-1969) ou encore des émissions produites et diffusées par les chaînes de Radio France (depuis 1902)¹⁵². Ces différents fonds de l'INA sont complétés par des fonds appartenant à des partenaires de l'INA ayant profité de l'expertise de l'INA peu de temps après l'instauration du PSN. Nous trouvons notamment le fonds des actualités cubaines (ICAIC)¹⁵³, le fonds Georges Brassens¹⁵⁴ confié par le neveu de l'artiste, le fonds du Comité International Olympique (CIO)¹⁵⁵, le fonds de la Fédération Française de Football (FFF)¹⁵⁶, le fonds de la chaîne privée TF1¹⁵⁷, le fonds de

¹⁵⁰ CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).

¹⁵¹ Chiffre donnée par le site web : < <https://www.INAmediapro.com/Editorial/Offres-et-services-INA-MEDIAPRO> >, (consulté le 13/08/22)

¹⁵² Fonds documentés sur le site web de INAMédia Pro consultable en ligne : < <https://www.INAmediapro.com/Editorial/Offres-et-services-INA-MEDIAPRO> >, (consulté le 13/08/22)

¹⁵³ Fonds ICAIC, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-Actualites-Cubaines-ICAIC> >, (consulté le 13/08/22)

¹⁵⁴ Fonds Georges Brassens, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-Brassens-Intime> >, (consulté le 13/08/22)

¹⁵⁵ Fonds CIO, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-CIO> >, (consulté le 13/08/22)

¹⁵⁶ Fonds FFF, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-FFF> >, (consulté le 13/08/22)

¹⁵⁷ Fonds TF1, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-TF1> >, (consulté le 13/08/22)

Médecins Sans Frontières (MSF)¹⁵⁸, le fonds Télérama¹⁵⁹ ou encore le fonds de l'UNESCO¹⁶⁰. Il s'agit donc d'une plateforme avec une forte diversité de programmes. Pour naviguer entre ces fonds, la plateforme propose un moteur de recherche capable d'interroger dans différents corpus et profitant, de la même manière que l'application « Totem » ou « HyperBase », de plusieurs champs pour affiner la recherche. Les documents disponibles sur la plateforme peuvent être annotés, sauvegardés, timecodés ou plus simplement téléchargés. L'utilisation de ce service se réalise avec l'accompagnement d'un documentaliste. Enfin la plateforme informe l'utilisateur sur les garanties juridiques propres aux archives. En conclusion, il s'agit d'un service professionnel de valorisation des archives profitant du tournant numérique de l'INA.

En 2006 c'est un autre portail qui est ouvert par l'INA. Il s'agit du site web « INA.fr »¹⁶¹. Il s'agit d'un portail tourné vers le grand public mettant des archives à disposition de tous et sans surcoût contrairement à « INAMédia Pro ». On y trouve des archives éclairées par un contenu multimédia. En effet, le site web propose chaque jour une archive de la télévision servant à décrypter l'actualité contemporaine. Ces archives sont sélectionnées et mises en contexte par des documentalistes de la structure. De plus, le site propose toutes les semaines une émission intitulée « INAttendu »¹⁶² présentée par le journaliste Nathanaël de Rincquesen. L'émission explique l'actualité grâce à l'aide d'archives mais aussi avec une revue de presse, un zapping et des interviews. L'ensemble de ces productions de l'INA est ensuite mis en ligne sur les réseaux sociaux que sont Twitter, Instagram, Tik Tok, Facebook et YouTube.

En 2010 l'INA lance « la revue des médias »¹⁶³, un magazine en ligne retraçant les évolutions des médias. Ce magazine s'intéresse en particulier à l'économie des médias, aux usages et aux contenus des médias français. Pour sa rédaction, le magazine s'appuie sur le monde journalistique et universitaire. Pour analyser ces questions, le magazine propose à la

¹⁵⁸ Fonds MSF, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-MSF-Medecins-sans-Frontieres> >, (consulté le 13/08/22)

¹⁵⁹ Fonds Télérama, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-Telerama> >, (consulté le 13/08/22)

¹⁶⁰ Fonds de l'UNESCO, consultable en ligne : < <https://www.INAmediapro.com/Collections/Collection-UNESCO> >, (consulté le 13/08/22)

¹⁶¹ INA.fr, consultable en ligne : < <https://www.INA.fr> >, (consulté le 13/08/22)

¹⁶² Présentation de l'émission, consultable en ligne : < <https://www.INA.fr/offres-et-services/l-INA-eclaire-l-actu> >, (consulté le 13/08/22)

¹⁶³ Site web de la « revue des médias », consultable en ligne : < <https://www.INA.fr/offres-et-services/la-revue-des-medias> >, (consulté le 13/08/22)

fois des articles, des interviews et des études. Enfin le magazine est gratuit sans publicité et s'adresse à un large public d'étudiants, universitaires et journalistes.

L'éducation n'est pas en reste dans les offres numériques de l'entreprise. Ainsi en 2019 l'INA et France Télévisions ont mis en place une plateforme éducative nommée « Lumni Enseignement »¹⁶⁴. L'offre propose des ressources audiovisuelles et pédagogiques dédiées à l'enseignement du primaire jusqu'au lycée. On y retrouve des ressources vidéo et audio, des cours rédigés par des enseignants et des supports d'exercice. Plusieurs quiz de différents niveaux sont proposés ainsi que le téléchargement des contenus. Cette plateforme à destination des élèves et des enseignants est une porte supplémentaire vers le monde des archives que propose l'INA.

Pour terminer, la dernière plateforme numérique de l'INA et sans doute la plus mise en avant par l'entreprise, au même titre que « INA Media Pro », est la plateforme « Madelen ». Il s'agit d'une offre de service en SVOD¹⁶⁵, à savoir une plateforme de vidéos par abonnement. La plateforme compte 14.000 programmes accessibles sur abonnement. On y trouve de la musique, des films, des séries, des pièces de théâtre, des documentaires ainsi que des émissions cultes de télévision. Ce service à destination d'un grand public, basé sur le même modèle que Netflix¹⁶⁶, vise à mettre en lumière les programmes longs que détient la structure.

En conclusion, nous observons que depuis 1999 l'INA prend un virage numérique et essaye autant que possible de valoriser ses archives sur le web par de nombreuses plateformes et sites web. L'offre est importante et vise à la fois un public de professionnel de l'audiovisuel, un public de chercheurs et d'étudiants, un public scolaire et également le grand public. Nous pouvons ainsi affirmer que l'INA essaye d'ouvrir ses archives sur d'autres espaces de consultations que sont « l'INA thèque » et les délégations. L'INA ouvre ses archives audiovisuelles mais n'entreprend encore rien de la même ampleur pour les archives du web.

C- L'INA et le web français

Même si de nombreuses structures archivent le web comme nous le verrons plus tard, la France n'a pas de politique institutionnalisée d'archivage du web. Ce n'est qu'en 2011 avec

¹⁶⁴ Site web de « Lumni Enseignement », consultable en ligne : < <https://www.INA.fr/offres-et-services/lumni> >, (consulté le 13/08/22)

¹⁶⁵ SVOD (subscription video on demand)

¹⁶⁶ Netflix est une entreprise américaine leader de la SVOD

l'application de la loi DADVSI¹⁶⁷ que l'archivage du web se fera à grande échelle sur le web français. Si la loi DADVSI date de 2006, elle n'est appliquée qu'en 2011 et avant cette date le dépôt légal du web est appliqué uniquement à titre d'expérimentation. Si cette date peut paraître proche, c'est parce que le web est récent. Depuis la création du système hypertexte en 1989 par Tim Berners-Lee dans les locaux du CERN jusqu'à la mise en place d'un archivage, le web a beaucoup évolué. Tout d'abord, l'année 1991 marque la création officielle du projet World Wide Web, un premier site devient consultable hors des murs du CERN. En 1993, le projet World Wide Web passe dans le domaine public, s'ensuivent des créations de plus en plus importantes aboutissant à la création du web 2.0. Ainsi nous pouvons citer la création de Yahoo en 1994¹⁶⁸, Google en 1998¹⁶⁹ et Facebook en 2001¹⁷⁰. Cela nous amène en 2004 à l'avènement du web 2.0 qui est considéré comme un web avec des interfaces innovantes ainsi qu'une facilité d'utilisation de l'outil web par les internautes. Les universitaires Luc Quoniam et Boutet Charles-Victor dans un article de la revue « *Document numérique* » datant de 2008 décrivent le web 2.0 en ces termes :

« Le web 2.0 est l'avènement du *writable Web* communément appelé web contributif. Le droit en écriture qui est l'un des fondements de l'univers informatique a permis d'opérer un décloisonnement révolutionnaire des autorisations d'accès aux ressources. »¹⁷¹

Les innovations et les créations ne cessent d'évoluer. Nous pouvons signaler la création de Twitter en 2005¹⁷², l'instauration du HTML 5 en 2007 ou la création du navigateur chrome en

¹⁶⁷ Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).

¹⁶⁸ ALBOUY Michel, « Yahoo! : l'entreprise qui valait 125 milliards de dollars en 2000. The Conversation, 16 août 2016 », dans *Chroniques de finance et de gouvernance d'entreprise*, sous la direction de ALBOUY Michel, Caen, EMS Editions, Caen, 2019, p. 123-124, consultable en ligne : < <https://www.cairn.info/--9782376872191-page-123.htm> >, (consulté le 13/08/22)

¹⁶⁹ Article du journal *Le monde*, consultable en ligne : < https://www.lemonde.fr/pixels/video/2018/09/14/comment-google-est-il-devenu-un-empire_5355009_4408996.html >, (consulté le 13/08/22)

¹⁷⁰ Article de la revue *Les échos*, consultable en ligne : < <https://www.lesechos.fr/tech-medias/hightech/les-dates-marquantes-qui-ont-fait-de-facebook-un-empire-en-15-ans-961653> >, (consulté le 13/08/22)

¹⁷¹ QUONIAM Luc, BOUTET Charles-Victor, « web 2.0, la révolution connectique », *Document numérique*, 2008, volume 11, p. 133, consultable en ligne : < <https://www.cairn.info/revue-document-numerique-2008-1-page-133.htm> >, (consulté le 13/08/22)

¹⁷² Article de la « revue des médias », consultable en ligne : < <https://larevuedesmedias.INA.fr/12-evenements-qui-ont-marque-lhistoire-de-twitter> >, (consulté le 13/08/22)

2008. Avec ces innovations et ces créations, le web devient de plus en plus important et régit de façon croissante nos pratiques et nos modes de vie. Ainsi la collecte du web pour garder une trace historique de nos sociétés et écrire une histoire contemporaine du XXIème siècle paraît de plus en plus évidente. L'archivage du web, c'est aussi le fait de permettre aux institutions de préserver leurs données. De plus, l'archivage du web peut servir de preuve lors d'un procès ou pour lutter contre un gouvernement autoritaire diffusant des fausses informations sur le web. L'impact de la collecte du web est très important sur les collections mais aussi pour comprendre les sociétés actuelles. L'INA, dans la continuité de sa mission d'archivage de l'audiovisuel, se charge des sites des stations de radio et des chaînes de télévision, en définitive de l'ensemble des communications audiovisuelles et la BNF s'occupe du domaine « .fr » dans son ensemble, exception faite des domaines déjà collectés par l'Institut. L'INA a lancé dès 2009¹⁷³ la collecte du périmètre du domaine du web français qui lui a été attribué. Il est intéressant de souligner que « Internet Archives » que nous analyserons plus tard, a aussi donné accès à ses collections à la BNF et à l'INA pour les périodes antérieures à leur propre travail d'archivage. Ainsi l'influence et l'impact de « Internet Archives » sont significatifs. Les fonds collectés par l'INA sont moins élevés que la BNF et représentent environ 16.000¹⁷⁴ sites web selon l'INA, cependant contrairement à la BNF, la collecte est quotidienne. Thomas Drugeon, coordinateur du dépôt légal de l'INA évoque les fonds en ces termes :

« Notre fond d'archives représente 4 pétaoctets (soit 4 000 000 Go) pour 53 milliards d'éléments, c'est-à-dire les URL, les images, les vidéos, etc. Chaque année, le fond s'accroît de 10 milliards d'éléments nouveaux à peu près »¹⁷⁵

Il est à noter que ces propos ont été recueillis en 2016 et que depuis le fonds s'accroît de plus en plus vite, notamment à cause de l'influence de réseaux sociaux et de la numérisation grandissante à cause de la Covid-19. A titre de comparaison, la BNF collecte sur environ 4,5

¹⁷³ FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016 < <https://larevuedesmedias.INA.fr/larchivage-du-web-un-outil-pour-comprendre-internet> > (consulté le 05/03/22).

¹⁷⁴ Site web INA : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> > (consulté le 05/03/22).

¹⁷⁵ FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016 < <https://larevuedesmedias.INA.fr/larchivage-du-web-un-outil-pour-comprendre-internet> > (consulté le 05/03/22).

milliards de domaines. La bibliothèque établit une liste des sites qui doivent être collectés sur l'année. Puis il y a des collectes ponctuelles liées à certains événements comme les attentats ou les Jeux olympiques. La collecte aussi bien pour l'INA que pour la BNF est entièrement robotisée. Les sites sont collectés grâce à un processus de moissonnage¹⁷⁶, les URL et les contenus sont stockés directement sur des serveurs.

Les fonds du web collectés par le l'INA s'étendent sur différentes thématiques. Il existe un fonds contenant les sites web des chaînes de télévision et de radio. Nous trouvons également un fonds de Web TV et de Web Radios. Une des thématiques d'un des fonds est les contenus édités par les acteurs de l'audiovisuel. Le dépôt légal du web collecte également les sites liés aux programmes diffusés par les chaînes comme les sites d'émissions, de séries ou encore des sites de fans sur un programme audiovisuel. Enfin un des fonds le plus important et récent est le fonds du flux des réseaux sociaux et notamment de Twitter. Ainsi ce fonds contient les hashtags liés aux programmes audiovisuels ou bien les comptes d'acteurs du milieu audiovisuel français. L'INA estime que dans ces fonds, il y a environ 16.000 comptes Twitter conservés au 31 décembre 2020¹⁷⁷ et que depuis 2014 l'INA a collecté environ 1,95 milliards de tweets. La délégation suit les évolutions des modes de publication sur le web, les réseaux sociaux prenant une place grandissante et d'autant plus importante pendant la période de pandémie, l'INA collecte en parallèle de cette évolution de plus en plus de comptes sur les réseaux sociaux. Cette collecte se fait d'une part sur Twitter, Facebook ou Instagram d'autre part sur les plateformes de publication comme YouTube et Dailymotion. L'impact de la collecte du web est important sur les collections et pour comprendre les sociétés actuelles. Laurent Vallet, Président de l'INA dans une interview accordée à un magazine de l'école Sciences-Po a déclaré au sujet des réseaux sociaux :

« À l'heure où les nouvelles générations ont accès à d'innombrables vidéos via les réseaux sociaux, notre but est de mettre en perspective, de montrer que l'Histoire est un éternel recommencement, que les images de migrants naufrageant au large de Lampedusa ressemblent, tragiquement et terriblement, à celles filmées en mer de Chine en 1979. La technique est efficace, notamment pour s'adresser de manière

¹⁷⁶ Site BNF Dépôt légal des sites web : Mode d'emploi : < <https://www.BNF.fr/fr/centre-d-aide/depot-legal-des-sites-web-mode-demploi#step02> > (consulté le 05/03/22).

¹⁷⁷ Site web INA : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> > (consulté le 05/03/22).

*journalistique aux moins de 30 ans, adeptes des réseaux sociaux et moins touchés par les remarquables documentaires historiques de 52 ou 90 min que produit ou coproduit l'INA depuis 40 ans pour les cases de deuxième partie de soirée des grandes chaînes hertziennes »*¹⁷⁸

En conclusion, l'INA a évolué pour devenir une institution de plus en plus numérique. Sa structure interne assez imposante et les différents objectifs fixés démontrent l'ambition de devenir un établissement numérique. Le virage numérique avec des offres de plus en plus tournées vers des archives présentes sur le web démontre le virage de l'entreprise. Enfin l'INA depuis quelques années collecte le web français. Il convient désormais de s'interroger sur l'évolution de l'archivage du web et comprendre la mise en place d'un dépôt légal du web.

¹⁷⁸ Site web du magazine de l'Ecole de Sciences Po : < <https://www.emilemagazine.fr/article/2017/11/22/laurent-vallet-a-lheure-des-reseaux-sociaux-les-enjeux-pour-lina-sont-passionnants> > (consulté le 05/03/22).

Partie 2 : De l'archivage web au dépôt légal du web, une lente évolution ?

A- Le dépôt légal comme photographie du foisonnement culturel français

Comme nous l'avons vu plus haut, la mise en place du dépôt légal est instaurée par François Ier dans le but de :

« Faire assembler en notre librairie toutes les œuvres de notre temps, dignes d'être vues, qui ont été ou qui seront faites, pour avoir recours aux dits livres, si, de fortune, ils étaient perdus de la mémoire des hommes »¹⁷⁹.

L'ordonnance du 28 décembre 1537 a pour but la constitution d'un patrimoine et la mise en place d'une surveillance de l'édition. Ainsi dès le XVIème, on peut voir transparaître les notions de collecte et de conservation. Cette idée de constitution d'un patrimoine pour les générations futures est présente dès l'instauration du dépôt légal comme nous en informe l'universitaire Frédéric Saby dans la revue « *Sociétés et représentations* »¹⁸⁰. Le dépôt légal sera par la suite affiné jusqu'à être défini dans le Code du Patrimoine à l'article L131-2¹⁸¹ :

« Les documents imprimés, graphiques, photographiques, sonores, audiovisuels, multimédias, quel que soit leur procédé technique de production, d'édition ou de diffusion, font l'objet d'un dépôt obligatoire, dénommé dépôt légal, dès lors qu'ils sont mis à la disposition d'un public. »

Dans cette définition, l'expression « dès lors qu'ils sont mis à disposition d'un public » permet d'exclure du dépôt légal tout ce qui n'a pas vocation à être publié¹⁸². On exclut alors du dépôt

¹⁷⁹ Ordonnance de Montpellier, 28 décembre 1537.

¹⁸⁰ Saby Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013, p. 15, consultable en ligne : < <https://www.cairn.info/revue-societes-et-representations-2013-1-page-15.htm> >, (consulté le 13/08/22)

¹⁸¹ Code du Patrimoine : < https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000020905828/ > (consulté le 05/03/22).

¹⁸² Saby Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013, p. 15, consultable en ligne : < <https://www.cairn.info/revue-societes-et-representations-2013-1-page-15.htm> >, (consulté le 13/08/22)

légal les documents imprimés mais pas publiés comme les cartes de visites ou les cartons d'invitation. C'est bien la publication, peu importe le support, qui fait entrer le document dans le champ du dépôt légal. Par ces lois successives, le dépôt légal progresse de manière à suivre les évolutions technologiques. Ainsi ce tableau issu des données de l'universitaire et conservateur Frédéric Saby montre l'ouverture de dépôt légal selon les supports :

Fig. 4 : Tableau des dates d'entrées dans le dépôt légal en fonction du support¹⁸³

Imprimés	1537
Estampes, cartes et plans	1648
Partitions musicales	1793
Photographies phonogrammes	1925
Affiches	1941
Vidéogrammes et documents multimédias composites	1975
Multimédias, logiciels, bases de données	1992
Internet	2006

A partir du 20 juin 1992 comme indiqué dans le tableau, le dépôt légal est élargi entre autres aux programmes de télévision et de radio. C'est l'INA thèque, un département créé en 1995 par l'INA pour répondre à ce besoin, qui va mettre en place le dépôt légal. Plusieurs dizaines de milliers d'heures de programmes vont ainsi être collectées chaque année. Cette collecte a évolué au fil des technologies, et aujourd'hui l'INA thèque grâce à un dispositif de captation numérique par liaisons satellites et fibres optiques collecte les programmes vingt-quatre heures sur vingt-quatre sur plus de 120 chaînes de télévision et de radio¹⁸⁴. Trois organismes ont été définis pour supporter la charge du dépôt légal, il s'agit de la BNF, le CNC pour le cinéma et la photo et l'INA pour la télé et la radio. Un conseil scientifique du dépôt légal est créé dans le but d'unifier les méthodes de collecte et de conservation. Le dépôt légal

¹⁸³ Données issues de l'ouvrage de SABY Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013, p. 15, consultable en ligne : < <https://www.cairn.info/revue-societes-et-representations-2013-1-page-15.htm> >, (consulté le 13/08/22)

¹⁸⁴ SABY Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013, p. 15, consultable en ligne : < <https://www.cairn.info/revue-societes-et-representations-2013-1-page-15.htm> >, (consulté le 13/08/22)

est élargi en 2002 aux chaînes du câble et du satellite, puis en 2005 à la télévision numérique terrestre (TNT). Le dépôt légal est à nouveau modifié le 1er août 2006 par la loi DADVSI¹⁸⁵ qui étend le dépôt légal au Web comme nous le verrons plus tard.

En conclusion dès le XVIème siècle, le pouvoir français a voulu se constituer une collection à vocation patrimoniale. C'est dans un objectif historique que le dépôt légal est né. La France, pionnière dans le domaine, a souhaité préserver ses documents en faisant évoluer de façon cyclique la législation. Le dépôt légal a permis la création d'une collection patrimoniale permettant aux chercheurs d'avoir des sources pour écrire l'histoire ou étudier un phénomène. Les critères précis et rigoureux, comme le précise Frédéric Saby¹⁸⁶ ont permis de garantir l'objectivité de jugement à laquelle le chercheur est tenu. Cependant la massification et l'internationalisation des données sur le web a induit une évolution du dépôt légal à partir de 2006.

B- Genèse de l'archivage web : des approches diverses pour un but commun

L'archivage web n'est pas né avec le dépôt légal du web. Les premières tentatives d'archivage du web remontent à 1996. C'est d'ailleurs cette année-là que lors de la conférence des directeurs des bibliothèques nationales (CDNL) Brian Lang a précisé que :

« [...] faute de quoi le passé ne laisserait plus de traces, et les recherches antérieures ne pourraient plus être retrouvées, comprises ou reproduites comme il se doit, pour entretenir le cycle continu de l'expérimentation qui fait progresser la connaissance »¹⁸⁷

Les archivistes vont recourir à plusieurs stratégies pour essayer de garder la mémoire d'un web éphémère. On y voit une stratégie d'échantillonnage du web, un archivage intégral ou encore un archivage plutôt exhaustif. De plus, il y a des collectes automatiques et des collectes

¹⁸⁵ Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).

¹⁸⁶ Saby Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013, p. 15, consultable en ligne : < <https://www.cairn.info/revue-societes-et-representations-2013-1-page-15.htm> >, (consulté le 13/08/22)

¹⁸⁷ LANG Brian, *Le dépôt légal des publications électroniques*, Paris : UNESCO, 1996, consultable en ligne : < https://unesdoc.unesco.org/ark:/48223/pf0000105504_fre >, (consulté le 14/08/22)

manuelles. Le premier projet datant de cette même année 1996 est celui de l'entrepreneur américain Brewster Kahle avec « Internet Archives ». Il s'agit ici d'une approche intégrale où la collecte se fait sur l'ensemble du web sans notion de valeur patrimoniale ou de sélection. Ainsi l'auteur Thomas Chaimbault en parle en ces termes :

« Ce qui n'était au départ qu'un simple projet de recherche va vite devenir une société basée à San Francisco, à l'origine dès juillet 1997 d'un outil commercial appelé *Alexa*. Cet outil permet de « butiner », rapatrier et indexer un nombre important de pages et de donner des indications sur leur fréquentation, le renouvellement, le nombre de liens, ... mais surtout il permet de donner accès aux versions précédentes des sites archivés par *Internet Archive* »¹⁸⁸

Le projet va permettre l'archivage de plusieurs centaines de milliards de pages web archivées comme le montre la page web¹⁸⁹ du projet toujours en activité. Cet archivage se fait sans considération de frontières, de nationalités ou encore d'autorisations. Cela peut poser problème aux entreprises vendant leurs archives, comme c'est le cas pour le journal *Le Monde*. Dès sa création, le projet « Internet Archive » coopère avec le Consortium pour la préservation d'internet (l'IIPC) qui travaille avec les bibliothèques nationales pour la conservation, le moissonnage et la diffusion des archives du web.

Un autre projet d'ampleur est créé cette même année 1996, « KulturarW » lancé par la bibliothèque royale de Suède. L'approche de ce projet est exhaustive, c'est-à-dire que le projet vise à collecter l'ensemble du web défini par un nom de domaine précis en l'occurrence « .se ». C'est donc un archivage d'un espace défini à l'avance. Cet archivage du web suédois se fait de façon automatisée comme le souligne Thomas Chaimbault¹⁹⁰.

Nous pouvons citer, toujours pendant l'année 1996, une dernière approche qui est une sélective visant à archiver une partie définie du web selon des critères sélectifs. Ce projet « Pandora » est mené par la bibliothèque nationale d'Australie. L'objectif est d'archiver le web australien reconnu pertinent d'un point de vue culturel et scientifique.

¹⁸⁸ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 29, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

¹⁸⁹ Site web « Internet Archives » : < <https://archive.org> >, (consulté le 14/08/22)

¹⁹⁰ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 31, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

Ces trois approches différentes de l'année 1996 marquent la diversité des projets, le foisonnement de l'idée d'archiver le web et indiquent un tournant dans l'histoire de l'archivage du web. En France, l'initiative de collecte du web est venue des institutions publiques. C'est effectivement la BNF qui a mené les premiers tests de collecte automatisée à partir de 1999¹⁹¹, dans le cadre d'un projet subventionné par la commission européenne de préservation des publications numériques intitulé NEDLIB¹⁹². L'approche de ce projet se veut à la fois automatisée tout en proposant un archivage ciblé ponctuel selon les besoins. Ainsi on peut parler d'une collecte mixte. L'INA, quant à elle, entame les tractations avec le ministère de l'Économie et des Finances pour archiver Internet, en parallèle l'État français lance le plan d'action gouvernementale pour la société de l'information (PAGSI)¹⁹³. Ce plan avait pour objectif de permettre le développement des technologies de l'information afin de combler le retard de la France dans ce domaine. Dans le cadre d'un partenariat avec le projet « Internet Archives », la BNF a réalisé deux instantanés du domaine « .fr ». Les deux instantanés sont réalisés à un an d'intervalle, le premier en 2004 et le second en 2005. De plus, la BNF a acquis les copies du web français réalisées par « Internet Archives » en 1996 et en 2005. Ces deux copies représentent 6 milliards de fichiers pour un volume de 60 téraoctets¹⁹⁴.

En 2003, une charte de l'UNESCO intitulée « Charte pour la conservation du patrimoine numérique »¹⁹⁵ explicite le besoin de collecte du web :

« Les documents numériques comprennent, parmi un large éventail de formats électroniques qui ne cesse de se diversifier, des textes, des bases de données, des images fixes et animées, des documents sonores et graphiques, des logiciels et des pages Web. Ils sont souvent éphémères, et leur conservation nécessite des mesures volontaires d'entretien et de gestion dès leur création.

¹⁹¹ FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016 < <https://larevuedesmedias.INA.fr/larchivage-du-web-un-outil-pour-comprendre-internet> > > (consulté le 05/03/22).

¹⁹² Projet européen NEDLIB (préservation des publications numériques) : < <https://cordis.europa.eu/project/id/LB5648/fr> > (consulté le 05/03/22).

¹⁹³ PAGSI : < <https://www.vie-publique.fr/eclairage/18925-e-administration-du-pagsi-au-programme-action-publique-2022> > (consulté le 05/03/22).

¹⁹⁴ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 35, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

¹⁹⁵ Charte pour la conservation du patrimoine numérique, consultable en ligne : < http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html >, (consulté le 14/08/22)

Beaucoup de ces documents ont une valeur et une importance durables et constituent par conséquent un patrimoine qui doit être protégé et conservé pour les générations présentes et futures. Ce patrimoine, qui ne cesse de grandir, peut exister dans n'importe quelle langue, n'importe quelle partie du monde et n'importe quel domaine de la connaissance ou de l'expression humaine. »¹⁹⁶

Au niveau mondial, les institutions prennent conscience de l'évolution rapide du web et du besoin de conservation pour des visées historiques. Enfin, en France, l'archivage du web est institutionnalisé en 2006 par la loi DADVSI qui élargit le champ de dépôt légal aux sites web français.

C- Contexte juridique et acteurs

C'est par la loi du 1er août 2006 DADVSI¹⁹⁷ qu'un dépôt légal de l'internet a été mis en place en France. La loi a permis l'extension du dépôt légal à :

« Les logiciels et les bases de données sont soumis à l'obligation de dépôt légal dès lors qu'ils sont mis à disposition d'un public par la diffusion d'un support matériel, quelle que soit la nature de ce support. Sont également soumis au dépôt légal les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique. »¹⁹⁸

Cette loi a permis de définir les contours de la mission confiée d'une part à l'INA et d'autre part à la BNF. Cette loi a répondu à l'émergence exponentielle des médias numériques dans le paysage audiovisuel français. C'est avec l'évolution des modes de diffusion que les institutions françaises ont dû réfléchir à des outils pour garder une trace du web et de l'impact

¹⁹⁶ Charte pour la conservation du patrimoine numérique, article 1, consultable en ligne : < http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html >, (consulté le 14/08/22)

¹⁹⁷ Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).

¹⁹⁸ Loi DADVSI titre IV : dépôt légal, article 39 : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).

grandissant de ces médias. La loi DADVSI est entrée en vigueur en décembre 2011¹⁹⁹ ; avant cette date, le dépôt du web est organisé à titre d'expérimentation. Il a fallu mettre en place la collecte des objets, inventer leurs modes de consultation et imaginer les usages qui en seraient faits²⁰⁰. Il s'agit donc d'une nouveauté dans la collecte et pour le dépôt légal. La loi DADVSI en 2006 mise en œuvre en 2011 a précisé le champ respectif des données collectées par les deux organismes : l'INA, dans la continuité de sa mission d'archivage de l'audiovisuel, se charge des sites des stations de radio et des chaînes de télévision, donc de toutes les communications audiovisuelles et la BNF s'occupe du domaine « .fr » dans son ensemble, exception faite des domaines déjà collectés par l'Institut. Il est à noter que la BNF archive également les extensions territoriales ainsi que les contenus produits par des français ou des auteurs domiciliés en France, dont les adresses sont en « .com » ou « .org » comme le soulignent les auteurs de l'ouvrage *« Qu'est-ce qu'une archive web »*²⁰¹. L'INA a lancé dès 2009²⁰² la collecte du périmètre du domaine du web français qui lui a été attribué. Les fonds collectés par l'INA sont moins élevés que la BNF et représentent environ 16.000²⁰³ sites web selon l'INA. Cependant contrairement à la BNF, la collecte est quotidienne. Thomas Drugeon, coordinateur du dépôt légal de l'INA décrit les fonds en ces termes :

« Notre fonds d'archives représente 4 pétaoctets (soit 4 000 000 Go) pour 53 milliards d'éléments, c'est-à-dire les URL, les images, les vidéos, etc. Chaque année, le fond s'accroît de 10 milliards d'éléments nouveaux à peu près »²⁰⁴

Il est à noter que les propos ont été recueillis en 2016 et que depuis le fonds s'accroît de plus en plus vite, notamment en raison de l'influence des réseaux sociaux et de la numérisation grandissante à cause de la Covid-19. A titre de comparaison, la BNF collecte sur environ 4,5

¹⁹⁹ Site web INA : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> > (consulté le 05/03/22).

²⁰⁰ Site web INA thème les sites web média : < <http://www.INAtheque.fr/fonds-audiovisuels/sites-web-media.html> > (consulté le 05/03/22).

²⁰¹ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 14.

²⁰² FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016 < <https://larevuedesmedias.INA.fr/larchivage-du-web-un-outil-pour-comprendre-internet> > (consulté le 05/03/22).

²⁰³ Site web INA : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> > (consulté le 05/03/22).

²⁰⁴ FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016 < <https://larevuedesmedias.INA.fr/larchivage-du-web-un-outil-pour-comprendre-internet> > (consulté le 05/03/22).

milliards de domaines. Enfin il est pertinent de signaler que certaines pages sont soumises au droit d'auteur comme les articles de journaux ou au droit privé comme les informations commerciales. Ainsi la reproduction de ces pages peut paraître difficile, d'autant plus que le web est international et dépend donc de plusieurs juridictions. En conclusion, si plusieurs projets ont pu impulser l'archivage du web, en France c'est grâce à la loi DADVSI que le web est archivé. Les deux acteurs principaux de l'archivage sont la BNF et l'INA qui se divisent le web à archiver. Malgré un archivage régulier et important, il convient de mettre en évidence les évolutions de l'archivage en France, les enjeux de gouvernance ainsi que les potentielles limites que l'archivage du web pourrait soulever.

Partie 3 : L'archive du web comme miroir de l'internet français ?

A- L'archive web : patrimoine fragile

Comme nous l'avons abordé plus tôt, l'archivage web se réalise grâce à plusieurs stratégies : la sélection d'archives, l'archivage de l'intégralité du web ou encore une approche plus exhaustive. Ne pouvant pas archiver tout ce qui se trouve sur le web, compte tenu de la masse des données, les acteurs se retrouvent dans l'obligation de faire des choix. Cependant les choix amènent inévitablement à mettre des archives de côté, et qui seront perdues à jamais. En effet, le web est un patrimoine très fragile car il est au cœur du progrès et des changements techniques. L'évolution d'internet et du web est telle que des pages peuvent devenir obsolètes en peu de temps, ces mêmes pages qui avaient un impact important sur les sociétés. Effectivement selon un article portant sur l'archivage web de l'INA²⁰⁵, l'espérance de vie d'une page web est de 44 jours, la permanence d'un lien est de 56h en moyenne pour un titre de presse selon le « *Chicago Tribune* »²⁰⁶. La durée de vie moyenne sur les réseaux sociaux est encore plus restreinte selon l'INA²⁰⁷ il s'agit de 4h04 pour un tweet, 14h42 pour un post Facebook et de seulement 21h36 pour une photo Instagram. Ainsi il est évident que l'ensemble du web ne peut pas être archivé compte tenu de l'aspect éphémère d'Internet, cependant ces durées doivent mettre en avant le fait qu'il est primordial de sauvegarder une mémoire collective. Une contrainte est alors évidente lors de la sauvegarde, il s'agit de prendre du recul pour essayer de comprendre quelle archive a une valeur historique. Si cela peut être fait de façon relativement aisée pour un document ancien, le fait que le web soit si proche de nous et de nos pratiques empêche d'avoir un certain recul. Ainsi nous pourrions perdre une partie des données importantes parce nous n'avons pas reconnu la valeur historique de l'archive. C'est par exemple le cas pour le site de partage de musique MySpace. A la suite d'une erreur de stockage en ligne des serveurs du groupe, l'entreprise a perdu 12

²⁰⁵ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

²⁰⁶ Site web de journal « Chicago tribune » : < <https://www.chicagotribune.com> >, (consulté le 14/08/22)

²⁰⁷ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

ans de données entre 2003 et 2015²⁰⁸. Les données en ligne de plus de 200 millions d'utilisateurs ont été perdues à la suite de cette erreur lors d'une migration de serveur. MySpace était pourtant un géant du web dans les années 2000, et malgré cela, les données ont été perdues. C'est d'autant plus préoccupant pour des sites patrimoniaux plus modestes. Jill Lepore, historienne américaine estime que :

« Facebook n'existe que depuis une décennie, il ne sera pas toujours là. Twitter²⁰⁹ est un cas rare : il s'est arrangé pour archiver tous ses tweets à la Bibliothèque du Congrès. Vous pourrez peut-être, un jour, y retrouver d'anciens messages, mais pas de sitôt : les archives de Twitter ne sont pas encore ouvertes à la recherche. En attendant, sur le Web, si vous cliquez sur un tweet supprimé, vous obtenez ce message : « Désolé, cette page n'existe pas ! » »

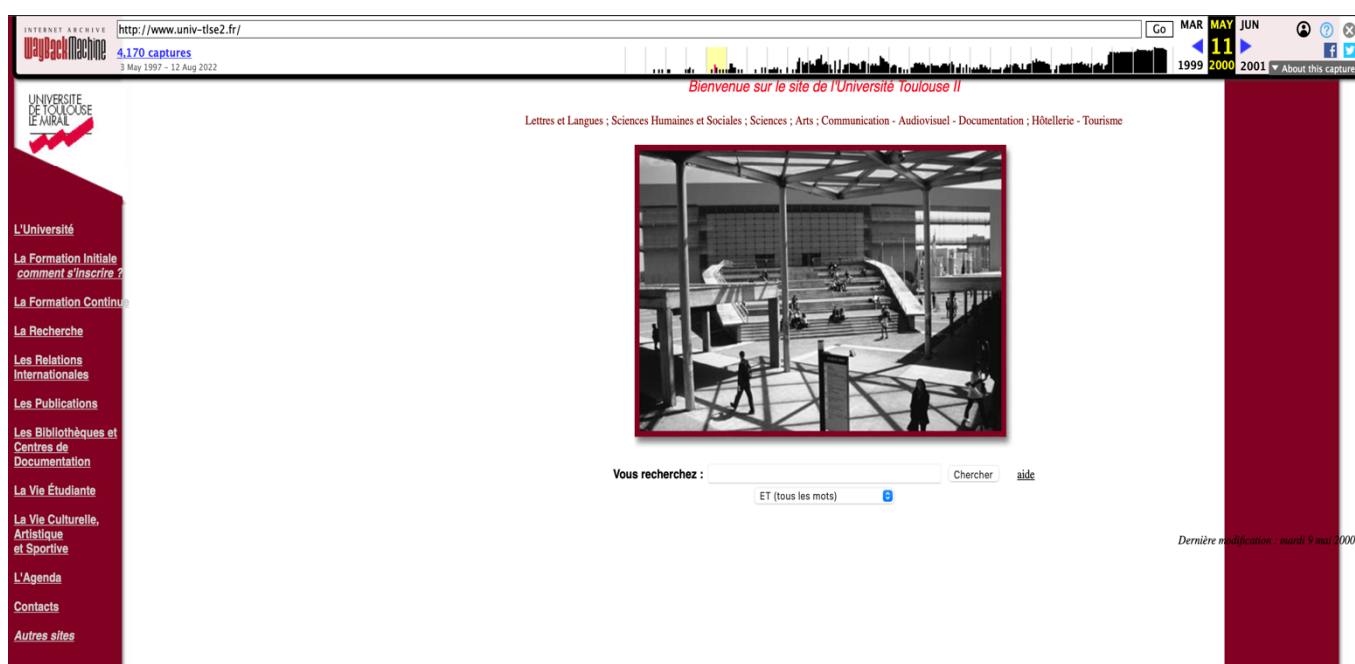
Ainsi le web apparaît comme instable éphémère et fragile, où une page peut être supprimée sans pouvoir en garder une trace. L'archivage web devient alors un élément primordial dans la sauvegarde d'une histoire sociétale et de la mémoire collective. L'archive web est d'autant plus fragile que les technologies numériques évoluent avec le temps ; l'archive web, elle se retrouve figée dans le temps. Rien n'affirme que l'archive puisse toujours être lu dans une dizaine d'années. Contrairement aux archives papiers qui n'évoluent pas et restent fixées dans le temps, l'archive du web, comme une archive audiovisuelle dépend d'un support pour être lisible. Il faut donc préserver les matériels au même titre que les archives. En plus de la suppression des fichiers sur Internet, la modification des pages web, des noms de domaines ou des URL peut empêcher un bon archivage. De plus, la modification des contenus d'une page ou des contenants comme la mise en page entraîne une fois de plus une fragilité de l'archive web. En utilisant l'outil mis en place par « Internet Archives » la « WayBack Machine », nous observons la fragilité de l'archive web. La « Wayback Machine » permet par une URL de naviguer dans les anciennes versions archivées de cette URL. Cet outil base sa communication sur l'invitation à un voyage dans le passé, cependant certains

²⁰⁸ Article « Le Parisien », consultable en ligne : < <https://www.leparisien.fr/high-tech/comment-myspace-a-efface-12-ans-de-musique-18-03-2019-8034222.php> >, (consulté le 14/08/22)

²⁰⁹ Propos de l'historienne Jill Lepore, dans le journal CNETFrance, consultable en ligne : < <https://www.cnetfrance.fr/news/pourquoi-il-faut-archiver-le-web-d-hier-et-d-aujourd-hui-39908597.htm> >, (consulté le 14/08/22)

éléments ne sont pas archivés et sont alors remplacés par une version plus récente. C'est particulièrement vrai dans le cas d'utilisation d'encart de publicité qui, lorsque on ouvre une archive des années 2000, propose d'intégrer à la page des publicités de 2022. De plus, il n'existe qu'une recherche par URL. Les archives ne permettent pas de naviguer via les liens hypertextes qui ne sont plus actifs. Ainsi voici l'archive web de la page de l'université Toulouse Jean-Jaurès datant du 11 mai 2000²¹⁰

Fig. 5 : Page de l'université Jean Jaurès datant du 11 mai 2000²¹¹



Si la navigation reste possible, l'archive ne permet pas d'aller profondément dans la ramification du site web, certaines pages se retrouvent alors inutilisables. De plus, il paraît évident que le site a des informations manquantes, comme l'atteste les grandes zones vides. Ces zones manquantes sont assez régulières dans les archives web. En utilisant la « Wayback Machine », il est possible de naviguer sur le site du journal « Le Parisien ». De façon encore plus parlante, nous observons l'aspect lacunaire de l'archive comme le montre la figure suivante :

²¹⁰ Archive du 11 mai 2000 trouvable avec l'outil « Wayback Machine », consultable en ligne : < <https://web.archive.org/web/20000511181204/http://www.univ-tlse2.fr/> >, (consulté le 14/08/22)

²¹¹ Archive du 11 mai 2000 trouvable avec l'outil « Wayback Machine », consultable en ligne : < <https://web.archive.org/web/20000511181204/http://www.univ-tlse2.fr/> >, (consulté le 14/08/22)

Fig. 6 : Page du journal « *Le Parisien* » datant du 5 mars 2000²¹²



En conclusion, étant donné la massification du web et ses évolutions rapides, en termes de technologie mais aussi en termes de contenu et contenant, un site web est très fragile, versatile et éphémère. Il convient alors de l'archiver de la meilleure des manières pour la conservation du patrimoine mondial, la recherche universitaire et également la préservation de la mémoire collective. Il incombe à la BNF et à l'INA de mener à bien ce travail. L'archive web peut également se lire comme étant au centre d'enjeux de gouvernance avec des limites qui lui sont propres.

²¹² Archive du 5 mars 2000 trouvable avec l'outil « Wayback Machine », consultable en ligne : < <https://web.archive.org/web/20000511181204/http://www.univ-tlse2.fr/> >, (consulté le 14/08/22)

B- Limites de l'archivage web

Comme nous l'avons étudié, du fait de la massification du web, il paraît impossible d'archiver l'ensemble des documents émanant du web. De plus, la fragilité et la volatilité du web entrave l'archivage. Cependant il y a d'autres problématiques et barrières à l'archivage du web. Nous pouvons citer l'archivage d'un lien mort, cela est d'autant plus dommageable lorsque ce lien sert d'entrée sur l'ensemble d'un site web archivé et qu'il s'agit d'un site institutionnel de référence. De plus, il va s'agir d'un lien inutile stocké par l'INA ou une autre institution. Le stockage des données est un élément primordial de l'archivage, il est donc plus dommageable d'avoir des liens morts inutilisables qui prennent de l'espace de stockage. En 2013, un lien mort a fait perdre à la cour suprême américaine plusieurs articles comme le met en évidence les auteurs de l'ouvrage « Qu'est-ce qu'une archive du web » :

« Un « lien mort » est d'autant plus dommageable qu'il a pu servir de référence, voire de garantie institutionnelle, comme en a témoigné l'affaire des articles disparus de la Cour suprême américaine révélée par le *New York Times* en 2013 – on parle alors de « *reference rot* ». Les liens et contenus web s'évanouissent au gré de la fermeture d'hébergeurs ou de plateformes, de la réorganisation de l'architecture d'un site, ou parce qu'un auteur a tout simplement choisi de supprimer un contenu, voire d'effacer complètement sa présence numérique, ce que l'on surnomme « *infosuicide* ». »²¹³

L'archivage web peut être difficile pour des raisons techniques. Dans un premier temps, parce que, pour des raisons techniques, l'archivage automatique pour capturer des contenus web mis en forme par un certain type de technologie peut s'avérer impossible si cette technologie ou ce dispositif n'est pas pris en charge ou simplement obsolète. Ainsi l'archivage du web va dépendre de l'utilisation des technologies. Effectivement certains langages web comme le Javascript peuvent poser des problèmes lors de l'archivage. Les auteurs de l'ouvrage « Qu'est-ce qu'une archive du web » l'expliquent en ces termes :

²¹³ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 32

« Suivant une logique de flux, le Web dynamique tend à encapsuler des contenus hébergés ailleurs, une page n'étant que de plus en plus rarement une unité homogène. Ainsi, ces dispositifs peuvent avoir tendance à reconstituer des pages « à trous ». Par exemple, le langage JavaScript permettant l'encapsulation de contenu a été l'un des premiers obstacles au moissonnage de données web par Heritrix, produisant des archives de pages web qui sont des coquilles vides. »²¹⁴

L'enchevêtrement de plusieurs types de langage et de contenu ou encore la superposition de codes peuvent rendre difficile ou impossible la mise en place d'un archivage, et c'est particulièrement le cas des forums se trouvant le web.

Une autre limite de l'archivage du web est le verrouillage par mot de passe de page web. C'est un problème rencontré notamment sur les réseaux sociaux ou sur les plateformes d'abonnement qui nécessitent une identification. Les réseaux sociaux étant des éléments impactant dans nos vies du XXIème siècle, il est difficile pour l'INA de ne pas l'archiver, cependant comme nous le verrons en détail plus tard, l'archivage des réseaux sociaux est assez complexe notamment parce que les plateformes changent souvent de technologie et de paramètre. Annick Le Follic chargée de collections numériques au département du dépôt légal de la BNF l'explique en ces termes dans un entretien datant du 21 mars 2016 :

« La limite de notre archivage des réseaux sociaux est technique : ces plateformes changent souvent de technologies et de paramètres, donc il nous faut donner à chaque fois une instruction manuelle à Heritrix pour qu'il capture bien les contenus qui nous intéressent. En particulier, les protocoles https nous posent parfois des problèmes, tout comme Facebook lorsqu'il utilisait des 'captcha'. Ces contraintes sont d'autant plus problématiques lors de collectes d'urgence, où le temps nous est compté. »²¹⁵

²¹⁴ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 32

²¹⁵ Entretien de Annick le Follic mené par Valérie Schafer, consultable en ligne : < <https://asap.hypotheses.org/168> >, (consulté le 14/08/22)

En conclusion, l'archivage est régi par des limites qui l'empêchent de s'ouvrir à l'ensemble du web. Au-delà de la massification qui rend un archivage global impossible, il existe plusieurs contraintes comme la mise en place de verrouillage par mot de passe sur certaines pages web et notamment sur les réseaux sociaux, l'utilisation de couches différentes de codes qui rend l'archive illisible ou encore plus simplement la problématique des liens morts. L'archivage est également dirigé par des enjeux de gouvernance nationale et internationale.

C- Enjeux de gouvernance

Il paraît évident que l'archivage du web répond à des enjeux de gouvernance. Cette gouvernance peut être étatique, institutionnelle ou encore mondiale. Effectivement l'archivage du web repose tout d'abord sur plusieurs organisations qui s'efforcent de travailler en coopération. Nous pouvons citer l'INA, la BNF, mais aussi des fondations comme « Internet Archives », « International Internet Preservation Consortium » (IIPC) ou encore le secteur privé comme Google. Ainsi comme le montrent les auteurs Musiani Francesca, Paloque-Berges Camille, Schafer Valérie et Thierry Benjamin²¹⁶, dans l'archivage du web se retrouvent les acteurs de la gouvernance d'Internet. Ainsi l'archivage n'échappe pas à des tensions notamment autour de la standardisation du web, et à des visions divergentes souvent autour des contraintes des formats propriétaires. L'archivage met en évidence également les tensions géopolitiques. L'exemple le plus parlant est le blocage de « Internet Archive » en Chine²¹⁷. Le modèle de l'Internet chinois étant très fermé, il est impossible de naviguer sur ce service notamment avec l'instauration du « Grand Pare-feu » comme l'indique Séverine Arsène politologue et sinologue :

« Avec la mise en place du célèbre "Grand Pare-feu", la Chine a été pionnière sur le filtrage et le blocage des données au niveau des principaux câbles et routeurs auxquels sont connectés les fournisseurs d'accès à Internet. Ce type de blocage, encore rare au début des années 2000, est en fait aujourd'hui pratiqué de manière courante et plus ouvertement

²¹⁶ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 45

²¹⁷ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 46

assumée, y compris dans les pays européens, à mesure que le contre-terrorisme, la lutte contre la pédopornographie et la protection de la cybersécurité ont pris place au cœur de l'agenda politique. »²¹⁸

D'autres tensions dans l'archivage web ont émergé après l'appel de Brewster Kahle, créateur de « Internet Archives » à la suite de la victoire de Donald Trump en 2017, qui demandait un financement participatif pour créer une copie complète des collections numériques de « Internet Archives » hors des États-Unis²¹⁹. Nous voyons donc que l'archivage du web est au centre des tensions internationales ou nationales car il est un enjeu de pouvoir des nations. Pour le sociologue du numérique Benjamin Loveluck, Internet au même titre que son archivage, est un enjeu d'administration.

« Au même titre que la plupart des technologies de communication qui l'ont précédée, l'informatique en réseau a initialement été célébrée comme une « révolution » des pouvoirs. Elle serait un moyen de diffusion des connaissances, de transparence démocratique, de renversement des hiérarchies et d'émancipation individuelle. »²²⁰

Avoir la main sur l'archivage du web est donc une force d'abord technologique mais ensuite d'affirmation du pouvoir des états. A ce titre, la présence des pays en voie de développement dans les archives du web n'est pas proportionnelle avec leur présence sur le web. Comme le soulignent les auteurs, Musiani Francesca, Paloque-Berges Camille, Schafer Valérie et Thierry Benjamin²²¹, il existe des régions du monde non archivées avec une forte présence sur le web, c'est par exemple le cas pour l'Inde, l'Amérique latine ou l'Afrique. Cela est dû à un manque de ressources techniques, légales et financières. En conclusion, l'archivage du web au sein des

²¹⁸ ARSENE Séverine, *La Chine et le contrôle d'Internet. Une cybersécurité ambivalente*, 2019, consultable en ligne : < <https://hal.archives-ouvertes.fr/hal-02166585/document> >, (consulté le 14/08/22)

²¹⁹ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?* Marseille, OpenEdition Press, 2019, p. 46

²²⁰ LOVELUCK Benjamin, « Les formes du pouvoir sur Internet », dans : Jean-François Dortier éd., *La Communication. Des relations interpersonnelles aux réseaux sociaux*. Auxerre, Éditions Sciences Humaines, « Synthèse », 2016, p. 324-335, consultable en ligne : < <https://www.cairn.info/--9782361063627-page-324.html> >, (consulté le 14/08/22)

²²¹ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 47

institutions françaises et à l'échelle du monde est au centre des enjeux de domination et de pouvoir de la géopolitique mondiale. Il y a des relations complexes entre les pratiques, les technologies et la concertation à l'échelle internationale. La mise en place d'un dépôt légal en France mais aussi dans le monde, les actions de l'IIPC, l'influence de la charte de l'UNESCO poussent à une gouvernance assez similaire des archives entre les pays, tout en activant des enjeux de pouvoirs.

Pour terminer, le web est une archive fragile qui évolue en fonction de la technologie mais aussi des changements de contenus. Une URL peut exposer des choses différentes d'une année sur l'autre. Il faut donc mettre en place des processus d'archivages réguliers et ciblés. Cependant nous avons vu plusieurs limites à l'archivage du web comme la multitude de domaines, la volatilité des pages web ou encore les barrières imposées par l'hébergeur ou les auteurs. Ainsi malgré cela l'INA et la BNF s'efforcent de mettre en place des politiques ambitieuses en matière d'archivage du web. Nous ne pouvons donc pas affirmer que les archives web sont le miroir du web français.

Chapitre 2 : Mise en pratique de l'archivage du web

Dans ce deuxième chapitre, nous allons analyser en détail le principe et le fonctionnement de l'archivage du web. Pour ce faire, nous allons étudier l'archivage semi-automatisé mis en place par l'INA ainsi que son périmètre. Nous mettrons en évidence les formes de collectes du web et d'indexation que l'INA a choisi de mettre en place. Puis nous nous pencherons sur la problématique inhérente aux archives sur tout support, à savoir le stockage et la pérennisation des archives. Dans une seconde partie, nous entrerons en détail dans les techniques et outils pour une collecte spécifique, en particulier la manière dont les robots de l'INA collectent le web et les métadonnées, ainsi que le format DAFF utilisé et développé par l'INA pour répondre aux défis de l'archivage. Nous verrons par la suite la collecte des UGC (User Generated Content). Il s'agit du contenu créé par les utilisateurs que l'INA propose d'archiver également. Enfin, nous nous pencherons sur le cas de l'archivage de Twitter par l'INA ainsi que l'influence des réseaux sociaux dans l'archivage. Dans une dernière partie, nous étudieront l'aspect documentaire de l'archivage comme la mise en place de veille et de critère de sélection, avant de nous questionner sur la consultation des archives. Pour terminer notre deuxième chapitre, nous évoquerons la mise en place de notice pour la consultation des archives du web.

Partie 1 : Principe et fonctionnement de l'archivage web

A- Définition du périmètre de l'archivage

Le périmètre de l'archivage du web est assez large pour l'INA, comme nous l'avons souligné plus haut : l'INA a la charge de l'archivage du web portant sur l'audiovisuel français. C'est dans la continuité de l'archivage audiovisuel que la structure se charge des sites web. Ainsi l'entreprise conserve l'ensemble des sites du web français se rapportant aux médias français. Plus précisément, comme le souligne le site web de la structure, l'INA collecte :

« Les sites émanant des services des médias audiovisuels (chaînes publiques et privées), y compris les Services de médias audiovisuels à la demande (SMAD).

Les Web TV et Web radios.

Les sites principalement consacrés aux programmes radio et télé (sites consacrés aux émissions de flux, aux séries, sites de fans).

Les sites des organismes de l'environnement professionnel et institutionnel du secteur de la communication audiovisuelle. »²²²

Plus précisément, l'INA collecte les sites des services de médias audiovisuels, les chaînes publiques et les chaînes privées. Ainsi par exemple le site web de TF1 et l'ensemble des sites sous-jacents du groupe se retrouvent archivés par l'entreprise. Les sites émanant du groupe France Télévision sont aussi archivés. C'est également le cas pour les SMAD (Service de Médias Audiovisuels à la demande) qui sont définis ainsi par l'ARCOM comme étant :

« La vidéo à la demande ou encore la télévision de rattrapage, permettant de visionner, sur le poste de télévision ou l'ordinateur, de manière gratuite ou payante, des programmes audiovisuels au moment choisi par l'utilisateur »²²³

²²² Site web de l'INA, consultable en ligne : < <https://www.ina.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²²³ Définition juridique donnée par l'ARCOM, consultable en ligne : < <https://www.csa.fr/Informer/Comment-recevoir-la-television-et-la-radio/Comment-regarder-la-television/Les-SMAD> >, (consulté le 15/08/22)

L'INA archive donc les services français de vidéos à la demande aussi appelé SVOD, nous pouvons citer la plateforme « SALTO »²²⁴ dirigé par France Télévision, TF1 et M6 ou encore la plateforme « Cinétek »²²⁵. L'intégralité des contenus disponibles sur ces sites web est archivé par l'INA. L'entreprise archive également les web TV et les web radios, ce sont des programmes audiovisuels ou radiophoniques disponibles uniquement le web. Nous pouvons citer la web TV « Made in Marseille »²²⁶ que l'INA archive. Au-delà de l'audiovisuel, l'INA archive également le web se rapportant aux médias audiovisuels comme les sites de programmes TV, des sites consacrés à une émission TV particulière. Enfin l'entreprise archive les sites professionnels et institutionnels du secteur de la communication audiovisuelle. Avec ce panel de site web, l'INA peut fournir un corpus large portant sur le paysage audiovisuel français. Cependant la structure collecte aussi des données sur des plateformes internationales.

En effet l'INA collecte les vidéos des plateformes d'échanges de vidéos entre particuliers comme YouTube ou Dailymotion. Si Dailymotion ne pose pas de problème juridique car l'entreprise est française, la collecte de YouTube peut s'avérer éloignée de la politique de l'INA à savoir collecter l'audiovisuel français. Pour ce faire l'INA collecte uniquement les comptes et vidéos des auteurs français. On dénombre environ 9.000²²⁷ comptes d'auteurs français actifs sur YouTube qui sont archivés par la structure. En plus des vidéos sur différentes plateformes, l'INA collecte l'intégralité des podcasts des chaînes de Radio France ainsi qu'une sélection de 5.300²²⁸ podcast indépendants ou issus de médias concurrents à Radio France. Enfin, la collecte du web se fait aussi sur les réseaux sociaux et en particulier sur Twitter. L'INA s'est donné pour mission²²⁹ d'archiver les tweets relatifs à l'audiovisuel ainsi qu'à l'actualité des médias ou encore aux grands événements médiatiques comme les Jeux Olympiques. C'est grâce à un suivi des publications de près de 15.000²³⁰

²²⁴ Site web de la plateforme : < <https://actu.salto.fr> >, (consulté le 15/08/22)

²²⁵ Site web de la plateforme : < <https://www.lacinetek.com/fr> >, (consulté le 15/08/22)

²²⁶ Site web de la web TV « Made in Marseille » : < <https://madeinmarseille.net/web-tv/> >, (consulté le 15/08/22)

²²⁷ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²²⁸ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²²⁹ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²³⁰ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

comptes et 2.500²³¹ hashtags que la collecte est effectuée. Au 31 décembre 2020, 15.637²³² comptes étaient collectés pour un total de 1,94²³³ milliards de tweets collectés depuis 2014. En conclusion, le dépôt légal du web couvre à l'INA plus de 16.000 sites web consacrés aux médias et à l'audiovisuel pour un total de 104 milliards de versions d'URL. Le périmètre d'archivage du web est donc assez large et imposant à l'INA ce qui oblige l'entreprise à concevoir une collecte semi-automatisée.

B- Collecter et indexer

La collecte et l'indexation du web se font donc sur le périmètre défini plus tôt. Compte tenu de ce périmètre large qui peut être agrandi lors d'événements médiatiques, l'INA a décidé de mettre en place une collecte à la fois automatique et ponctuelle, dans une stratégie d'échantillonnage du web et non dans une stratégie de collecte de l'intégralité du web comme c'est le cas pour « Internet Archives ». La collecte semi-automatisée est organisée au moyen de robots. C'est sous l'autorité de la cellule de veille que les robots de l'INA vont collecter le web défini à l'avance. C'est à l'équipe de « Recherche et Développement » de développer les outils et les méthodes de collectes. Cette même équipe définit les formats et les outils de stockage. Enfin c'est à ce service de l'INA qu'incombe la fouille vers les archives du web. Pour archiver le web, en plus des outils et logiciels propriétaires de l'INA, la structure utilise des logiciels et applications en « open-source », c'est-à-dire des logiciels dont le code est consultable et modifiable par tous. Ainsi l'INA utilise le logiciel « Elastic Search »²³⁴, un logiciel permettant l'indexation et la recherche de données. Le logiciel fournit un moteur de recherche et d'analyse des données. En complément du logiciel « Elastic Search », la structure utilise le logiciel « Kibana »²³⁵ qui permet d'obtenir une interface pour naviguer et utiliser les données du logiciel « Elastic Search ». Enfin pour le stockage et le traitement de large volume, l'entreprise utilise le logiciel « Hadoop »²³⁶.

²³¹ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²³² Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²³³ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²³⁴ Site web de « Elastic Search » : < <https://www.elastic.co/fr/elasticsearch/> >, (consulté le 15/08/22)

²³⁵ Site web de « Kibana » : < <https://www.elastic.co/fr/kibana/> >, (consulté le 15/08/22)

²³⁶ Site web de « Hadoop » : < <https://hadoop.apache.org> >, (consulté le 15/08/22)

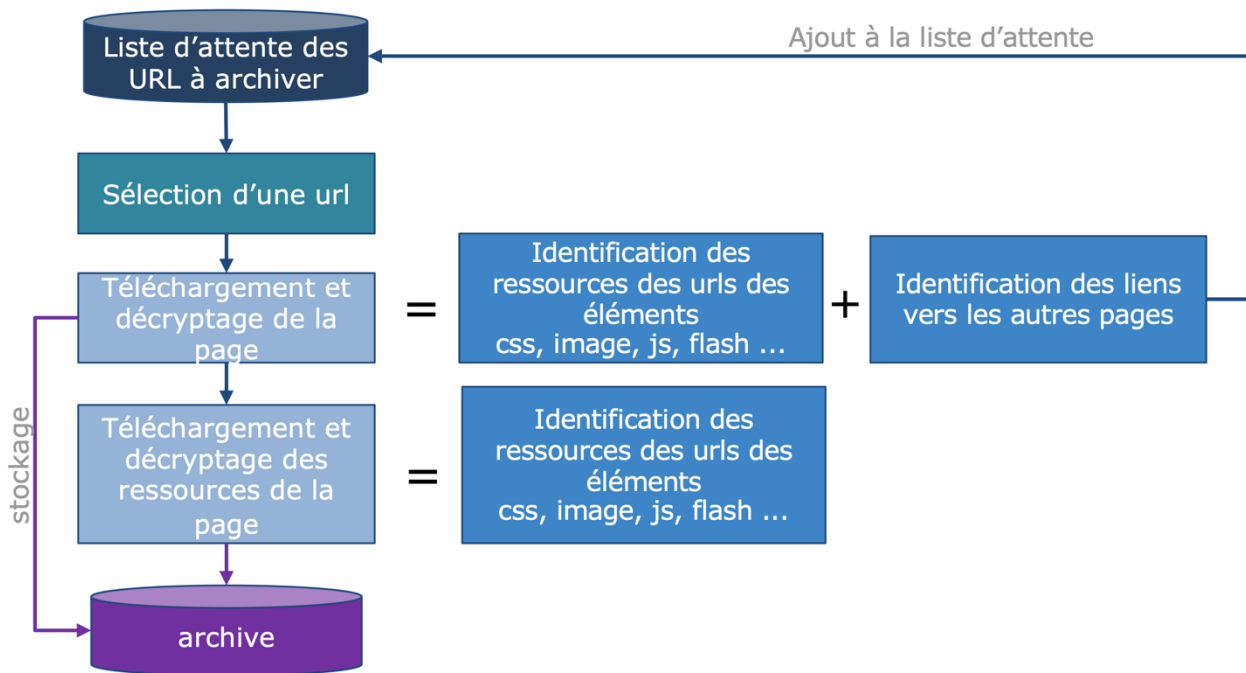
Du point de vue plus technique, l'INA utilise des robots pour explorer le web, le collecter et l'indexer. Les robots explorent automatiquement internet et collectent l'ensemble des informations contenues dans une page web. Le robot peut collecter à la fois des contenus textuels, des images, des vidéos et peut lire plusieurs langages de code. On utilise le terme de « crawler » pour désigner ces robots. Un « crawler » va parcourir une URL, il va automatiquement parcourir l'ensemble des liens HTML d'un site et les ressources vers lesquelles mènent les liens. Le « crawler » va alors simuler l'ensemble des interactions entre un poste client et le serveur, le tout en pouvant lire à la fois des formats en Javascript et également en HTML5.

Un robot fonctionne en deux phases : une phase d'analyse et une autre d'enregistrement. Dans un premier temps en parcourant l'ensemble d'un site, le robot va rassembler le maximum d'informations comme la structure de la page web, sa profondeur, son adresse, les liens pointant vers le site, les contenus ou encore la fréquence des mises à jour. Puis dans une seconde phase, le robot va indexer automatiquement l'archive, la classer et l'enregistrer dans une base de données. Le robot « crawler » va agir selon 4 principes mis en lumière par l'INA.²³⁷ Le premier est le principe de sélection qui définit quelles pages vont être téléchargées. Le second principe est celui de la re-visite qui définit la fréquence de vérification sur un site web s'il y a des changements. Le troisième principe est le principe de politesse qui détermine comment éviter la surcharge des pages web, souvent en délimitant l'accès aux crawler dans des plages horaires moins fréquentées par les utilisateurs. Enfin le dernier principe est celui de la coordination qui spécifie comment les robots vont se coordonner entre eux. Ainsi en utilisant ces principes, les robots de l'INA archivent le périmètre du web archivable par la structure. Cependant les robots sont incapables d'enregistrer certains contenus comme celle demandant une interaction avec un moteur de recherche ou encore les contenus demandant un paramètre de navigation. L'INA a développé ses propres robots « crawler » mais ils existent plusieurs robots comme les robots « Heritrix » pour « Internet Archives », « Googlebot » pour Google, « Applebot » pour Apple.

²³⁷ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

Le schéma ci-dessous reproduit d'un article²³⁸ de l'INA portant sur l'archivage du web met en lumière le fonctionnement des robots de collecte.

Fig. 7 : Fonctionnement des robots de collectes.²³⁹



Après identification d'une URL à archiver, le robot télécharge et décrypte la page. Pour cela il identifie les ressources des éléments des URL ainsi que sur l'ensemble des codes en différents langages. Puis, il recherche les différents liens vers d'autres pages et les place en liste d'attente. Enfin le robot télécharge l'ensemble des ressources de la page comme les images, les vidéos ou tout contenu multimédia.

Pour terminer, les robots de collecte sont gérés par un ordonnanceur. Il peut être humain ou numérique. Il est à noter que chaque robot archive un site à la fois. Plusieurs types de robots existent pour répondre à des besoins différents. Il y a des robots assez légers et simples à mettre en œuvre, des robots lourds utilisés uniquement pour des pages complexes et des robots dédiés utilisés pour un site particulier ou un type de sites, comme les sites

²³⁸ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

²³⁹ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

demandant beaucoup d'interactions ou encore les réseaux sociaux. Le robot le plus utilisé est un robot « phagosite »²⁴⁰ qui archive un maximum de pages web simples. Le second robot utilisé est le robot « crocket »²⁴¹, il s'agit d'un robot spécialisé capable d'exécuter du langage Javascript et apte à détecter les événements utilisateurs comme le clic dans un menu ou encore le défilement du page. Enfin est utilisé, dans de plus rares cas, le robot « Lap », un robot manuel commandé par un humain qui est utilisé pour archiver les sites demandant une interactivité très forte comme les sites interactifs ou encore le web documentaire.

En conclusion, l'INA grâce à l'utilisation des robots mis en place par une cellule de veille archive le web. Thomas Chaimbault²⁴² met en évidence la chaîne de traitement différencié selon les informations du site web ainsi :

« L'INA collecte quant à elle également les sites en utilisant des robots sous l'autorité d'une cellule de veille. La copie de chaque site est ensuite préparée pour une indexation séparée des contenus du site et de leur structure, les différentes images qui composent une page, la page elle-même, les films, les musiques, etc. étant traités indépendamment. »²⁴³

L'ensemble des éléments d'une page se retrouve archivé et collecté puis indexé immédiatement par les robots de l'INA. Après la réception et la vérification de l'archive qui prend la forme de vérification de l'intégrité, de la conformité des contenus, des volumes ainsi que des formats, l'INA établit un contrôle de qualité des archives avant de les stocker dans un système de stockage adapté. Ce système d'archivage par échantillonnage semi-automatisé doit par la suite faire l'objet d'une pérennisation de son stockage.

²⁴⁰ Article de l'INA, consultable en ligne : < https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA_-_L_archivage_du_Web_-_Cours_Master_HAV_-_Octobre_2019_I_0.pdf >, (consulté le 14/08/22)

²⁴¹ Article de l'INA, consultable en ligne : < https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA_-_L_archivage_du_Web_-_Cours_Master_HAV_-_Octobre_2019_I_0.pdf >, (consulté le 14/08/22)

²⁴² CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 36, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

²⁴³ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 36, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

C- Stocker et pérenniser

Comme pour toutes les archives, l'une des problématiques de l'archive web est le stockage de celle-ci. Si pour les formats papiers, une bonne conservation suffit à conserver les archives pendant plusieurs années dans de bonnes conditions, il est plus difficile de maintenir une archive web dans de bonnes conditions. Le spécialiste de la conservation du numérique Howard Besser²⁴⁴ identifie cinq problèmes techniques liés à la conservation :

« Le problème majeur est l'installation et la maintenance d'une infrastructure et de l'expertise nécessaire pour rendre les objets numériques lisibles.

Le décodage de logiciels de compression ou de services de protection des pages web.

La préservation du contexte qui rend aux informations leur signification, tels que les liens hypertextuels.

La définition de standard de bonnes pratiques et de politiques d'acquisition qui définissent les limites de la collection, sa provenance et son authenticité.

Des problèmes de migration à propos de la façon dont l'expérience des pages web change lorsqu'elles sont transférées sur de nouveaux appareils. »²⁴⁵

Dans un premier temps stocker une archive web nécessite une infrastructure et une maintenance importantes. Dans le cas de l'INA, l'entreprise détient une infrastructure moderne et étendue ainsi qu'une équipe dédiée au monitoring et l'analyse des serveurs de stockage. Une équipe est également chargée d'établir la stratégie de préservation et de consultation sur le long terme. Les équipes de l'INA s'efforcent également de mettre en contexte l'archive via l'instauration de notices comme nous le verrons plus tard. Cependant comme l'indique Howard Besser, la problématique de migration de l'archive sur des nouveaux appareils au fur et à mesure des évolutions peut s'avérer problématique pour l'INA, en ce sens

²⁴⁴ Howard Besser est spécialiste de la préservation numérique, des bibliothèques numériques et de la préservation des films et des vidéos. Il est professeur d'études cinématographiques et directeur fondateur d'un programme d'études supérieures de la Tisch School

²⁴⁵ BESSER Howard, *Digital Longevity*, dans SITIS Maxine *Handbook for Digital Projects : a Management Tool for Preservation*, Andover, Mass : Northeast Document Conservation Center, 2000 cité dans Lyman Peter, *Archiving the world wide web* dans In Council on library and Information resources, *Building a National Strategy for Preservation : Issues in Digital Media Archiving*, Berkeley, 2002, consultable en ligne : < <https://www.clir.org/pubs/reports/pub106/Web/> >, (consulté le 15/08/22)

la structure s'efforce de maintenir du matériel plus ancien pour la lecture d'archives anciennes. Par ces problèmes mis en lumière par Howard Besser, il apparaît qu'au-delà de simplement préserver une page web, il convient de maintenir la possibilité de lecture de ces pages. Cette problématique est d'autant plus importante dans un monde dans lequel les produits numériques ne sont pas faits pour durer. Ainsi pour l'INA, la pérennisation de l'archive web passe par la préservation de l'archive mais aussi de son environnement matériel et logiciel pour pouvoir toujours la lire. Il est souligné par Thomas Chaimbault²⁴⁶ que la mise en place de documentation décrivant les configurations requises sont mises en avant par l'INA dans cette perceptive de pérenniser l'archive.

Dans une perspective plus technique, l'INA a collecté plus de 100 milliards²⁴⁷ de ressources web avec un rythme d'accroissement de 10 milliards par an, représentant environ dix pétaoctets de stockage. Un pétaoctet est une unité de mesure de la capacité de stockage de données. Un pétaoctet équivaut à dix puissance quinze octet. En d'autres termes un pétaoctet égal plus d'un million de gigaoctet. Il s'agit d'une masse importante de stockage que l'INA doit déployer pour collecter et pérenniser ces archives du web. C'est lors d'un agrandissement de 4.0000 mètres carrés des bâtiments de l'INA à Bry-sur-Marne, terminé en 2021, que le président directeur général de l'entreprise Laurent Vallet a tenu ses propos :

« Ce projet immobilier incarne pleinement les ambitions d'une entreprise pionnière du digital et de la data. Il symbolise la transformation numérique de l'INA autant qu'il vient confirmer son fort ancrage territorial et sa volonté de jouer un rôle pivot pour l'émergence d'un futur pôle dédié à l'image et au son »²⁴⁸

Dans le cadre de ce projet, le stockage des archives du web a été agrandi et amélioré. Pour stocker les archives du web, l'INA bénéficie d'un centre de données aussi appelé « data

²⁴⁶ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 36, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

²⁴⁷ Vidéo promotionnelle de l'INA, mise en ligne le 18 décembre 2020, consultable en ligne : < <https://www.youtube.com/watch?v=Aji9HpPWhQY> >, (consulté le 16/08/22)

²⁴⁸ Propos de Laurent Vallet rapporté par Florent Bascoul dans l'article du quotidien d'information « Citoyens », consultable en ligne : < <https://94.citoyens.com/2019/premiere-pierre-symbolique-pour-l'extension-de-l'INA-a-bry-sur-marne,26-02-2019.html> >, (consulté le 16/08/22)

center ». Ce centre permet un stockage de plusieurs pétaoctets de façon entièrement sécurisée. De la même façon que dans un magasin d'archive, le centre de données a une hygrométrie et une température régulées et vérifiées en permanence. Ce centre permet un stockage massif des archives du web et assure un faible risque environnemental ou de sécurité des données. La figure suivante met en évidence le centre des données.

Fig. 8 : Centre de donnée de l'INA à Bry-sur-Marne²⁴⁹



Avec ce bâtiment et avec une politique de pérennisation des archives du web par la préservation des matériels nécessaires à la lecture des archives, l'INA a développé une stratégie rigoureuse dans le stockage des archives du web. Il paraît pertinent de signaler que la stratégie de stockage de l'INA devra évoluer au fur et à mesure des avancées technologiques pour toujours garantir une stabilité des supports de conservation numérique qui sont par nature de durée limitée. Ainsi les enjeux de stockage seront toujours d'actualité à l'INA. Si aujourd'hui le stockage ne semble pas être un frein à l'archivage du web par l'INA cela peut s'avérer être le cas dans les années ou décennies à venir. Il convient dès à présent d'aborder les outils et la collecte spécifiques des User Generated Content (UGC) et du réseau social Twitter.

²⁴⁹ Centre de données de l'INA à Bry-sur-Marne, issue d'une vidéo promotionnelle, mise en ligne le 13 octobre 2021, consultable en ligne : < <https://www.youtube.com/watch?v=FIASGqFxLOk&list=PLG7FWgNsg9e-9a1u1SEft96eNVtLUSMYc&index=19> >, (consulté le 16/08/22)

Partie 2 : Techniques, outils et mise en place de collectes spécifiques

A- Utilisation des métadonnées et du format DAFF

Lors de la collecte des archives du web, il est évident que l'INA collecte un maximum de métadonnées fournissant des informations supplémentaires. Les métadonnées sont définies comme des données informant sur une autre donnée. Il s'agit d'un ensemble structuré d'informations décrivant une ressource. Ainsi l'auteur Jérôme Poirot dans l'ouvrage collectif « *Dictionnaire du renseignement* » définit les métadonnées en ces termes :

« Une métadonnée est une donnée sur une donnée, ou encore une information caractérisant une autre information. « Meta », en grec, signifie ce qui « dépasse » ou ce qui « englobe ». »²⁵⁰

Une métadonnée permet alors d'enrichir les informations sur une donnée. Nous pouvons citer l'exemple d'un ouvrage, dans ce cas les données sont le contenu de l'ouvrage tandis que les métadonnées sont : l'éditeur, l'auteur, la date et le lieu d'édition ou encore le numéro ISBN. Il en va de même pour l'archive web qui en plus de son contenu peut contenir des métadonnées. Les métadonnées vont varier selon l'archive web collectée. Ainsi dans le cas d'un archivage d'une image numérique récupérée sur un site web, l'image va contenir, s'il s'agit d'une image prise avec un smartphone ou un appareil photo numérique, des métadonnées du standard EXIF²⁵¹ qui fournissent des informations sur l'image comme les dimensions, la focale ou encore les canaux de couleur. La photo fournit également des métadonnées appartenant au standard IPTC²⁵² décrivant le contenu de la photo comme la

²⁵⁰ POIROT Jérôme, « Métadonnées – Data mining », dans MOUTOUH Huges, *Dictionnaire du renseignement*, Paris, Perrin, 2018, p. 545, consultable en ligne : < <https://www.cairn.info/dictionnaire-du-renseignement--9782262070564-page-545.htm?contenu=article> >, (consulté le 16/08/22)

²⁵¹ EXIF (Exchangeable image format file) ensemble de métadonnées prise automatiquement par l'appareil lors de la prise de vue

²⁵² IPTC (International Press Telecommunications Council) permet à l'utilisateur d'ajouter manuellement des métadonnées à une image

date de prise de vue, l'auteur ou encore l'emplacement. Dans le cas de métadonnées intégrées dans une vidéo numérique, les informations sont la durée, les sous-titres enregistrés ou encore les dimensions de l'image. Enfin les métadonnées encodées dans une page web sont la date de création de la page mais d'un point de vue technique, nous pouvons trouver les informations sur l'encodage, la transmission de la page ou encore sur la compression de la connexion HTTP. Ainsi la collecte de l'ensemble de ces métadonnées permet de fournir les données techniques sur l'archive mais aussi de contexte historique de l'archive. De plus, il est à noter que les métadonnées des archives du web peuvent également contenir des informations juridiques sur l'archive ou encore sur la composition de la page. Ces informations peuvent venir compléter l'archive. Le web étant une entité en évolution constante, il est pertinent de capturer l'instabilité du web pour rendre compte fidèlement de son évolution. Pour ce faire, la collecte des métadonnées en même temps que l'archive permet de capturer à un moment précis du web. De plus lors de l'indexation, les robots « crawler » marquent leur passage par des métadonnées. Les métadonnées permettent alors également une mise en contexte de l'archivage de l'archive web. La collecte des métadonnées peut s'apparenter à l'utilisation d'un outil lors de l'archivage. C'est également le cas du format DAFF mis en place par l'INA.

Le format DAFF est une format propriétaire de l'INA utilisé pour l'archivage du web. Le format DAFF (Digital Archiving File Format) est un format « container » c'est-à-dire qu'il s'agit d'un format permettant d'intégrer plusieurs flux de fichiers dans un seul fichier. Au même titre qu'un fichier en format « ZIP » permettant de le compresser et d'incorporer plusieurs fichiers et informations tout en réduisant sa taille ou encore le format « AVI » contenant du son, des images et des sous-titres, le format DAFF permet d'encoder plusieurs informations en prenant le moins d'espace possible. En effet, la grande force du format DAFF est de ne prendre que très peu d'espace de stockage, et il prend encore moins de place que le format ARC compressé utilisé par « Internet Archive » pour stocker le web par la structure américaine. Le format DAFF permet une écriture automatique des métadonnées incorporées lors du traitement de l'archive. De plus, le format DAFF permet le contrôle d'intégrité d'une fonction de hachage. Le hachage de l'archive permet de calculer une empreinte grâce à des données, ainsi il est possible d'identifier de manière unique un contenu numérique. L'INA

utilise le plus souvent un hachage en « SHA256 »²⁵³ Le hachage du contenu numérique permet par la suite de comparer deux empreintes d'archives du web ce qui sera utile pour la structure notamment pour le stockage des données. Effectivement si une signature d'un fichier est présente, alors la structure sait pertinemment que le fichier est présent dans sa base de données. Il n'est alors pas nécessaire d'enregistrer une nouvelle fois l'élément. Ce hachage est pertinent dans le sens où si un fichier est trop volumineux comme c'est le cas pour les archives web, il est plus rapide de comparer les empreintes que d'analyser l'ensemble du document. Le format propriétaire de l'INA a été conçu pour répondre aux exigences de l'INA en termes d'archivage. Pour l'auteur Mathieu Gallet, ancien président directeur-général de l'INA de mai 2010 à février 2014 :

« Le format de stockage DAFF développé par l'INA permet ainsi un facteur d'économie en stockage à 91%. Grâce à ces outils développés en interne, en toute autonomie vis-à-vis d'Internet Archive, l'INA est ainsi en mesure de s'adapter avec souplesse aux évolutions du web, tout en s'assurant le maintien d'une comptabilité avec les autres formats d'archivage en vigueur dans le monde. »²⁵⁴

En conclusion, l'INA collecte les métadonnées pour fournir des informations supplémentaires lors de l'archivage. Ces métadonnées peuvent être créées directement par l'entreprise lors de l'archivage. Les robots « crawler » en archivant permettent automatiquement d'ajouter des informations de contexte de la collecte. Enfin grâce à des outils mis en place par la structure comme le format DAFF, l'entreprise a la main sur l'ensemble des aspects, de l'archivage jusqu'à l'adaptation du format de stockage aux besoins de l'entreprise. Il convient à présent d'étudier la collecte spécifique des UGC.

²⁵³ Article de l'INA, consultable en ligne : < https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA_-_L_archivage_du_Web_-_Cours_Master_HAV_-_Octobre_2019_I_0.pdf >, (consulté le 14/08/22)

²⁵⁴ GALLET Mathieu, « La France et l'Europe face à l'enjeu du *Big Data*. L'exemple de la collecte et du traitement des données médiatiques », *Géoeconomie*, 2014, numéro 69, p. 7-23, consultable en ligne : < <https://www.cairn.info/revue-geoeconomie-2014-2-page-7.htm> >, (consulté le 16/08/22)

B- La collecte des UGC

Les UGC ou « User Génératif Content » sont des contenus mis en ligne par des utilisateurs. Les UGC peuvent prendre des formes très différentes, et être simplement des commentaires postés par des utilisateurs d'un site web ou d'une plateforme sur une vidéo ou une publication. Ainsi les commentaires de vidéos YouTube ou les commentaires d'une publication sur les réseaux sociaux peuvent être considérés comme des UGC. Les UGC ne sont pas uniquement des contenus sous formes d'écrit. Ils peuvent intervenir sous forme vidéo ou sonore. C'est notamment le cas pour la plateforme YouTube que l'INA collecte. Effectivement comme nous l'avons vu, l'INA collecte environ 9.000²⁵⁵ comptes d'auteurs français actifs sur YouTube. Ainsi depuis 2010 l'INA a collecté plus de 24 millions²⁵⁶ de contenus issus des UGC. On y trouve notamment des contenus provenant de YouTube, Facebook, Twitter, Vimeo, Vine, SoundCloud, Dailymotion, Wat, France TV Pluzz ou encore ArteConcert. Même si l'INA peut archiver des commentaires ou des messages, elle se concentre particulièrement sur la production multimédias que les utilisateurs mettent en ligne sur ces plateformes. Plusieurs problèmes apparaissent lors de la collecte de ces fichiers. Tout d'abord, il s'agit de plateformes et de sites différents avec des systèmes d'encodage de l'information, cela peut poser des problèmes de lecture aux « crawler ». De plus les vidéos sont souvent dans des formats différents ce qui peut également compliquer la collecte. Ces contenus peuvent être volumineux et prendre du temps à être traités. Enfin ce genre de contenu n'est pas mis à jour car étant destiné à une utilisation immédiate. La plus grande limite à la collecte des UGC est que ces contenus sont difficiles à détecter pour les robots de l'INA comme pour les documentalistes de l'entreprise. Le flux d'une chaîne de télévision ou d'un web radio est d'autant plus simple à capter que l'on met en place des protocoles. Dans le cas de contenus créés par les utilisateurs, l'INA suit plusieurs milliers de comptes actifs postant du contenu sur le web et a mis en place des protocoles pour archiver ses comptes régulièrement. Cependant dans les UGC, il y a des contenus mis en ligne de manière ponctuelle par des utilisateurs non

²⁵⁵ Site web de l'INA, consultable en ligne : < <https://www.INA.fr/institut-national-audiovisuel/collections-audiovisuelles/le-web-media> >, (consulté le 15/08/22)

²⁵⁶ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

identifiés par la structure. Il est donc assez difficile²⁵⁷ pour l'entreprise de détecter l'ensemble des contenus susceptibles d'être archivés. Pour collecter les UGC, l'INA a plusieurs canaux. Tout d'abord le canal de collecte des UGC par la collecte d'une page web. L'ensemble des ressources d'une page web, dont les UGC, vont être collectées par les robots. Le second canal suit l'archivage des tweets. Ainsi lors de la captation de Twitter, les UGC sont collectés. Enfin le dernier canal utilise des « API » de YouTube ou Dailymotion. Une « API » est un logiciel qui permet de connecter un logiciel à un autre logiciel pour échanger des données ou des informations. La CNIL (Commission nationale de l'informatique et des libertés) définit ainsi le terme « API » :

« Une API (*application programming interface* ou « interface de programmation d'application ») est une interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités. Les API offrent de nombreuses possibilités, comme la portabilité des données, la mise en place de campagnes de courriels publicitaires, des programmes d'affiliation, l'intégration de fonctionnalités d'un site sur un autre ou l'*open data*. »²⁵⁸

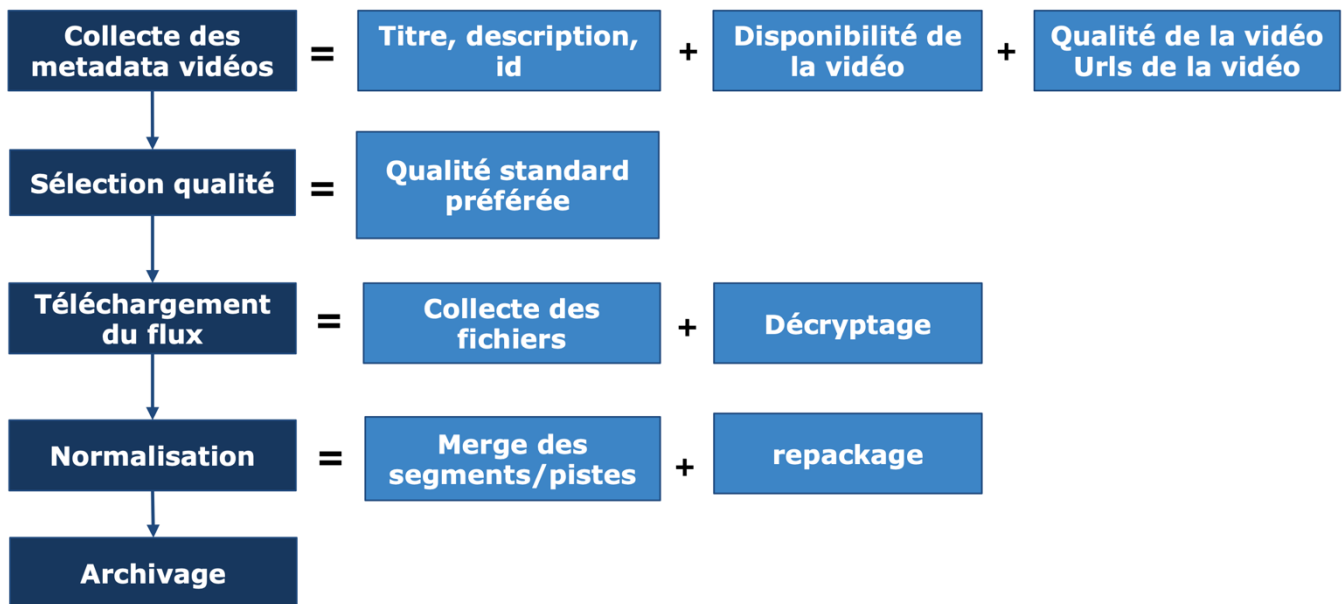
Les « API » permettent donc de faire communiquer certaines informations à l'INA ou à d'autres structures. Cela permet alors à l'INA de toujours être tenu au courant lors des mises en ligne de vidéos.

Pour collecter les UGC, l'INA utilise un protocole défini dans la figure suivante. Tout d'abord les robots collectent les métadonnées d'une vidéo, cela peut être le titre, la description mais aussi la qualité de la vidéo ou encore l'URL et l'auteur. Le robot, comme le fera un utilisateur, sélectionne la qualité de la vidéo, puis la collecte du contenu est faite, c'est lors de cette phase que le téléchargement de la vidéo est effectué. Enfin le robot passe à la phase de normalisation. C'est lors de cette phase que le robot fractionne et segmente l'image pour la calculer avant de fusionner l'image pour en rendre une vidéo utilisable et visionnable.

²⁵⁷ Article de l'INA, consultable en ligne : < https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA_-_L_archivage_du_Web_-_Cours_Master_HAV_-_Octobre_2019_I_0.pdf >, (consulté le 14/08/22)

²⁵⁸ Définition d'une API donnée par la CNIL, consultable en ligne : < <https://www.cnil.fr/fr/definition/interface-de-programmation-dapplication-api> >, (consulté le 16/08/22)

Fig. 9 : Principe de collecte des UGC²⁵⁹



Les UGC sont assez spécifiques lors de l'archivage du web car ne dépendent pas des groupes audiovisuels. Les UGC constituent tout de même un contenu considérable pour l'INA qui s'efforce de rendre compte du web. Depuis 2014, l'INA archive de façon systématique des contenus UGC sur la plateforme Twitter.

C- L'archivage de Twitter

Twitter au même titre que Facebook ou Instagram prend une place toujours plus importante dans notre société du XXIème siècle. A ce titre, depuis 2014, l'INA archive les tweets et comptes de la plateforme qui portent sur des contenus audiovisuels. L'ampleur de Twitter est tellement importante que la bibliothèque du Congrès aux Etats-Unis archive l'ensemble de Twitter. En effet, comme le soulignent les auteurs Musiani Francesca, Paloque-Berges Camille, Schafer Valérie Et Thierry Benjamin²⁶⁰ :

« La Bibliothèque du Congrès (LoC) aux États-Unis a ainsi passé un accord en 2010 avec l'entreprise Twitter pour récupérer tous les tweets émis depuis 2006 et

²⁵⁹ Article de l'INA, consultable en ligne : < <https://isor.panthéonsorbonne.fr/sites/default/files/inline-files/INA - L archivage du Web - Cours Master HAV - Octobre 2019 I 0.pdf> >, (consulté le 14/08/22)

²⁶⁰ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 39

poursuivre cette conservation. Reste qu'à ce jour, cette collection n'est pas encore accessible pour les chercheurs et soulève diverses questions, amenant même la LoC à revenir sur son projet d'exhaustivité pour se concentrer sur un périmètre plus restreint et sélectif de collecte. »²⁶¹

La bibliothèque du congrès américaine, en voulant archiver l'ensemble des tweets, s'est heurté à un problème de logiciel et de budget. Pour pallier ce problème, la bibliothèque a décidé d'archiver uniquement les tweets sur critères sélectifs. Ainsi les tweets archivés seront thématiques et mis en relation avec des événements majeurs comme par exemple les élections présidentielles. Ces objectifs revus à la baisse sont la conséquence d'une explosion du nombre de tweet entre 2010 et 2012 passant de 140 millions²⁶² en février 2010 à 500 millions²⁶³ en octobre 2012 comme le soulignent les auteurs de l'ouvrage « *Qu'est-ce qu'une archive du web* ». Comme évoqué précédemment, l'INA archive une partie de Twitter mais uniquement sur son périmètre de collecte, l'audiovisuel. Ainsi depuis 2014 l'INA a archivé plus de 1 milliard²⁶⁴ de tweets. Si l'INA est d'abord présente sur Twitter et seulement depuis 2018 sur Facebook alors qu'il est plus utilisé par les Français et dispose d'une diversité sociologique plus forte, c'est parce que les contenus ne sont pas visibles de tous. L'archivage de Facebook impose des difficultés techniques. De plus grâce à l'« API » publique de Twitter, l'INA peut davantage récupérer les informations et données de Twitter de façon automatique, cela est utilisé par l'INA pour faire de la veille. Cependant cette « API » est limitée, elle ne calcule que 400 hashtags, 5.000 utilisateurs et seulement 1% des tweets publiés dans les dernières minutes et 3.200 derniers tweets. De plus l'« API » ne permet d'avoir des données sur que sur les sept derniers jours. L'ouverture de cette « API » permet tout de même à l'INA d'avoir des données de manière plus simplifiée que sur les réseaux sociaux du groupe « Meta » (Facebook, Instagram et WhatsApp). Lors de grands événements médiatiques, l'INA met en place de grandes collectes de tweets, cela a été le cas pour les attentats de 2015 en France.

²⁶¹ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 39

²⁶² MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 39

²⁶³ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 39

²⁶⁴ MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie et THIERRY Benjamin, *Qu'est-ce qu'une archive du web ?*, Marseille, OpenEdition Press, 2019, p. 39

Dans un entretien réalisé par Valérie Schafer en 2017, Zeynep Pehlivan, membre de l'équipe du dépôt légal du web à l'INA, explique l'archivage à l'occasion d'évènements en ces termes :

« Nous avons poursuivi les collectes sur les attentats après 2015, par exemple à Nice à l'été 2016. Nous avons aussi des archives relevant d'attentats qui ont eu lieu en Europe, à Bruxelles, Londres ou Manchester. En effet s'ils ne se sont pas passés en France, ils ont été profondément relayés par les médias français et sont entrés rapidement dans les *trends* de Twitter, car les Français ont réagi. Ces tweets font partie intégrante du contexte médiatique et permettent en outre au chercheur de mettre en perspective les tweets de notre cœur de corpus du dépôt légal. Par contre, on ne fait pas des collectes pour tous les attentats dans le monde, seulement pour ceux qui ont un écho fort en France, en particulier dans le monde de l'audiovisuel, qui est notre périmètre dans le cadre du dépôt légal du Web »²⁶⁵

Aujourd'hui l'INA a mis en place un archivage automatique de Twitter, lorsqu'un hashtag arrive en tendance, il est automatique archivé. Il est à noter que le nombre de retweets n'est pas archivé par Twitter. Lors de la collecte d'un tweet, la majorité des informations collectées n'est pas le contenu du tweet mais l'ensemble des métadonnées accompagnées. L'auteur Raffi Krikorian, employé de l'« API » de Twitter, a mis en évidence dans une infographie²⁶⁶ les métadonnées comprises dans un tweet. Nous trouvons donc un identifiant unique à chaque tweet, la date de création du tweet, le nom de l'auteur du tweet, l'URL du compte de l'auteur, la localisation de l'auteur, la date de création du compte, le nombre de followers, la langue sélectionnée par l'auteur ou encore le terminal de l'envoi du tweet. Toutes ces informations peuvent donc être archivées par l'INA.

En conclusion, par la mise en place d'outils spécifiques comme le format DAFF ou encore la collecte des métadonnées, l'INA peut mettre en place une stratégie de collecte du web. Cette stratégie de collecte permet d'archiver les contenus mis en ligne par des utilisateurs (UGC). De plus, la structure archive les tweets depuis 2014, et cet archivage permet aux chercheurs d'étudier des phénomènes de société sur les réseaux sociaux ou

²⁶⁵ Entretien réalisé par Valérie Schafer en 2017, *Gazette des archives*, numéro 250

²⁶⁶ Infographie disponible en ligne : < https://affordance.typepad.com/mon_weblog/2010/05/anatomie-dun-tweet-.html >, (consulté le 16/08/22)

encore l'histoire de l'utilisation des réseaux sociaux. Cependant cet archivage n'est pratiqué que sur le périmètre de l'INA, le monde de l'audiovisuel français. Il convient à présent d'analyser la veille mise en place par l'entreprise ainsi que la création d'une notice pour les archives du web.

Partie 3 : Les aspects documentaires de l'archivage web : de la veille à la création d'une notice

A- Mise en place d'une veille

Pour mener à bien ses missions d'archivage de l'ensemble des contenus de l'audiovisuel français, l'INA met en place une veille documentaire pour sélectionner les sources à collecter. Cette veille prend la forme d'une veille informationnelle. Jean Pierre Lardy, professeur à l'Unité de Régionale de Formation à l'Information Scientifique et Technique (URFIST) propose la définition suivante pour le concept de veille informationnelle :

« La veille informationnelle est l'ensemble des stratégies mises en place pour rester informé, en y consacrant le moins de temps possible, par l'utilisation de processus de signalement automatisés dont les flux RSS. »²⁶⁷

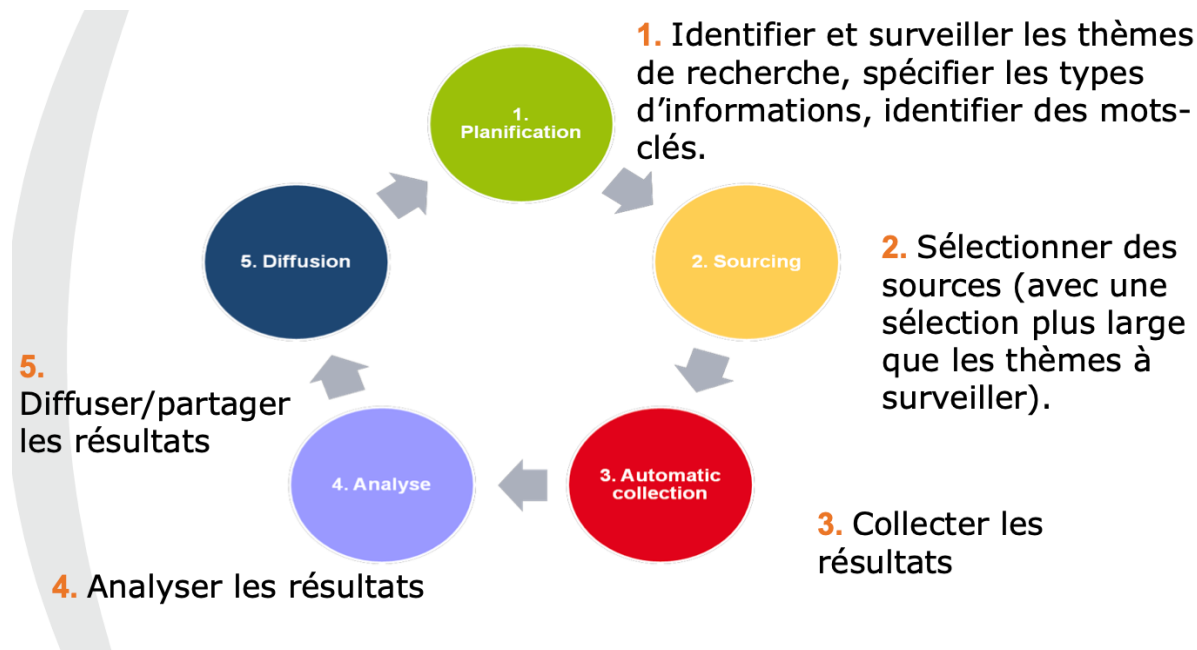
La mise en place d'une veille par l'INA va permettre à l'entreprise de rester informée sur les nouvelles sources à collecter. Il s'agit pour l'entreprise d'être toujours à la pointe de l'information. Organiser une stratégie de veille va permettre en plus d'obtenir des informations, de faire remonter l'information jusqu'à l'INA et ainsi de gagner du temps, dans la recherche d'information et dans la collecte. L'INA va s'appuyer sur les 5 étapes d'une veille documentaire définie par Isabelle Rey, dans son article *« Documentary watch on the web »*²⁶⁸ publié en novembre 2011. La première étape est la planification qui permet à la structure d'identifier et de surveiller les thèmes à collecter. La seconde étape est la sélection de différentes sources probables à partir desquelles l'INA pourra aller collecter des archives. La troisième étape est la collecte des résultats qui seront analysés dans une quatrième étape. Enfin dans la dernière, l'information de la demande d'archive pourra être communiquée. Cette grille de lecture de la veille informationnelle est utilisée principalement dans la

²⁶⁷ Citation mise en avant par le site de la bibliothèque de l'université d'Aix-Marseille, consultable en ligne : < <https://bu.univ-amu.libguides.com/veille> >, (consulté le 17/08/22)

²⁶⁸ REY Isabelle, *Documentary watch on the web*, novembre 2021

recherche scientifique mais elle s'applique également à la veille mise en place par l'INA. Ainsi le schéma ci-dessous illustre le processus en cinq étapes.

Fig. 10 : Les cinq étapes de la veille informationnelle²⁶⁹



L'INA utilise deux méthodes de veille, la première méthode est celle de la veille en « Pull ». Il s'agit d'une recherche d'information sur le web par des humains. Cette méthode permet d'écarter facilement les résultats non pertinents cependant elle demande beaucoup de moyens humains, elle est longue et chronophage. Le principal point positif de cette méthode est que l'ensemble du processus est maîtrisé car cette méthode ne peut pas être automatisée. La structure met en place également une veille avec la méthode de « Push », le principe est alors que l'information va être poussée automatiquement vers l'INA qui n'aura plus qu'à la traiter. Cela permet un gain de temps et de ressources, tout en proposant en permanence une actualité actualisée. Il convient tout de même de mettre en place des critères précis pour que le flux d'information ne soit pas parasité par des informations non pertinentes. Pour mener à bien cette stratégie, l'INA utilise des logiciels comme « Hootsuite »²⁷⁰ ou « Talkwalker Alert »²⁷¹. Ces logiciels vont permettre la gestion de plusieurs flux notamment sur les réseaux sociaux. Cependant des points négatifs sont inhérents à l'utilisation de ces logiciels. Le premier

²⁶⁹ REY Isabelle, *Documentary watch on the web*, novembre 2021

²⁷⁰ Site web « Hootsuite » : < <https://www.hootsuite.com/fr> >, (consulté le 17/08/22)

²⁷¹ Site web « Talkwalker Alert » : < <https://www.talkwalker.com/fr/alerts> >, (consulté le 17/08/22)

est l'obligation d'un paramétrage précis des outils pour éviter au maximum le bruit et les résultats non pertinents. De plus l'INA, en choisissant de se fier à ces logiciels, n'a qu'une faible maîtrise sur le processus de recherche. Enfin cette veille en « Push » est utilisée sur les réseaux sociaux et en particulier sur les comptes importants du milieu de l'audiovisuel français, comme ceux de chaînes de télévision, des personnalités de l'audiovisuel français, des émissions ou encore les comptes de fans. Le processus est également efficace pour la surveillance des « tendances ». En conclusion, l'INA met en place une politique de veille informationnelle pour être à la pointe de l'information et ainsi ne passer à côté d'aucun phénomène. La veille est quotidienne à l'INA. Ce système fait gagner du temps et des ressources en externalisant la veille avec des logiciels mais en ne maîtrisant pas la technique et le processus de recherche. La veille est primordiale dans le traitement documentaire de l'archive du web, au même titre que la mise en place d'une notice.

B- Permettre la consultation d'une archive web

Après que les robots « crawler » ont collecté le web, le serveur central, que l'on appelle ordonnanceur, récupère une copie du site web en séparant le contenu du site comme le texte, les vidéos, le son ou l'image, de sa structure qui est l'adresse URL et les métadonnées. Lors de la consultation d'une archive web, l'enjeu est de rendre vivante l'archive en essayant de restructurer l'interactivité du site. C'est grâce à la séparation entre le contenu et la structure que l'interaction est rendue disponible. Si la séparation n'est pas réalisée, lors d'une modification de la structure de la page, il peut apparaître des dissociations dans un site web archivé. Lorsqu'un utilisateur fait appel à une archive du web, le serveur de consultation²⁷² met en relation la base de données contenant l'ensemble de la structure du site et la base des contenus qui contient les textes, vidéos, sons et images. Il est à noter que les feuilles de style CSS ainsi que la structure de la page en HTML ou en Javascript sont archivées dans la base de données de la structure de la page web.

Ainsi Sophie Gebeil, spécialiste de l'histoire du web, dans son ouvrage « Website Story, Histoire, mémoires et archive du web » explique en ces termes le principe de consultation d'une archive web :

²⁷² GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Editions, 2021, p. 65

« La consultation de l'archive est l'aboutissement de la chaîne de traitement. La séparation entre les contenus et les informations structurelles permet de reconstituer l'interactivité d'un site à une date donnée. Ainsi via un navigateur de consultation, l'utilisateur saisit sa recherche. Celle-ci est transmise au serveur de consultation qui interroge la base de données (informations sur la structure du site et de la date de capture) et la banque de contenus. Il ou elle obtient alors une reconstitution des sites à la date requise, avec les hyperliens et les contenus présents sur la page. »²⁷³

La consultation d'une archive du web se fait donc de manière interactive. Ce n'est pas uniquement une page web qui est reconstituée mais l'ensemble ou du moins le maximum d'interactions que l'on pouvait retrouver sur la page. Ainsi l'archivage des liens hypertextes ainsi que de la profondeur du site permet de retrouver une interactivité forte. Cette interactivité forte est un point unique de l'archive web. Si une archive papier ou audiovisuelle peut être considérée comme une unité intellectuelle fixe, une archive web a une limite bien plus floue. En effet, un site peut renvoyer à un autre site. Ainsi nous ne sommes plus dans le cas d'une unité intellectuelle fixe et délimitée. L'archive du web essaye au maximum d'engranger un maximum d'interactivité et de réponses vers différents sites pour maintenir une historicité. L'auteur Thomas Chaimbault explique que la description d'une archive est floue en ces termes :

« Si l'on considère ainsi qu'en moyenne, une page web contient une quinzaine de liens vers d'autres pages, et environ cinq objets d'origines diverses (sons, images, code, films...), la description technique d'une page demeure ambiguë et floue »²⁷⁴

²⁷³ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 65

²⁷⁴ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 16, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf> >, (consulté le 14/08/22)

²⁷⁴ Site web « Internet Archives » : < <https://archive.org> >, (consulté le 14/08/22)

Dans le cas où un site archivé a un lien vers un site ou une requête interrogeant le web profond, le lien du site archivé ne pourra jamais fonctionner car le web archivé est uniquement dans le web visible. Le web visible est défini comme étant les pages web accessibles facilement par les usagers ainsi que les robots crawler. A l'inverse, le web invisible ou profond est défini comme les pages non accessibles par un moteur de recherche, comme les bases de données ou des sites protégés par un mot de passe. L'auteur Thomas Chaimbault définit le web profond en ces termes :

« Le « web invisible » (*deep web, hidden web*) désigne cette partie du web non accessible aux moteurs de recherche classiques. Il comprend des bases, banques de données et bibliothèques en ligne gratuites ou payantes... Difficilement accessible, ce web comprend des pages protégées par un mot de passe, des pages orphelines, des documents ou des sites trop volumineux, des formats mal reconnus ou des pages générées dynamiquement, c'est-à-dire uniquement en réponse à une requête sur leur moteur interne. Il n'existe pas alors d'URL (adresse) statique de ces pages. »²⁷⁵

Ainsi l'archivage et aussi la consultation du web profond est actuellement impossible. Cela est d'autant plus dommageable que selon une étude de 2001 de Michael Bergman, le web profond serait 400 à 550 fois plus volumineux que le web visible. Ainsi l'archivage reste encore en surface du web. En conclusion, la consultation d'une archive du web appelle l'utilisation d'un ordonnanceur qui va mettre en relation la structure et le contenu d'un site. Il s'agit d'un outil supplémentaire pour l'archivage mais aussi pour le traitement documentaire de l'archive. Il convient à présent d'étudier comment l'INA présente l'archive web via les notices documentaires.

²⁷⁵ CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008, p. 16, consultable en ligne : < <https://www.enssib.fr/bibliotheque-numerique/documents/1730-l-archivage-du-web.pdf>. >, (consulté le 14/08/22)

²⁷⁵ Site web « Internet Archives » : < <https://archive.org> >, (consulté le 14/08/22)

C- Mise en place d'une notice

De la même façon qu'une archive audiovisuelle, l'archive web bénéficie d'une notice documentaire. La notice documentaire mise en place à l'INA communique des informations sur l'archive comme les URL, la date de collecte ou encore des éléments de contexte. Les notices des archives du web ont une double utilité : une utilité de catalogage qui permet de décrire un élément dans un catalogue et une utilité de collecte car la notice contient des informations spécifiques pour les robots « crawler ». La notice d'une archive web est composée de cinq onglets fournissant des informations diverses. Les onglets sont nommés ainsi : « général », « source », « crawl », « documentation » et « stratégie ».

Le premier onglet nommé « général » regroupe les renseignements liés aux informations communes du site. Nous retrouvons donc le logo du site sous forme d'URL ainsi que le nom de l'entrée du site. Il s'agit ici du nom de l'archive. Le second onglet « source » informe sur l'identifiant du site. L'identifiant utilisé est l'URL du site. Le troisième onglet « crawl » comporte les informations pertinentes qu'utilisent les robots « crawler ». Ainsi cet onglet rassemble l'ensemble des URL des pages du site à archiver. De plus, l'onglet permet de signaler aux robots les pages du site qui sont exclues de la collecte. Enfin l'onglet dispose d'une entrée « paramètre » permettant de configurer précisément les « crawler » archivant le site. Le quatrième onglet nommé « documentation » est le plus fourni. Il contient le nom de l'éditeur du site, le secteur d'activité et la fonction du site. Parmi les fonctions du site, on trouve des entrées comme blog, forum, guide TV, site marchand ou encore site institutionnel. L'onglet permet également de renseigner la couverture du média (locale, nationale ou internationale) ainsi que le média représenté par l'archive web (télévision, radio ou web). De plus, l'onglet permet de renseigner les acteurs de l'archive ainsi que leur fonction. On trouve des entrées portant sur le type de programme (chronique, long-métrage, documentaire...), le public cible, la nature du programme, le statut du site, la date de la première et de la dernière collecte et enfin un champ pour des notes libres. L'ensemble des informations de l'onglet « documentation » sera utilisé dans le cadre d'une recherche. Enfin le dernier onglet est celui de la stratégie de collecte. On retrouve des informations comme la priorité de collecte, la fréquence de collecte, le type de robot utilisé et le nombre de pages à collecter. Ainsi l'ensemble de ces informations est consultable dans les notices mises à disposition dans les

bornes de consultations de l'INA. Le tableau suivant met en évidence l'ensemble des informations les plus pertinentes repérables par onglet.

Fig. 11 : Tableau des informations consultables dans une notice documentaire d'une archive web par onglet

Onglets	Informations consultables
Général	Nom de l'entrée Logo du site
Source	Identifiant du site
Crawl	URL des pages du site à collecter URL des pages du site à ne pas collecter
Documentation	Nom de l'éditeur Secteur d'activité Fonction du site Type de média représenté Couverture du média Personnalité Type de programmes Public cible du site Nature du site Statut du site Date de la première et dernière capture Note libre
Stratégie	Priorité de collecte Fréquence de collecte Type de robot Nombre de page à collecter

Pour terminer, la notice d'une archive web est assez complète, avec des informations sur le contenu de l'archive et sur la structure et la façon d'archiver le site en question. Dans cette partie, nous avons pu analyser les outils mis en place pour une bonne gestion documentaire des archives du web. La veille et ses stratégies permettent de rendre compte du paysage à archiver, la consultation d'une archive est rendue possible grâce à la technologie des ordonnanceurs, enfin la notice permet de fournir des informations aux utilisateurs et aux personnels de l'INA pour une future collecte. Il convient dans le dernier chapitre de s'interroger sur l'utilisation et la diffusion des archives du web de l'INA.

Chapitre 3 : Diffusion et enjeux des archives du web

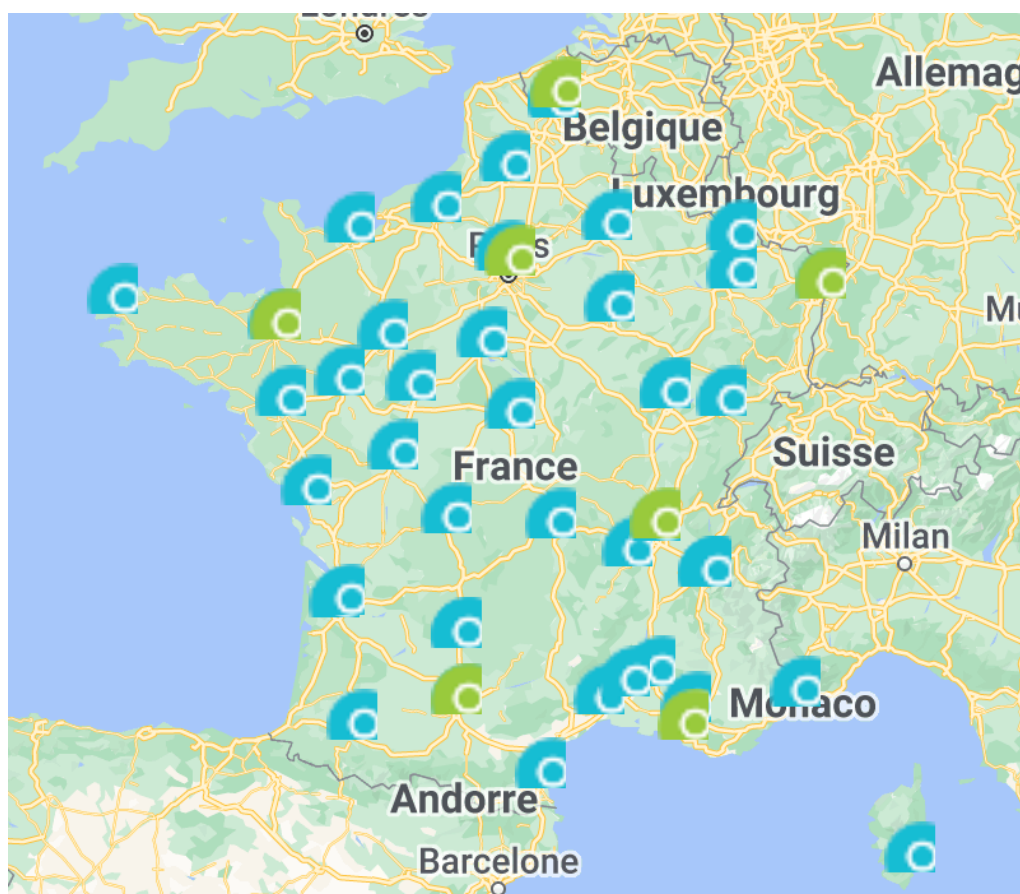
Ce dernier chapitre vise à analyser la diffusion ainsi que les enjeux historiques dans l'utilisation d'une archive du web. Pour diffuser les archives web, l'INA se doit de permettre la mise en consultation des archives. Ainsi notre première partie a pour objectif de mettre en évidence la mise en consultation des archives du web. Il convient dans un premier temps d'étudier les lieux de consultations des archives du web mais aussi de comprendre à qui s'adressent les archives. Ensuite, nous évoquerons les critères de recherche que l'INA instaure lors de la consultation. Enfin, nous dénombrerons les multiples outils mis en place par l'INA pour permettre une recherche détaillée dans la base de données des archives des réseaux sociaux et notamment de Twitter. Notre deuxième partie étudiera les formes de valorisation d'une archive du web. De plus, nous verrons que la valorisation des archives se fait en rupture avec la valorisation des archives papiers et audiovisuelles. Enfin dans une troisième, nous nous pencherons plus en détail sur les enjeux historiques d'une archive du web. En effet, nous analyserons la construction d'un regard d'historien sur les archives du web, qui nécessitent un travail et une analyse différente que le travail sur les archives papiers. Il convient alors de construire un regard critique et des outils d'analyses de l'archive. Enfin nous aborderons, le nouveau rapport à l'archive (reborn digital heritage) qui est induit par la nature de l'archive.

Partie 1 : Mise en consultation des archives du web

A- Pour qui et où consulter ?

Les archives web comme les archives audiovisuelles de l'INA sont mises en consultation par l'entreprise dans des centres de consultations. L'INA dispose d'un maillage de centres de consultation sur le territoire métropolitain mais aussi en Guadeloupe, en Martinique, en Guyane, à La Réunion, en Polynésie française ainsi qu'à Nouméa. Sur le territoire français, il y a 58²⁷⁶ lieux de consultations. La carte suivante met en évidence l'ensemble de ces lieux de consultations en France métropolitaine.

Fig. 12 : Carte des lieux de consultation des fonds des archives web de l'INA²⁷⁷



²⁷⁶ Donnée du site web de l'INA Thèque, consultable en ligne : < <http://www.INAtheque.fr/consultation.html?location=Aix%20en%20Provence> >, (consulté le 18/08/22)

²⁷⁷ Carte provenant du site web de l'INA Thèque, consultable en ligne : < <http://www.INAtheque.fr/consultation.html?location=Aix%20en%20Provence> >, (consulté le 18/08/22)

L'INA dispose de deux catégories de lieux de consultation : la consultation experte et la consultation autonome. La consultation experte est possible uniquement dans les 6 délégations auxquelles vient s'ajouter l'INA Thèque à la BNF. La consultation experte est définie en ces termes par l'INA :

« La consultation dans les centres INA permet de profiter de l'accompagnement des équipes, d'outils d'aide à l'analyse et d'un environnement de travail numérique personnel. »²⁷⁸

La consultation experte est guidée par les documentalistes de l'INA. Dans le cadre d'une consultation experte dans les délégations ou à l'INA Thèque à la BNF ils interviennent pour garantir une bonne prise en main de l'interface de consultation. De plus, ils dispensent une formation et une assistance aux outils d'analyse de l'INA. Lors de cette consultation, il est possible d'utiliser les outils d'analyse, et dans le cas d'archive web, il est possible d'utiliser des outils que l'on détaillera plus tard. La consultation se fait uniquement sur les postes SLAV (station de lecture audiovisuelle). Il s'agit de poste de travail avec un navigateur de consultation des fonds de l'INA, les outils de l'INA ainsi que la possibilité de stocker les données personnelles de l'utilisateur et de conserver une trace de ses consultations. La recherche experte est ouverte uniquement par accréditation, qui s'obtient en informant précisément l'infrastructure d'accueil de la recherche effectuée sur les bornes de consultation. Cette accréditation est plus facilement accessible lorsque la demande vient d'un étudiant, d'un chercheur ou d'un professionnel de l'audiovisuel, cependant il est possible d'effectuer des recherches personnelles à l'INA en informant les structures. La consultation experte est plus avantageuse lors d'une recherche au long court. Il est possible de consulter les archives de l'INA de façon moins contraignante grâce à la consultation autonome. Cette consultation s'effectue dans les centres culturels comme des cinémathèques ou des bibliothèques ayant fait la demande de borne de consultation de l'INA. La consultation se fait via des postes donnant l'accès à l'ensemble des notices documentaires de tous les fonds de l'INA. L'interface de recherche est plus simple que sur les postes de consultation experte, la recherche apparaît simplifiée avec uniquement des critères basiques de recherche comme le

²⁷⁸ Site web de l'INA Thèque, consultable en ligne : < <http://www.INAtheque.fr/consultation/services-de-consultation.html> >, (consulté le 18/08/22)

titre, le canal de diffusion, l'URL ou encore la date. La consultation se fait en autonomie bien que des menus d'aides, une FAQ²⁷⁹ ainsi qu'un glossaire soient à la disposition des utilisateurs. La consultation autonome ne permet pas de sauvegarder ces recherches et de créer un espace personnel. Ainsi la consultation sur les postes de consultation autonome est utilisée pour les recherches ponctuelles qui ne demandent pas un investissement de temps conséquent. La consultation des archives du web est réalisée par des publics différents selon la structure choisie. Si la conservation du web a d'abord une visée patrimoniale et d'enrichissement de la mémoire collective, le public cible de ces archives sont les chercheurs et les étudiants de tout domaine qui bénéficient des outils de travail lors d'une consultation experte. Une partie du public qui consulte ces archives est constituée de professionnels du domaine de l'audiovisuel ou du domaine web dans le cadre d'une mission professionnelle. Enfin, les individus curieux du monde de fonctionnement de l'INA et des archives web forment la deuxième partie du public que l'on retrouve plutôt dans les centres culturels pour une consultation autonome. Il paraît pertinent de se pencher sur l'analyse de la recherche lors d'une consultation experte à l'INA.

B- Critères de recherche

La consultation experte a trait à la recherche ou à une visée professionnelle notamment par l'accréditation que la consultation demande. Dans une visée de recherche, l'INA met à disposition des chercheurs des outils pour effectuer des recherches dans les fonds du web archivé. Tout d'abord, la recherche sur les stations SLAV peut se faire par 3 entrées : une entrée par URL, une entrée par thématique et une entrée par recherche textuelle. Sophie Gebeil, historienne spécialiste de l'archivage du web le définit en ces termes :

« L'utilisateur accède au web archivé par un navigateur intégré aux postes de consultation : les archives de l'internet de la BNF ont été développées à partir d'un modèle d'IA tandis que le web Média de l'INA a été conçu en interne. Si chaque interface a son propre fonctionnement, la « navigation » dans le web archivé s'effectue selon 3 modalités : par l'URL d'un site ou d'une page, par recherche

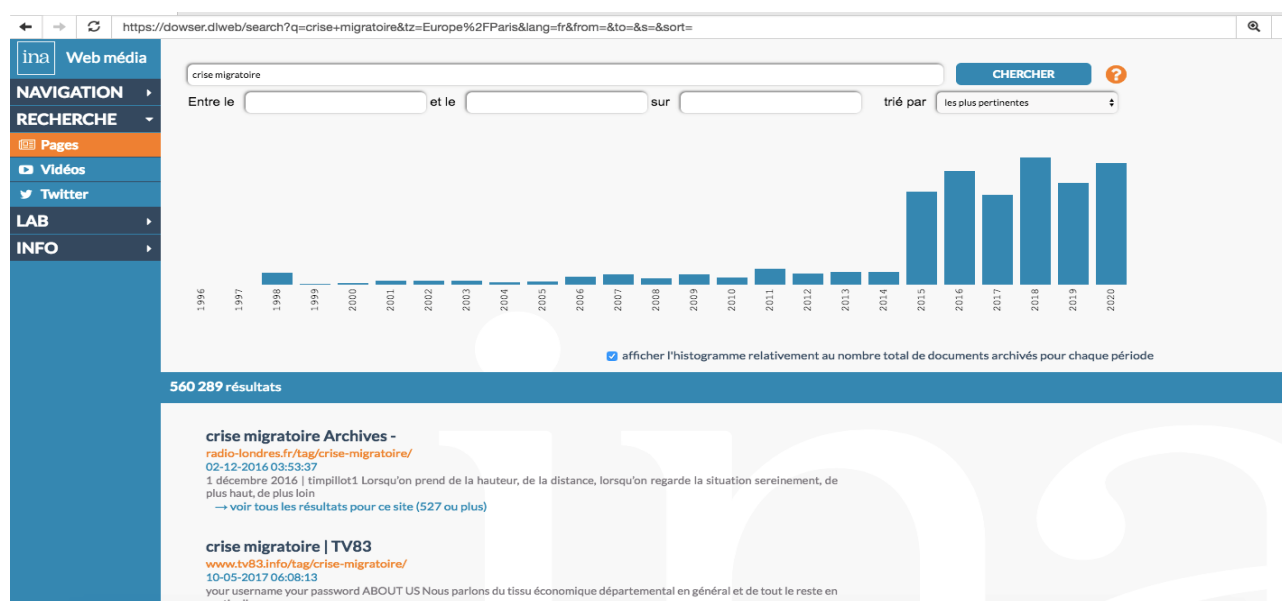
²⁷⁹ FAQ : Foire aux questions

textuelle, et enfin par thématique ou type de contenus. L'URL est disponible via la barre de recherche du navigateur de consultation, à laquelle s'ajoute désormais la possibilité de consulter les bases archivées par l'INA et la BNF à distance, ce qui permet de gagner un temps précieux avant de se déplacer. Le navigateur de l'INA donne accès, en salle de consultation et depuis plusieurs années, à la base de données des sites collectées qui peuvent être interrogés au moyen de plusieurs critères. »²⁸⁰

Ainsi la consultation d'archive se fait à l'INA par trois entrées différentes. Les postes de consultation offrent la possibilité d'ajouter des critères de sélection comme la date de la collecte ou encore le nom du domaine. Si plusieurs entrées de recherche existent pour cibler un site, l'entrée la plus pertinente reste la recherche par URL. Elle permet de fournir à l'utilisateur l'ensemble des collectes d'un site web. Lors d'une recherche par URL, l'utilisateur a accès à la répartition du site, à des versions du site par année, ainsi qu'au nombre de versions du site disponible. L'utilisateur peut alors choisir la version du site qui l'intéresse ; dès lors que la version du site est choisie, la navigation dans l'archive se fait uniquement sur la version demandée. Depuis 2015 l'INA ajouté la possibilité de faire des recherches en « plein texte ». La figure suivante illustre une recherche faite en « plein texte », elle est issue de l'ouvrage : « Website Story, Histoire, mémoires et archive du web »

²⁸⁰ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 66-67

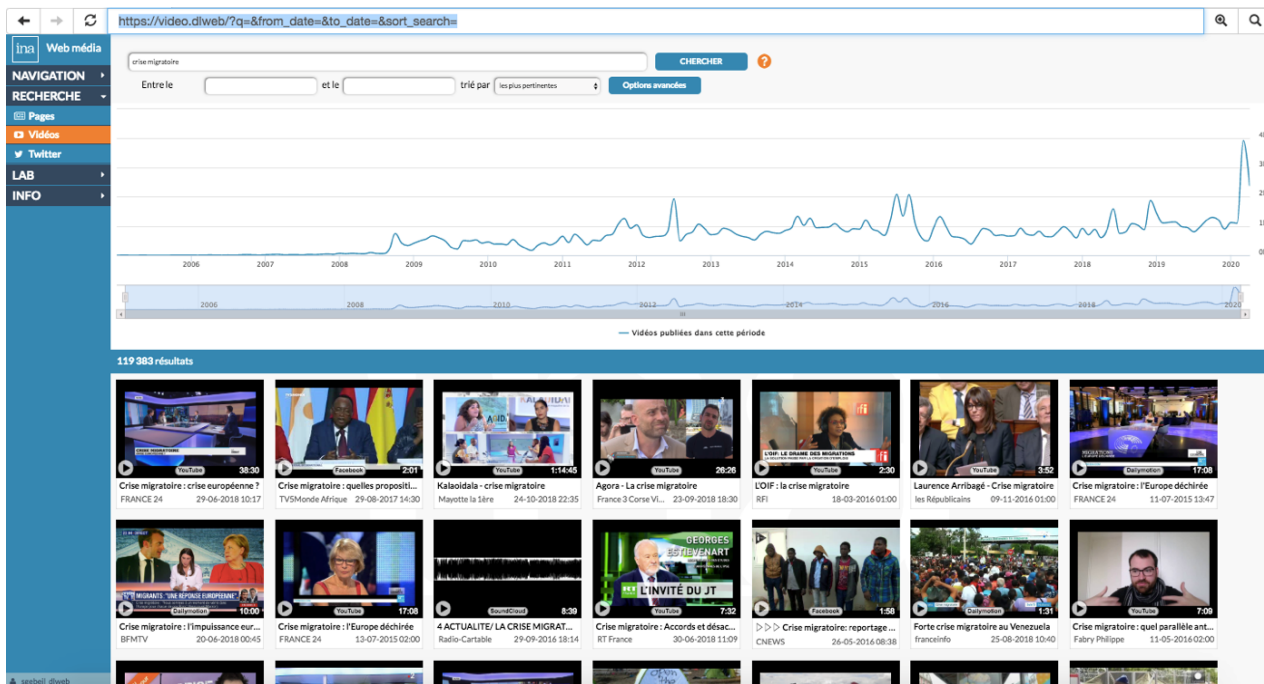
Fig. 13 : Une recherche en « plein texte »²⁸¹



La méthode qui consiste à faire correspondre une requête textuelle avec les contenus d'un site permet de cibler une thématique ou un texte précis sur l'ensemble du web archivé. Cependant il convient de cumuler cette archive avec des bornes chronologiques et différents critères pour éviter au maximum le bruit lors de la recherche. En plus de fournir les résultats de la recherche, le navigateur permet de fournir un histogramme montrant le nombre d'occurrences trouvées par année. Cela permet par exemple à un chercheur de se rendre compte d'un phénomène. De plus l'INA dispose depuis 2013 de la possibilité de restreindre la recherche aux contenus vidéo ou audios, cela permet aux chercheurs d'étudier un corpus spécifique d'œuvres. La figure suivante qui est issue de l'ouvrage : « *Website Story, Histoire, mémoires et archive du web* », montre la recherche par vidéos :

²⁸¹ Figure issue de l'ouvrage GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 68

Fig. 14 : Une recherche par contenus vidéo²⁸²



De la même manière que pour une recherche en « plein texte », l'INA fournit une courbe du nombre d'occurrences par année. Ainsi la consultation d'archives est aidée lors d'une consultation par les outils mis en œuvre pour faciliter la recherche. Lors de la consultation comme nous l'avons dit, les sites sont reconstruits de façon dynamique avec une interactivité qui est maintenue. Il est à noter, comme signalé par Sophie Gebeil²⁸³, que sans se rendre dans des centres de consultation, il est possible d'effectuer des recherches sur les fonds du web de l'INA depuis le site de l'INA Thèque, cependant n'est accessible qu'une version simplifiée des notices. La recherche est également simplifiée, il s'agit ici de fournir des indications en amont d'une recherche sur les bornes de consultation de l'INA. La figure suivante met en évidence la recherche simplifiée sur le site web de l'INA Thèque.

²⁸² Figure issue de l'ouvrage GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 68

²⁸³ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 66-67

Fig. 15 : Une recherche sur le site web de l'INA Thèque²⁸⁴

The screenshot shows the search interface of the INA Thèque website. At the top, there are three tabs: 'TV-RADIO', 'ÉCRIT', and 'WEB-MEDIA', with 'WEB-MEDIA' being the active tab. Below the tabs, a green instruction reads 'Saisissez le nom ou l'adresse d'un site (ex : tf1.fr)'. A large, empty light blue search bar is positioned below this. Underneath the search bar, the section 'Critères de recherche avancée' (Advanced search criteria) is visible. It includes a 'Statut' (Status) dropdown menu currently set to 'Choisissez une valeur...'. Below that, there are fields for 'Première capture' (First capture) and 'Operateur booléen' (Boolean operator). The 'Première capture' field has a date range selector with an equals sign, a date input field, and an 'à' (to) separator, followed by another date input field. An example 'Ex : 7/02/2009' is provided. The 'Operateur booléen' field has a dropdown menu currently set to 'Et' (And). Below these fields, the 'Tri des résultats' (Sort results) section is shown, with a 'Trier par' (Sort by) dropdown menu set to 'Aucun' (None) and a 'Résultats par page' (Results per page) dropdown menu set to '50'. At the bottom of the search criteria section, there are two buttons: a pink 'Effacer' (Clear) button with a right arrow and a black 'Rechercher' (Search) button with a right arrow.

En conclusion, la recherche sur les bornes de consultation experte de l'INA permet d'utiliser des outils supplémentaires comme les différents types de recherche dans le catalogue, une recherche par URL, par thématique ou encore par recherche textuelle. De plus viennent s'ajouter des outils pour affiner la recherche comme la recherche en « plein texte » ou la recherche uniquement sur les contenus vidéos. L'ensemble de ces outils permet de fournir aux chercheurs des dispositifs facilitant la consultation de la base de données de l'INA. Ces outils efficaces pour la recherche sur des sites web ne sont pas aussi pertinents pour des archives provenant des réseaux sociaux. Il convient dès à présent d'étudier la recherche sur les réseaux sociaux.

C- De multiples outils pour la recherche détaillée sur Twitter

Les réseaux sociaux et notamment Twitter sont consultables sur les fonds du web archivé de l'INA. Comme nous l'avons évoqué, Twitter est archivé depuis 2014 et Facebook depuis

²⁸⁴ Capture d'écran du site web de l'INA Thèque, consultable en ligne : < <http://INAtheque.INA.fr/index/WEB-MEDIA/> >, (consulté le 19/08/22)

seulement 2018. Nous prendrons l'exemple des archives de Twitter qui proposent des dispositifs nombreux pour faciliter la recherche. En effet, les archives de Twitter, archivées depuis plusieurs années, disposent d'outils de consultation. Tout d'abord la recherche sur ces archives se réalise de manière simple par un mot-clef. Lors d'une recherche de ce type, la base de données effectue une recherche du mot demandé dans l'ensemble des tweets archivés. A la suite d'une recherche simple par mot-clef, de la même manière que pour le fonds des sites web archivés, l'INA propose une courbe représentant la répartition des occurrences par années. Il est à noter que la courbe peut être affinée par mois, par jour et par heure. L'INA dispose également d'une recherche détaillée. Il est possible de définir la langue du tweet, de rechercher par hashtags, par nom d'utilisateur, par mention ou dans un sous-corpus défini à l'avance par l'INA. De plus, il est possible de naviguer entre 4 collections de tweets distincts. La première collection appelée « Médiasphère » est celle des tweets liés à l'audiovisuel français. Il s'agit de la plus imposante collection de tweets. La seconde collection est appelée « Attentats », nous retrouvons l'ensemble des tweets portant sur les attentats en janvier 2015 et en novembre 2015. Cette collection a été mise en place après une grande collecte de tweets visant à sauvegarder la mémoire collective de web français à la suite de ces événements. La troisième collection, nommée « Élections 2017 » fournit l'ensemble des tweets se rapportant aux élections présidentielles de 2017 en France. Il est à noter que la collection « Élections 2022 » n'est pas encore prête à être mise sur la base de données de l'INA. Enfin, la dernière collection est appelée « Trends », il s'agit des tweets tendances qui ont pour vocation de décrire un phénomène du web ou un événement à forte visibilité du web. Il est donc possible, dans une recherche détaillée, d'affiner la recherche pour trouver plus rapidement un tweet ou un compte Twitter.

Après avoir lancé une recherche, l'INA offre la possibilité d'afficher les métadonnées liées aux tweets. Ainsi il est possible de trouver la date, le nombre de retweets, la langue ou encore la localisation du compte. La consultation experte permet également l'utilisation de sept outils complémentaires pour l'analyse des résultats. Le premier outil permet d'obtenir la statistique de distribution par pays, ville et langue des tweets archivés contenant le terme demandé. Ce dispositif permet de dresser une cartographie ou de limiter la recherche dans une zone géographique. Le deuxième outil permet de mettre en évidence les hashtags les plus utilisés dans les tweets contenant le terme demandé. Le troisième permet d'observer les

mentions les plus présentes dans les tweets contenant le terme demandé. Le quatrième dispositif agit de la même manière pour les liens les plus partagés. Il en va de même pour le cinquième, sixième et septième outil qui permettent de mettre en évidence les utilisateurs les plus représentés, les médias les plus partagés et les émoji les plus utilisés dans les tweets contenant le terme demandé. Ces dispositifs permettent de faciliter l'analyse géographique et comparative des tweets archivés par l'INA. Il est possible d'obtenir des diagrammes et des courbes montrant les évolutions dans le temps ou la répartition selon les outils utilisés par l'utilisateur.

En conclusion, les archives de Twitter bénéficient d'outils facilitant la recherche, il en va de même pour les sites web archivés qui grâce à plusieurs dispositifs bénéficient d'une consultation intuitive. L'ensemble de ces outils sont à destination des chercheurs dans le cadre d'une consultation experte dans les délégations ou à l'INA Thèque. Après avoir évoqué la mise en consultation des archives, il convient d'aborder la valorisation des archives du web.

Partie 2 : Valorisation des archives du web

A- Des valorisations en rupture avec les archives papiers et audiovisuelles

Du fait de la nature des archives du web, il n'est pas possible de fournir des valorisations sur le même modèle que les papiers. Tout d'abord la valorisation des archives signifie mettre en évidence un document pour le rendre disponible aux yeux de tous. La valorisation est un moyen pour mettre en évidence une archive, évitant ainsi que l'archive ne reste dans une base de données ou un magasin. Il s'agit ainsi de faire vivre l'archive pour une diffusion la plus large possible. La valorisation doit permettre à la fois d'éduquer une population, de témoigner d'évènements du passé et de cultiver une mémoire collective. La valorisation peut prendre de nombreuses formes. Il peut s'agir d'une exposition dans un service d'archives ou dans un musée comme l'INA le propose avec la délégation marseillaise de l'INA, grâce à un partenariat avec le Mucem, et fournit des archives audiovisuelles pour enrichir des expositions. La valorisation peut prendre la forme d'un projet scolaire comme le fait l'INA avec la mise en place d'un site web à destination des scolaires enrichi par des archives audiovisuelles, le site web « Sudorama »²⁸⁵ regroupant plus de 500 archives régionales de l'INA contextualisées par des universitaires. Il s'agit alors d'une valorisation avec une ambition éducative. La valorisation peut également s'orienter vers la recherche avec des partenariats ou des projets de recherche financés par l'INA voir même la mise en place de prix. Nous pouvons citer les prix décernés par l'INA Thèque qui récompensent les recherches menées avec les archives de l'INA et qui enrichissent les connaissances sur l'audiovisuel et le numérique. Les prix sont définis en ces termes par le site web de l'INA Thèque :

« Les Prix de l'INA Thèque, décernés par un jury rassemblant chercheurs et professionnels, distinguent des travaux qui contribuent à enrichir la réflexion contemporaine sur les évolutions de l'écosystème audiovisuel et numérique au cœur duquel s'exercent les métiers de l'INA. Ces Prix constituent l'une des nombreuses actions (débat, colloques, publications...) menées par l'Institut pour favoriser

²⁸⁵ Site web « Sudorama » consultable en ligne : < < <https://fresques.ina.fr/sudorama> > (consulté le 05/03/22).

l'exploitation scientifique de ses collections à travers plus de 50 centres de l'INA Thèque ouverts en France. »²⁸⁶

Ainsi ces prix encouragent le développement des travaux scientifiques à partir des fonds de l'INA. Il existe deux prix, le prix de la recherche qui récompense un travail de recherche de niveaux doctorat ou plus et le prix de l'encouragement qui distingue un travail de mémoire. Les lauréats sont récompensés par une bourse ainsi qu'un soutien à la diffusion de leur recherche. Ainsi la valorisation des archives peut prendre plusieurs formes dont également la forme de séminaires ou conférences organisés en lien avec les archives, ou encore la mise en place de service de SVOD comme « Madelen »²⁸⁷ qui permet une mise à disposition d'une sélection d'archives de l'INA.

Il apparaît que les archives de l'INA papiers ou audiovisuelles sont valorisées sur de nombreux supports, à la fois physiquement dans des expositions, sur le web avec des fresques historiques ou lors de recherche. La forme la plus courante de valorisation reste les expositions en lien avec des structures muséales ou dans un service d'archives. Il est important de nous questionner sur la faculté d'une archive du web à fournir assez de matière pour une exposition en physique. L'archive web n'étant pas matérielle, il est assez difficile d'envisager une exposition portant uniquement sur les archives du web mais cela pourrait être organisé par des musées ou institutions bénéficiant d'un budget important pour mettre à disposition un grand nombre d'écrans et de vidéoprojecteurs. Ainsi l'archive web dépend du matériel, de la même manière qu'une archive audiovisuelle. Si des expositions avec des archives audiovisuelles existent, faire de même avec des archives du web n'est pas pratiqué. Les valorisations sont donc assez en rupture avec les valorisations des archives papiers ou audiovisuelles. Cela est dû également à la considération faible qu'ont les archives du web vis-à-vis de papiers anciens ou de vieux films qui grâce à leur ancienneté ont une vocation historique facilement lisible par le grand public. Les archives web n'ont pas encore atteint le statut de patrimoine historique pour le grand public, c'est également le cas pour les chercheurs, qui comme nous le verrons dans notre prochaine partie, doivent aiguïser leur

²⁸⁶ Prix de l'INA Thèque, consultable en ligne : < <http://www.INAtheque.fr/publications-evenements/prix-INAtheque.html> >, (consulté le 19/08/22)

²⁸⁷ Site web du service de SVOD « Madelen » : < <https://madelen.INA.fr> >, (consulté le 19/08/22)

regard sur l'archive du web. Les archives du web doivent donc trouver des perspectives différentes pour pouvoir être valorisées.

B- Exemples de valorisation des archives du web

Si aujourd'hui, à ma connaissance, il n'existe pas de valorisation uniquement des archives du web organisé par l'INA, il convient de mettre en évidence les formes de valorisation des archives du web. Hormis la mise en avant de ces archives par le biais de la recherche qui pourra se faire dans les années futures avec l'augmentation des recherches sur les questions numériques, l'INA avec les bornes de consultation experte et autonome fournit des accès vers les collections du web archivées. Les postes de consultation autonome, par leur interface intuitive et facile d'accès, sont une invitation pour tout curieux à naviguer dans les archives du web. Il s'agit ici d'un premier pas pour la valorisation et la diffusion des archives du web. Cependant cela reste assez restreint, les archives ne sont pas contextualisées, il n'y a pas de plus-value apportée par la mise en place de ces postes de consultation. A l'inverse, la BNF qui archive également le web a mis en place, de la même façon que l'INA, dans des bibliothèques partenaires, un accès aux archives du web. Ainsi les collections du dépôt légal du web sont consultables depuis une application nommée « Archives de l'internet ». L'ouverture de cette application au même titre que la consultation experte à l'INA doit s'accompagner d'une accréditation²⁸⁸. L'application est disponible dans les salles de recherche de la BNF et dans les différentes bibliothèques du réseau. La visite est également disponible depuis un ordinateur personnel grâce au portail d'accès « AVEC »²⁸⁹. La recherche sur ces plateformes, de la même façon qu'à l'INA, se fait à partir d'une URL ou en « plein texte ». De plus, la BNF met à disposition des personnes accréditées dans « parcours guidés ». Il s'agit d'une sélection de sites archivés et de ressources portant sur un sujet précis comme « le web lorrain »²⁹⁰, « le

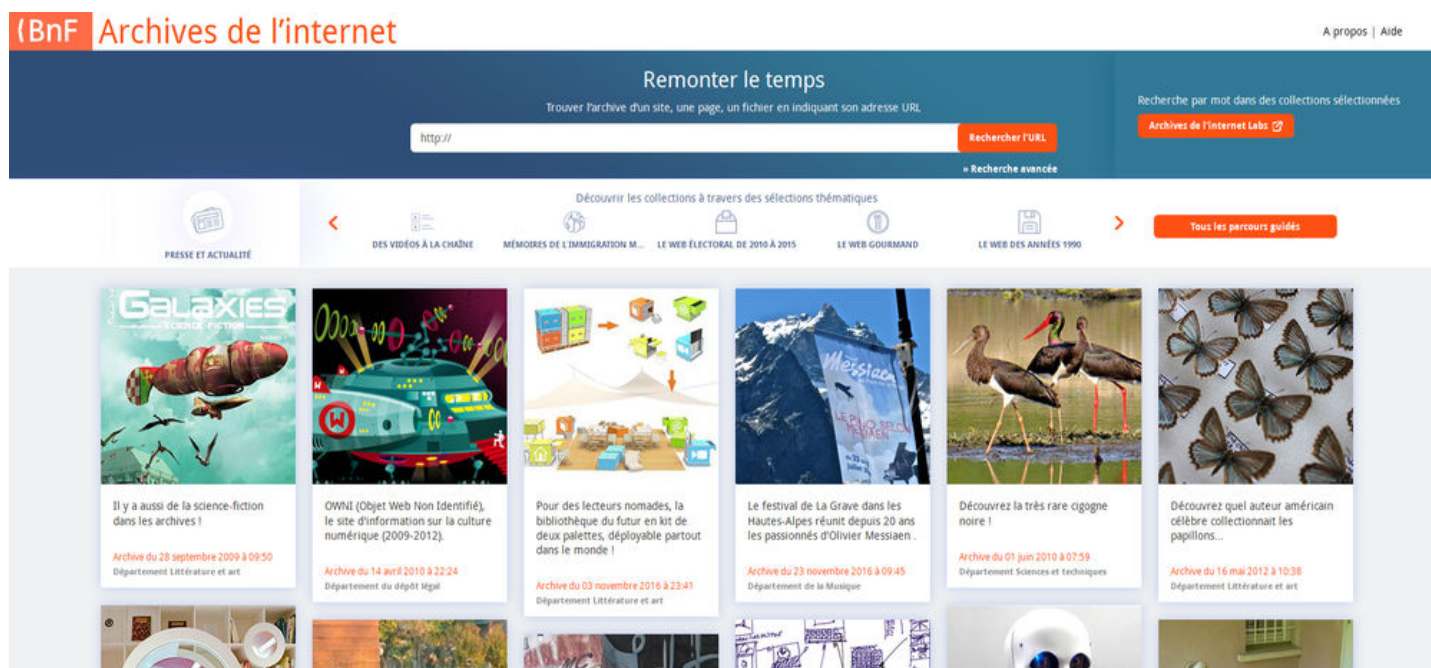
²⁸⁸ Article provenant du site web de la BNF consultable en ligne : < <https://www.BNF.fr/fr/archives-de-linternet> >, (consulté le 19/08/22)

²⁸⁹ Article provenant du site web de la BNF consultable en ligne : < <https://www.BNF.fr/fr/utiliser-des-ordinateurs-et-internet-la-bnf> >, (consulté le 19/08/22)

²⁹⁰ Article datant de septembre 2021 définissant le contour et les ressources disponible dans le parcours « le web lorrain », consultable en ligne : < https://www.BNF.fr/sites/default/files/2021-09/web_lorrain_parcours.pdf >, (consulté le 19/08/22)

web des années 90 »²⁹¹ ou encore « le web scientifique »²⁹². Ces sujets sont choisis et élaborés par des bibliothécaires et des chercheurs pour inviter à la découverte. En ce sens, nous pouvons définir cette démarche comme une démarche de valorisation. Il est à noter que dans une démarche de transparence et de l'accès des sources, la liste des sites sélectionnés est fournie sur le site « Data.gouv.fr »²⁹³ qui regroupe l'ensemble des ressources de la BNF. L'application « Archives de l'internet » est à destination des chercheurs mais a une politique éditoriale de mise en avant de certaines sources précises, ce n'est pas le cas pour l'INA qui lors de sa consultation ne fournit que l'accès au catalogue. La figure suivante illustre la page d'accueil de l'application « Archives de l'internet »

Fig. 16 : Application « Archives de l'internet »²⁹⁴



²⁹¹ Article datant de novembre 2016 définissant le contour et les ressources disponible dans le parcours « le web des années 90 », consultable en ligne : < https://www.BNF.fr/sites/default/files/2018-11/parcours_web90%20final.pdf >, (consulté le 19/08/22)

²⁹² Article datant d'août 2014 définissant le contour et les ressources disponible dans le parcours « le web scientifique », consultable en ligne : < https://www.BNF.fr/sites/default/files/2018-11/le_web_scientifique_parcours.pdf >, (consulté le 19/08/22)

²⁹³ Site web regroupant les ressources de la BNF : < <https://data.BNF.fr> >, (consulté le 19/08/22)

²⁹⁴ Capture provenant de l'article du site web de la BNF consultable en ligne : < <https://www.BNF.fr/fr/utiliser-des-ordinateurs-et-internet-la-BNF> >, (consulter le 19/08/22)

La structure « Internet Archive » dans la mise en place de son dispositif « Wayback Machine », qui permet de naviguer dans le web archivé, met en avant le concept de voyage temporel. Il s'agit ici d'une forme de valorisation des archives du web dans le sens où la façon de présenter l'outil insiste sur le fait de voyager soi-même dans le web du passé. Nous avons affaire à une scénographie de la recherche avec cet outil. Le dispositif invite au voyage dans le temps comme le ferait une exposition dans une institution muséale, mais ici la recherche d'archives est réalisée par l'utilisateur. Pour conclure, l'INA ne met pas en avant la valorisation de ses archives du web, et peu d'établissements le font. La BNF produit une valorisation assez limitée en public, et « Internet Archive » ne le fait qu'avec son outil de la « Wayback Machine ». Ainsi il y a encore de nombreux progrès à fournir par l'INA et les acteurs de l'archivage du web pour rendre faciles d'accès les recherches mais aussi l'accès à ses archives pour le grand public. En ce sens, une exposition avec des archives du web projetées ou affichées sur des écrans pourrait permettre de mettre plus de lumière sur ce patrimoine. Il n'existe à ma connaissance aucune exposition de ce type en France. La nature de ces archives est un frein à toutes valorisations. Si la valorisation est assez réduite pour ces archives, c'est aussi parce que les enjeux historiques et la construction du regard de l'historien sur ces archives sont encore à mettre en place.

Partie 3 : Enjeux historiques

A- Construction d'un regard d'historien sur les archives du web

Avec la mise en place à l'INA d'outils pour faciliter la recherche historique sur les archives du web, la structure contribue à la création d'un regard de l'historien. Ces archives longtemps peu ou pas utilisées par les historiens sont de plus en plus faciles d'accès pour les chercheurs. Sophie Gebeil l'analyse en ces termes dans son ouvrage « *Website Story, Histoire, mémoires et archive du web* » :

« La création d'outils adaptés aux corpus issus des archives du web constitue un enjeu considérable pour les institutions en charge du dépôt légal du web à la BNF comme à l'INA. L'objectif est double : promouvoir l'intérêt de ces matériaux, trop longtemps restés des « archives sans historiens », et en faciliter l'exploitation en répondant aux besoins des chercheur.e.s »²⁹⁵

Ainsi l'INA permet aux chercheurs de travailler sur des sources du web. Cependant il convient pour l'historien de se former à l'utilisation de ces archives qui sortent de l'ordinaire dans la perspective d'un travail de recherche. C'est d'abord une problématique de méthodologie à laquelle vont devoir faire face les chercheurs. Effectivement les historiens, lors d'un travail sur un fonds, vont essayer de l'analyser pour en comprendre la structuration ainsi que les biais à prendre en considération comme le souligne Valérie Schafer²⁹⁶. Les fonds sont souvent lacunaires pour les archives du web. Il peut manquer des éléments dans des pages ou des pages entières, cela constitue une difficulté sur laquelle le chercheur doit passer. Si cette problématique d'un fonds lacunaire est bien connue des historiens dans le cas d'une archive du web, l'aspect éphémère du web accentue cette caractéristique. Il est alors indispensable pour l'historien d'étudier dans son corpus l'ensemble des sources complémentaires comme des communiqués de presse, des rapports annuels, des descriptions de sous-sites ou encore

²⁹⁵ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 100

²⁹⁶ SCHAFFER Valérie, « Internet historians and the "risks" of web archive », AG de l'IIPC, Bibliothèque Nationale de France, le 26 mai 2014, Paris, consultable en ligne : < <http://netpreserve.org/general-assembly/2014/presentations> >, (consulté le 19/08/22)

des entretiens avec les acteurs de la recherche. L'archivage d'un site web ne produisant pas une copie exacte du site, l'historien doit donc le prendre en considération cette situation. L'historien doit, avoir un esprit critique lors de l'étude de ces sources. Ainsi il est important de prendre en considération l'ensemble des versions d'un site et non pas une capture d'un site à un instant défini. L'une des difficultés dans le travail d'historien réside dans le processus de référence et de citation à une site web. Si les références et citations sont normées dans un processus de recherche dans le cas d'un site web, il n'y a aucune garantie que le site web soit retrouvé plus tard²⁹⁷. En effet, la page peut n'être jamais retrouvée car elle n'est pas archivée ou parce qu'elle est archivée mais que le processus d'archivage a transformé la page. Ainsi les historiens doivent se confronter à des problématiques nouvelles lors de l'utilisation des sources du web dans le cadre d'une recherche. L'INA en 2009²⁹⁸ a mis en place des ateliers à destination des chercheurs pour les aider dans l'étude de sources du web, à rechercher sur le web et à définir les usages d'une source du web. Ces ateliers ont étudié le web comme une ressource pour l'histoire en dégagant des problématiques comme les logiques de diffusion, les espaces destinés aux publics, les enjeux pédagogiques ou encore la gestion de la data. L'atelier a donc précisé les contours des archives ainsi que les enjeux inhérents à ce type d'archive. Cela a permis une réflexion collective sur les archives du web. Ces réflexions ont permis de mener une enquête en ligne entre le 5 et 30 mai 2011 intitulée « Un web archivé pour quoi faire ? »²⁹⁹. L'enquête diffusée dans un cercle de chercheurs a permis de cerner les attentes. Ainsi il apparaît que l'usage des sources du web s'inscrit dans les disciplines des sciences humaines et sociales et les questionnements rejoignent les sciences et technologies de l'information et de la communication. L'enquête montre que les chercheurs soulignent l'instabilité du web et la volatilité des données publiées, ainsi que la difficulté à traiter des données hétérogènes en croissance. Les chercheurs sont également en demande d'outils d'analyses. Il est à noter que plusieurs réponses ont été amenées par l'INA notamment par la mise à disposition d'outils et un archivage de plus en plus fiable. Par la dissociation entre

²⁹⁷ BRUGGER Niels, « L'historiographie de sites Web : quelques enjeux fondamentaux », *Le Temps des médias*, 2012 (n° 18), p. 159-169, consultable en ligne : < <https://www.cairn.info/revue-le-temps-des-medias-2012-1-page-159.htm> >, (consulté le 20/08/22)

²⁹⁸ MUSSOU Claude, « Et le Web devint archive : enjeux et défis », *Le Temps des médias*, 2012 (n° 19), p. 259-266, consultable en ligne : < <https://www.cairn.info/revue-le-temps-des-medias-2012-2-page-259.htm> >, (consulté le 20/08/22)

²⁹⁹ MUSSOU Claude, « Et le Web devint archive : enjeux et défis », *Le Temps des médias*, 2012 (n° 19), p. 259-266, consultable en ligne : < <https://www.cairn.info/revue-le-temps-des-medias-2012-2-page-259.htm> >, (consulté le 20/08/22)

support et contenu, les archives du web contribuent à la disparition de la « notion d'originale »³⁰⁰. C'est une notion à prendre en compte lors d'une recherche historique. En conclusion, le regard de l'historien dans le processus de recherche a été construit par l'historiographie sur le sujet, les différentes enquêtes et les ateliers de l'INA. L'historien peut s'emparer pleinement des archives du web et de ses questionnements notamment le concept de « patrimoine nativement numérique »³⁰¹ (reborn digital heritage)

B- Le nouveau rapport à l'archive (reborn digital heritage)

L'utilisation des archives du web pousse l'historien à s'interroger sur les conditions d'élaboration des archives. Le web et ses contours sont indéfinissables³⁰², ainsi les archivages du web ne prennent en compte qu'un périmètre défini dans l'immensité du web. De plus l'archivage n'est possible que sur le web visible, la majorité du web étant dans le web profond, l'archivage est par nature toujours lacunaire. L'INA et la BNF ne tendent pas à atteindre un archivage exhaustif mais plutôt à rendre consultable une partie du web français. Si l'historien est habitué comme le souligne Arlette Farge à : « faire parler les archives pour en interroger les silences »³⁰³, il est important lors de l'utilisation des sources du web pour l'historien d'identifier les choix institutionnels ou socio-culturels expliquant le choix d'archiver ou non un site web. Sophie Gebeil spécialiste de web l'explique en ces termes :

« L'une des priorités pour l'historien.ne face aux archives du web est donc d'identifier les manques, de les expliquer, d'étudier, en somme, le sens de l'absence pour en déjouer les pièges. En effet, l'expérience de navigation génère l'illusion d'être en face d'un miroir du web vivant, une sorte de copie fidèle malgré quelques incomplétudes. Or, la nature des archives du web est plus complexe »³⁰⁴

³⁰⁰ MUSSOU Claude, « Et le Web devint archive : enjeux et défis », *Le Temps des médias*, 2012 (n° 19), p. 259-266, consultable en ligne : < <https://www.cairn.info/revue-le-temps-des-medias-2012-2-page-259.htm> >, (consulté le 20/08/22)

³⁰¹ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 70

³⁰² GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 70

³⁰³ FARGE Arlette, *Le goût de l'archive*, Paris, Seuil, 1989

³⁰⁴ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 71

L'utilisation pour l'historien d'un patrimoine nativement numérique ou « reborn digital heritage » induit donc un rapport nouveau à l'archive dans le sens où le chercheur doit réfléchir aux biais de fabrique de l'archive. L'utilisation du patrimoine nativement numérique implique de prendre en compte l'aspect volatil et lacunaire des archives et la problématique de l'interactivité des sites. Comme nous l'avons évoqué, si des sites archivés par l'INA le sont en profondeur, ce n'est pas le cas de tous et la navigation peut s'arrêter à la page d'accueil. Lors de l'étude des réseaux sociaux grâce aux archives du web comme avec les archives de Twitter, l'historien doit avoir en tête plusieurs biais. Le premier est rappelé par Sophie Gebeil³⁰⁵, la fréquence d'archivage trop faible exerce une influence sur la représentativité de certains éléments dans les archives. Enfin le nouveau rapport à l'archive est aussi induit par le processus de consultation des archives. En effet, comme nous l'avons vu, le contenu d'une archive et la structure se retrouvent séparés lors de l'indexation, pour ensuite au moyen d'une clé d'identification être assemblés lors de la consultation pour former une unité cohérente. L'archive n'est donc plus une archive originale, il faut alors que l'historien en prenne compte dans son rapport à l'archive. L'archive apparaît comme une unité figée et isolée dans un web qui se veut vivant et en constante évolution. Ainsi les archives du web peuvent être analysées comme des archives « reconstruites »³⁰⁶, cette idée est à l'origine du terme de « reborn digital heritage », lui-même repris de la notion de patrimoine nativement numérique porté par l'UNESCO en 2003. L'ensemble du processus d'archivage a pour vocation à faire ressembler le plus possible l'archive à une page web vivante. Cela crée alors une incohérence que l'historien doit comprendre et garder en tête lors d'un travail de recherche. En conclusion, l'utilisation d'une archive du web pousse le chercheur à avoir un rapport différencié à l'archive en comparaison avec les archives papiers ou audiovisuelles. Ce n'est qu'en prenant en compte des biais lors de l'archivage et de la notion de « reborn digital héritage » que l'historien pourra fournir une recherche pertinente.

³⁰⁵ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 73

³⁰⁶ GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Éditions, 2021, p. 76

Conclusion

Cette étude sur le traitement des archives de l'INA s'est efforcée de traiter les archives du web selon l'angle de vue de l'INA. Cet établissement public a la charge de l'archivage du web par le dispositif du dépôt légal. Le périmètre du dépôt légal du web a été défini par l'ARCOM et divise le travail d'archivage entre la BNF et l'INA. L'INA a alors pour mission d'archiver l'ensemble des sites de médias français, comprenant les sites émanant des services des médias audiovisuels, les services de médias audiovisuels à la demande, les web tv et radio, les sites consacrés aux programmes de radio et de télévision ainsi que les sites des organismes professionnels et institutionnels du secteur de la communication audiovisuelle. L'archivage à l'INA ne s'arrête pas ici, et la structure archive également les comptes ainsi que les vidéos et podcasts des créateurs français sur les plateformes de partage de contenu comme YouTube. De plus l'INA archive les réseaux sociaux et en particulier Twitter, toujours dans la perspective de rendre compte des grands événements médiatiques et de l'actualité des médias.

Nous avons pu répondre lors de cette étude à la question du traitement des archives du web par l'INA, l'évolution des pratiques et l'organisation de la chaîne de traitement d'une archive du web ainsi nous avons observé l'ensemble du processus mis en place par l'entreprise, la collecte du web dans le cadre du dépôt légal, l'indexation automatique proposée par les robots de l'INA, les stratégies de stocks et de pérennisation des archives et enfin la mise en consultation, la valorisation et l'utilisation des archives dans le cadre de la recherche. Cela a également permis de répondre aux questionnements sur les enjeux des archives du web. Étant des archives nouvelles, les chercheurs doivent prendre en compte leur nature et leur originalité pour les étudier de façon pertinente.

Pour répondre à ces questions, nous avons proposé un plan en trois chapitres. Notre première partie vise à décrire les évolutions de l'archivage du web des années 1990 à nos jours. Pour ce faire nous avons évoqué dans un premier temps l'évolution numérique de l'INA, ainsi nous avons étudié la structure actuelle de l'INA pour comprendre comment fonctionne une imposante structure. Nous avons décrit le virage numérique de l'INA et enfin nous avons mis en perspective l'INA et le web français. Cette première partie nous a permis de

comprendre les rouages de la structure ainsi que les évolutions qui se font durant les dernières années à l'INA. Notre second chapitre vise à analyser l'archivage web et le dépôt légal. Ainsi le dépôt légal a été étudié depuis sa création jusqu'à la mise en place d'un dépôt légal du web. Nous avons analysé les différentes approches dans l'archivage du web ainsi que la mise en perspective du contexte juridique et l'instauration d'un dépôt légal du web par la loi DADVSI. Dans une troisième partie, nous avons étudié en détail la nature d'une archive web. Nous avons démontré que l'archive web est un patrimoine fragile, qui comporte des limites comme le manque, l'aspect lacunaire, enfin nous avons décrit l'archive web comme étant au centre d'enjeux de gouvernance. Dans un second chapitre, nous avons étudié d'un point de vue plus technique l'archivage du web à l'INA. Pour ce faire, il a d'abord fallu définir le périmètre d'archivage proposé à l'INA, puis nous avons étudié les principes de collecte et d'indexation par les robots « crawler ». Puis nous avons mis en évidence les problématiques de stockage et de pérennisation de ces archives fragiles. Dans une seconde partie, nous nous sommes attardés sur les techniques et les outils mis en place lors de collectes. Il a fallu aborder l'utilisation des métadonnées dans le processus de collecte ainsi que la création d'un format propriétaire : le DAFF. Nous avons analysé la collecte des UGC et enfin l'archivage de Twitter qui depuis 2014 prend de plus en plus de place. Pour terminer ce second chapitre, la question de l'aspect documentaire de l'archivage a été posée. Nous nous sommes penchés sur la construction d'une veille informationnelle par l'INA pour gérer la collecte, puis la question de la mise en consultation et de la création de notices documentaires a été notre centre d'intérêt. Enfin notre dernier chapitre a étudié la diffusion et les enjeux inhérents aux archives du web. Dans un premier temps, la mise en consultation a été analysée, nous avons abordé les lieux de mise en consultation ainsi que les différents principes de consultation. La consultation à l'INA s'accompagne d'outils mis à disposition des chercheurs que nous avons décrits. Dans un second temps, nous avons mis en lumière la valorisation des archives du web qui se veut en rupture avec la valorisation d'archives papiers et audiovisuelles. De plus, nous avons pu observer des exemples de valorisation des archives du web. Pour terminer, dans une dernière partie, nous avons axé la réflexion sur les enjeux historiques qu'imposent les archives du web. Nous avons étudié la construction d'un regard d'historien sur ces archives ainsi que le rapport à l'archive qui évolue du fait de la nature de l'archive du web.

L'étude nous a donc permis de décrire l'ensemble du processus de traitement des archives du web et de comprendre les enjeux autour de ces nouvelles archives qui bouleversent les pratiques de l'archiviste mais également des chercheurs. Ainsi dans la perspective d'une étude plus large il conviendrait d'étudier l'impact des archives du web sur la production scientifique et de comprendre l'impact sur le monde scientifique de l'archivage du web par l'INA. Il paraît alors pertinent de questionner les zones d'ombres de l'archivage du web comme le web profond.

Partie III :

Projet professionnel

À la suite du Master Valorisation Numérique du Patrimoine : Archives et Images, je souhaite m'insérer professionnellement dans le domaine de la valorisation du patrimoine et de la recherche. Au mois de juin, durant mon stage à l'INA, j'ai candidaté à une offre d'emploi relative à un contrat de recherche en contrat à durée déterminée portant l'analyse de la médiation de l'histoire et de la valorisation des activités de l'université d'Aix-Marseille.

Grace à ma formation initiale en Histoire complétée par les enseignements du Master VALARI et avec l'expérience acquise durant le stage réalisé à l'INA et le sujet du mémoire théorique sur l'archivage du web, ma candidature a été retenue sur un contrat de recherche de 13 mois au sein du centre méditerranéen de sociologie, de sciences politiques et d'histoire (MESOPOLHIS). Ce centre de recherche, sous la tutelle de l'Université d'Aix-Marseille, de Sciences Po Aix et du CNRS a une dynamique pluridisciplinaire, croisant les disciplines en sociologie, sciences politiques, histoire, anthropologie, démographie, géographie et urbanisme. Mes missions qui démarrent au 1^{er} septembre s'inscrivent dans un projet de recherche intitulé « Les médiations de l'histoire en ligne depuis le début des années 2000 sur le territoire d'Aix-Marseille Université. ». Ainsi il m'a été confié de dresser un état des lieux des contenus de médiation de l'histoire publiés sur le web et les réseaux socionumériques par des acteurs du territoire d'Aix-Marseille Université afin de mieux comprendre les rapports entre sciences et société des années 2000 à nos jours. Cette recherche permettra d'analyser l'évolution des nouvelles formes d'écriture et de médiation de l'histoire nées avec l'essor du web et des médias sociaux à l'échelle du territoire marseillais et aixois. Je vais devoir analyser les différents types de dispositif, de pratiques de médiations et identifier les acteurs à l'origine des contenus et leurs publics. De plus, une étude de la réception des médiations devra être apportée. Ce choix de postuler sur ce projet de recherche découle de mon souhait de travailler avec des archives de tout support. Mes connaissances et compétences en archivistique, en informatique ainsi qu'en traitement documentaire que acquises durant le Master VALARI seront un atout nécessaire pour mener à bien ce contrat de recherche.

A la suite de ce contrat, j'envisage de poursuivre par un doctorat. J'aimerais proposer un projet de thèse pluridisciplinaire englobant l'archivistique, l'information, la communication et l'histoire. J'ai acquis une solide formation en Histoire contemporaine, par une Licence en Histoire et par deux années de recherche en Master 1 et 2 en Histoire. La recherche est donc un domaine qui m'attire particulièrement et me motive. Le DDAME ne fournit pas d'offre de doctorat, cependant d'autres universités, notamment l'Université d'Angers dispose d'une école doctorale spécialisée en archiviste, l'Université d'Aix-Marseille dispose également d'un pôle pluridisciplinaire développé.

Enfin selon la réception de ce projet de thèse, je souhaiterais m'insérer professionnellement dans une structure culturelle proposant des valorisations d'archives audiovisuelles ou numériques. J'ai par ailleurs eu l'occasion d'échanger avec Monsieur Leroy, lui-même documentaliste audiovisuel, intervenant dans le Master VALARI dans le cadre d'un cours intitulé : « Traitement de l'image animée ». Nous avons échangé sur son parcours en tant que documentaliste audiovisuel au sein de la rédaction de France 3 Région Occitanie et également sur sa formation et ses relations dans le monde de la documentation audiovisuelle. C'est d'ailleurs Monsieur Leroy qui m'a conseillé de candidater à l'INA pour réaliser un stage afin de me familiariser avec la documentation audiovisuelle et les images. En définitive, mon projet professionnel s'inscrit dans le domaine de la documentation et de la recherche.

Bibliographie du mémoire théorique

I- Ouvrages généraux sur les médias

- BALLE Francis, *Les médias*, Paris, PUF, 2004 [2ème ed. 2006]
- BALLE Francis, *Médias et sociétés : édition, presse, cinéma, radio, télévision, internet*, Paris, Domat Politique, 1994
- BALNAVES Marc, DONALD James et HEMELRYK DONALD Stéphanie, *Atlas des médias dans le monde*, traduit de l'anglais par PERRON Sylvie, Paris Autrement, 2001
- BARBIER Dominique et BERTHO-LAVENIR Catherine, *Histoire de médias : De Diderot à Internet*, Paris, U Histoire, 2009 [3ème ed. 2009]
- BROCHAND Christian, *Histoire générale de la radio et de la télévision en France*, Paris, La Documentation Française, volume 2, Paris, La documentation Française, 1994
- BROCHAND Christian, *Histoire générale de la radio et de la télévision en France*, Paris, La Documentation française, volume 3, 2006
- CLAVERT Frédéric et NOIRET Serge, *L'histoire contemporaine à l'ère numérique*, Bruxelles, Peter Lang, 2013
- DELPORTE Christian et D'ALMEIDA Fabrice, *Histoire des médias en France de la Grande Guerre à nos jours*, Paris, Champs Histoire, 2003 [ed. 2010]
- FERRO Marc et PLANCHAIS Jean, *Les médias et l'histoire*, Paris, Seuil, 1997
- HOOG Emmanuel, *La télé, une histoire en direct*, Paris, Découvertes Gallimard, 2010
- JEANNENEY Jean-Noël, *Une histoire des médias des origines à nos jours*, Paris, Seuil, 1996
- JOST François (dir.), *50 fiches pour comprendre les médias*, Paris, Bréal, 2012
- JOST François, *Comprendre la télévision et ses programmes*, Paris, Armand Colin, 2005 [2ème ed. 2009]
- JOST François, *Introduction à l'analyse de la télévision*, Paris, Ellipses, 1999 [3ème ed. 2007]

II- Ouvrages généraux sur les archives

- BABELON Jean-Pierre, BARBAT Philippe, FORLIVESI Luc, GILLET Patricia, *Les Archives Mémoire De La France*, Paris, Découvertes Gallimard Culture Et Société, 2008
- COEURE Sophie, DUCLERT Vincent, *Les Archives. Nouvelle édition Entièrement Refondue Et Mise à Jour ed. 2011*, Paris, Collection Repères Histoire, 2011
- FARGE Arlette, *Le goût de l'archive*, Paris, Seuil, 1989
- FAVIER Jean, *Les Archives, 7e édition Mise à Jour ed. 2001*, Paris, Que Sais-je ? 2001.
- GUIET-MONCHAL, Lydiane (dir.), *Abrégé d'archivistique Principes et pratiques du métier d'archiviste*, 4ème édition refondue et augmentée, Paris : Association des archivistes français, 2020

III- Ouvrages spécialisés sur les archives audiovisuelles

- GUYOT Jacques et ROLLAND Thierry, *Les archives audiovisuelles, Histoire, culture, politique*, Paris, Armand Colin, 2011
- HOOG Emmanuel, *L'INA*, Paris, PUF, 2006
- LABONNE Sophie, BRAEMER Christine, *Les archives audiovisuelles*, Paris : Association des archivistes français, 2013
- STOCKINGER Peter, *Les Archives Audiovisuelles Description, Indexation Et Publication*, Paris, Hermès Science Lavoisier, 2011
- STOCKINGER Peter, *Nouveaux Usages Des Archives Audiovisuelles Numériques*, Paris, Hermès Science Lavoisier, 2011

IV- Ouvrages spécialisés sur les archives du web

- BACHIMONT Bruno, *Patrimoine et numérique : technique et politique de la mémoire*, Bry-sur-Marne, INA-Éditions, 2017
- BLANDIN Claire, ROBINET François, SCHAFER Valérie (dir.), *Penser l'histoire des médias*, Paris, CNRS Editions, 2019
- CHAIMBAULT Thomas, *L'archivage du web*, Paris, ENSSIB, 2008
- GEBEIL Sophie, *Website Story, Histoire, mémoires et archive du web*, INA Editions, 2021
- LANG Brian, *Le dépôt légal des publications électroniques*, Paris : UNESCO, 1996

- LYMAN Peter, Archiving the world wide web dans In Council on library and Information resources, *Building a National Strategy for Preservation : Issues in Digital Media Archiving*, Berkeley, 2002
- MUSIANI Francesca, PALOQUE-BERGES Camille, SCHAFER Valérie, THIERRY Benjamin G., *Qu'est-ce qu'une archive du web ?* OpenEdition Press, 2019
- SCHAFER Valérie, *Archives : comment le Web devient patrimoine*, The Conversation, 24 avril 2017

V- Ouvrages spécialisés sur le web et internet

- ALBOUY Michel (dir.), *Chroniques de finance et de gouvernance d'entreprise*, Caen, EMS Editions, 2019
- ARSENE Séverine, *La Chine et le contrôle d'Internet. Une cybersécurité ambivalente*, Paris, Annuaire français des relations internationales, 2019
- DORTIER Jean-François (dir.), *La Communication. Des relations interpersonnelles aux réseaux sociaux*. Auxerre, Éditions Sciences Humaines, « Synthèse », 2016
- DUFOUR Arnaud, GHERNAOUTI-HELIE Solange, *Internet*, Paris, PUF, « Que sais-je ? », 2006
- MOUTOUH Huges, *Dictionnaire du renseignement*, Paris, Perrin, 2018
- SCHAFER Valérie et SERRES Alexandre, *Histoire de l'Internet et du Web*, Infoclio, 2016

VI- Articles scientifiques

- BIGA Maria-Giovanna, CHARPIN Dominique, DURAND Jean-Marie, « Paolo Matthiae, une appréciation », *Revue d'assyriologie et d'archéologie orientale*, 2012 (Vol. 106), p. 1-4
- BRUGGER Niels, « L'historiographie de sites Web : quelques enjeux fondamentaux », *Le Temps des médias*, 2012, numéro 18
- CLAVERT Frédéric, MULLER Caroline (dir.), « Le goût de l'archive à l'ère du numérique », *La Gazette des archives*, n°253, 2019
- DUCHEIN Michel, « Le respect des fonds en archivistique : principes théoriques et problèmes pratiques », *La Gazette des Archives*, 2e trimestre 1977, p. 71-96.

- FOATELLI Alexandre, « L'archivage du web : un outil pour comprendre internet », dans *La Revue des médias*, 2016
- GALLET Mathieu, « La France et l'Europe face à l'enjeu du *Big Data*. L'exemple de la collecte et du traitement des données médiatiques », *Géoéconomie*, 2014, numéro 69,
- MUSSOU Claude, « Et le Web devint archive : enjeux et défis », *Le Temps des médias*, 2012, numéro 19
- OUERFELLI Tarek, « Archives audiovisuelles et valorisation du patrimoine à l'ère du numérique », dans *Les Cahiers du numérique*, 2015 (Vol. 11)
- QUONIAM Luc et BOUTET Charles-Victor, « *Web 2.0, la révolution connectique* », *Document numérique*, 2008, volume 11
- SABY Frédéric, « Approche historique du dépôt légal en France », *Sociétés & Représentations*, numéro 35, 2013
- SCHAFER Valérie, « Internet historians and the “risks” of web archive », AG de l'IIPC, Bibliothèque Nationale de France, le 26 mai 2014, Paris

VII- Article de presse

- GOUDAILLIER Christophe, « Comment Myspace a effacé 12 ans de musique », *Le Parisien*, mars 2019
- LELOUP Damien et TROUVE Pierre, « Comment Google est devenu un empire », *Le Monde*, septembre 2018
- RICHAUD Nicolas, « Les 15 dates qui ont fait de Facebook un empire », *Les échos*, février 2019
- SOYEZ Fabien, « Pourquoi il faut archiver le web d'aujourd'hui et d'hier », *CENT France*, août 2020

VIII- Entretien

- Entretien de Annick le Follic mené par Valérie Schafer, consultable en ligne : < <https://asap.hypotheses.org/168> >, (consulté le 14/08/22)

Sitographie

I- Juridique

- Code du Patrimoine : < https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000020905828/ > (consulté le 05/03/22).
- CSA : < <https://www.csa.fr/Cles-de-l-audiovisuel/Connaitre/Histoire-de-l-audiovisuel/La-sauvegarde-de-la-memoire-audiovisuelle-par-l-INA> > (consulté le 05/03/22).
- Loi DADVSI : < <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000266350/> > (consulté le 05/03/22).
- Loi du 21 juin 1943 : < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006071152/> > (consulté le 05/03/22).
- Loi du 4 août 1974 : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000333539> > (consulté le 05/03/22).
- Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication < <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068930/> > (consulté le 05/03/22).
- Déclaration des Droits de l'Homme et du Citoyen adopté en 1789 : < <https://www.legifrance.gouv.fr/contenu/menu/droit-national-en-vigueur/constitution/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789> >, (consulté le 12/08/22)
- Charte UNESCO portant sur la conservation du numérique : < http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html >, (consulté le 12/08/22)

- Décret du 20 mai 2020 portant sur le renouvellement du mande de Laurent Vallet : < <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041898341> >, (consulté le 13/08/22)

II- INA et BNF

- Rapport d'activité INA, p.21 : < <https://www.INA.fr/sites/default/files/2021-06/RA-2020.pdf> > (consulté le 05/03/22).
- Site web de INA sup : < <https://www.INA-expert.com/enseignement-superieur> > (consulté le 05/03/22).
- Site web de INAMédiaPro : < <http://www.INAmediapro.com> > (consulté le 05/03/22).
- Site web de l'INA : < <https://www.INA.fr> > (consulté le 05/03/22).
- Site web délégation INA Méditerranée : < <https://www.INA.fr/institut-national-audiovisuel/en-regions/INA-mediterranee> > (consulté le 05/03/22).
- Site web Madelen : < <https://madelen.INA.fr> > (consulté le 05/03/22).
- Site BNF Dépôt légal des sites web : Mode d'emploi : < <https://www.BNF.fr/fr/centre-d-aide/depot-legal-des-sites-web-mode-demploi#step02> > (consulté le 05/03/22).
- Site web de la BNF utilisation d'internet : < <https://www.BNF.fr/fr/utiliser-des-ordinateurs-et-internet-la-BNF> >, (consulter le 19/08/22)
- Site web regroupant les ressources de la BNF : < <https://data.BNF.fr> >, (consulté le 19/08/22)
- Site web « Sudorama » : < < <https://fresques.INA.fr/sudorama> > (consulté le 05/03/22).

III- Lieux de consultation INA rattaché à l'INA Méditerranée

- Site web bibliothèque Ceccano : < <https://bibliotheques.avignon.fr/in/faces/details.xhtml?id=816093b1-a9ef-42d4-8430-3c370fe36fd3> > (consulté le 05/03/22).
- Site web bibliothèque Louis Nucéra : < <https://bmvr.nice.fr> > (consulté le 05/03/22).
- Site web cinémathèque Casa di Lume : < <https://www.casadilume.corsica> > (consulté le 05/03/22).
- Site web de la MMSH : < <https://www.mmsh.fr> > (consulté le 05/03/22).

IV- Archivage du web

- International Internet Preservation Consortium : < <https://netpreserve.org> >, (consulté le 12/08/22)
- Site web de la « Wayback Machine » : < <https://archive.org/web/> >, (consulté le 12/08/22)
- Site web « Internet Archives » : < <https://archive.org> >, (consulté le 14/08/22)
- Projet européen NEDLIB (préservation des publications numériques) : < <https://cordis.europa.eu/project/id/LB5648/fr> > (consulté le 05/03/22).

V- Plateforme audiovisuelle archivée par l'INA

- Site web de la plateforme Salto : < <https://actu.salto.fr> >, (consulté le 15/08/22)
- Site web de la plateforme Lacinetek : < <https://www.lacinetek.com/fr> >, (consulté le 15/08/22)
- Site web de la web TV « Made in Marseille » : < <https://madeinmarseille.net/web-tv/> >, (consulté le 15/08/22)

VI- Outils utilisés par l'INA

- Site web de « Elastic Search » : < <https://www.elastic.co/fr/elasticsearch/> >, (consulté le 15/08/22)
- Site web de « Kibana » : < <https://www.elastic.co/fr/kibana/> >, (consulté le 15/08/22)
- Site web de « Hadoop » : < <https://hadoop.apache.org> >, (consulté le 15/08/22)
- Site web « Hootsuite » : < <https://www.hootsuite.com/fr> >, (consulté le 17/08/22)
- Site web « Talkwalker Alert » : < <https://www.talkwalker.com/fr/alerts> >, (consulté le 17/08/22)

Table des illustrations

Figure 1 : Organigramme du secrétariat général de l'INA	p. 93
Figure 2 : Organigramme de la direction générale déléguée	p. 94
Figure 3 : Organigramme de la direction générale campus.....	p. 95
Figure 4 : Tableau des dates d'entrées dans le dépôt légal en fonction du support	p. 105
Figure 5 : Page de l'université Jean Jaurès datant du 11 mai 2000.....	p. 114
Figure 6 : Page du journal « <i>Le Parisien</i> » datant du 5 mars 2000.....	p. 115
Figure 7 : Fonctionnement des robots de collectes	p. 126
Figure 8 : Centre de donnée de l'INA à Bry-sur-Marne	p. 130
Figure 9 : Principe de collecte des UGC».....	p. 136
Figure 10 : Les cinq étapes de la veille informationnelle	p. 141
Figure 11 : Tableau des informations consultables dans une notice documentaire d'une archive web par onglet.....	p. 146
Figure 12 : Carte des lieux de consultation des fonds des archives web de l'INA	p. 148
Figure 13 : Une recherche en « plein texte ».....	p. 152
Figure 14 : Une recherche par contenus vidéo.....	p. 153
Figure 15 : Une recherche sur le site web de l'INA Thèque	p. 154
Figure 16 : Application « Archives de l'internet ».....	p. 160

Annexes

Annexe 1 : logo de l'INA

Consultable en ligne : < <https://www.ina.fr> > (consulté le 11/08/22)



Annexe 2 : Organigramme de l'INA

