



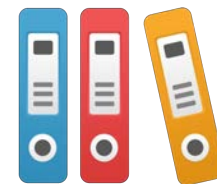
# Dépôt Légal du Web de l'Ina

novembre 2025



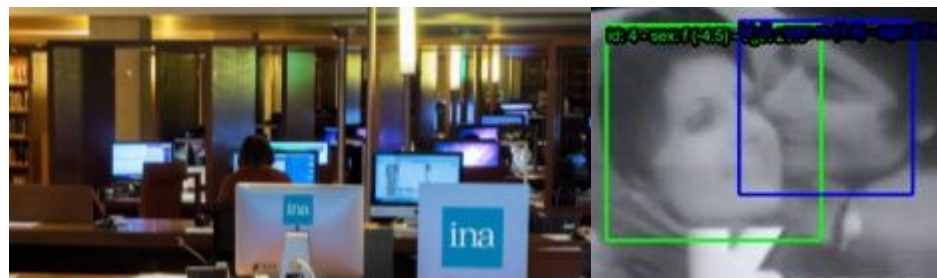
ina

# Institut National de l'Audiovisuel



Conservation des archives  
audiovisuelles, Formation, et  
Recherche

Recherche sur  
les médias



1974



1992  
Dépôt légal

1999  
Numérisation

1995  
INA thèque

2004  
INA mediapro

2009  
Archivage du web

2020  
mediaclip

2022  
le lab

L'INA, producteur et  
éditeur de contenus

2006

Ina.fr Youtube | Facebook | Twitter | Instagram madelen data.ina.fr

2020



# Dépôt Légal de l'Internet : Loi DADVSI, 2006

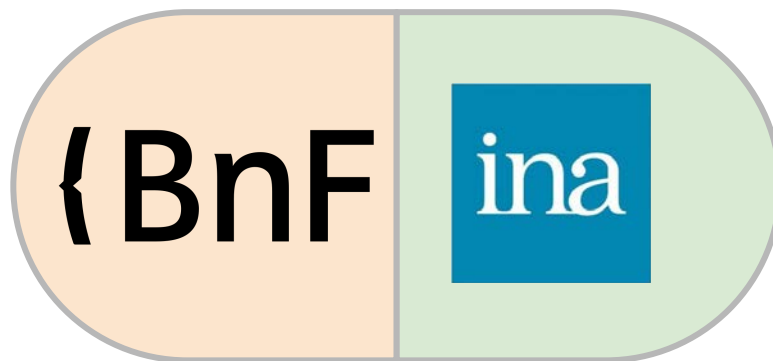
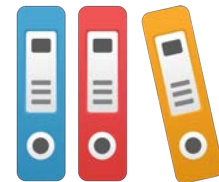


Extension du dépôt légal à l'Internet

Transposition de la Directive Européenne 2001/29/EC sur l'harmonisation des droits d'auteurs et droits voisins

*Extension du dépôt légal aux «signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique»*

# Dépôt Légal de l'Internet : Décret, 2011



- Le périmètre de l'Ina, parmi les sites français, correspond aux :
- Sites émanant des services des médias audiovisuels ;
- Web TV et Web radios ;
- Sites principalement consacrés aux programmes radio et télé ;
- Sites des organismes de l'environnement professionnel et institutionnel du secteur de la communication audiovisuelle.

La BnF archive les autres sites français.



# Dépôt Légal INA : Consultation

**Consultation  
experte**

**7 lieux  
Inathèque  
Délégations  
régionales**

**Consultation  
autonome**

**50 PCMs  
Bibliothèques  
partenaires**

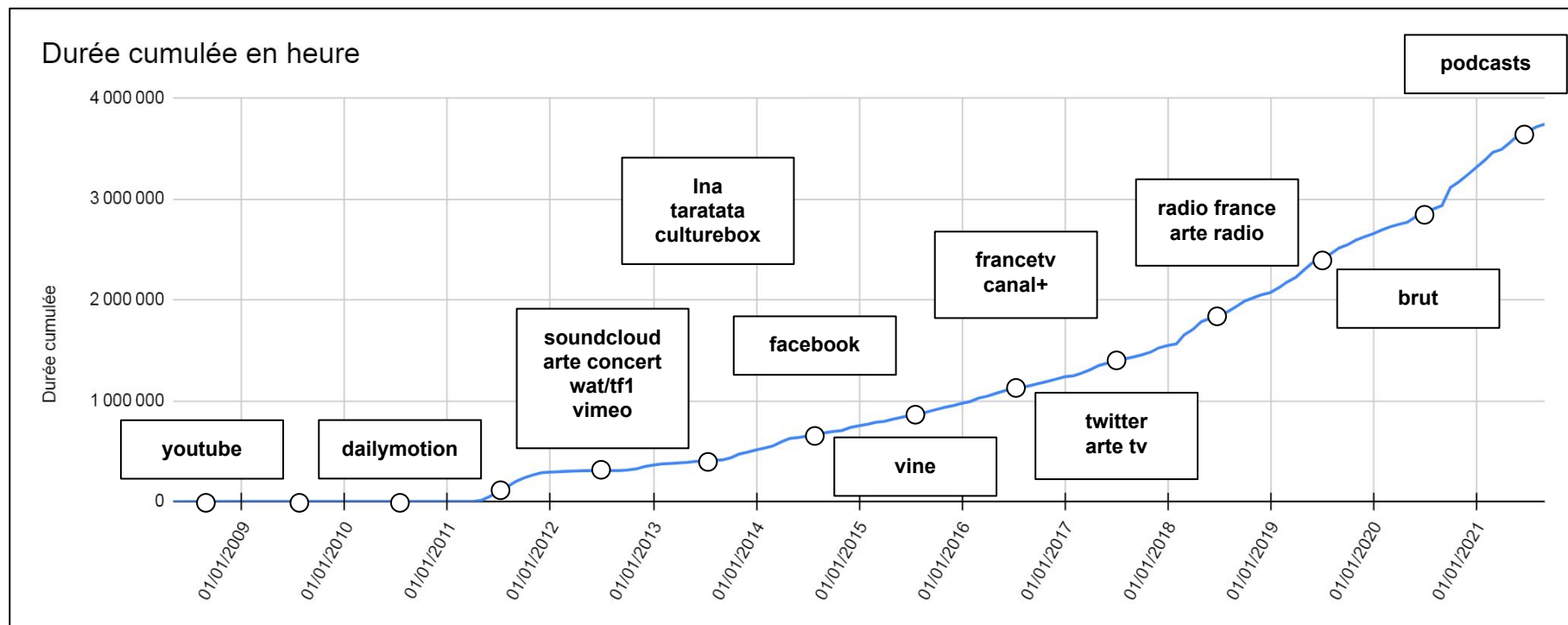
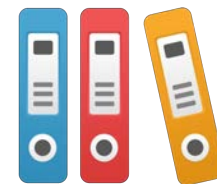


# Périmètre : nomination et documentation



- Périmètre :
  - En lien avec la Télévision et la Radio en France
  - Médias audiovisuels natifs Français
  - Publications liées à des événements à forte résonance en France
- Type de sources :
  - Sites web
  - Chaînes vidéo
  - Comptes et hashtags de réseaux sociaux
  - Programmes de Replay TV et vidéos natives
  - Emissions de Replay Radio et podcasts natifs

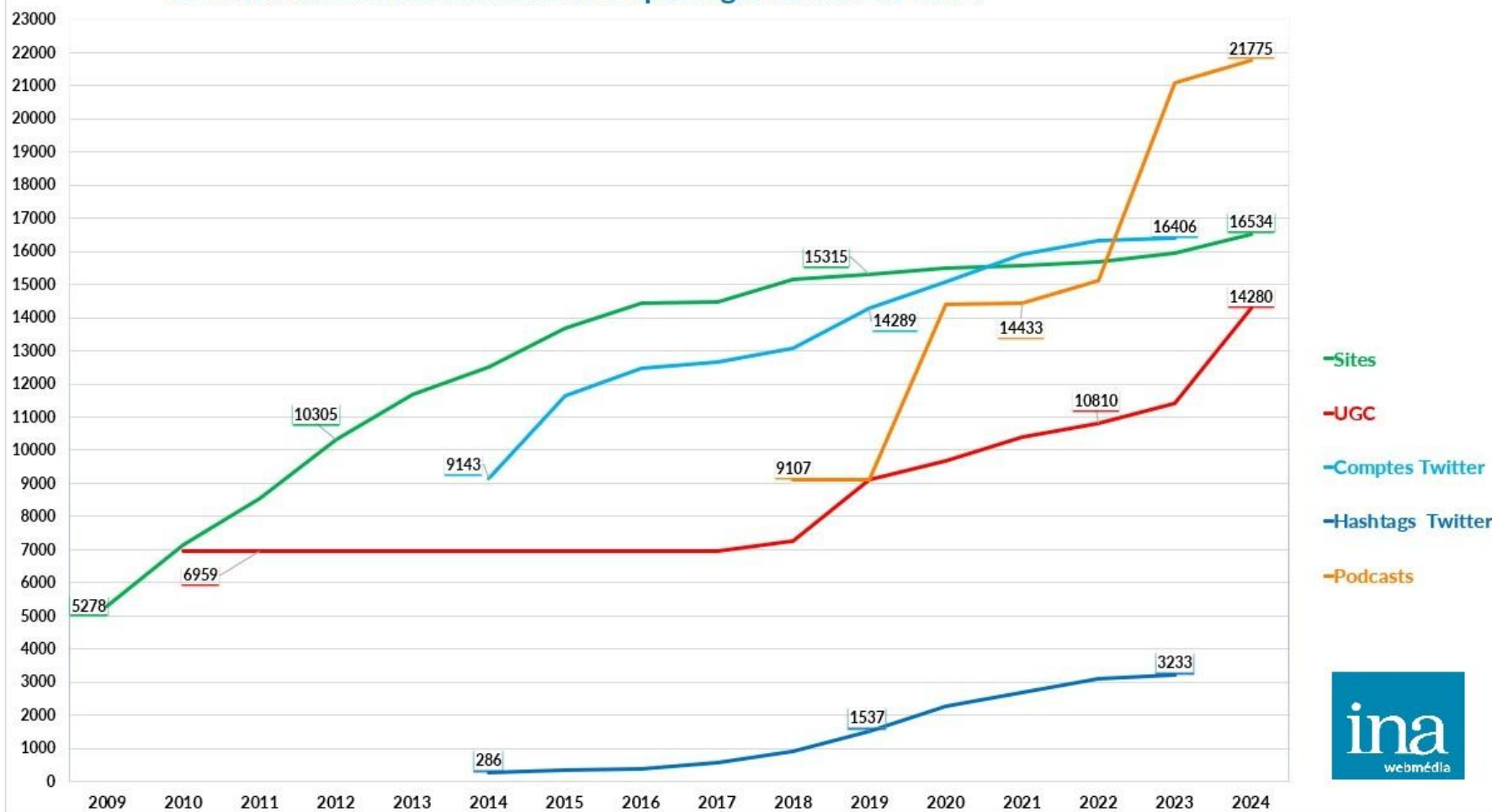
# Du web des sites au web des plateformes



# Evolution des sources de la collecte

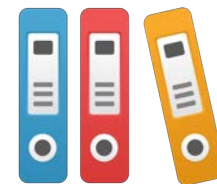


Évolution annuelle du nombre d'objets web inscrits à la collecte et consultables dans l'archive du Dépôt légal du web de l'INA



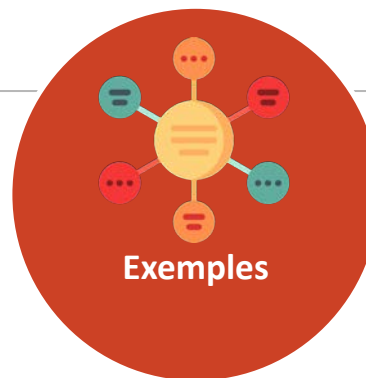


# Archive des sites



170 milliards d'urls, dont 20 milliards de pages à partir de 29 000 sites web :

- éditeurs TV/Radio
- sites professionnels
- sites & blogs de fans



france•tv

| MY TFI |

6play

Brut.

Chérie FM

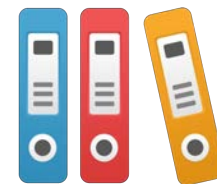
TLM

S

GRENOUILLE  
88.8fm  
grenouille888.org

ina

# Archive des vidéos



## Vidéos

42 millions de vidéos  
sur 15 000 comptes ou  
chaînes sur 15 plateformes  
- TV/Radio replay/exclus  
- Média web natifs  
- Youtubeurs populaires

5 millions d'heures

2 millions d'heures

## plateformes



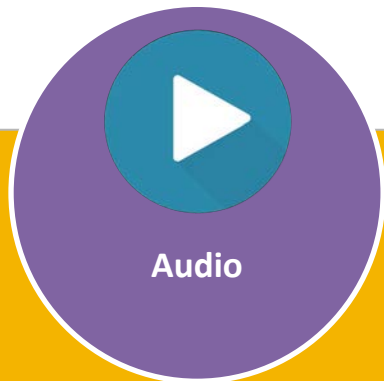
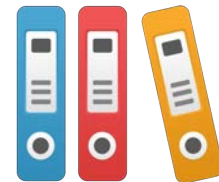
Collecte YouTube limitée depuis septembre  
2024

## Chaînes



ina

# Archive des audios & podcasts



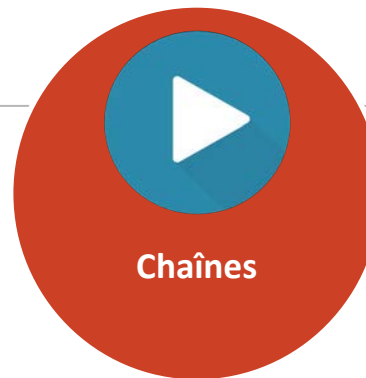
Audio

5 millions de sons  
sur 30 000 chaînes ou  
émissions

- Replay Radio
- podcasts indépendants et natifs

2 millions d'heures

5 millions d'heures



Chaînes



arte

nova

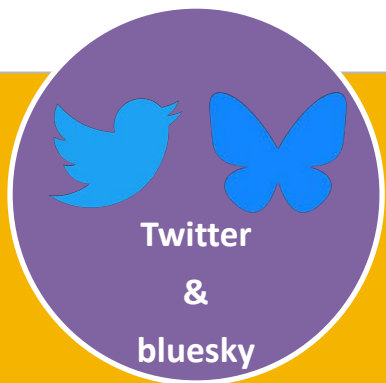
radiofrance



:le billet sciences



# Archive des réseaux sociaux



## Twitter :

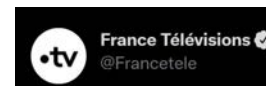
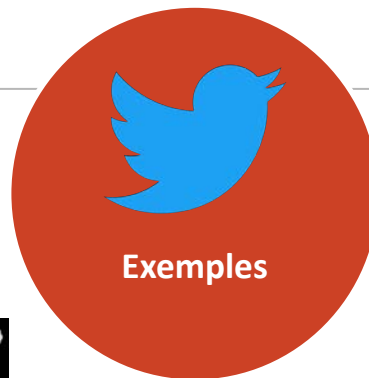
3.1 milliards de tweets à partir  
de 13 500 comptes & 2 700  
hashtags :

- chaînes, émissions et personnalités TV/Radio
- actualités des médias
- grands événements
- comptes @elysee, PR

## Bluesky :

Comptes des TVs et Radios

Comptes des TVs et Radios  
Bluesky :



#jesuischarlie

#NuitDebout

#tpmp

#13hfoot

#Covid19

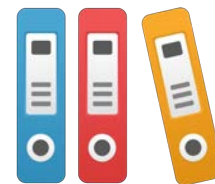
#GiletsJaunes

#KohLanta

Collecte Twitter stoppée en juillet 2023  
Collecte Bluesky démarrée en juin 2025



# Dépôt Légal et grandes plateformes



Aujourd'hui, la loi de dépôt légal ne s'applique pas aux acteurs bénéficiant du statut d'hébergeur.

Ce qui exclut pour l'heure les grandes plateformes numériques de devoirs envers le dépôt légal en France.

L'INA n'a pas de facilité particulière d'accès aux données de ces plateformes, et peut-être empêché de collecter.

# Projets en cours



- Collecte de Contenus (S)VOD :
  - Automatisation de screen recording
  - problématiques de fiabilité, fidélité et qualité à estimer et parer si possible
  - Test de contournement des DRMs ?
- Collecte de Vidéos Live :
  - Live Twitch, Youtube
  - Fast TV, ...
- Mise à jour des applications et moteurs de recherche
  - Consultation de l'archive des médias du web
  - Consultation de l'archive des réseaux sociaux
  - Intégration des données Bluesky à la consultation



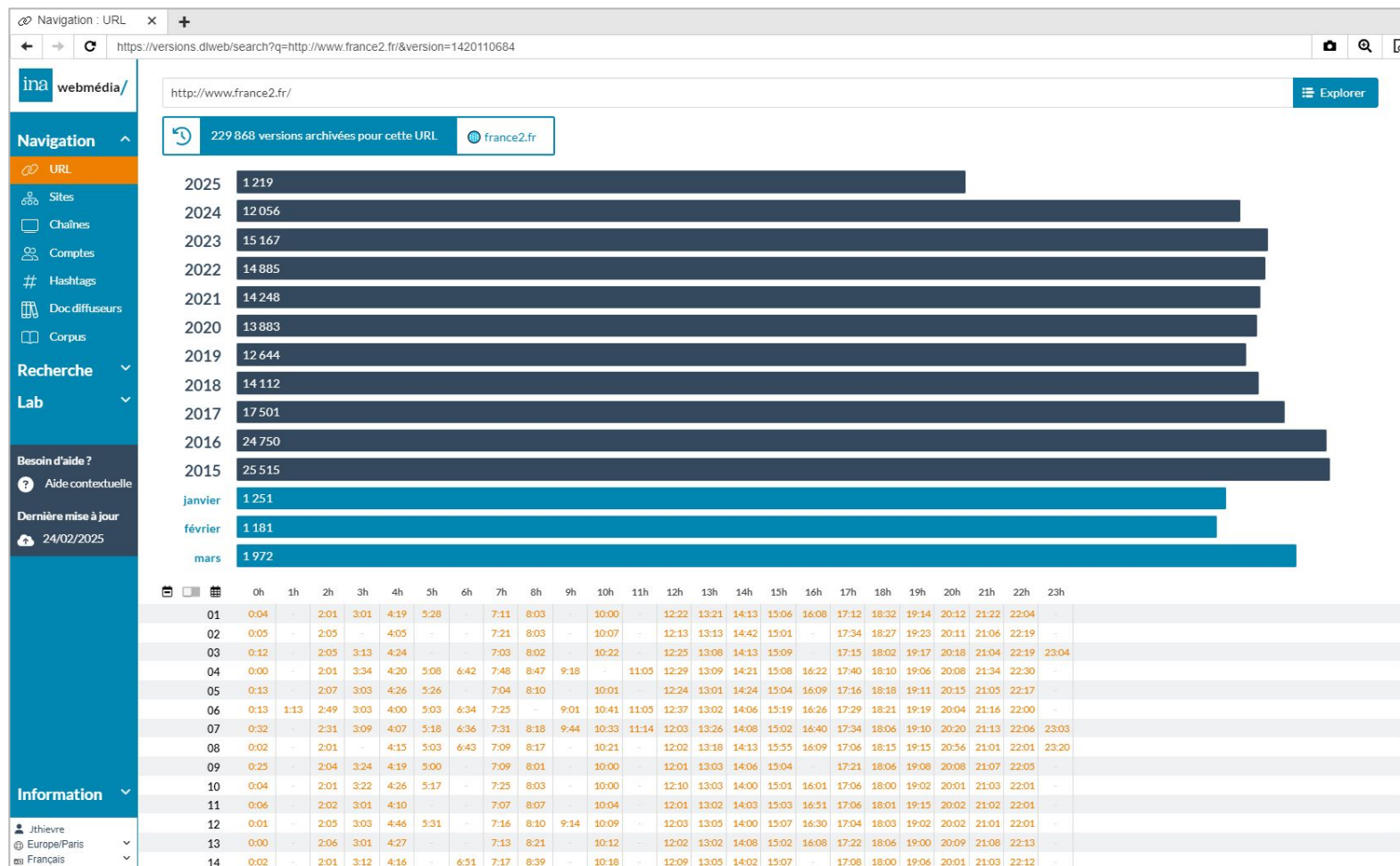
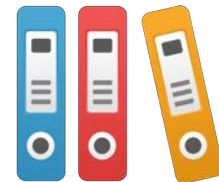
# Démo

Webmédia



ina

# Consultation des sites & pages web





# Consultation des sites & pages web



**ina webmédia/**

**Navigation** ▾  
**Recherche** ▴  
Pages  
Vidéos  
Twitter  
Social beta  
**Lab** ▾

Besoin d'aide ?  
Aide contextuelle  
Dernière mise à jour  
23/11/2023

**Information** ▾  
Anonyme  
Europe/Paris  
Français

ukraine guerre

Chercher

Entre le dd-mm-yyyy et le dd-mm-yyyy sur ex: tf1.fr trié par les plus pertinentes ▾

☒ valeurs relatives

Année	Nombre de résultats (approximatif)
1996	0
1997	0
1998	0
1999	0
2000	0
2001	0
2002	0
2003	0
2004	0
2005	0
2006	0
2007	10
2008	0
2009	5
2010	10
2011	10
2012	5
2013	10
2014	20
2015	20
2016	20
2017	20
2018	15
2019	15
2020	15
2021	15
2022	100
2023	80
2024	0
2025	0

4 220 296 résultats

<https://vl-media.fr/tag/guerre-ukraine/>

**guerre Ukraine -**

vl-media.fr

26-03-2022 13:32:25 — totalement dégénéré ces dernières heures entre la Russie et l'Ukraine, focus sur les cinq questions à se poser sur cette guerre "surprise". Que s'est-il passé ces derniers jours ? L'hypothèse d'une... Soyez Informés Catégories (...) 9 mars 2022 En plein contexte de guerre, Alexis Corbière s'est retrouvé dans une situation délicate après avoir évoqué partiellement le contenu des (...) livraisons d'armes françaises à l'Ukraine en direct. Alexis Corbière était l'invité de Marc Fauvelle... 24 février 2022 Alors que la situation a

voir tous les résultats pour ce site (1 513 ou plus)

<http://www.france-ukraine.com/La-guerre-n-a-pas-un-visage-de.html>

**'La guerre n'a pas un visage de femme', de Svetlana Alexievitch - France-Ukraine.com**

28-09-2010 16:57:58 — 'La guerre n'a pas un visage de femme', de Svetlana Alexievitch Le 20 novembre 2004 Svetlana Alexievitch est née en Ukraine, en 1948. Journaliste

voir tous les résultats pour ce site (47 970 ou plus)

<https://www.cnews.fr/guerre-en-ukraine>

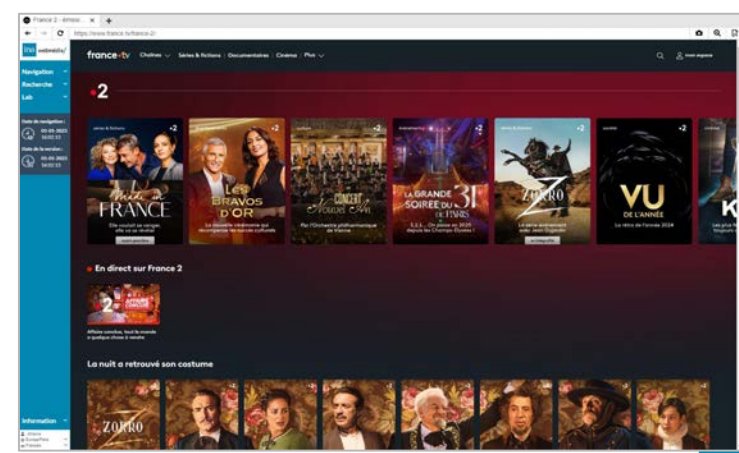
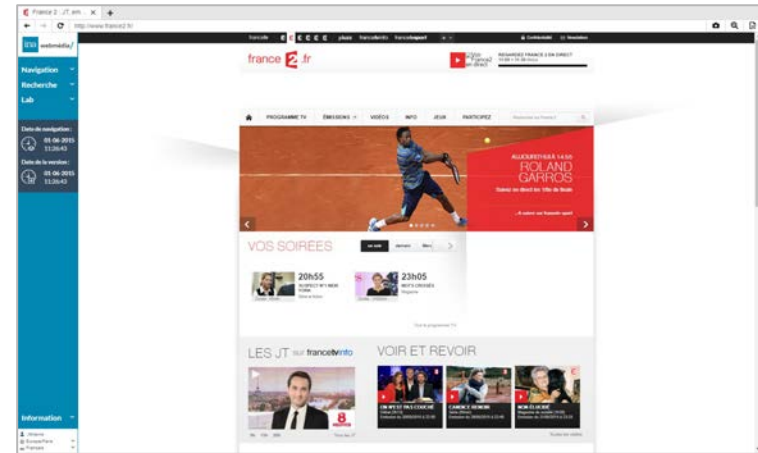
**Guerre en Ukraine | CNEWS**

cnews.fr

02-12-2022 02:19:40 — Guerre en Ukraine Retrouvez toute l'actualité autour de la guerre en Ukraine sur CNEWS

voir tous les résultats pour ce site (22 214 ou plus)

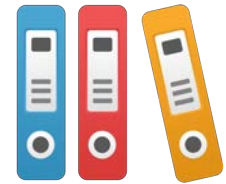
# Consultation des sites & pages web







# Consultation des vidéos/podcasts



ina webmédia/

**Navigation**

**Recherche**

Pages

**Videos**

Twitter

Social beta

**Lab**

Besoin d'aide ?

Aide contextuelle

Dernière mise à jour

24/02/2025

**Information**

Anonyme

Europe/Paris

Français

1:35

TikTok

"M. Birch ne travaille pas pour le ch...

24-02-2025 09:03

1:29

TikTok

#SixNationsRugby | Élu meilleur...

France tv sport 24-02-2025 09:01

0:20

TikTok

Procès de Mohamed Amra : "J'espè...

RTL 24-02-2025 08:48

0:42

TikTok

FIGHT WEEK POUR GOMIS À l'...

RMC Sport Combat 24-02-2025 08:35

5:54

france+tv

Les dossiers d'ICI : La police munic...

France 3 24-02-2025 08:26

31:59

france+tv

ICI 19/20 édition toutes régions : É...

France 3 24-02-2025 08:22

08:04

SOUNDCLOUD

"Tropical Fever" vol.133 "Tropical ...

radiomega 24-02-2025 08:21

**Dossier**

19/20 Edition Toutes Régions

0:02 / 5:54

5s 1 1 5s

Les dossiers d'ICI : La police municipale recrute

France 3

france.tv / Pluzz

00:06:54

24-02-2025 08:26

webdav:france3tv-0970915

Immersion, coulisses, décryptage : la rédaction vous emmène chaque jour au cœur de nos territoires.

En semaine, un dossier de fond pour mieux comprendre les enjeux de l'actualité.

Le week-end, un grand reportage pour suivre au plus près les acteurs de la société française et de ses changements.

1:00:46

SOUNDCLOUD

"Tropical Fever" vol.132 "caralbes ...

radiomega 24-02-2025 08:16

2:04

TikTok

Quel rôle jouent les réseaux sociaux...

france3hautsdefra... 24-02-2025 08:15

2:17

TikTok

Le dernier établissement de la myt...

BFMTV 24-02-2025 08:13

10:18

Europe 1

LE JOURNAL DE LA RÉDACTION

Podcast

Le journal de 8h du 24/02/2025

Europe 1-Le jour... 24-02-2025 08:13

10:18

Europe 1

LE JOURNAL DE LA RÉDACTION

Podcast

Le journal de 8h du 24/02/2025

Europe 1-L'info en ... 24-02-2025 08:13

10:22

france+tv

Journal 08h00 : Édition du lundi 24...

France 2 24-02-2025 08:10

6:51

TikTok

Louis Boyard contre Cyril Hanouna...

TPMP Officiel 24-02-2025 08:09

RTM 9:55

Les infos

3:01

Les infos

3:01

Les infos

3:00

Les infos

3:31

Les infos

3:00

2:00

# Consultation des publications RS



ina webmédia/

Navigation

Recherche

Pages

Vidéos

Twitter

Social Beta

Lab

Besoin d'aide ?

Aide contextuelle

Dernière mise à jour

18/07/2023

Information

Anonyme

Europe/Paris

Français

Recherche

Visualisations

805 832 publications (44ms)

☐ Texte 661 635

☐ Photo 108 214

☐ Vidéo 20 058

☐ Gif 15 325

Langue

Hashtags

Rechercher par Hashtags

☐ Elonmusk 119 558

☐ Ukraine 63 824

☐ Covid19 52 403

☐ Mahsaamini 50 978

☐ Crypto 31 409

☐ Nft 26 134

☐ Vr 24 000

☐ Metaverse 20 977

☐ Bitcoin 19 054

☐ Cryptocurrency 18 581

Voir plus

Emojis

17 962

11 306

10 685

9 566

7 573

6 520

5 587

4 751

4 003

3 682

3 536

3 533

3 531

3 517

3 014

2 939

2 751

2 666

2 546

2 494

2 464

1 962

1 306

1 068

956

757

652

558

1 / 67 153

Les plus anciennes

Filtres

Terme : elonmusk

SpaceX @SpaceX  
30/04/2009 17:02:35 (Europe/Paris)  
@ElonMusk Elon Musk does not have a Twitter account. Please discontinue this account immediately. Your fraud has been reported.  
6.6 K 30.7 K 0 0

Ted Patrick @\_led\_  
14/02/2013 18:32:55 (Europe/Paris)  
Dammit, now I want a car with an API, thanks @elonmusk  
0 1 0 0

mathieu Kassovitz @kassovitz1  
19/03/2013 13:47:20 (Europe/Paris)  
@elonmusk Sir. I have a S Signature.it's time to make it fun let us tweak with the cpu driving modes should be incorporated in the system.  
0 0 0 0

mathieu Kassovitz @kassovitz1 a répondu  
23/03/2013 03:03:11 (Europe/Paris)  
@elonmusk like... Super super dead ?  
0 0 0 0

Mark Rechlin @MarkRechlin  
09/04/2013 02:37:41 (Europe/Paris)  
You say you want a revolution? Well, you know, Elon wants to change the world.  
http://t.co/7OjKKRaUqq @TeslaMotors @ElonMusk  
32 17 0 0

Dominique Delport @domdelport  
30/05/2013 08:03:41 (Europe/Paris)  
ELECTRIC : #ElonMusk: #Tesla Supercharger network to triple by the end of next month  
http://t.co/2yPYqD365  
0 0 0 0



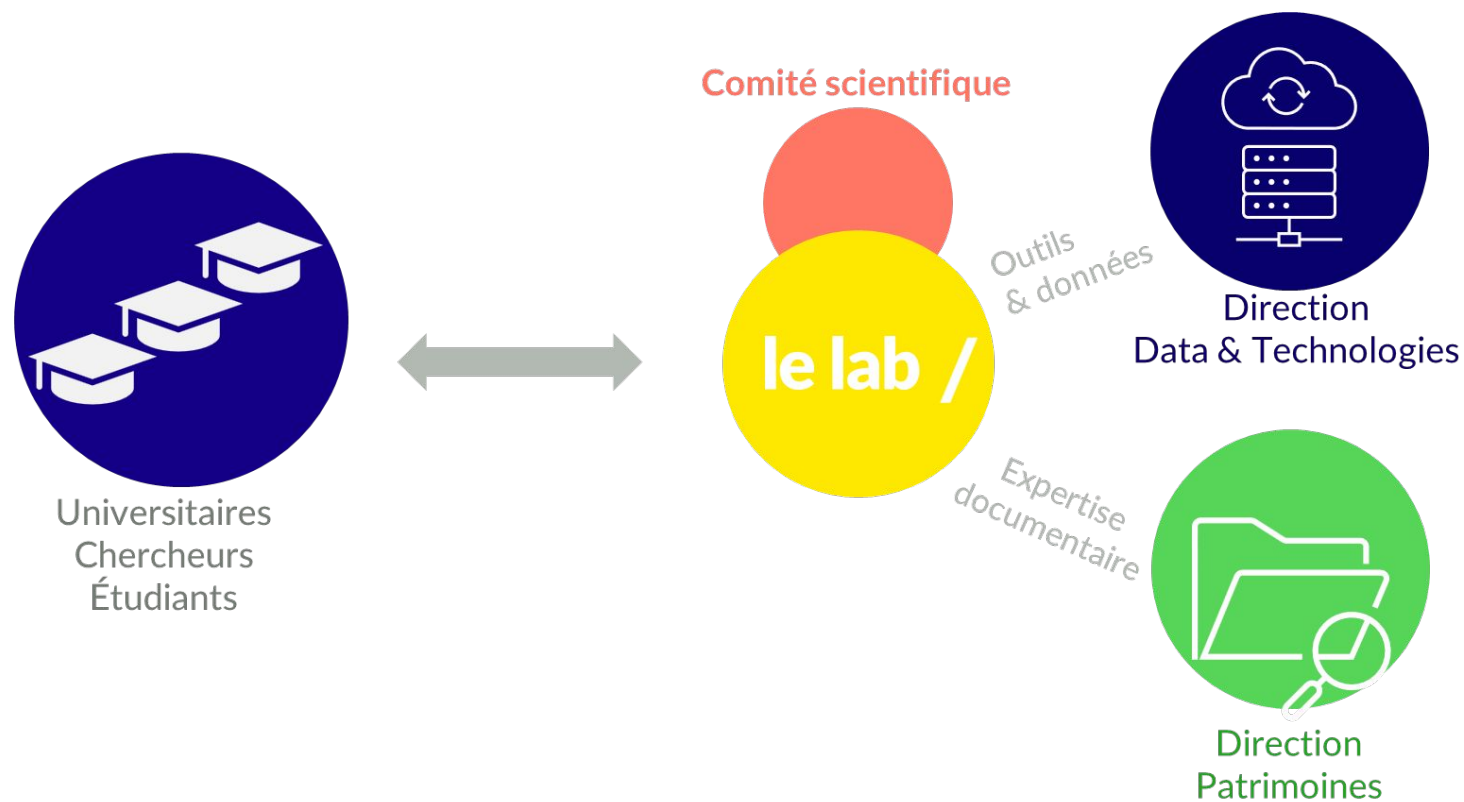
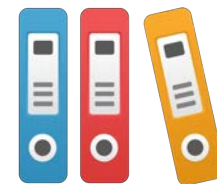
# Ina / le lab

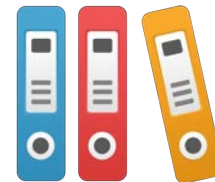
accompagnement des chercheurs



ina

# INA / Le lab





## Une offre à facettes

- **Accompagnement méthodologique**
  - Permanence sur RDV
  - Sprints (1-2 mois)
  - Projets en résidence (AAP annuel)
- **Mise à disposition de données et d'outils**
  - **Animation scientifique**
  - Séminaire
  - Ateliers



# INA / le lab : missions

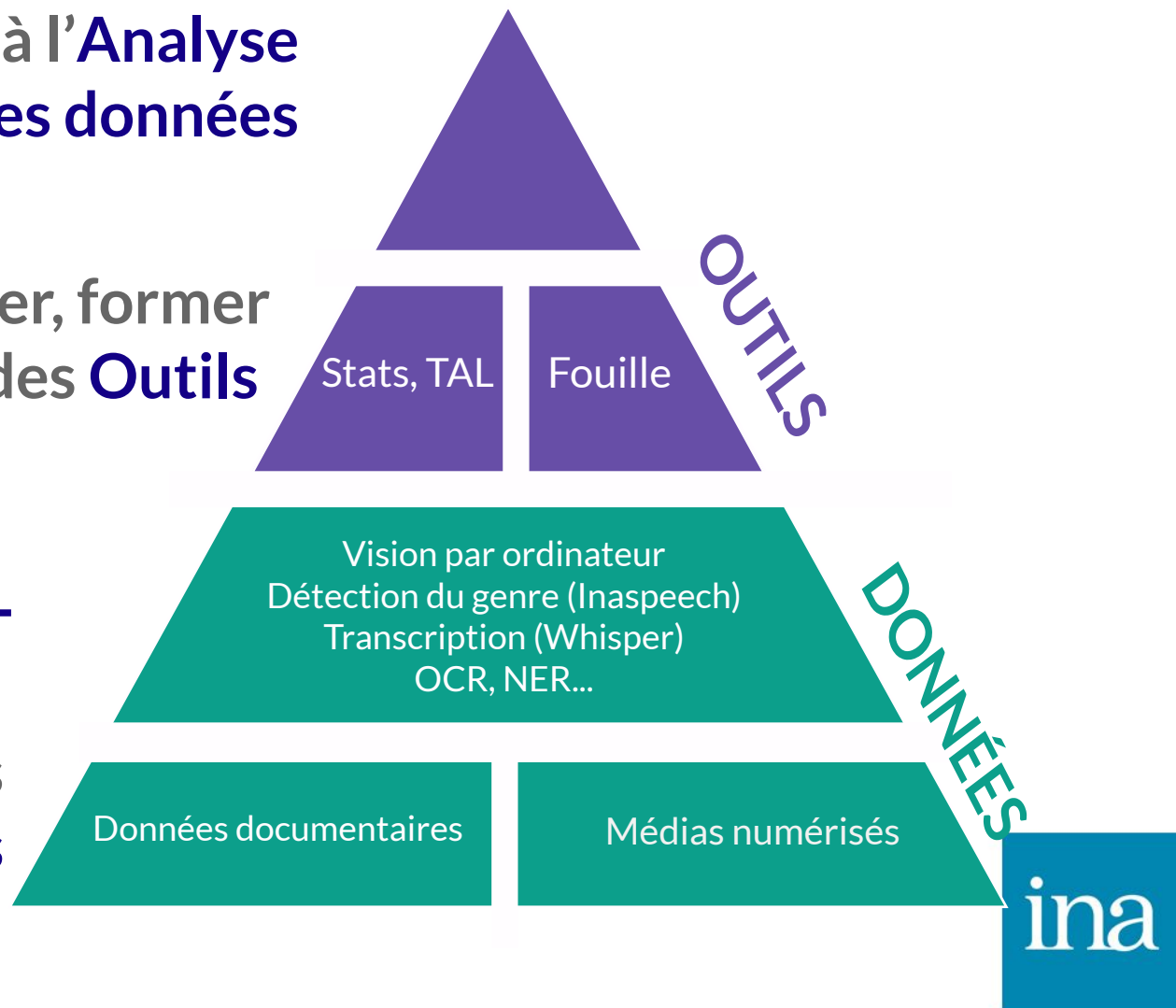


Participer à l'**Analyse**  
des données

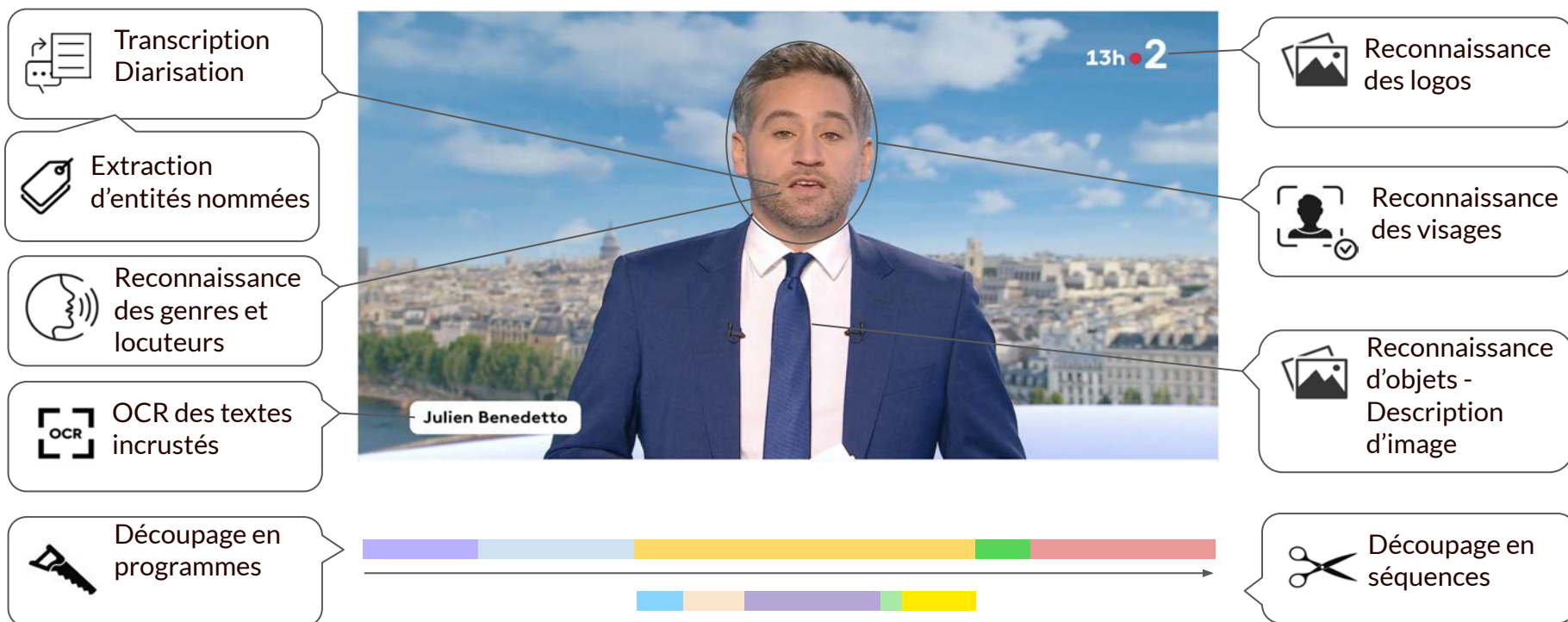
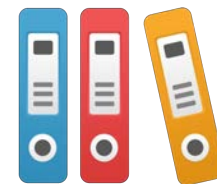
Déployer, former  
à des **Outils**

Réaliser des  
**Traitements IA, ML**

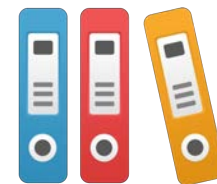
Fournir des corpus  
d'**Archives**



# INA / Le lab : traitements & outils



# INA / Le lab : accompagnement de projets



**105** chercheurs rencontrés depuis 2022

**84** projets soutenus

dont **54** l'année passée

## Par type d'approche

**44%** de thèse

**25%** de recherche individuelle

**10%** de projet collectif

**10%** travaux de master

## Par discipline

**27%** en information et communication

**20%** en sociologie

**17%** en économie

**8%** en informatique

**3%** en histoire



## Données utilisées (2024-2025)

### Par type d'approche

**85%** télévision

**10%** web

**5%** radio

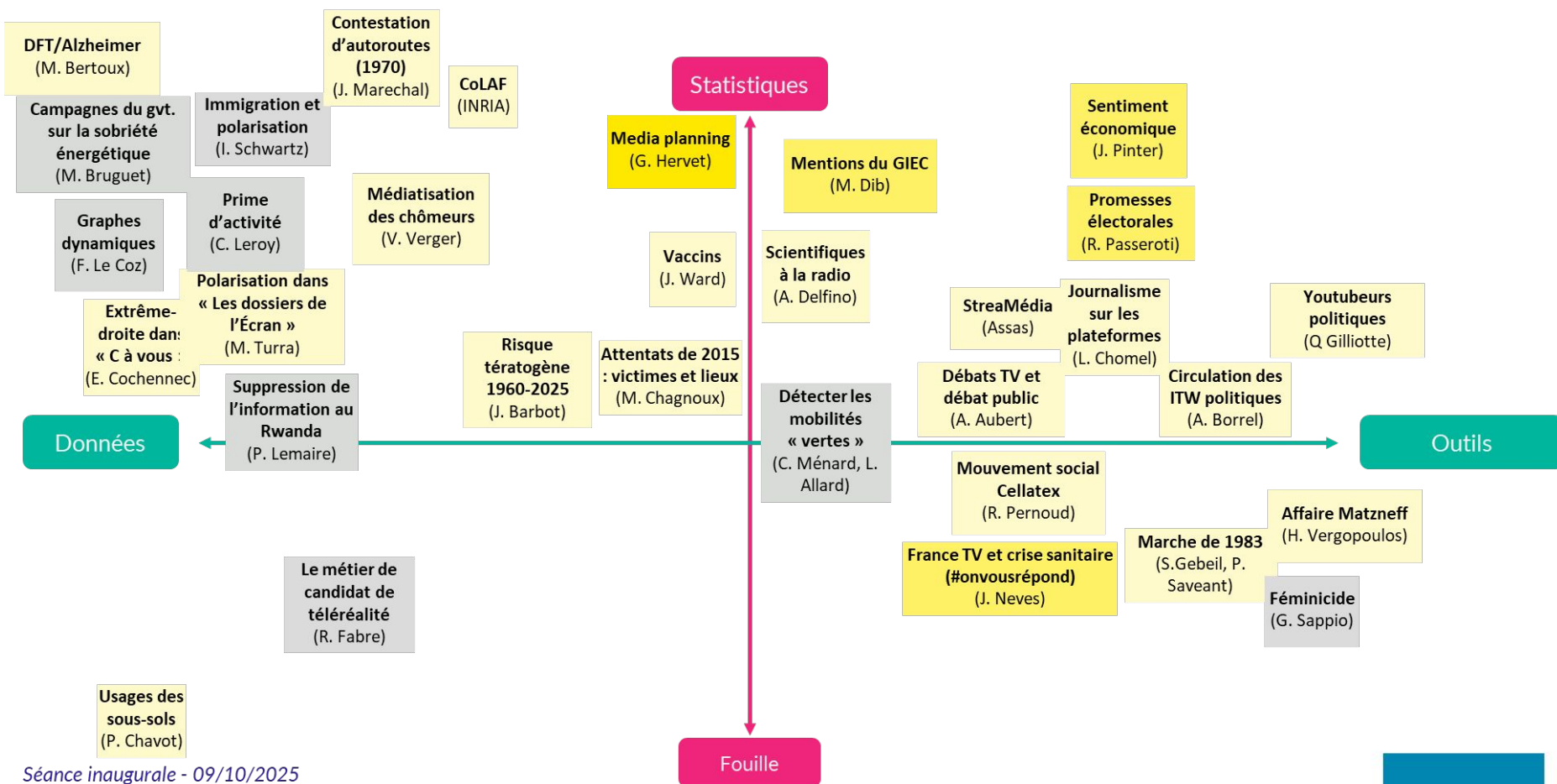
### Par discipline

**75%** utilisent la transcription

**22%** utilisent uniquement les notices  
documentaires et métadonnées

**8%** utilisent le média en lui même

# INA / Le lab : accompagnement de projets



Séance inaugurale - 09/10/2025

# INA / Le lab : accompagnement de projets



**Janine BARBOT**, Directrice de recherche en **sociologie** à l'INSERM

« *La trajectoire du risque tératogène dans les médias français depuis l'affaire de la thalidomide (1960-2025)* »

**Marie CHAGNOUX**, Maîtresse de conférences en **SIC** à l'Université Paris 8 Vincennes-Saint-Denis

« *Médiagénie des actants : appréhender en dynamique le rôle des médias dans la construction mémorielle des attentats de novembre 2015* »

**Philippe CHAVOT**, Maître de conférences en **SIC** à l'Université de Strasbourg

« *La construction des représentations publiques des usages des ressources du sous-sol* »

**Laurent CHOMEL**, Docteur et chercheur associé en **SIC** à l'Université Sorbonne-Paris-Nord

« *Reconfigurations du journalisme critique sur les plateformes : analyse comparée de 6 médias alternatifs vidéographiques français (2021-2025)* »

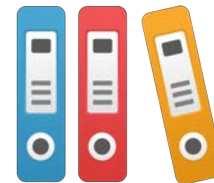
**Anouck DELFINO**, Doctorante en **SIC** à Sorbonne Université

« *Les scientifiques à l'antenne : étude de la visibilité et du traitement médiatique des chercheurs dans les médias sonores français (radio et podcasts natifs) entre 2018 et 2025* »

**Rémi PASSEROTI**, Doctorant en **Sciences politiques** à l'Université Grenoble Alpes

« *La trajectoire des promesses électorales : dynamiques médiatiques & stabilité programmatique lors de la présidentielle de 2022* »

# INA / Le lab : accompagnement de projets



Séminaire 2025-2026

Replays : [www.youtube.com/@inarecherche](https://www.youtube.com/@inarecherche)

Jeudi 9 octobre 2025    **Séance inaugurale:**  
Critique de l'économie politique de l'espace public

Nikos Smyrnaio

14 novembre 2025    Santé

Body Capital (Solene Lellinger et Christian Bonah)  
Jeremy Ward et Senmiao Yang  
Marius François et Mohamed Hassan

16 janvier 2026    Écologie et climat

Quota Climat (Jean Sauvignon)  
Charles Menard et Laurence Allard  
Samuel Gache

13 mars 2026    Emploi et précarité

Michael Zemmour  
Vincent Verger  
Claire Leroy

22 mai 2026    Mouvements sociaux

Paola Sedda  
Raphaël Pernoud  
Thibault Grison

19 juin 2026    Intervention politique et information

TBD  
Julien Boyadjian  
Pauline Lemaire

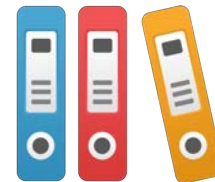




# Impacts de l'IA



# L'IA à l'INA



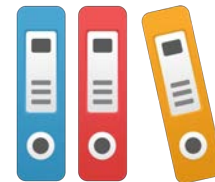
## L'équipe 2IA : Intégration de l'IA

Industrialisation de traitements à des archives Radio, TV et web

### **Analyses image/vidéo, audio et textes**

- transcription de la parole en texte
- segmentation paroles/musiques & hommes/femmes & locuteurs
- reconnaissance des locuteurs
- référentiel des visages des personnalités publiques
- analyse des génériques
- classification thématique
- recherche multi-modale : plein-texte / vectorielle

# L'IA à l'INA

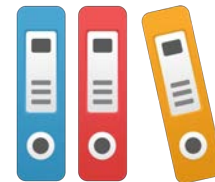


## Le service de la recherche

La description des contenus audiovisuels est au coeur de la création des activités de recherche de l'INA

- Gender Equality Monitor : Rapport annuel ARCOM/INA
- Pantagruel (ANR) / ArGIMI (BPI) : création et évaluation de LLMs souverains
- STREAMédia (ANR) : reconnaissance et classification des genres éditoriaux des vidéos d'actualité

# L'IA à l'INA



## Applications au dépôt légal du web

### Traitements par lots de médias du web en fonction :

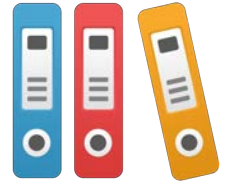
- besoins internes
- demandes de chercheurs

### Principalement des traitements audio aujourd'hui

- transcription de la parole en texte
- segmentation paroles/musiques & hommes/femmes & locuteurs
- reconnaissance des locuteurs

360 000 heures en cours de traitements

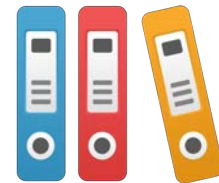
# Usages des données dépôt légal & IA



Toutes demandes d'export de données pour des projets de recherche passent une analyse scientifique et juridique.

Les droits et devoirs des chercheurs d'un projet font l'objet d'une licence d'utilisation de données.

# Impacts de l'IA sur la collecte du Web



Les grands modèles de langage se sont nourris des données présentes sur le web.

Depuis quelques années nous observons une fermeture des APIs permettant l'accès aux données des plus grandes plateformes.

Ces mêmes plateformes se protègent plus efficacement du scraping, ce qui impacte aussi l'archivage du web.

C'est aussi le cas des sites produisant des contenus de valeurs comme la presse et les plateformes de TV et Radios.



Merci



ina