

## Séminaire d'étude

# 20 ans de web électoral (2002-2022)

À l'occasion du 20<sup>e</sup> anniversaire des collectes d'archives du web électoral qu'elle effectue avec des bibliothèques partenaires, la Bibliothèque nationale de France (BnF), en partenariat avec le groupe Science politique du numérique de l'Association française de Science politique, le CEVIPOF (Centre de recherches politiques de Sciences Po) et le médialab Sciences Po, organise un séminaire d'étude de deux jours.

Les ateliers organisés le 9 novembre au BnF DataLab permettront de découvrir et d'explorer collectivement ces collections de dépôt légal du web, source unique pour l'étude de la vie politique française des vingt dernières années, au prisme de différents outils.

La journée de conférences du 10 novembre sera consacrée aux évolutions qui ont marqué depuis 2002 le web électoral, les données et les méthodes permettant d'en faire l'étude. Le web a profondément bouleversé la communication partisane et les modalités du « faire campagne » ; les réseaux sociaux numériques et les grandes plateformes de médias ont un poids grandissant dans la circulation des contenus et la production de données en période électorale. Le séminaire questionnera ces évolutions et la complémentarité des matériaux qui permettent de les étudier : collections constituées par la BnF dans le cadre du dépôt légal du web, corpus de sites, de pages ou de posts issus des réseaux sociaux numériques constitués par les équipes de recherche, ou encore données générées par les grandes plateformes de médias et réseaux sociaux, devenus des acteurs majeurs dans la collecte et la production de données. Quels sont les conditions d'un usage croisé de ces sources ? Comment les méthodologies pour les étudier ont-elles évolué ?

### Jeudi 9 novembre 2023

9h30 – 17h00

**BnF | François-Mitterrand  
BnF DataLab**

Quai François-Mauriac, Paris 13<sup>e</sup>

### Vendredi 10 novembre 2023

9h00 – 17h30

**BnF | Richelieu  
Salle des conférences**

5, rue Vivienne, Paris 2<sup>e</sup>

### Inscription obligatoire :

<https://framaforms.org/inscription-au-seminaire-detude-20-ans-de-web-electoral-2002-2022-les-9-et-10-novembre-2023>

Il est possible de s'inscrire uniquement à l'une des journées.

Image : Allec Gomes, Unsplash



# Journée du 9 novembre 2023 : Explorer collectivement 20 ans d'archives du web électoral au prisme de différents outils

Ateliers, BnF DataLab

9 h 30

**Accueil au BnF DataLab**

10 h 00

**Atelier : découverte des archives du web électoral conservées à la BnF**

Effectuées par la BnF et un réseau de bibliothèques réparties sur l'ensemble du territoire à l'occasion de chaque grande échéance électorale, ces collectes permettent de capturer les débats qui agitent la toile à travers l'archivage des sites des formations politiques, des candidats, d'organisations de soutien, de médias, d'analyse politique, de syndicats, d'association, de militants ou de simples observateurs, et constituent une source unique pour l'étude de la vie politique et sociale française des vingt dernières années.

12 h 30

**Pause méridienne**

14 h 00

**3 ateliers au choix**

▪ **Atelier 1 : Usages avancés des outils de recherche et de consultation**

Cet atelier vous permettra d'approfondir l'exploration des archives du web électoral grâce à l'utilisation de la recherche plein texte dans l'application Archives de l'internet Labs, les fonctions de recherche avancée dans l'application Archives de l'internet et la manipulation des listes des sites sélectionnés à l'occasion des collectes électorales et publiées en open data sur les sites [api.bnf.fr](http://api.bnf.fr) et [data.gouv.fr](http://data.gouv.fr).

▪ **Atelier 2 : Outils d'exploration enrichie et aide à la fouille de texte et de données sur la collection "élections 2002"**

Cet atelier vous permettra d'explorer la plus ancienne collection d'archives du web électoral, la collection relative à l'élection présidentielle et aux élections législatives de 2002, via un ensemble d'outils déployés à titre expérimental. Parmi ces outils, vous trouverez :

- SolrWayback, outil de recherche plein texte, de recherche d'images et de data-visualisation développé par la Bibliothèque royale du Danemark,
- des indicateurs et données dérivées qui permettent une première lecture distante de la collection,
- des Jupyter Notebooks qui permettent de faire des analyses sur des échantillons (extraction de la liste des images ou du plein texte des pages, par exemple).

▪ **Ateliers 3 : Explorer les archives du web électoral avec Hyphe**

Hyphe est un outil de constitution, d'exploration et de catégorisation de corpus web développé par le médialab de Sciences Po. En crawlant les web entités (des pages web individuelles, des sites web ou des agrégats de ces éléments) avec un niveau de profondeur défini, Hyphe permet de construire itérativement et de visualiser un réseau des liens hypertextes existants entre les web entités et de découvrir de nouvelles web entités potentiellement intéressantes à ajouter au corpus. Hyphe-BnF permet de constituer de tels corpus directement à partir des archives du Web.

# Journée du 10 novembre 2023 : 20 ans de web électoral (2002-2022)

9 h

Accueil

9 h 30

**Ouverture**

**Frédéric Manfrin** (chef du service Histoire, BnF)

09 h 45

**20 ans de web électoral, perspectives diachroniques**

Collecter l'éphémère : 20 ans de web électoral vus à travers les collections d'archives du web de la BnF, **Anaïs Crinière-Boizet** (service du Dépôt légal numérique, BnF)

Analyser l'usage du web en campagne électorale en France: retour sur 10 ans d'enquête **Anaïs Théviot** (ARENES, Université catholique de l'Ouest)

Les stratégies de communication des partis politiques dans les médias sociaux et la polarisation des opinions de leurs sympathisants

**Julien Figeac** (LISST, Université de Toulouse - Jean-Jaurès)

10 h 45

Échanges avec la salle

11 h 05

Pause

11 h 20

**Sources et données pour l'analyse du web électoral : de la collection au corpus**

Cette session portera sur les types de sources et de données pouvant être mobilisées pour comprendre le web électoral, ainsi que sur les méthodes de collecte de ces données éphémères. Quels sont les choix scientifiques, organisationnels et techniques qui président à la constitution de ces corpus et quelle est l'incidence des outils utilisés ? Comment passe-t-on de la collection à un corpus de recherche permettant une exploration thématique ?

Constituer les archives du web électoral

**Isabelle Copin** (BnF, chargée de collections Presse Département Droit, Economie, Politique),

**Anaïs Crinière-Boizet** (BnF, service du Dépôt légal numérique), **Laurence Le Bellec**

(Médiathèque d'Orléans, responsable du Dépôt légal imprimeur), **Catherine Soulé-Sandic**

(Bibliothèque nationale universitaire de Strasbourg, cheffe du département Arts, Langues, Littératures et Aires culturelles)

Argumenter en ligne sur ce qui n'existe pas : le cas du clonage humain dans l'élection de 2002

**Guillaume Levrier** (LISIS, CNRS)

12 h 10

Echanges avec la salle

12 h 30

Pause méridienne

14 h 00

**Sources et données pour l'étude du web électoral (suite) : stratégies et outils de collecte, comprendre et archiver le fonctionnement de Youtube en période électorale**

Les grandes plateformes de médias et réseaux sociaux sont devenus des acteurs majeurs de la collecte et de la production de données en période électorale : comment rendre compte de leur rôle dans la circulation des contenus ?

L'évolution des recommandations Youtube en période électorale : le cas des présidentielles de 2017

**Guillaume Chaslot** (PeREN)

Mesurer les recommandations algorithmiques sur Youtube lors de la campagne présidentielle 2022

**Erwan Le Merrer** (INRIA) et **Gilles Tredan** (LAAS, CNRS)

14 h 40

Échanges avec la salle

15 h 05

Pause

15 h 15

**Méthodes pour l'analyse du web électoral**

Session animée par **Julien Boyadjian** (CERAPS, Université de Lille)

Cette session portera sur l'évolution des méthodes d'analyse des données et des corpus, et notamment sur les perspectives ouvertes par l'adoption de méthodes quantitatives : analyses lexicales et sémantiques, analyse de corpus massifs, perspectives comparatistes, qu'elles soient longitudinales ou transnationales, ou encore cartographie du web. L'adaptation récente par le médialab de Hyphe, un outil de constitution et de cartographie de corpus web, au web archivé ouvre en effet des perspectives stimulantes et novatrices : analyse diachronique des campagnes électorales, des réseaux partisans ou militants, ou encore études thématiques ciblées à partir des archives du web électoral.

Méthodes quantitatives dans l'analyse du web électoral

**Julien Boyadjian** (CERAPS, Université de Lille)

Les élections européennes sur le web en France et en Allemagne : vers un espace politique transnational ?

**Morgane Le Boulay** (SAGE, Université de Strasbourg), **Céline Monicolle** (SAGE, Université de Strasbourg)

Etude de la structuration des communautés politiques autour des candidats à la présidentielle - l'exemple des communautés autour de Jean-Luc Mélenchon (2012, 2017, 2022)

**Fabienne Greffet** (IRENEE, Université de Lorraine)

L'analyse des opinions politiques sur Twitter (titre provisoire)

**Marta Severo** (Dicen-IdF, Université de Nanterre)

16 h 50

Échanges avec la salle