

# Impresso – Media Monitoring of the Past

*Mining 200 years of historical newspapers.*

## Lead

Comment les archives de journaux peuvent-elles nous aider à comprendre le passé ? Comment les explorer ? Le projet interdisciplinaire « Media Monitoring of the Past », auquel la section d'histoire est associée, vise à développer un cadre méthodologique et technologique permettant de nouveaux modes d'exploitation des archives de presse numérisées et de nouvelles approches pour aborder des questions historiques.

## Contexte du projet de recherche

Conservés depuis des décennies sur les rayons de bibliothèques et d'archives, les archives de journaux font actuellement l'objet d'importantes campagnes de numérisation. Des millions de fac-similés et leurs équivalents textuels (acquis grâce à la reconnaissance optique de caractères) sont ainsi accessibles en ligne via de nombreux portails. Si cela représente une avancée majeure en termes de préservation et d'accès aux documents, de nombreux défis restent cependant à relever en vue de fournir un accès étendu et élaboré au *contenu* de ces ressources numériques. Ces défis concernent, entre autres, des modes hétérogènes de disponibilité et d'accessibilité aux collections, une qualité encore modeste des techniques de reconnaissance optique de caractères et un manque d'outils conformes aux impératifs de la recherche historique pour le traitement et l'exploration.

Soutenu par un réseau de partenaires associés mettant en vedette des bibliothèques, des archives, des éditeurs de journaux et des historiens, un consortium composé de linguistes informaticiens, de spécialistes de sciences humaines et de designers travaille en étroite collaboration pour relever ces défis et faire ainsi progresser la recherche dans les domaines émergents de l'histoire numérique et des humanités digitales.

## Objectifs du travail de recherche

Lancé en septembre 2017, le projet « Media Monitoring of the Past » est financé par le programme SINERGIA du Fonds national suisse de la recherche scientifique. Il est porté par le Digital Humanities Laboratory (DHLAB) de l'École Polytechnique Fédérale de Lausanne, le Luxembourg Centre for Contemporary and Digital History (C<sup>2</sup>DH) et l'Institut für Computerlinguistik de l'Université de Zurich, la section d'histoire de l'UNIL figurant au sein des partenaires associés.

Le programme de recherche poursuit trois objectifs. Premièrement, le développement et l'évaluation systématique de plusieurs composants de traitement automatique du langage naturel pour l'extraction d'information à partir de textes historiques. L'analyse automatique de 200 ans d'archives de presse suisses et luxembourgeoises permettra la construction d'une base de connaissances structurée, dont le contenu sera entièrement traçable et interopérable. Deuxièmement, la réalisation d'interfaces de visualisation innovantes pour l'exploration de vastes quantités de données historiques. Ces interfaces

seront mises au point conjointement par les historiens, les bibliothécaires et les informaticiens (technique de *co-design*), en fonction des besoins identifiés en début de projet et réévalués par la suite. Enfin, l'évaluation active et continue des outils produits avec l'exploration d'un cas d'étude historique - la résistance à l'Europe -, complétée par une réflexion commune sur l'utilisation des outils numériques dans les sciences historiques en prenant en compte des aspects méthodologiques, épistémologiques et pédagogiques.

Pour en savoir plus sur ce projet et suivre son avancement, voir <http://impresso-project.ch/>