

# Introduction

**Lyne Da Sylva, Pascal Cuxac**

DANS **DOCUMENT NUMÉRIQUE 2017/2 Vol. 20** , PAGES 7 À 9  
ÉDITIONS **JLE**

**ISSN 1279-5127**

**ISBN 9782746248533**

**Date de mise en ligne : 10/09/2018**

**Article disponible en ligne à l'adresse**

<https://stm.cairn.info/revue-document-numerique-2017-2-page-7?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...  
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



**Distribution électronique Cairn.info pour JLE.**

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur [Cairn.info/copyright](http:// Cairn.info/copyright).

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

## INTRODUCTION

---

Les 8 et 9 mai 2017 s'est tenu le colloque « Analyser la science : les bibliothèques numériques comme objet de recherche » à l'Université McGill de Montréal, dans le cadre du 85<sup>e</sup> congrès de l'ACFAS (Association francophone pour le savoir).

Ce colloque<sup>1</sup> s'insérait dans une dynamique émergente sur l'analyse de corpus d'articles scientifiques et voulait faire la promotion de la recherche francophone. Il a permis d'aborder des thématiques variées faisant cohabiter des disciplines différentes autour de la problématique de l'analyse de contenu des bibliothèques, archives et musées numériques, d'un point de vue théorique ou pratique.

À l'ère d'internet, l'accès à un volume conséquent de publications impacte le travail des chercheurs, leur positionnement dans la communauté internationale, l'évaluation de la recherche et les modèles économiques sous-jacents. Le développement des données en libre accès (OpenData), les bibliothèques, archives et musées en ligne ne sont que quelques exemples illustrant l'évolution de la notion de texte ou document, sa diversité et sa prolifération. Des collections issues de bibliothèques traditionnelles sont maintenant accessibles librement : Gallica, Digital Public Library of America. À ce type de bibliothèques numériques s'ajoutent les publications savantes, qui sont une part non négligeable des publications numériques. Des acteurs francophones importants ont vu le jour aussi bien au Canada qu'en France (Persée, CAIRN, HAL, Erudit) et de récentes initiatives nationales ont permis le développement d'archives scientifiques (ISTEX en France, SwissBib en Suisse, GBV en Allemagne, Scholars Portal en Ontario) ou muséales (Europeana, Musée virtuel du Canada).

Parallèlement, les méthodes automatiques de traitement se diversifient et se spécialisent : fouille de textes, méthodes d'apprentissage profond (deep learning) pour traiter des données massives, transformation d'informations en données liées. Cependant, de nombreuses questions restent en suspens, notamment en ce qui concerne la gestion de gros corpus multithématiques. Pouvoir disposer d'outils d'analyse efficaces, capables de s'adapter à de gros volumes de données, souvent de nature hétérogène, pas toujours structurés, dans des langues variées et des domaines très spécialisés, reste un défi.

Le traitement des données numériques implique de multiples domaines (traitement automatique des langues, intelligence artificielle, linguistique, statistique, informatique),

---

1. <http://dasyuva.ebsi.umontreal.ca/ACFAS2017/accueil.html>.

permettant des applications diversifiées : recherche d'information, veille, évaluation de programmes de recherches, prospective... Nous avons voulu couvrir tous ces aspects ainsi que les usages et pratiques autour des bibliothèques numériques.

Dans ce numéro de la revue *Document numérique*, nous proposons sept articles sélectionnés parmi les meilleures communications et illustrant différents aspects abordés pendant les deux jours du colloque.

Le premier article, signé Éric Kergosien, Marie-Noëlle Bessagnet, Maguelonne Teisseire, Joachim Schöpfel, Amin Farvardin, Stéphane Chaudiron, Bernard Jacquemin, Annig Le Parc Lacayrelle, Mathieu Roche, Christian Sallaberry et Jean Philippe Tonneau, illustre parfaitement le sujet du colloque en présentant une étude de cas faisant intervenir des approches de fouille de textes afin d'identifier des fronts de recherche en relation avec des territoires d'études, en exploitant des bibliothèques numériques de textes scientifiques.

Joseph Mariani, Gil Francopoulo et Patrick Paroubek se focalisent sur le traitement automatique de la langue en présentant le corpus NLP4NLP et présentent une analyse bibliométrique du domaine.

Annaïg Mahé et Camille Prime-Claverie abordent l'étude des pratiques communicationnelles en sciences humaines et sociales dans le domaine de l'« Open Access ».

L'article de Nicolas Rollet, Valérie Beaudouin et Isabelle Garron aborde l'usage d'une bibliothèque numérique comme Gallica grâce à une approche vidéo-ethnographique.

Éléonore Alquier, Jean Carrive et Steffen Lalande nous présentent le traitement de données documentaires audiovisuelles et les évolutions pressenties dans le domaine.

Maxime Sainte-Marie, Jean-François Gauvin et Vincent Larivière nous entraînent dans le monde des collections muséales en croisant les catalogues de collections, la littérature connexe, les données ouvertes liées et les réseaux sociaux.

Finalement, l'article de François Dominic Laramée nous présente l'analyse de documents anciens numérisés dans le but de caractériser les représentations géographiques dans l'encyclopédie de Diderot.

En tant que responsables de ce numéro spécial, nous remercions chaleureusement les auteurs pour leur contribution ainsi que les membres du comité de lecture pour leur aide précieuse, les articles soumis ayant fait l'objet d'une phase d'évaluation réalisée par au moins deux relecteurs qui, par leurs commentaires, ont participé à la qualité des articles retenus.

LYNE DA SYLVA  
EBSI, université de Montréal, Montréal, Canada

PASCAL CUXAC  
INIST-CNRS, Vandœuvre-lès-Nancy, France

*Comité de lecture*

Marc Bertin, université de Toulouse, Toulouse, France

Guillaume Cabanac, université de Toulouse, Toulouse, France

Patrick Drouin, université de Montréal, Montréal, Canada

Nicolas Dugué, université du Maine, Le Mans, France

Benoît Epron, ENSSIB, Lyon, France

Dominic Forest, université de Montréal, Montréal, Canada

Jean-François Gauvin, BanQ, Montréal, Canada

Luc Grivel, université Paris 1-Panthéon-Sorbonne, Paris, France

Evelyne Jacquey, ATILF, Nancy, France

Guy Lapalme, université de Montréal, Montréal, Canada

Audrey Laplante, université de Montréal, Montréal, Canada

Charles Lenay, université de technologie de Compiègne, Compiègne, France

David Reymond, I3M, Toulon, France