

LES MANUSCRITS MAGHRÉBINS

ET LES HUMANITÉS NUMÉRIQUES EN FRANCE

COLLECTIONS, TECHNOLOGIES ET NOUVEAUX USAGES SCIENTIFIQUES

2, 3 ET 4 MAI 2022

BULAC &

CAMPUS CONDORCET

Organisation et contact

GIS Moyen-Orient et
Mondes musulmans
Antoine Perrier (CNRS,
IREMAM)

[antoine.perrier\[at\]cnrs.fr](mailto:antoine.perrier[at]cnrs.fr)

Les manuscrits originaux du Maghreb sont particulièrement nombreux dans les collections des bibliothèques françaises, une partie d'entre eux est de plus numérisée et accessible en ligne : ils restent toutefois sous-exploités par la recherche et l'enseignement. Les nouveaux outils de philologie numérique intègrent depuis plusieurs années les textes écrits en caractères arabes : ils constituent un des moyens à la disposition des chercheurs pour s'emparer de ces collections. Pourtant, rares sont les projets scientifiques associant les technologies de reconnaissance des caractères imprimés (OCR) ou manuscrits (HTR), ou encore les techniques de traitement ou d'encodage des textes, avec les disciplines concernées par les manuscrits historiques (histoire, linguistique, codicologie, philosophie, littérature...).

L'une des raisons de ce manque, particulièrement prononcé en France, est l'absence d'identification des usages scientifiques potentiels de ces outils numériques ainsi que la présence d'une communauté d'acteurs (étudiants, chercheurs, ingénieurs et bibliothécaires) nombreux mais dispersés. Dans le cadre d'une démarche au long cours du groupement d'intérêt scientifique Moyen-Orient et Mondes Musulmans (GIS MOMM, UAR du CNRS 2999), cette école de printemps se donne comme objectifs :

- de **valoriser les collections de manuscrits des bibliothèques françaises**,
- de **fédérer une communauté d'intérêt** autour des usages numériques du traitement des manuscrits,
- de contribuer à la **formation** des acteurs de cette communauté aux technologies actuellement développées pour le traitement des manuscrits,
- d'identifier les **usages et besoins de la communauté scientifique** travaillant sur des manuscrits maghrébins vis-à-vis des humanités numériques.

Contexte

Les études maghrébines et les humanités numériques ont été identifiées parmi les priorités de la politique scientifique du [GIS MOMM](#), mettant en œuvre les axes stratégiques du CNRS et du ministère de l'Enseignement supérieur et de la recherche en matière de sciences humaines et sociales (plans SHS 2020 et 2021). Parmi les initiatives destinées à structurer le champ des études sur le Maghreb, les humanités numériques constituent un volet important pour compenser deux retards identifiés par le livre blanc de septembre 2020 ([Vers la science ouverte ? La transition numérique et la recherche sur le Moyen-Orient et les mondes musulmans en France](#)) : le référencement des ressources documentaires sur le Maghreb présentes en France et le développement des techniques de reconnaissance des caractères manuscrits (HTR) pour les langues non latines, en particulier l'arabe.

 **Moyen-Orient et
Mondes Musulmans**
Groupement d'Intérêt Scientifique



CAMPUS 
CONDORCET
Paris-Aubervilliers

BULAC
[بۇلەك] [بۇلەك] [بۇلەك] [بۇلەك]
Bibliothèque universelle
des langues et civilisations

Le projet [DigiMagh](#), débuté en 2020, répond au premier objectif tandis qu'un [premier hackathon](#) organisé au sein de la BULAC par Noémie Lucas, en partenariat avec [Calfa](#), a abouti une preuve de concept pour un modèle de reconnaissance automatique du texte de manuscrits maghrébins du fonds de la BULAC. Un second hackathon, actuellement en cours, est destiné à renforcer la polyvalence et la robustesse de ce premier outil, à partir du [dataset Rasam](#) déjà établi en 2021. La résidence numérique de Noémie Lucas à la BULAC a abouti à la rédaction d'un rapport sur la reconnaissance automatique des écritures en alphabet arabe ; celui-ci souligne la rareté des projets scientifiques intégrant des techniques d'HTR ou d'OCR en France, notamment autour des manuscrits historiques. Il met en lumière le double besoin de formation et de formalisation d'une demande scientifique pour ces technologies qui se développent très rapidement dans le monde.

Déroulement de l'école de printemps

La journée aura lieu durant trois jours :

**Lundi 2 mai 2022 : les collections de manuscrits maghrébins en France
En hybride - amphithéâtre de la BULAC (65 rue des Grands Moulins, 75013, Paris)**

Les représentants des principales institutions conservant ou travaillant sur des manuscrits maghrébins en France (**BULAC, BnF, MMSH**) dresseront un état des lieux de fonds de manuscrits en caractères arabes, en insistant sur la variété matérielle et thématique de ces fonds. Ces présentations seront associées à celles de projets portant sur l'histoire, l'indexation et l'identification de ces collections par **Augustin Jomier** (INaLCO), **Noémie Lucas** (University of Edinburgh), **Muriel Roiland** (CNRS, IRHT) et **Mehdi Sakatni** (GIS MOMM, projet DigiMagh).

**Mardi 3 mai 2022 : les usages scientifiques des humanités numériques (hybride)
En hybride - amphithéâtre de la BULAC (65 rue des Grands Moulins, 75013, Paris)**

Cette deuxième journée propose une réflexion sur les potentiels ouverts par les humanités numériques aux différentes disciplines travaillant sur les manuscrits arabes du Maghreb. À partir de leurs propres corpus, qui ne se limiteront pas aux collections françaises, les chercheurs évoqueront l'apport réel ou potentiel des outils numériques, la façon dont ils modifient les façons de lire ou d'analyser le document manuscrit, ainsi que les limites possibles de ces technologies. Les nouvelles pratiques induites par la numérisation des collections, en particulier dans les pays du Maghreb, et par les technologies HTR et OCR, seront abordées dans le cadre de projets en cours ou de projets futurs, par des spécialistes d'histoire, d'histoire de l'art, de linguistique, de littérature, ou de philosophie.

Avec les interventions de **Hiba Abid** (New York Université), **Claire Grégoire** et **Jean-Christophe Peyssard** (projet LiPoL), **Jawdath Jabeur** et **Sarah Maloberti** (projet PhASIF), **Elhoussaine Oussiali** (CNRS, Archimède), **Nuria Martinez de Castilla** (EPHE), **Clément Salah** (Université de Lausanne), **Elise Voguet** (CNRS, IRHT) et **Ismail Warscheid** (CNRS, IRHT).

Mercredi 4 mai 2022 : ateliers de formation**En petit groupe –Campus Condorcet Bâtiment Sud- salles 1.122 et 1.023
(5 cours des Humanités 93300 Aubervilliers)**

Cette dernière journée sera structurée en ateliers en petits groupes autour de deux moments :

- Une matinée consacrée aux **technologies d'HTR**, en partenariat avec **Calfa**, dans le prolongement des deux premiers hackathons de la BULAC. Étudiants, enseignants-chercheurs, chercheurs, ingénieurs, bibliothécaires sont invités à se former, lors d'une séance de transcription collective, à l'usage de la plateforme de Calfa et des modèles jusqu'à présent développés pour l'arabe maghrébin.
- Un après-midi dédié au **traitement automatique des langues (TAL)**, en partenariat avec **PostLab**. L'identification des problématiques de recherche, des besoins d'édition ou d'indexation des textes permettra d'envisager les réponses offertes par les outils de TAL (alignement et classification des textes, du vocabulaire, indexation et comparaison des entités nommées, encodage, etc.) pour déterminer les apports, pour la recherche et les bibliothèques, de textes enrichis.

Modalités de participation

La journée est ouverte à tout public intéressé (étudiants de master, doctorants et post-doctorants, chercheurs et enseignants-chercheurs, ingénieurs, bibliothécaires et professionnels des bibliothèques, etc.) pour les deux premières journées.

L'accès aux ateliers de la troisième journée est limité aux participants arabisants, travaillant sur des manuscrits et intéressés par leur traitement numérique.

Les personnes souhaitant participer sont invitées à [remplir le formulaire suivant](#) en ligne avant le **1er mars 2022.**