

# PROJET

## Convention de coopération de recherche

### Programme d'étude Madame Bovary – Transcription des manuscrits

Entre :

La VILLE DE ROUEN

Dont le siège est : Place Général de Gaulle 76 037 ROUEN

N° SIRET : 217 605 401 000 17

CODE NAF : 76540

Représentée par son Maire : Monsieur Pierre ALBERTINI en application d'une délibération du Conseil Municipal du .....

Ci-après désignée par la VILLE

Et :

L'UNIVERSITE DE ROUEN

Etablissement Public à Caractère Scientifique Culturel et Professionnel

Dont le siège est : 1, rue Thomas Becket – 76821 MONT SAINT AIGNAN cedex

N° SIRET : 197 619 042 000 17

CODE NAF : 803 Z

Représentée par son Président, Monsieur Jean Luc NAHEL

Ci-après désignée par l'UNIVERSITE

L'UNIVERSITE et la VILLE sont ensemble ci-après désignées par les PARTIES

Il est convenu ce qui suit :

#### **Article 1      Objet de la convention**

La VILLE et l'UNIVERSITE ont décidé d'engager un programme de coopération scientifique visant la valorisation des fonds patrimoniaux de la Bibliothèque Municipale de Rouen.

Cette volonté s'est traduite par la signature de la convention générale de coopération et de recherche du ..... (date). La présente convention, relative au programme particulier "Madame Bovary", vient préciser et compléter les articles stipulés dans le cadre de la convention générale.

La VILLE et l'UNIVERSITE décident de coopérer dans le cadre d'un programme intitulé : "Analyse automatique de documents manuscrits: Contribution à la transcription du dossier "Madame BOVARY""

Le programme, ci-après désigné PROGRAMME, est décrit dans l'annexe technique jointe.

La VILLE autorise l'UNIVERSITE à indiquer cette convention de coopération de recherche, ainsi que son objet et le nom de la VILLE et sa Bibliothèque dans ses documents d'information ou rapports d'activités.

## **Article 2      Description générale du programme**

La VILLE conserve tous les scénarios, brouillons et manuscrits de "Madame Bovary" de Gustave Flaubert.

Le programme associant l'Université et la Ville vise à permettre la conception et le développement d'une édition hypertextuelle de la totalité des manuscrits avec leur transcription. Les manuscrits de "Madame Bovary" numérisés seront proposés au monde entier, accessibles depuis le site Flaubert de l'UNIVERSITE (<http://www.univ-rouen.fr/flaubert>) et depuis le portail web en cours de définition de la VILLE.

Le laboratoire Perception Systèmes Information (PSI), à travers notamment le travail de l'équipe de recherche du Pr. Thierry PAQUET a une expertise particulière dans la reconnaissance automatique de l'écriture. La contribution du laboratoire PSI concerne les opérations de transcription des folios des manuscrits (le détail et les objectifs du programme figurent en annexe technique).

## **Article 3      Exécution du programme – responsables scientifiques**

Les travaux seront réalisés par le Laboratoire Perception Systèmes Information (PSI) ci-après désigné par le LABORATOIRE, dirigé par Monsieur le Professeur Jean-Pierre PECUCHET .....

Le Responsable Scientifique du programme est Monsieur le Professeur Thierry PAQUET .....

Les travaux de numérisation seront réalisés par la VILLE, au sein de la Bibliothèque Municipale de Rouen, dirigée par Madame Françoise LEGENDRE, Conservateur général .....

Le correspondant pour la VILLE est Monsieur Pierre-Yves CACHARD, Conservateur .....

## **Article 4      Réunions – rapports**

Des réunions de travail entre les responsables scientifiques auront lieu tous les six (6) mois ou à la convenance des PARTIES.

Un rapport final de synthèse sera remis à la VILLE dans le mois qui suit l'expiration de la coopération scientifique (voir Article 9 Durée)

## **Article 5      Utilisation des documents numérisés**

Pour la mise en place du PROGRAMME, des fichiers haute résolution seront confiés temporairement à l'UNIVERSITE pour permettre la réalisation des travaux prévus, mais devront être restitués à la VILLE à échéance de l'étude.

L'UNIVERSITE veillera à ce qu'aucune copie haute résolution ne soit ensuite conservée par les équipes de recherche après la fin de l'étude.

La VILLE conservera la propriété exclusive des images produites dans le cadre d'un programme d'étude associé à cette convention. Toute démarche de publication, sur quelque support que ce soit, fera l'objet d'une demande d'autorisation préalable à déposer auprès de la VILLE, et la mention légale suivante devra systématiquement figurer :

*Coll. Bibliothèque Municipale de Rouen  
Photo : [Nom du photographe]*

Le nom du photographe sera communiqué lorsque la demande aura été autorisée.

Les images hautes résolutions demandées à la VILLE par des EDITEURS et concernant un corpus étudié feront l'objet d'une facturation en prestation dans le cadre des tarifs appliqués habituellement aux éditeurs. Dans certains cas (publications scientifiques), des exonérations pourront être concédées par la VILLE.

Toute exploitation commerciale des images de la VILLE en dehors de son accord ou consentement est interdite. Tout manquement entraînera la résiliation de la convention dans les conditions décrites à l'article 10.

Chacune des PARTIES s'engage à tout mettre en œuvre pour éviter à l'autre PARTIE tout recours ou action qui pourrait être formé à un titre quelconque et par un tiers quelconque en raison de l'exploitation des informations fournies.

## **Article 6      Secret – Publications**

La VILLE s'engage à ne pas publier, ni divulguer de quelque façon que ce soit les informations scientifiques, techniques ou économiques appartenant à l'UNIVERSITE dont elle pourrait avoir connaissance à l'occasion de l'exécution de la présente convention. Cet engagement reste en vigueur pendant dix (10) ans à compter de la date de signature de la présente convention, nonobstant la résiliation ou l'arrivée à échéance de cette dernière.

Chacune des PARTIES s'engage à informer par ailleurs préalablement l'autre PARTIE de toute action de publication ou communication relative aux différents programmes d'études mis en œuvre, et à y faire systématiquement mention du partenariat.

## **Article 7      Exploitation des résultats**

Chacune des PARTIES conservera l'intégralité des droits d'exploitation pour les éléments dont elle a eu la responsabilité de réalisation, programmes informatiques compris.

Dans le cas où le développement aura été réalisé conjointement, les droits seront répartis au prorata de la participation de chacun.

Toutefois, les dispositions du présent article ne pourront faire obstacle :

- ni à l'obligation qui incombe aux enseignants-chercheurs des établissements d'enseignement supérieur d'établir un rapport d'activité périodique ou en cas d'information ayant un haut caractère de confidentialité, un rapport confidentiel au Président de la Commission de Spécialistes concernée et au Président de la Section concernée du Conseil National des Universités, dans la mesure où cette communication ne constitue pas une divulgation au sens des lois sur la propriété industrielle ;
- ni à l'obligation qui incombe aux autres fonctionnaires de l'UNIVERSITE participant au PROGRAMME de faire état de leurs travaux dans leur rapport d'activité ou en cas d'information ayant un haut caractère de confidentialité, dans un rapport confidentiel au Directeur concerné de l'UNIVERSITE ;
- ni à la soutenance de thèse des chercheurs dont l'activité scientifique est en relation avec l'objet de la présente convention, cette soutenance étant organisée chaque fois que nécessaire de façon à garantir, tout en respectant la réglementation universitaire en vigueur, la confidentialité de certains résultats.

## **Article 8      Propriété intellectuelle**

Les résultats même portant sur le domaine du PROGRAMME, mais non issus directement des travaux exécutés dans le cadre de la présente convention, appartiennent à la partie qui les a obtenus. L'autre partie ne reçoit sur les brevets et le savoir-faire correspondants, ainsi que sur les logiciels préexistants, aucun droit du fait de la présente convention.

Le sort des résultats du PROGRAMME, brevetable ou non, protégé par droit d'auteur ou non, sera déterminé dans un avenant particulier.

Par ailleurs, l'UNIVERSITE et la VILLE s'engagent :

- à ce que les noms des inventeurs ou co-inventeurs soient associés (à moins qu'ils ne s'y opposent) à la mention de l'UNIVERSITE, en accord avec les dispositions légales en vigueur dans les demandes de brevet que l'un ou l'autre déposera,
- à ce que leurs chercheurs respectifs, cités comme inventeurs, donnent toutes signatures et accomplissent toutes les formalités nécessaires pour le dépôt, le maintien et la défense desdits brevets,
- à ce que les noms des auteurs ou co-auteurs soient associés (à moins qu'ils ne s'y opposent) à la mention de l'UNIVERSITE, en accord avec les dispositions légales en vigueur, notamment dans les demandes de dépôt que l'un ou l'autre réalisera.

## **Article 9      Durée**

La présente convention est conclue à compter de sa signature et de sa notification aux parties pour une durée de 27 mois : la coopération scientifique durera 24 mois ; au cours du 25ème mois le rapport final sera remis aux PARTIES, les deux derniers mois permettront éventuellement de régler entre les PARTIES la bonne fin d'achèvement ou de prorogation de la convention. Elle pourra être prorogée à la fin de cette période par un avenant qui précisera l'objet et la durée de cette prolongation ainsi que les modalités.

Nonobstant l'échéance de la convention ou sa résiliation anticipée dans les cas prévus à l'article "RESILIATION". :

- les dispositions prévues à l'article : "SECRET - PUBLICATIONS" restent en vigueur pour les durées fixées au dit article,
- les dispositions prévues aux articles : "PROPRIETE INTELLECTUELLE" et "EXPLOITATION DES RESULTATS", restent en vigueur.

## **Article 10      Résiliation**

La présente convention peut être résiliée de plein droit par l'une des PARTIES en cas d'inexécution par l'autre d'une ou plusieurs obligations contenues dans ses diverses clauses. Cette résiliation devient effective trente (30) jours après l'envoi par la partie plaignante d'une lettre recommandée avec accusé de réception exposant les motifs de la plainte, à moins que dans ce délai la partie défaillante n'ait satisfait à ses obligations ou n'ait apporté la preuve d'un empêchement consécutif à un cas de force majeure.

## **Article 11      Litige**

En cas de difficulté sur l'interprétation ou l'exécution de la présente convention ou d'une convention particulière relevant de cette convention, les PARTIES s'efforceront de résoudre leur différend à l'amiable. En cas de désaccord persistant, les Tribunal administratif de Rouen sera seuls compétent.

Fait à Rouen, le , en 2 exemplaires

Pour l'Université de Rouen,  
le Président Jean-Luc NAHEL

Pour le Maire de ROUEN,  
Par délégation,

**FICHE D'APPROBATION**  
**à la Convention de Coopération de Recherche**  
**entre l'Université de Rouen et La BMR**

**POUR LE LABORATOIRE : PSI (UNIVERSITE DE ROUEN).**

**“ ANALYSE AUTOMATIQUE DE DOCUMENTS MANUSCRITS:**

**CONTRIBUTION À LA TRANSCRIPTION DU DOSSIER "MADAME BOVARY" ”**

*Vu,*

*Le Directeur du LABORATOIRE, Monsieur Jean-Pierre PECUCHET*

*Vu,*

*Le Responsable Scientifique du LABORATOIRE, Monsieur Thierry PAQUET*

*Vu,*

*La Directrice des bibliothèques municipales, Madame Françoise LEGENDRE*

*Vu,*

*Le Responsable Scientifique des Bibliothèques, Monsieur Pierre-Yves CACHARD*

*Vu,*

*Le Chargé de Mission à la Valorisation de la Recherche, Monsieur Michel MAYER*

**ANNEXE TECHNIQUE**

**à la Convention de Coopération de Recherche**

**entre l'Université de Rouen**

*POUR LE LABORATOIRE : PSI (UNIVERSITE DE ROUEN)*

**et la Ville de ROUEN**

*POUR LA BIBLIOTHEQUE MUNICIPALE*

***”ANALYSE AUTOMATIQUE DE DOCUMENTS MANUSCRITS:***

***CONTRIBUTION À LA TRANSCRIPTION DU DOSSIER "MADAME BOVARY" ”***

***RENSEIGNEMENTS SCIENTIFIQUES ET TECHNIQUES***

**1. Situation actuelle du programme proposé :**

Le programme de recherche proposé s'inscrit dans le cadre de la valorisation du patrimoine et vise plus particulièrement la valorisation des fonds anciens de la bibliothèque municipale de Rouen. Au plan national cette démarche fait l'objet du soutien du ministère de la culture qui a entamé un vaste programme de numérisation des collections auprès des bibliothèques. Parmi ceux-ci on peut citer notamment le programme « Gallica » de la bibliothèque nationale dont l'objectif est de proposer un accès multimédia sur le web à des ressources documentaires allant du Moyen Age au XXe siècle.

La bibliothèque municipale de Rouen dispose pour sa part d'un fonds documentaire d'une richesse de premier plan et entame un programme de numérisation avec le soutien du ministère de la culture. Elle disposera très prochainement d'un système de numérisation très performant (acquisition haute résolution et en couleur) permettant une restitution de très haute qualité, sous forme image, du patrimoine documentaire dont elle dispose. L'un des premiers objectifs de ce programme est la numérisation du dossier « Madame Bovary » composé de près de 5000 feuillets manuscrits originaux constituant la genèse de l'œuvre : états successifs des brouillons de l'auteur. A terme, l'objectif de ce programme est de rendre accessible ce fonds numérisé en proposant une édition hypertextuelle de l'ensemble du dossier, et de proposer ainsi une consultation à distance de ce dossier qui suscite, de par le monde, bon nombre de travaux de recherche dans le domaine de l'analyse littéraire et de la génétique des textes.

Pour parvenir à cet objectif, un certain nombre de difficultés restent à surmonter notamment en ce qui concerne l'indexation des documents numérisés qui seront prochainement disponibles. Ce travail colossal fait l'objet des travaux de recherche des chercheurs spécialistes en génétique des textes anciens et est effectué le plus souvent en ayant recours à des moyens rudimentaires (manuels) pour ce qui concerne les techniques d'annotation.

La mise à disposition d'un corpus numérisé permet d'envisager d'autres modes de travail assisté par ordinateur. L'apport du numérique offre à ce niveau un potentiel qui est encore aujourd'hui peu exploré mais qui doit naturellement bénéficier des techniques de traitement d'images développées dans le domaine de l'analyse et de la reconnaissance de documents ainsi que des nouveaux standards de représentation et de manipulation des documents électroniques proposés par le W3C (XML, XSLT,...).

Le laboratoire Perception Système Information possède une solide expérience dans le domaine de l'analyse des documents. Il développe depuis une quinzaine d'années des algorithmes de traitement automatique d'images de documents et de reconnaissance de l'écriture. Son expertise dans le domaine permettra de couvrir le volet scientifique du programme de numérisation de la Bibliothèque Municipale de Rouen en participant en tant que de besoin à la définition des programmes de recherche qui pourront accompagner le programme de numérisation déjà engagé. Il faut souligner que le corpus qui sera prochainement disponible de façon numérique (unique en son genre) ouvre de nouveaux champs d'investigation à l'Analyse Automatique de Documents qu'il est opportun d'initier dès à présent de par le potentiel que l'équipe du PSI est susceptible d'apporter au projet d'une part, et d'autre part par l'aspect tout à fait original de la recherche que peut susciter le corpus numérisé. La présente convention a donc pour objectif d'initier la collaboration entre le PSI et la BMR.

## 2. Plan de recherche :

### - Objectif général des travaux

Dans le cadre de la présente convention, le Laboratoire PSI apportera son expertise dans le domaine de la numérisation des documents, de l'analyse automatique des documents numérisés (analyse de la mise en page, indexation par le contenu, reconnaissance de l'écriture), pour assister la BMR dans les programmes de recherche et/ou de déploiement qu'elle jugera nécessaire de développer autour du programme de numérisation déjà engagé. Le Laboratoire PSI évaluera pour sa part les possibilités d'exploitation du corpus numérisé en l'état actuel des techniques d'analyse dont il a la maîtrise et des moyens humains dont il dispose. Il sera à même de proposer un programme de recherche qui pourrait être jugé nécessaire pour atteindre les objectifs d'indexation que pourrait se fixer la BMR dans le cadre du programme de numérisation qui débute.

Ce premier travail d'expertise aura donc pour principal objectif de formaliser les « points durs » du programme d'indexation du dossier « Madame Bovary », d'accompagner la BMR dans son programme de numérisation, et de proposer un programme de recherche argumenté pouvant notamment fédérer des travaux de recherche pluridisciplinaires impliquant les Sciences Humaines et les Sciences et Technologies de l'Information et de la Communication sur le thème du document manuscrit numérique.

### - Programme des travaux

En fonction des objectifs généraux du programme il ne peut être actuellement défini de programme scientifique contractuel précis.

En revanche ce programme cadre permettra, par la mise à disposition par la BMR des premiers corpus numérisés, d'évaluer les difficultés posées par l'indexation des documents manuscrits.

A ce stade du programme, plusieurs réunions de travail devront permettre de préciser les objectifs de la BMR tant en ce qui concerne les outils d'indexation que les formats de représentation des documents indexés compatibles avec les modes de consultation et de navigation souhaités.

### - Conséquences attendues (au plan scientifique, technique, ou socio-économique)

Les techniques développées pour l'analyse de l'écriture l'ont été jusqu'à présent dans le cadre d'applications limitées. Le présent programme permettra d'étudier la problématique de l'indexation des documents manuscrits ce qui constitue en l'état actuel de l'état de l'art du domaine une étape encore loin d'être résolue. Le programme permettra de proposer des techniques de pré-traitement spécifiques pour les documents en couleurs, de faire progresser les techniques d'analyse de la mise en page pour guider la détection des composants de l'écriture (lignes, mots, annotations, ratures, signes distinctifs du scripteur,...).

En ce qui concerne la reconnaissance de l'écriture, le programme permettra de tester à grande échelle une technique d'adaptation d'un système de reconnaissance à une écriture spécifique en recourant à des algorithmes d'apprentissage. On étudiera également à ce niveau l'apport de techniques d'apprentissage semi-supervisé permettant de limiter le volume des données étiquetées pour l'apprentissage.

Sur le plan technique, la nature même des documents rendra nécessaire d'utiliser des formats d'indexation composites intégrant texte ascii et représentation sous forme image afin de restituer les représentations graphiques non textuelles telles que les ratures, les lignes de sur-lignage, les blocs d'écriture posant des difficultés à la transcription automatique. De même la restitution des documents originaux ou des parties originales des documents nécessaires à l'expertise littéraire nécessiteront des formats composites.

Plus généralement le programme envisagé permettra de faire évoluer la problématique de l'indexation des documents anciens, de proposer de nouveaux outils d'indexation. La méthodologie proposée et les outils qui pourraient être développés à cette occasion auront naturellement des applications pour l'analyse littéraire et la génétique des textes anciens. Cette première expérience Rouennaise et Flaubertienne sera naturellement applicable à d'autres corpus en voie de numérisation également. On pense notamment aux fonds Stendhal (Grenoble) ou Montesquieu (Lyon)