

---

## DMP du projet "RessOURces tExtuelles Numérisées"

*Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).*

### Renseignements sur le plan

<b>Titre du plan</b>	DMP du projet "RessOURces tExtuelles Numérisées"
<b>Livable</b>	Livable D9.5
<b>Version</b>	Version finale
<b>Objet/périmètre du plan</b>	Analyse d'un corpus de fanfictions sur la mythologie gréco-latine.
<b>Langue</b>	fra
<b>Date de création</b>	2021-10-26
<b>Date de dernière modification</b>	2021-10-26
<b>Identifiant</b>	MythFan
<b>Type d'identifiant</b>	URL
<b>Licence</b>	Creative Commons Attribution Non Commercial Share Alike 4.0 International
<b>Documents (publications, rapports, brevets....), sites web associés</b>	<ul style="list-style-type: none"><li>• Wattpad</li><li>• Fanfiction</li><li>• Archive of Our Own</li><li>• Webtoons</li></ul>

### Renseignements sur le projet

**Titre du projet** RessOURces tExtuelles Numérisées

**Acronyme** ROUEN

**Résumé** Analyse des procédés didactiques d'écriture numérique d'un corpus de fanfictions afin d'envisager la création d'un cours en Education au Média et à l'Information.

**Sources de financement**

- Agence Nationale de la Recherche : ROUEN-100122

**Date de début** 2021-01-10

**Date de fin** 2022-01-10

**Partenaires**

- University of Rouen ()

<b>Référence du plan expérimental</b>	<b>Titre</b>	Fanfictions
	<b>Identifiant</b>	Fan-00
	<b>Type d'identifiant</b>	URL

**Produits de recherche :**

1. Fanfictions (Texte)

**Contributeurs**

Contributeurs	Rôles
Mylène LANGLOIS	<ul style="list-style-type: none"> <li>• Coordinateur du projet</li> <li>• Personne contact pour les données</li> <li>• Responsable du plan de gestion de données</li> </ul>

# DMP du projet "RessOURces tExtuelles Numérisées"

---

## 1. Description des données et collecte ou réutilisation de données existantes

### 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les nouvelles données textuelles seront extraites des sites Wattpad.com, Fanfictions.net, Archiveofourown.org et Webtoons.com et recueillies dans Open Office Writer. Les titres des chapitres et les noms propres seront ensuite extraits dans un fichier Open Office Calc, afin de réaliser des statistiques sur la récurrence des noms des personnages évoqués. Puisque les fanfictions sont protégées par le droit d'auteur, il s'agira de ne faire que des citations courtes et d'analyser les procédés d'écriture, la démarche et la réflexion de chaque auteur.

---

### 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données textuelles collectées seront stockées dans un format .odt et les données produites dans un format .ods, car ce sont des formats ouverts disponibles gratuitement et compatibles avec Windows, Linux, Appel Mac OS, etc. Le but est de faciliter l'accès aux données et de permettre la modification des données produites, dans le cadre de l'ouverture des données de la recherche universitaire. Le corpus étant réduit à un vingtaine de fanfictions, les fichiers Texte ne seront pas très volumineux et ne prendront pas beaucoup de place dans l'espace de stockage.

---

## 2. Documentation et qualité des données

### 2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les métadonnées qui accompagneront les fanfictions seront : le nombre de vues, de "j'aime" et de commentaires de chaque chapitres et de l'œuvre dans son ensemble, ainsi que la date à laquelle les chapitres ont été publiés et les mots-clés qui permettent de rechercher l'histoire dans le moteur de recherche de la plateforme en ligne. Elles seront collectées dans une base de données en SQL grâce à un programme codé en Python. Ces informations seront obtenues et enregistrées dans la base de données avec des liens vers des fichiers "Read me" codés en Markdown.

---

### 2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Etant donné que le corpus d'une vingtaine de fanfictions sera méticuleusement composé par la chercheuse dans le cadre de son mémoire de recherches, en fonction des thèmes mythologiques abordés, du nombre de chapitres, de l'avis des lecteurs, etc., le contrôle de la qualité et de la conformité de la collecte des données sera dès lors assuré.

---

### 3. Stockage et sauvegarde pendant le processus de recherche

#### 3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données seront stockées sur l'ordinateur personnel de la chercheuse, sur son espace de stockage universitaire en ligne et sur un espace de stockage externe (clé USB), pour assurer des conditions pérennes de stockage. La sauvegarde sera effectuée à chaque fois que de nouvelles données seront collectées ou produites, manuellement ou automatique selon les systèmes de stockage.

---

#### 3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Puisque les données sont sauvegardées grâce à trois systèmes de stockage différents, en cas d'incident, il suffira à la chercheuse de récupérer les données sur un des deux systèmes restants. L'ordinateur et l'espace de stockage sont protégés par un mot de passe et l'ordinateur possède un anti-virus performant. En ce qui concerne la protection des données, les données collectées sont disponibles par tous sur les sites de fanfictions en question. Il ne s'agit pas de données sensibles, mais de textes soumis au droit d'auteur.

---

### 4. Exigences légales et éthiques, codes de conduite

#### 4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Il sera demandé à chaque auteur le droit de traiter leurs données dans le cadre d'un travail de recherche universitaire. Les données ne seront pas publiées dans le projet de recherche, mais renverront aux liens des sites concernés. Seules quelques citations courtes seront utilisées pour analyser la démarche de l'auteur.

---

#### 4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Comme les données textuelles concernent des fanfictions, elles sont soumises au droit d'auteur. Ainsi, théoriquement, dans le cadre d'un travail universitaire, il faut nécessairement faire référence à l'auteur et faire des citations courtes du texte pour procéder à leur analyse. Il n'est pas interdit selon la loi de diffuser le corpus si les auteurs sont mentionnés, puisque les textes ne seront pas modifiés.

#### **4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?**

Pour une question d'éthique, la chercheuse demandera tout de même l'accord des auteurs pour publier, sur une autre plateforme que celles qu'ils utilisent, les textes du corpus dans le cadre du projet ROUEN. En effet, cela pourrait éviter qu'ils l'attaquent pour atteinte à leur réputation, bien que le projet n'est pas mené pour décrédibiliser de travail de ces auteurs-amateurs. Ainsi, s'ils l'y autorisent, les données pourront être accessibles aux membres de la recherche universitaire, voire au grand public.

---

### **5. Partage des données et conservation à long terme**

#### **5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?**

Les données pourront être retrouvées et partagées sur l'archive ouverte HAL, qui permet la diffusion d'articles scientifiques. Les fichiers textes sont envoyés au CINES pour garantir leur accessibilité à long terme. Les délais de publication dépendent des délais de modération de la plateforme, qui accepte ou non la publication des articles.

---

#### **5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?**

Les données ne seront pas divulguées sans l'accord de leur auteur. Elles seront utilisées sous forme de statistiques ou de citation pour analyser la démarche de l'auteur-amateur dans le cadre de la mobilisation de ses connaissances. Les données seront déposées dans Nakala mis en place par la TGIR Huma-Num, qui aide à valoriser la recherche scientifique.

---

#### **5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?**

Les utilisateurs potentiels auront besoin d'avoir accès à Open Office qui est gratuit et largement compatible avec les systèmes d'exploitation. Ils pourront aussi télécharger MySQL Workbench pour avoir accès aux bases de données. Les données seront partagées sur Nakala, ce qui offre une accessibilité et une citabilité des données sur le long terme.

---

#### **5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?**

Les données pourraient être réutilisées dans le cadre d'un autre projet de recherche plus large sur les fanfictions. Nakala attribuera donc un DOI pour chaque document, puisqu'il s'agit d'un entrepôt pérenne de confiance.

---

## 6. Responsabilités et ressources en matière de gestion des données

**6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?**

La chercheuse sera responsable de la gestion des données, épaulée par les membres de l'IRHIS de l'université de Rouen, pour réaliser son premier projet en sciences humaines.

---

**6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?**

La gestion des données dans le cadre du projet ROUEN se fera sur un an. Peu de frais de stockage sont à prévoir, à part l'achat d'une clé USB et le coût de réparation d'un ordinateur portable si un incident advient. Il faut également prévoir le salaire de la chercheuse, que l'on peut estimer à environ 700 euros par mois. Aucun frais d'entrepôt et d'archivage n'est engagé, car Nakala est un service gratuit.