

EMAN

*construction d'une plateforme
d'édition numérique de
manuscrits et de fonds d'archives
modernes*

Richard Walter

Séminaire Hubert de Phalese, février 2019

Nous partîmes cinq cents ; mais par un prompt renfort,

Nous nous vîmes trois mille en arrivant au port,

Tant, à nous voir marcher avec un tel visage,

Les plus épouvantés reprenaient de courage !

La plate forme
EMIAN

Un projet

Tout projet a une histoire (évolutive)

Tout projet a un contexte (laboratoire, besoins)

2014 : 2 projets

2015 : expérimentation / ouverture

2016 : + de projets / + de fonctions

2017 : 18 projets + CoPil

2018 : des outils et des données

2019 : 28 projets

De multiples évolutions : scientifiques, techniques, équipe...

Unique stabilité : ... du contenu, un ingénieur

Un projet



Plateforme d'édition de manuscrits modernes
francophones

Cette plateforme a pour vocation de mettre à disposition de la communauté des chercheurs mais aussi d'un public plus large les documents d'archives relatifs aux auteurs majeurs de la francophonie littéraire.

Mouloud Feraoun



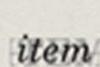
Mouloud Feraoun est un écrivain algérien, né le 8 mars 1913 à Tizi Hibel en Haute Kabylie et assassiné à Alger le 15 mars 1962.

Jean-Joseph Rabearivelo



Jean-Joseph Rabearivelo (1903-1937) est l'un des principaux fondateurs de la littérature contemporaine malgache.

[Soutiens du projet](#) [Contact](#) [Mentions légales](#) [OAI-PMH](#) [Connexion](#)



Un projet

Premier âge :

EdiManusLitt

Version travail 2.0, 21-02-2014

Plate-forme de gestion et d'édition de textes et manuscrits hispaniques et africains

*Titre ??? On garde texte plutôt que manuscrit comme le nom de l'ITEM
Des textes et manuscrits modernes = trop ITEM en général
Faire un sous-titre avec acronyme avec littérature africaine et hispanique*

Les corpus de manuscrits et/ou de textes sont des objets numériques spécifiques où il n'y a pas un texte unique à traiter mais différents états de celui-ci avec donc différents niveaux de transcription à éditer et à aligner : image du texte, transcription diplomatique et traduction, encodage différents états génétiques (ratures, ajouts, suppressions, etc.), sans parler des annotations et des commentaires des éditeurs.

Un constat émerge :

- Les projets gèrent des contenus textuels de plus en plus exponentiels.
- Les équipes doivent gérer des corpus multiples pour des finalités multiples.
- On répète souvent les mêmes procédures.
- Les résultats se retrouvent éparpillés sur des sites, des disques et des formats différents.
- Il faut de plus en plus relier les documents, comparer image et texte transcrit.
- On veut intégrer dans la démarche projet l'exploitation numérique du contenu.

Pour répondre à ces besoins, différentes étapes sont nécessaires :

- Encodage et signalement des données, en respectant normes et standards du domaine ;
- Publication des données recueillies ;
- Stockage des des fichiers sources et des publications ;
- Archivage avec inventaire.

*Pourquoi que 2 équipes ? Les 2 équipes ont commencé séparément à encoder XML/TEI/EAD des textes et dans une démarche de publication de corpus.
On monte une plate-forme avec ces 2 types de littérature, si cela marche, on peut étendre*

Un projet

Premier âge :

⇒ Convergence des besoins :

« Créer un outil de publication interopérable pour :

- Éditer les manuscrits numérisés avec des métadonnées qui respectent les normes et standards internationaux
- Associer au sein de la plateforme les documents entre eux grâce à des liens logiques, temporels et génétiques
- Associer le manuscrit numérisé à sa transcription enrichie et à d'autres manuscrits pour constituer un dossier génétique. »

Un projet

Actuellement :

« EMAN est une plate-forme de publication numérique pour la diffusion et l'exploitation de manuscrits et de fonds d'archives modernes. Il est basé sur le logiciel Omeka et a développé des extensions pour l'édition numérique savante. EMAN héberge plus d'une vingtaine de projets scientifiques d'édition de corpus, aussi variés que l'édition numérique de thésors de la Renaissance, des corpus de presse ou de cours de l'ENS, comme des sources génétiques de l'opéra ou de la bande dessinée jusqu'à l'édition génétique d'un conte de Paul Valéry ou de l'autobiographie de Vittorio Alfieri, ou la correspondance de Gaspard Monge ou d'Émile Zola. La plupart sont encore en cours de production.

Le projet vise à créer un outil de publication interopérable favorisant l'exploitation numérique de documents. »

⇒ Un « cher disparu »...

Un projet

Pourquoi cela marche ?

Ou pas...

Savoir choisir l'outil : une plate-forme

Une collection de projets ?

Savoir faire / défaire une gestion de projet

Une roue de secours : les chantiers...

Les chaînes pour rouler en tout terrain : les ateliers...

Une plateforme

Cahier des charges « simple » :

La publication en ligne d'images, de textes et de diverses sources numériques dans un dispositif éditorial répondant aux exigences scientifiques d'une édition critique et génétique propre à exploiter les contenus proposés pour produire une analyse des processus de création.

En « clair » :

- métadonnées avec normes et standards internationaux (W3C)
- associer les documents avec des liens multiples
- associer le document à d'autres ressources dans un ensemble
- associer le document numérisé à sa transcription enrichie
- interopérabilité des données pour diffusion, archivage et « moissonnage » orienté web sémantique.

À moyen constant...

Une plateforme

Plateforme générique et modulaire

Choix d'Omeka

Des plugins et des templates

Choix du Dublin Core

En 15 champs, décrire une donnée...

... Exercice violent et salutaire ...

⇒ « Y plus qu'à »

Construction d'un outil d'édition

On tord l'outil de plus en plus, jusqu'à quel point ?



Une plateforme

Construction d'un outil d'édition

Affirmation de base :

L'édition numérique ne doit être considérée comme la simple reproduction numérique d'un document papier. Il y a une réflexion intellectuelle et éditoriale qui aboutit à l'acte de numérisation ou d'éditer numériquement.

- ⇒ « Numérisée » n'est pas « numérique »
- ⇒ Une démarche éditoriale savante et numérique

Une plateforme

Une démarche éditoriale savante et numérique

Parti-pris de base :

Une édition numérique sert surtout à retracer la vie d'un document, de sa genèse jusqu'à sa diffusion. Il faut donc en publier, à côté de sa reproduction et relié à elle, son dossier documentaire/génétique.

- ⇒ Publication des sources et des variantes
- ⇒ Changements quantitatifs et qualitatifs

- ⇒ Cumuler dans un même espace virtuel l'édition des versions successives du document, mais aussi présenter les « environs »

Une collection de projets ?

Multiplicité de projets

Variabilité des projets (temps, volumétrie, granularité)

⇒ de la « microédition » sur un corpus très réduit
à de la publication en masse de documents très variés

Des projets ouverts / des projets en cours de production

⇒ Ce n'est pas une collection, c'est une plate-forme

Une collection de projets

Au 09 octobre 2017, sur la plate-forme EMAN et ses dix-huit sites projets

7.439 notices

581 collections

93.084 fichiers

123 comptes utilisateurs

soit

437,58 notices en moyenne par projet

34,17 collections en moyenne par projet

5475,52 fichiers en moyenne par projet

8,07 comptes utilisateurs par projet

12,80 notices par collection en moyenne

12,51 fichiers par notices en moyenne

Une collection de projets

Au 1er février 2019, sur la plate-forme EMAN et ses vingt-six sites projets

18.448 notices

1.493 collections

140.030 fichiers

206 comptes utilisateurs

soit

683,25 notices en moyenne par projet

55,29 collections en moyenne par projet

5186,29 fichiers en moyenne par projet

7,62 comptes utilisateurs par projet

12,35 notices par collection en moyenne

7,59 fichiers par notices en moyenne

15 sites sur 26 ont augmenté ou diminué leur corpus

Gestion de projet

Nouvelles démarches éditoriales et scientifiques

⇒ Démarche éditoriale souple et itérative

Dualité du chercheur dans ce type de projet numérique :

A/ éditer un objet scientifique

B/ interpréter des données

⇒ Abondance de possibilités

⇒ Quand on publie ?

⇒ Ou on continue de travailler sans cesse ?

En plus tous les problèmes se posent en même temps

Encodage, transcription, relations, traduction, etc.

Gestion de projet

Choix radical :

On édite le document avec une base minimale de codage

On complexifie au fur et à mesure

Tous les outils ne sont pas prêts, toutes les fonctions ne sont pas opérationnelles mais la plate-forme existe, publie, diffuse et travaille...

On travaille par étape. On produit par strate

Gestion de projet

Modélisation & granularité

Couche supplémentaire permise par Omeka :

Métadonnées personnalisables

- ⇒ Personnalisables en fonction du projet ou des documents
- ⇒ Déclinables à l'infini
- ⇒ Regroupées par type
- ⇒ Créées et gérées ***par*** le chercheur...

Changement de paradigme :

On ne part pas de l'objet fini vers son indexation simplificatrice

Mais de sa description documentaire à sa présentation scientifique

Gestion de projet

Modélisation & granularité

Pour chaque projet :

Imaginer un modèle de traitement et de visualisation de tous les types de données pouvant être intégrées au projet

Des questions de base :

Quelle unité de base ?

La notice porte sur quoi ?

Les notices sont dans une collection par quel critère ?

Les liens entre les notices ?

Etc. Etc. Etc.

⇒ Réunionite aigüe

⇒ « En revenir sans cesse aux données »

Gestion de projet

Dispositif éditorial et outillage

On a un thème de base

On a un modèle éditorial des données

On a des outils spécifiques

⇒ On expérimente / On confronte / On modélise

⇒ On partage, on copie, on améliore

On met à disposition de la communauté

⇒ Que des verbes d'action...

Gestion de projet

Dispositif éditorial et outillage

Mutualisation des coûts et des risques :

⇒ Multiplication des demandes et des rustines...

Imaginer un modèle de gestion de la plate-forme avec une mise à disposition de tous les développements pour tous les projets en respectant leurs propres contraintes...

Gestion de projet

Accompagnement de projets

Nécessité de se structurer

Comment répondre à 27 projets en même temps

⇒ Organisation / Structuration

⇒ On fait un comité

... Un Comité de pilotage

... Des journées d'études

... Des ateliers communs

... *Une communauté d'intérêt... (nouvelle réalité?)*

Interopérabilité / Mutualisation...

Mixité / stabilité ...

Gestion de projet

Accompagnement de projets

Le « coût prohibitif » de la plate-forme

Nécessité de nombreux outils d'accompagnement

⇒ Pour expliquer

⇒ Pour fédérer

- Guide de recommandations

- Guide de saisie

- Espace collaboratif interne (forum, actualités, gestionnaire de docs)

Bulletin d'information hebdomadaire

- Carnet de recherche :

- Astuces d'EMAN :

Chantiers

Transcription (module Transcript)

Objectif : *Pouvoir encoder le maximum de phénomènes avec un apprentissage technique minimum et en prévoyant l'image du document à côté*

Visualisation (module Graphe)

Objectif : *Avoir une représentation graphique des multiples relations possibles d'un document ou au sein d'une collection*

Le dur apprentissage du « *sujet, prédicat, objet* »

D'autres besoins : référentiels, exports, multilingue, etc.

À suivre...

Conclusions ?

Un outil générique : Omeka

Une plate-forme de projets : EMAN

Des fonctions et des outils propres

⇒ Un modèle d'édition génétique numérique

Changement d'échelle (d'époque ?)

On peut publier sans limite tous les états d'un même document, avec leurs transcriptions, accompagné de commentaires et d'analyses et avec le dossier de genèse qui éclaire le processus de création

⇒ Une réelle édition multi-usages

L'âge de la retraite est à 67 ans...

12 vols Dactylographis à partir des notes des cahiers.

est 1 recopie
 Notes éparpillés à l'intérieur des cahiers

		Les deux feuillets correspondent à une même et unique frappe dactylographiée				Les deux séries correspondent - au f. 38 près - à une même et unique frappe dactylographiée		
f. 43 (feuilles séminales)	f. 26 (dactylographie avec corrections et ajouts ms)	f. 27 (dactylographie avec corrections et ajouts ms)	f. 29-31 (série de dactylographies avec corrections et ajouts ms)	f. 32-34 (série de dactylographies avec corrections et ajouts ms)	Page tête-bêche du cahier « Robinson » (f. 24 page de gauche dans la foliotation de la BnF)	f. 35-38 (série de dactylographies avec corrections et ajouts ms)	Cahier « Robinson » (à l'endroit)	« Robinson » dans l'édition posthume des <i>Histoires brisées</i> (1950)
		f. 28 (dactylographie avec corrections et ajouts ms)				f. 39-41 (série de dactylographies avec corrections et ajouts ms)		

15/23
 média

1
 a 1 influence
 est en écho
 rétroactive
 [pointillés]
 unilatérales

3
 est 1 état
 obtenu par
 résultat de
 (idem posthume)

5
 est 1 état
 obtenu
 par
 l'édition
 posthume
 de

Scan
 unique
 à en-01K

4 643 a 1 influence sur

Le carburant...

« Chaque chose faite n'est plus à faire »

« Tout document = une notice »

« Faire et défaire est toujours travailler »

« In EMAN we trust »

« Ad maiorem EMAN gloriam »