

Edition numerique d'un document archival : un cas d'étude

Lou Burnard Consulting

2016-05-10

Dans cet exercice, nous allons :

- discuter la modélisation d'un document d'archive
- expérimenter son balisage en XML-TEI

Notre cas d'étude concerne quelques documents extraits des archives de l'OULIPO.

Pour ceux qui ne le connaissent pas, l'OULIPO ou Ouvroir de littérature potentielle est un groupe d'écrivains un peu particulier qui (depuis les années 1960) se réunissent tous les jeudis discuter de leur méthodes et de leurs productions. Ils ont ainsi créé une archive importante, notamment composée des comptes rendus de leur réunions, actuellement disponible à la BnF, et qui fait l'objet d'un projet de numérisation financé par l'ANR.

Voici une liste des fichiers sur lesquels nous allons travailler:

cr610417	3	xml
cr610428	6	xml
cr611219	5	txt
cr620911	3	xml
cr650308	5	txt

1 Quels sont les objets d'intérêt à encoder ?

Vous êtes d'abord invités à considérer les documents présentés en mode image, par ex cr650308_1.jpg. Ce fichier représente la première page du compte rendu (cr) de la réunion du 8 mars 1965 (650308). Quels sont, d'après vous, les objets d'intérêt que vous pourriez souhaiter encoder dans ce document ? Comment est-il structuré ? Faites une liste des concepts qui vous intéressent.

2 Sélection des éléments XML-TEI

Ensuite, décidez comment représenter ces concepts en utilisant des balises TEI. Par exemple, si vous considérez que les remarques des participants sont présentées comme si il s'agissait d'une représentation dramatique, voire théâtrale, vous aurez alors probablement besoin des balises `<sp>`, `<speaker>`, `<stage>`. Si par contre, vous considérez que cette présentation de forme dramatique est moins intéressante ou moins cohérente qu'une présentation en série de paragraphes, dont quelques-uns portent un label, vous n'aurez besoin que des balises `<label>` et `<p>`. Si cela vous intéresse, peut-être faudra-t-il faire quelque chose avec les noms de personnes, de lieux, des œuvres qui pourront être balisés avec `<name>`, `<title>`, etc. Et *quid* des corrections, des ajouts etc. ? En tout les cas, vous aurez besoin des éléments `<div>`, `<pb>`, `<hi>`, etc.

3 Documentation par un Schéma ODD avec Roma

Documenter ensuite votre sélection, en utilisant Roma comme nous l'avons fait précédemment.

4 Production de la transcription sous le contrôle d'un schéma

Vous devez maintenant réaliser votre transcription sous le contrôle d'un Schéma. Pour ce faire, générez un schema RelaxNG à partir de votre document ODD ou de Roma, et associez le à un nouveau document XML.

4 PRODUCTION DE LA TRANSCRIPTION SOUS LE CONTRÔLE D'UN SCHEMA

Pour faciliter le travail de transcription vous trouverez dans le fichier `cr650308.txt` une première transcription en texte brut.