

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

Lou Burnard Consulting

2016-05-08

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 sur le site Gallica de la BnF, et qui fait un des objets d'un projet de recherche à l'université de Paris III, financé par l'ANR (voir Site Difdepo).

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 ? Faîtes une liste des concepts qui vous intéressent.

Ensuite, décider 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'un représentation dramatique, voire théâtrale, vous aurez alors probablement besoin des balises `<sp>`, `<speaker>`, `<stage>`. Si par contre, vous considérez que cette présentation en forme de drame 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, ajouts etc. ? En tout les cas, vous aurez besoin des `<div>`, `<pb>`, `<hi>`, etc.

Documenter ensuite votre sélection, en utilisant Roma comme nous l'avons fait précédemment. Enfin, réalisez votre transcription sous le contrôle du schéma correspondant. Pour faciliter le travail de transcription vous trouverez dans le fichier `cr650308.txt` une première transcription en texte brut.