

FORMATION TEI Niveau avancé

- **Quoi** : une formation approfondie pour les consortiums Cahier, Corpus Ecrits et IRCOM de la TGIR Corpus
- **Quand** : 19 - 22 novembre 2012
- **Où** : Fédérations de Recherche en linguistique (ILF et TUL), 44 rue de l'Amiral Mouchez, 75014 PARIS

Responsables : Lou Burnard (TEI), assisté de Lauranne Bertrand (BVH, CESR), et par Alexandre Gefen (Sorbonne). La participation d'autres experts français est aussi envisagée selon leur disponibilité.

Cette formation s'articulera autour de 3 thèmes:

- La modélisation des ressources et la sélection des traits signifiants
- L'encodage et l'explicitation TEI des structures modélisées
- L'exploitation et l'analyse des ressources structurées

Chaque journée sera découpée en 4 sessions que feront alterner des présentations et des sessions de travaux pratiques.

Prérequis :

Connaissances basiques de l'encodage XML et des essentiels de la TEI, par exemple issue de la formation Patrimoine Numérique du CESR.

En l'absence d'une salle pre-equipée d'ordinateurs, les participants seront invités à travailler sur leurs portables personnels (PC, Mac, ou Linux). Les logiciels à installer leur seront communiqués avant la formation.

Public :

Les participants doivent être porteurs d'un projet en cours et seront invités à présenter des échantillons des sources sur lesquelles ils travaillent et à discuter leur encodage éventuel.

Programme :

Lundi 19 novembre (9h30 - 17h30)

- Introduction ; présentation des participants et de leurs intérêts ; objectifs de la formation
- Session de rappels pratiques: éditer avec oXygen, quelques projets TEI actuels
- Présentation du paysage TEI
- Construction de schéma pour ses propres textes en se servant de Roma

Mardi 20 novembre (9h30 - 17h30)

- Présentation du langage ODD
- Création d'un ODD à partir d'un corpus
- Présentation du TEI Header et considérations sur les métadonnées
- Création d'un header ; export de métadonnées pour Isidore

Mercredi 21 novembre (9h30 - 17h30)

- Encodage TEI des corpus oraux avec analyses linguistiques
- Exercice sur l'encodage d'un texte oral
- Encodage des entités nommées avec analyses littéraires
- Exercice "data mining" dans un corpus de documents historiques

Jeudi 22 novembre (9h30 - 17h30)

- Encodage TEI des sources manuscrites
- Exercice sur l'encodage d'un manuscrit
- Stockage, affichage et analyse des corpus TEI: discussion sur les outils
- Affichage des textes TEI avec des logiciels tels que teichi ou teiboilerplate ; analyse des textes TEI avec des logiciels tels que txm

Nombre de places: 20 participants, tous porteurs d'un projet en cours.

- 10 participants relevant d'un laboratoire membre du consortium CAHIER : frais de transport et de séjour pris en charge par le consortium
- 5 participants relevant d'un laboratoire membre du consortium Corpus Ecrits : frais de transport et de séjour pris en charge par le consortium
- 5 participants relevant d'un laboratoire membre du consortium IRCOM : frais de transport et de séjour pris en charge par le consortium