



École  
nationale  
des  
chartes

# **Master Humanités numériques et computationnelles**

**Université PSL**

Année universitaire 2023 - 2024

**Syllabus**





## Maquette de la première année

Intitulé de l'enseignement	Langue	Volume horaire	ECTS
<b>S1 - Enseignements obligatoires</b>		<b>152</b>	<b>23</b>
Semaine intensive de présentation du master	Fr/En	20	
Initiation aux humanités numériques	Fr	20	3
Introduction aux mathématiques	Fr	20	3
Fondamentaux de l'informatique			
<i>Systèmes d'exploitation, algorithmique, Linux</i>	Fr/En	12	3
<i>Expressions régulières</i>	Fr	4	
<i>Bonnes pratiques de développement, Git</i>	Fr	6	
Méthodes quantitatives en SHS	Fr	10	2
Algorithmique et programmation pour les humanités : Introduction à R	Fr	15	3
Algorithmique et programmation pour les humanités : introduction à Python	Fr	20	4
Méthodologie et mémoire			
<i>Méthodologie de la recherche en SHS computationnelles (ateliers)</i>	Fr	8	4
<i>Outils pour la rédaction du mémoire</i>	Fr	12	
Séminaire de recherche du master	Fr/En	5	1
<b>S2 – Enseignements obligatoires</b>		<b>65</b>	<b>17</b>
Modélisation et structuration des données			
<i>Structuration XML et XML/TEI</i>	Fr/En	20	6
<i>Traitement et Analyse de données relationnelles</i>	Fr	10	
Algorithmique et programmation pour les humanités : Python avancé	Fr	20	4
Méthodologie et mémoire			
<i>Méthodologie de la recherche en SHS computationnelles (ateliers)</i>	Fr	10	6
Séminaire de recherche du master	Fr/En	5	1
<b>S1 et 2 - Enseignements optionnels</b>		<b>env. 100h</b>	<b>20</b>
<p>·Parcours pédagogique individualisé, composé de cours et séminaires de recherche en sciences humaines et sociales</p> <p>Séminaires de SHS des 4 établissements – en règle générale, 1 ECTS par 5h de cours (comportant validation finale et travail personnel)</p> <p style="text-align: center;">Cours de méthodologie de votre discipline (très recommandé)</p> <p>Cours de recherche bibliographique – 2 ECTS (obligatoire, évalué par la bibliographie du mini-mémoire)</p> <p style="text-align: center;">PSL week(s) – 2 ECTS par PSL week</p> <p style="text-align: center;">UE Sport optionnelle (1 ECTS)</p>			
<b>TOTAL</b>		<b>317</b>	<b>60</b>

## Maquette de la deuxième année

Intitulé de l'enseignement	Langue	Volume horaire	ECTS
<b>S3 - Enseignements obligatoires</b>			
Fondamentaux de mathématiques et Python pour DataScience	Fr	10	2
Approches quantitatives et modélisation mathématique en SHS	Fr	20	4
Langue vivante (au choix)	-	10	1
Séminaire de recherche du master	Fr/En	5	1
Stage de découverte en laboratoire	Fr/En	4 à 8 sem.	3
<b>S3 - Enseignements optionnels</b>			
<b>Spécialisation de SHS computationnelles</b>			
<i>soit</i> Philologie computationnelle	Fr	25	5
<i>soit</i> Traitement et analyse de l'information spatialisée	Fr	25	5
Séminaire du directeur de recherche	Fr	26	5
Enseignement d'ouverture (à choisir: séminaire, cours, PSL week, ...)	Fr	10	2
<b>S3 - Mémoire et bilan d'étape</b>			
<b>S4 - Enseignements obligatoires</b>			
<b>Apprentissage machine</b>			
Algorithmique et programmation pour les humanités : Python avancé	Fr	10	2
Python avancé	Fr	20	4
Séminaire de recherche du master	Fr/En	5	1
<b>S4 - Enseignements optionnels</b>			
<b>Spécialisation de SHS computationnelles</b>			
<i>soit</i> Traitement automatique de la langue et analyse sémantique	Fr	25	5
<i>soit</i> Traitement automatique de l'image	Fr	25	5
<i>soit</i> Analyse de réseaux	Fr	25	5
Séminaire du directeur de recherche	Fr	26	5
<b>S4 – Mémoire et soutenance</b>			
<b>TOTAL</b>			
		<b>192</b>	<b>60</b>

# Sommaire

<b>Master 1.....</b>	<b>8</b>
<b>Semestre 1 : tronc commun – 23 ECTS.....</b>	<b>9</b>
Semaine intensive de rentrée.....	9
Initiation aux humanités numériques (EPHE 100).....	9
Introduction aux mathématiques (ENC 2313).....	10
Fondamentaux de l'informatique (ENC 2435).....	10
<i>Contenu 1 : Séminaire intensif / install Party.....</i>	<i>10</i>
<i>Contenu 2 : Linux et Shell.....</i>	<i>10</i>
<i>Contenu 3 : Expressions régulières.....</i>	<i>10</i>
<i>Contenu 4 : Git et bonnes pratiques de dev.....</i>	<i>10</i>
Méthodes quantitatives en SHS (ENC 2310).....	11
Algorithmique et programmation pour les humanités : introduction à R .....	12
Algorithmique et programmation pour les humanités : introduction à Python.....	12
Méthodologie et mémoire (ENC 2314).....	12
<i>Cours 1 : Ateliers d'initiation à la recherche (ENC 2314-1).....</i>	<i>13</i>
<i>Cours 2 : Outils pour la rédaction du mémoire (ENC 2314-2).....</i>	<i>14</i>
Séminaire de recherche du master (ENC 2316).....	14
<b>Semestre 2 : tronc commun – 17 ECTS.....</b>	<b>16</b>
Modélisation et structuration des données (ENC 2437).....	16
<i>Cours 1 - Structuration XML et XML/TEI (ENC 2437-1).....</i>	<i>16</i>
<i>Cours 2 - Traitement et Analyse de données relationnelles (SQL et Dataiku) (ENC 2437-2).....</i>	<i>18</i>
Algorithmique et programmation pour les humanités : Python avancé (ENC 2320).....	18
Méthodologie et mémoire (ENC 2314).....	19
<i>Contenu 1 : Ateliers d'initiation à la recherche - suite (ENC 2314-1).....</i>	<i>19</i>
<i>Contenu 2 : Mémoire (ENC 2413-3).....</i>	<i>20</i>
Séminaire de recherche du master (ENC 2316).....	20
<b>Semestres 1 et 2 : enseignements de spécialité – 20 ECTS.....</b>	<b>21</b>
1. Bibliographie et recherche documentaire (ENC 1403) - obligatoire.....	21
2. PSL Week – au choix – 2 ECTS par PSL Week – optionnel.....	21
3. Cours de méthodologie de la recherche – ECTS selon module choisi.....	21
4. Séminaires et cours de spécialité de SHS.....	21
5. UE Sport de PSL – 1 ECTS – optionnel.....	21

<b>Master 2.....</b>	<b>22</b>
Stage de découverte – 3 ECTS.....	23
Mémoire – 20 ECTS sur l'année.....	23
<b>Semestre 3 : Enseignements fondamentaux – 11 ECTS.....</b>	<b>25</b>
Fondamentaux de mathématiques et Python pour DataScience (ENC 2325).....	25
Approches quantitatives et modélisation mathématique en SHS (ENC 2322).....	25
Langue vivante (au choix).....	26
Séminaire de recherche du master.....	26
<b>S3 – Enseignements optionnels – 12 ECTS.....</b>	<b>27</b>
Spécialisation de SHS Computationnelles (au choix) – 5 ECTS.....	27
<i>Philologie computationnelle – 24h, 4 ECTS (ENC 2323).....</i>	<i>27</i>
<i>Traitement et analyse de l'information spatialisée (EHESS 103).....</i>	<i>28</i>
Séminaire du directeur de recherche.....	29
Enseignement d'ouverture.....	29
<b>Mémoire – bilan d'étape – 5 ECTS.....</b>	<b>29</b>
<b>Semestre 4 : Enseignements fondamentaux – 7 ECTS.....</b>	<b>30</b>
Apprentissage machine (ENC 2321).....	30
Algorithmique et programmation pour les humanités : Python avancé (ENC 2320).....	30
Séminaire de recherche du master.....	30
<b>S4 – Enseignements optionnels – 10 ECTS.....</b>	<b>31</b>
Spécialisation de SHS Computationnelles (au choix) – 5 ECTS, au choix parmi les trois suivants.....	31
<i>Traitement automatique de la langue et analyse sémantique (ENS N104).....</i>	<i>31</i>
<i>Traitement automatique de l'image (EPHE P102).....</i>	<i>32</i>
<i>Analyse de réseaux (ENC 2324).....</i>	<i>32</i>
Séminaire du directeur de recherche.....	32
<b>Mémoire et soutenance – 15 ECTS.....</b>	<b>32</b>



## Master 1

En première année, le programme se subdivise entre:

- un tronc commun de méthodes computationnelles en SHS (40 ECTS) ;
- un programme pédagogique individualisé, composé essentiellement d'enseignements méthodologiques et thématiques propres à leur discipline: histoire, philologie, lettres, etc. (20 ECTS)

Ils réalisent en outre un mini-mémoire, qui compte pour 1/6e de la moyenne de l'année.

Les cours du **tronc commun humanités numériques** se concentrent sur les jeudis et vendredis durant les deux semestres.



## Semestre 1 : tronc commun – 23 ECTS

### Semaine intensive de rentrée

La première semaine de l'année est occupée par des présentations des enseignants du master, visant à présenter leur domaine, leurs recherches actuelles, les problèmes qu'elles adressent, et les possibles sujets de recherches futures (notamment les sujets de mémoire qui pourraient être traités par les étudiants du master). L'objectif est de plonger d'emblée les étudiants et étudiantes dans la recherche la plus actuelle en SHS computationnelles, et de leur faire entrevoir ce qu'ils ou elles pourraient y contribuer.

### Initiation aux humanités numériques (EPHE 100)

- **Enseignant** : Peter Stokes (EPHE)
- **Mode d'évaluation** : dossier à rendre en janvier
- **Volume horaire** : 20h
- **ECTS** : 3

### Programme

3h par séance, sauf 2h la dernière

1. Qu'est-ce que ce sont les HN. Principes de l'ordinateur; langages de programmation; séance pratique en programmation basique.
2. Le livre numérique ; HTML, PDF et ePub ; numérisation de textes et de documents éditions et représentations; séance pratique en HTML et TEI (Oxygen).
3. Humanités visuelles : images ; séance pratique 2D et élaboration d'images.
4. Données structurées; bases des données, ontologies et RDF ; séance pratique en BDD
5. TAL; authorship attribution; stylometry; distant reading; séance pratique en concordances (Wordsmith et TMX) et expressions régulières.
6. Intelligence Artificielle et Apprentissage de la Machine
7. Modélisation, prototypage et maquettes

### Éléments bibliographiques

- MOUNIER, Pierre, dir. (2012). *Read/Write Book 2 : Une introduction aux humanités numériques*. Nouvelle édition [en ligne]. Marseille : OpenEdition Press. doi :10.4000/books.oep.226
- CLIVAZ, Claire (2019). *Ecritures Digitales : Digital Writing, Digital Scriptures*, esp. Sauf le nom : nommer les Digital Humanities (p. 44–81). Leiden : Brill. doi :10.1163/9789004402560\_\_006
- MCCARTY, Willard (2004). Modeling : A Study in Words and Meanings. In Siemens, R. & Schreibman, S. *A Companion to Digital Humanities*. Oxford : Blackwell, p. 254–70. <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-3-7&toc.depth=1&toc.id=ss1-3-7&brand=default>
- FLANDERS, Julia, et JANNIDIS, Fotis (2015). Data Modeling. In Schreibman, S., Siemens, R. and Unsworth, J. (dir.), *A New Companion to Digital Humanities*. Chichester, UK : John Wiley & Sons, pp. 229–37 doi :10.1002/9781118680605.ch16

## Introduction aux mathématiques (ENC 2313)

- Enseignant : Michel Capot (SPIDEO)
- Mode d'évaluation : participation
- Volume horaire : 20h
- ECTS : 3

Introduction générale aux mathématiques : le langage symbolique et ses notations, éléments de logique, fonctions et dérivées, algèbre linéaire et éléments de probabilité et de statistique.

### Éléments bibliographiques

- Patrick Juola et Stephen Ramsay, *Six Septembers : Mathematics for the Humanist*, 2017, url : <http://digitalcommons.unl.edu/zeabook/55>.

## Fondamentaux de l'informatique (ENC 2435)

- Enseignants : Marc Bui (EPHE), Vincent Jolivet (ENC), Jean-Victor Boby (ENC).
- Mode d'évaluation : devoir sur table
- Volume horaire : 20h
- ECTS : 3

Contenu 1 : Séminaire intensif / install Party

Contenu 2 : Linux et Shell

Contenu 3 : Expressions régulières

Contenu 4 : Git et bonnes pratiques de dev

Ce module permet de découvrir les fondamentaux du fonctionnement du système d'exploitation Linux, du mode texte, des réseaux et des serveurs Web d'une part, de la programmation d'autre part, ce dernier point grâce à une introduction à l'algorithmique ponctuée d'exercices. Il propose également une première approche à la fouille de texte via l'utilisation d'expressions régulières, ainsi qu'aux bonnes pratiques du développement collaboratif, par une initiation à Git.

### Éléments bibliographiques

#### Linux et Bash

- Mathieu Nebra, «Reprenez le contrôle à l'aide de Linux», *OpenClassrooms*, en ligne : <https://openclassrooms.com/courses/reprenez-le-controle-a-l-aide-de-linux>.
- Documentation Ubuntu Francophone, notamment «Initiez-vous à GNU/Linux avec Ubuntu», en ligne : <https://doc.ubuntu-fr.org/initiation>.
- Eric Sanchis, *Introduction à la programmation en Bash*, en ligne : [http://aral.iut-rodez.fr/fr/sanchis/enseignement/IntroProgBash\\_2015-07-16.pdf](http://aral.iut-rodez.fr/fr/sanchis/enseignement/IntroProgBash_2015-07-16.pdf).

- Mendel Cooper, *Guide avancé d'écriture des scripts Bash*, en ligne : <https://abs.traduc.org/abs-5.1-fr/abs-5.1.04.pdf>.

### **Algorithmique**

- Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein, *Algorithmique*, Paris, 2010.
- Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, *Introduction to Algorithms*, 3e édition, MIT, 2009, en ligne : [http://ressources.unisciel.fr/algoprogram/s00aaroot/aa00module1/res/%5BCormen-AL2011%5DIntroduction\\_To\\_Algorithms-A3.pdf](http://ressources.unisciel.fr/algoprogram/s00aaroot/aa00module1/res/%5BCormen-AL2011%5DIntroduction_To_Algorithms-A3.pdf).
- Ebel Franck et Rohaut Sébastien, *Algorithmique. Techniques fondamentales de programmation. Exemples en Python*, Edition ENI, 2014.

## **Méthodes quantitatives en SHS (ENC 2310)**

- **Enseignants** : Florian Cafiero (CNRS), Jean-Baptiste Camps (ENC), Marie Puren (CNRS)
- **Mode d'évaluation** : exposé portant sur la lecture d'un article de recherche et devoir sur table.
- **Volume horaire** : 10h
- **ECTS** : 2

Ce cours est conçu comme une première approche à l'utilisation de méthodes quantitatives et statistiques dans les sciences humaines. Son objectif principal est de permettre la lecture et la compréhension autonomes de travaux de recherche se situant dans le champ des méthodes quantitatives en sciences humaines et des humanités numériques, en donnant les clefs et concepts nécessaires pour accéder à la partie mathématique du raisonnement. Il doit notamment permettre de pouvoir apprécier en connaissance de cause les partis pris méthodologiques ou les résultats présentés et leur significativité. On présentera également la démarche nécessaire à l'application de ces méthodes à une recherche académique, en mettant en lumière leur potentiel comme leurs limites. Dans un premier temps, ce cours se concentrera sur quelques notions et concepts fondamentaux, et sur les clefs de lecture les plus indispensables. Dans un second temps, il sera consacré à la lecture critique d'articles et l'étude de cas concrets d'application.

### **Programme indicatif**

1. Introduction et clefs de lecture ; SHS et données
2. Analyse de réseaux
3. Analyse de données textuelles
4. Lecture distante
5. Modèle de sujet
6. Stylométrie
7. Analyse causale
8. Systèmes complexes et modélisation mathématique
9. intelligence artificielle

## Algorithmique et programmation pour les humanités : introduction à R (EPHE P101)

- **Enseignant** : Daniel Stockholm (EPHE)
- **Mode d'évaluation** : devoir d'analyse de données à rendre.
- **Volume horaire** : 15h
- **ECTS** : 3

Ce cours est dédié à l'apprentissage de l'analyse et de la visualisation de données à partir de textes (littéraires ou autres). Il est conçu pour des étudiants sans expérience en programmation avec un niveau en mathématique de type bac littéraire.

Les techniques enseignées vous permettront d'importer des données issues de corpus de textes numériques accessibles par exemple via internet, de les organiser et de les analyser afin d'en extraire des informations quantitatives.

### Éléments bibliographiques

- Pierre-André Cornillon, Arnaud Guyader et François Husson, *Statistiques avec R*, Rennes, 2012.
- Matthew Jockers, *Text Analysis with R for Students of Literature*, Springer, 2014 (Quantitative methods in the Humanities and Social Sciences).

## Algorithmique et programmation pour les humanités : introduction à Python (ENC 2436)

- **Enseignant** : Jean Barré (ENC).
- **Mode d'évaluation** : contrôle continu (1/2) et devoir maison final (1/2)
- **Volume horaire** : 20h
- **ECTS** : 4

Ce module a pour objet de donner les connaissances de base en programmation informatique. Il revient sur les fondamentaux de l'algorithmique, mis en œuvre autour d'un langage très usité dans la recherche en SHS, Python.

### Détails sur le mode d'évaluation

$$\text{Note finale} = (A + B) / 2$$

où A (note sur 20) = Somme des 4 meilleures notes des devoirs à rendre au cours du semestre (6 à rendre). Chaque devoir produit une note sur 5, chaque devoir non rendu équivaut à une note de 0.

### Éléments bibliographiques

- MAILLE Vincent, *Python les Bases de l'Algorithmique et de la Programmation*, Ellipses, 2015.
- AMINI Massih-Reza, *Data Science*, Eyrolles, 2018.

## Méthodologie et mémoire (ENC 2314)

- **Enseignant** : Enseignants du master.

- **Mode d'évaluation** : mini-mémoire à rendre le **1ER JUIN (PREMIER JUIN, [CALENDIS JUNII])**.
- **Volume horaire** : 20h
- **ECTS** : 4

#### **Cours 1 : Ateliers d'initiation à la recherche (ENC 2314-1)**

- **Volume horaire** : 8h

Ce module de tutorat permet à chaque étudiant de se lancer dans l'application de méthodes et outils des humanités numériques à ses problématiques de recherche et à son mémoire.

À l'issue de la première année de master, les étudiants devront rendre un mini-mémoire de dimensions raisonnables (une trentaine de pages), accompagnées d'annexes numériques (données, scripts, etc.) permettant la vérification et la reproduction des analyses présentées.

Ce mini-mémoire doit être conçu comme un premier pas vers une démarche de recherche approfondie en M2, pour permettre d'en vérifier la faisabilité. Partant d'une question de recherche et d'un état de l'art du sujet, il doit se concentrer sur un sous-ensemble du corpus envisagé, et proposer un premier protocole d'analyse, ainsi que quelques pistes d'interprétation des résultats.

Il est impératif **d'identifier des directeurs**, de **définir le sujet** et de **commencer l'état de l'art et les recherches** aussi tôt que possible dans l'année.

**N'hésitez pas à solliciter les conseils de l'équipe pédagogique** : la démarche de préparation du mémoire vous est propre, doit être faite en autonomie, et **il n'appartient pas aux directeurs de recherche de venir vous chercher, mais à vous des les solliciter** quand vous en éprouvez le besoin, ainsi que pour faire des points réguliers.

Réaliser un mémoire peut (doit) être une entreprise passionnante, et c'est le travail où vous mettez le plus de vous-mêmes. Son succès dépend avant tout de votre investissement et de votre intérêt pour la problématique. Il dépend toutefois aussi de la faisabilité et de la pertinence de l'approche, points sur lesquels vous aurez besoin des conseils de l'équipe pédagogique, ainsi que du test grandeur nature fourni par le mini-mémoire de M1.

Dans ce contexte, le premier semestre devra être le moment de débiter ce travail en :

- déterminant le **sujet**, avec l'accord de directeurs de recherche (directeur de spécialité et directeur «numérique») et élaborant un **questionnement** et une **problématique**;
- faisant **l'état de l'art**;
- collectant **les données**.

#### **Ateliers du premier semestre**

1. Qu'est-ce qu'un mémoire? Construire une démarche de recherche.
2. Les données.
3. Le protocole d'analyse.
4. Résultats et interprétation.

#### **Éléments bibliographiques**

*Forme*

- Jean-Baptiste Camps, biblatex-enc, *BibLaTeX style files for the bibliographic standards of the École nationale des chartes*, en ligne : (<https://ctan.org/pkg/biblatex-enc>), <https://github.com/Jean-Baptiste-Camps/biblatex-enc>.
  - Imprimerie nationale, *Lexique des règles typographiques en usage à l'Imprimerie nationale*, Paris, 2002
- Fond*
- Mémoires des années précédentes, prochainement en ligne sur : <https://zenodo.org/communities/dhma-psl/?page=1&size=20>.
  - *A companion to digital humanities*, dir. Susan Schreibman, Raymond Georges Siemens, et John Unsworth, Oxford, 2004. <http://www.digitalhumanities.org/companion>. Une version assez largement révisée de cet ouvrage est parue en 2016 : *A new companion to digital humanities*, dir. Susan Schreibman, Raymond Georges Siemens et John Unsworth, Malden, etc., 2016.
  - Jim Gray, « Jim Gray on eScience : A transformed scientific method ». In *The fourth paradigm : Data-intensive scientific discovery*, édité par Tony Hey, Stewart Tansley, et Kristin Tolle, xviii. Washington, 2009.
  - *Digital Scholarship in the Humanities* (olim *Literary and Linguistic Computing*), <https://academic.oup.com/dsh/>.
  - *Digital Humanities Quarterly*, <http://www.digitalhumanities.org/dhq/>.
  - Actes des conférences annuelles *Digital Humanities*.

#### Cours 2 : Outils pour la rédaction du mémoire (ENC 2314-2)

- **Enseignante** : Énimie Rouquette (ÉNC)
- **Volume horaire** : 12h

Présentation de quelques outils pour la constitution de bibliographies (Zotero), pour la rédaction du mémoire ( $\LaTeX$ ) et de bonnes pratiques (sauvegardes fréquentes, organisations des fichiers, etc.).

#### Séminaire de recherche du master (ENC 2316)

- **Coordinateurs** : J.B. Camps, Marie Puren, et les enseignants du master.
- **Mode d'évaluation** : participation et discussion (M1); présentation de 15 minutes (M2)
- **Volume horaire** : 5h
- **ECTS** : 1

Ce séminaire, qui s'étend sur les quatre semestres de la formation, permet aux étudiants de première année de découvrir les recherches les plus actuelles dans le domaine des humanités numériques, notamment celles menées par des étudiants plus avancés. Aux étudiants de deuxième année, elle fournit également l'occasion de confronter leurs expériences et leurs idées, et de s'initier à la présentation de leurs travaux.

Le séminaire est évalué de la sorte :

- pour les étudiants de M2, ils doivent une fois dans l'année réaliser une présentation de leurs recherches de 15 minutes;

- pour les étudiants de M1, ils doivent endosser une fois dans l'année le rôle de discutant d'une des présentations, et animer les débats autour de celle-ci.

### **Éléments bibliographiques**

Les compte-rendus des années précédentes sont disponibles sur le carnet de recherche du séminaire :

- Master HN, *Données et sciences humaines : Carnet de recherche du master Humanités numériques de l'Université Paris, Sciences & Lettres*, dir. Jean-Baptiste Camps & Marie Puren, <https://humanodon.hypotheses.org/>, 2019-...

## Semestre 2 : tronc commun – 17 ECTS

### Modélisation et structuration des données (ENC 2437)

Ce module propose une initiation à deux modes de structuration des données d'emploi très courant dans les humanités numériques, le métalangage XML, qui permet notamment de structurer des données textuelles, et les bases de données relationnelles.

#### Cours 1 - Structuration XML et XML/TEI (ENC 2437-1)

- **Enseignant** : Patricia O'Connor, Lucence Ing (ENC)
- **Mode d'évaluation** : dossier d'édition électronique
- **Volume horaire** : 20h
- **ECTS** : 4

Séance	Titre
Séance 1	Introduction à XML
Séance 2	Présentation de TEI
Séance 3	Structuration générale d'un fichier TEI
Séance 4	Les TEI guidelines
Séance 5	Initiation à l'édition scientifique et description des sources manuscrites
Séance 6	Encoder une édition scientifique
Séance 7	Les schémas XML (1) : Les différents formats
Séance 8	Les schémas XML (2) : Personnalisations avancées de ODD
Séance 9	Les schémas XML (3) : Mise en pratique
Séance 10	Initiation à Xpath
Séance 11	Transformer et interroger son XML
Séance 12	Retour d'expérience

•

#### Modalité d'évaluation

Évaluation finale : proposer un encodage XML-TEI pertinent et argumenté pour un corpus donné. Le fichier XML devra être régi par un schéma ODD et accompagné d'une documentation.

#### Éléments bibliographiques

#### Ouvrages à consulter



BUREAU, Bruno, « Quelques réflexions sur la notion de littérarité à partir de l'édition numérique de commentateurs anciens », *Interférences. Ars scribendi*, janvier 2012, En ligne : <http://journals.openedition.org/interferences/186>.

BURNARD, Lou et BURGHART, Marjorie, *Qu'est-ce que la Text Encoding Initiative ?*, 2015, en ligne : <http://books.openedition.org/oep/1237>.

DUVAL, Frédéric, « Pour des éditions numériques critiques. L'exemple des textes français », in *Le texte à l'épreuve du numérique*, Saint-Denis, France, Presses universitaires de Vincennes, 2017, p. 1330.

HAROLD, Elliotte Rusty, MEANS, W. Scott, ENSARGUET, Philippe [et al.], *XML en concentré*, Paris, O'Reilly, 2005.

KAY, Michael R., *XPath 2.0 programmer's reference*, Indianapolis, IN, Wrox Press, 2004.

PALLUAULT, Florent, « Informatiser des descriptions complexes : l'utilisation de l'EAD et de la TEI pour les manuscrits et les livres anciens en France », *IFLA 2012 (Helsinki)*, en ligne : <http://conference.ifla.org/past/2012/212-palluault-fr.pdf>.

PIERAZZO, Elena, *Digital scholarly editing : theories, models and methods*, Farnham Burlington (Vt.), Ashgate, 2015.

RAHTZ, Sebastian et BURNARD, Lou, « Reviewing the TEI ODD System », *ACM*, 2013, p. 193 196, En ligne : <http://dx.doi.org/10.1145/2494266.2494321>.

ROBINSON, Peter, « Where We Are with Electronic Scholarly Editions, and Where We Want to Be », *Jahrbuch für Computerphilologie*, vol. 5 / 5, 2003, p. 126146.

Ron Van den Branden, Melissa Terras, et Edward Vanhoutte, *TEI by Example*, <http://teibyexample.org/TBE.htm>.

### Ressources numériques

BURGHART, Marjorie, PIERAZZO, Elena et al., *Online course : Digital Scholarly Editions : Manuscripts, Texts, and TEI Encoding – Digital Editing of Medieval Manuscripts*, en ligne : <https://www.digitalmanuscripts.eu/digital-editing-of-medieval-texts-a-textbook/>.

*TEI guidelines*, en ligne : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

*XML Tutorial*, en ligne : <https://www.w3schools.com/xml/default.asp>.

*XPath Tutorial*, en ligne : [https://www.w3schools.com/xml/xpath\\_intro.asp](https://www.w3schools.com/xml/xpath_intro.asp).

*XSLT Tutorial*, en ligne : [https://www.w3schools.com/xml/xsl\\_intro.asp](https://www.w3schools.com/xml/xsl_intro.asp).

### Éditions à consulter

*Beta maṣāḥəft : Manuscripts of Ethiopia and Eritrea*, <http://betamasaheft.eu>.

*Geste : un corpus de chansons de geste*, <http://dev.chartes.psl.eu/elec/geste/>.

*L'année 1437 dans la pratique de Pierre Christofle, notaire du Châtelet d'Orléans*, <http://elec.enc.sorbonne.fr/christofle/index.html>.

*Perseus Digital Library*, <http://www.perseus.tufts.edu/hopper/>.

*The Shelley-Godwin Archive*, <http://shelleygodwinarchive.org/sc/oxford/frankenstein/volume/i/#/p1/mode/rdg>.

## Cours 2 - Traitement et Analyse de données relationnelles (SQL et Dataiku) (ENC 2437-2)

- **Enseignants** : Vincent Jolivet (ENC)
- **Mode d'évaluation** : devoir de traitement et d'analyse de données non textuelles.
- **Volume horaire** : 10h
- **ECTS** : 2

1. Initiation aux bases de données relationnelles;
2. Manipulation et traitement de données; outils du web sémantique;

### Éléments bibliographiques

- Christian Soutou, *Apprendre SQL avec MySQL, avec 40 exercices corrigés*, Paris, 2006.
- Georges Gardarin, *Bases de données : objet et relationnel*, Paris, 1999.
- Dominique Nanci, Bernard Espinasse, *Ingénierie des systèmes d'information : Merise*, Paris, 2001.

## Algorithmique et programmation pour les humanités : Python avancé (ENC 2320)

- **Enseignant** : Chahan Vidal-Gorène (ENC).
- **Mode d'évaluation** : contrôle continu (1/2) et devoir maison final (1/2)
- **Volume horaire** : 20h
- **ECTS** : 4

Ce cours est consacré au renforcement des compétences en Python et en mathématiques, notamment en mathématiques appliquées à l'apprentissage machine (algèbre linéaire, mesures, métriques, clustering, ...). **Ce cours est commun avec les M2.** Le programme alterne une année sur deux.

### Détails sur le mode d'évaluation

$$\text{Note finale} = (A + B) / 2$$

où A (note sur 20) = Somme des 4 meilleures notes des devoirs à rendre au cours du semestre (6 à rendre). Chaque devoir produit une note sur 5, chaque devoir non rendu équivaut à une note de 0.

### Éléments bibliographiques

#### *Statistiques et Probabilités*

- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R., *An introduction to statistical learning : with applications in R*, Springer, 2017.
- MÉLÉARD, S., *Aléatoire, introduction à la théorie et au calcul des probabilités*, Palaiseau, Ecole Polytechnique, 2010.

#### *Application python*

- VANDERPLAS, J. T., *Python Data Science Handbook Essential Tools for Working with Data*, O'reilly, 2017.  
*Algèbre Linéaire*
- GRIFONE, J., *Algèbre Linéaire*, Cepadues, 2018.

### **Méthodologie et mémoire (ENC 2314)**

- **Enseignant** : Enseignants du master.
- **Mode d'évaluation** : mini-mémoire à rendre le **1er juin**
- **ECTS** : 6

#### **Contenu 1 : Ateliers d'initiation à la recherche - suite (ENC 2314-1)**

- **Volume horaire** : 10h

Suite du module du premier semestre. Pour rappel :

À l'issue de la première année de master, les étudiants devront rendre un mini-mémoire de dimensions raisonnables (une trentaine de pages), accompagnées d'annexes numériques (données, scripts, etc.) permettant la vérification et la reproduction des analyses présentées.

Ce mini-mémoire doit être conçu comme un premier pas vers une démarche de recherche approfondie en M2, pour permettre d'en vérifier la faisabilité. Partant d'une question de recherche et d'un état de l'art du sujet, il doit se concentrer sur un sous-ensemble du corpus envisagé, et proposer un premier protocole d'analyse, ainsi que quelques pistes d'interprétation des résultats.

Il est impératif **d'identifier des directeurs**, de **définir le sujet** et de **commencer l'état de l'art** et les recherches aussi tôt que possible dans l'année. **N'hésitez pas à solliciter les conseils de l'équipe pédagogique** : la démarche de préparation du mémoire vous est propre, doit être faite en autonomie, et **il n'appartient pas aux directeurs de recherche de venir vous chercher, mais à vous des les solliciter** quand vous en éprouvez le besoin, ainsi que pour faire des points réguliers.

Réaliser un mémoire peut (doit) être une entreprise passionnante, et c'est le travail où vous mettez le plus de vous-mêmes. Son succès dépend avant tout de votre investissement et de votre intérêt pour la problématique. Il dépend toutefois aussi de la faisabilité et de la pertinence de l'approche, points sur lesquels vous aurez besoin des conseils de l'équipe pédagogique, ainsi que du test grandeur nature fourni par le mini-mémoire de M1.

*Le second semestre est le moment de l'année consacré essentiellement à l'analyse des données, l'obtention de résultats, leur interprétation, ainsi qu'à leur présentation rédigée, claire et argumentée, dans un mini-mémoire.*

#### **Programme**

[Sera précisé ultérieurement]

#### **Éléments bibliographiques**

*Forme*

- Jean-Baptiste Camps, biblatex-enc, *BibLaTeX style files for the bibliographic standards of the École nationale des chartes*, en ligne : (<https://ctan.org/pkg/biblatex-enc>), <https://github.com/Jean-Baptiste-Camps/biblatex-enc>.
- Imprimerie nationale, *Lexique des règles typographiques en usage à l'Imprimerie nationale*, Paris, 2002  
*Fond*
- Mémoires des années précédentes, prochainement en ligne.
- *A companion to digital humanities*, dir. Susan Schreibman, Raymond Georges Siemens, et John Unsworth, Oxford, 2004. <http://www.digitalhumanities.org/companion>. Une version assez largement révisée de cet ouvrage est parue en 2016 : *A new companion to digital humanities*, dir. Susan Schreibman, Raymond Georges Siemens et John Unsworth, Malden, etc., 2016.
- Jim Gray, « Jim Gray on eScience : A transformed scientific method ». In *The fourth paradigm : Data-intensive scientific discovery*, édité par Tony Hey, Stewart Tansley, et Kristin Tolle, xviixxxi. Washington, 2009.
- *Digital Scholarship in the Humanities* (olim *Literary and Linguistic Computing*), <https://academic.oup.com/dsh/>.
- *Digital Humanities Quarterly*, <http://www.digitalhumanities.org/dhq/>.
- Actes des conférences annuelles *Digital Humanities*.

## Contenu 2 : Mémoire (ENC 2413-3)

### Séminaire de recherche du master (ENC 2316)

- **Coordinateurs** : J.B. Camps, Marie Puren, et les enseignants du master.
- **Mode d'évaluation** : participation et discussion (M1); présentation de 15 minutes (M2)
- **Volume horaire** : 5h
- **ECTS** : 1

Ce séminaire, qui s'étend sur les quatre semestres de la formation, permet aux étudiants de première année de découvrir les recherches les plus actuelles dans le domaine des humanités numériques, notamment celles menées par des étudiants plus avancés. Aux étudiants de deuxième année, elle fournit également l'occasion de confronter leurs expériences et leurs idées, et de s'initier à la présentation de leurs travaux.

Le séminaire est évalué de la sorte :

- pour les étudiants de M2, ils doivent une fois dans l'année réaliser une présentation de leurs recherches de 15 minutes;
- pour les étudiants de M1, ils doivent endosser une fois dans l'année le rôle de discutant d'une des présentations, et animer les débats autour de celle-ci.

### Éléments bibliographiques

Les compte-rendus des années précédentes sont disponibles sur le carnet de recherche du séminaire :

- Master HN, *Données et sciences humaines : Carnet de recherche du master Humanités numériques de l'Université Paris, Sciences & Lettres*, dir. Jean-Baptiste Camps & Marie Puren, <https://humanodon.hypotheses.org/>, 2019-...

## **Semestres 1 et 2 : enseignements de spécialité – 20 ECTS**

Un total de 20 ECTS, provenant des enseignements dispensés dans les établissements partenaires, selon un programme pédagogique validé par le(s) directeur(s) de recherche et le responsable pédagogique. **Aucun cours qui n'aurait pas été au préalable validé par l'équipe pédagogique ne pourra être suivi.**

Il comporte:

### **1. Bibliographie et recherche documentaire (ENC 1403) - obligatoire**

- **Coordinateurs** : Dégez-Selves Camille, Mathis Rémi
- **Mode d'évaluation** : bibliographie du mini-mémoire
- **Volume horaire** : 16h (8h au S1, 8h au S2)
- **ECTS** : 2

Séminaire qui s'étend sur les deux semestres, visant à donner les clefs pour la création de bibliographies et de découvrir les outils pour la recherche et la gestion documentaire.

### **2. PSL Week – au choix – 2 ECTS par PSL Week – optionnel**

Enseignements d'ouverture interdisciplinaire ou de renforcement en méthodes computationnelles. Liste complète sur le site de PSL :

<https://psl-week.psl.eu/LISTE-DES-PSL-WEEKS.php>

### **3. Cours de méthodologie de la recherche en... (votre discipline) – ECTS selon module choisi**

À sélectionner dans les cours des 4 établissements (ENC, ENS, EPHE, EHESS)

### **4. Séminaires et cours de spécialité de SHS**

**Programme de cours et séminaires de votre discipline**, sélectionné en cohérence avec votre projet de mémoire et vos directeurs/directrices de recherche.

À sélectionner dans les cours des 4 établissements (ENC, ENS, EPHE, EHESS)

**Valorisation** : sauf exception, 1 ECTS par 5h de cours (comportant obligatoirement une **validation finale notée** et du travail personnel)

*Voir les listes par établissement:*

- École nationale des chartes : voir la liste des cours disponibles sur le site de l'École, <http://www.chartes.psl.eu/fr/recherche/bundle/cours/>.
- École normale supérieure : se renseigner auprès des départements.
- École pratique des hautes études : <https://www.ephe.fr/formations/conferences/>.
- EHESS : <https://enseignements.ehess.fr/>

### **5. UE Sport de PSL – 1 ECTS – optionnel**

UE de sport proposée par PSL, programme et inscription à <https://psl.eu/vie-de-campus/paris-est-notre-campus/psl-sport>

## Master 2

La deuxième année de master se construit sur un **tronc commun**, permettant d'approfondir l'initiation reçue en première année (méthodes quantitatives, programmation, intelligence artificielle, science des données), tandis que des **cours et séminaires optionnels** permettent à chaque étudiant d'approfondir sa spécialisation dans les méthodes propres à son champ disciplinaire. Les étudiants suivent également plusieurs **séminaires de recherche**. Un séminaire du master donne l'occasion aux étudiants de deuxième année de présenter leurs travaux aux étudiants de première année.

La deuxième année est validée pour 1/3 par l'écriture et la soutenance d'un mémoire de recherche, encadré par un co-directeur ou co-directrice, enseignant du tronc commun du master, ainsi qu'un co-directeur ou co-directrice dans la spécialité SHS retenue. Le mémoire devra mettre les méthodes computationnelles (données structurées, algorithmes, méthodes quantitatives, ...) au service de la recherche, via une démarche de test d'hypothèse appuyée sur des données, et une bonne connaissance de l'état de l'art (tant en humanités numériques qu'en SHS traditionnelles).

Pour les étudiants visant une poursuite en doctorat, le choix des séminaires de recherche optionnels est crucial. De même, pour les étudiants envisageant une professionnalisation dès la sortie du master 2, le choix des cours et séminaires doit refléter cette ambition.

## Stage de découverte – 3 ECTS

- **Volume horaire** : de 4 à 8 semaines
- **ECTS** : 3

Chaque étudiant devra, au cours de la deuxième année, réaliser un stage de découverte d'un laboratoire, d'une durée **maximale de deux mois**. Ce stage pourra avoir lieu soit en alternance durant l'année universitaire, soit en continu hors période des cours, et peut prendre place durant l'été **précédant le début des cours ou suivant le rendu du mémoire**. **Ce stage peut se voir substituer une mobilité internationale ou un séjour de recherche à l'étranger.**

## Mémoire – 20 ECTS sur l'année

Mémoire de recherche

- donnant lieu à un bilan d'étape à la fin du 1<sup>er</sup> semestre (5 ECTS) ;
- présenté devant la communauté des M1, M2 et enseignants dans le séminaire du master ;
- rendu avant le **15 juin** (première session) ou le **4 septembre** (deuxième session). Tout retard injustifié aura des conséquences.
- défendu dans une **soutenance publique** (rendu et soutenance, 15 ECTS).

Pour obtenir son diplôme, l'étudiant devra réaliser et soutenir un mémoire de recherche, mettant en œuvre ses compétences disciplinaires et numériques, et qui sera suivi par un directeur de recherche spécialiste du sujet et un spécialiste des humanités numériques. Le mémoire devra mettre les technologies numériques (données structurées, algorithmes, méthodes quantitatives, ...) au service de son sujet de recherche, et accompagner son mémoire de fichiers techniques.

Comme en M1, il ne faut **pas hésiter à solliciter les conseils des encadrants** : ici encore, la démarche de préparation du mémoire vous est propre, doit être faite en autonomie, et **il n'appartient pas aux directeurs de recherche de venir vous chercher, mais à vous de les solliciter** quand vous en éprouvez le besoin, ainsi que pour faire des points réguliers.

Réaliser un mémoire peut (doit) être une entreprise passionnante, et c'est le travail où vous mettez le plus de vous-mêmes. Son succès dépend avant tout de votre investissement et de votre intérêt pour la problématique.

Entreprise de dimensions plus importantes que le mini-mémoire, ce mémoire de M2 doit capitaliser sur les expériences menées au cours du M1. Le travail qui lui est consacré ne doit pas se concentrer sur la fin de l'année, mais avoir démarré dès la rentrée, voire avant (pause estivale).

Pour les M2, le mémoire est l'aboutissement de vos études de master – voire le premier pas vers une thèse de doctorat – et il est également le travail dans lequel vous avez le plus d'autonomie et de liberté. Il importe néanmoins de faire un point régulier sur l'état d'avancement de vos travaux avec vos directeurs.

Les mémoires ne sont pas jugés sur la quantité de pages qu'ils contiennent, mais sur la qualité et rigueur de leur démarche scientifique et numérique, tant du point de vue des sujets abordés que de la méthode (données, outils d'analyse, etc.). En M2, 80 pages peuvent suffire, si

elles sont claires, bien construites et fondées sur une démarche pertinente. Il s'agit là toutefois d'indications et non pas de règles absolues.

Le mémoire doit être rédigé dans une langue claire et rigoureuse, évitant les anglicismes ou le jargon. Il s'adresse à un public relativement large et a vocation à pouvoir être diffusé en ligne. À ce titre, d'une part, la qualité de son contenu vous engage autant que PSL et, d'autre part, le texte du mémoire doit être clair et compréhensible, tant que faire se peut, par des chercheurs même non spécialistes du sujet traité. Les concepts, normes, outils... mentionnés devront être définis et présentés. Tous les tableaux et illustrations devront être légendés (et repris dans la table des illustrations).

Les mémoires de master 2 donnent lieu à **soutenance publique**. Le jury est composé de trois membres, dont le ou les directeurs de recherche et un président de jury. Durant cette soutenance, l'étudiant dispose de 15 minutes pour présenter sa démarche et ses recherches, les résultats et limites de celle-ci et les perspectives qu'elle ouvre, par un exposé oral aidé ou non d'un diaporama, et qui peut être étayé par une démonstration technique. Son exposé laisse ensuite la place à une discussion, incluant les remarques de chaque membre du jury et des questions. La durée maximale de la soutenance est d'1h30.



## Semestre 3 : Enseignements fondamentaux – 11 ECTS

### Fondamentaux de mathématiques et Python pour DataScience (ENC 2325)

- **Enseignant** : Chahan Vidal-Gorène
- **Mode d'évaluation** : contrôle continu (1/2) et devoir maison final (1/2)
- **Volume horaire** : 10h
- **ECTS** : 2

Le cours vise à revoir les bibliothèques essentielles pour la manipulation et la visualisation des données textuelles et non textuelles (image). Il comporte pour moitié des rappels en mathématiques pour l'analyse numérique, le traitement de l'image, ainsi que de l'analyse (étude de fonctions).

#### Détails sur le mode d'évaluation

Note finale =  $(A + B) / 2$

où A (note sur 20) = Somme des 4 meilleures notes des devoirs à rendre au cours du semestre (6 à rendre). Chaque devoir produit une note sur 5, chaque devoir non rendu équivaut à une note de 0.

#### Éléments bibliographiques

- VANDERPLAS, J. T., *Python Data Science Handbook Essential Tools for Working with Data*, O'reilly, 2017.

### Approches quantitatives et modélisation mathématique en SHS (ENC 2322)

- **Enseignants** : Jean-Pierre Nadal (EHESS), Julien Randon-Furling (Univ. Paris 1), Annick Vignes (ENPC).
- **Mode d'évaluation** : exposé à présenter sur la lecture d'un article de recherche.
- **Volume horaire** : 20h
- **ECTS** : 4

Le cours, pluri- et trans-disciplinaire, porte sur une palette d'approches quantitatives pour l'analyse et la modélisation mathématique de données : analyse multi-variée; méthodes économétriques; modélisation multi-agent; modèles dynamiques (dynamiques de population...), théorie des jeux... Le cours est structuré de manière à présenter les concepts, méthodes, modèles, dans le cadre d'une démarche globale d'analyse et de modélisation. Les notions et techniques sont systématiquement illustrées sur des cas concrets d'intérêt en SHS. Le cours demande une forte implication des étudiants : mise en œuvre des théories ou méthodes sur des données réelles, lecture critique d'articles. La validation se fait sur un travail personnel s'appuyant sur un article scientifique.

#### Éléments bibliographiques

*Cours et ressources d'apprentissage en ligne*

*Statistiques*, <http://wikistat.fr/>.

Cours *Statistique et Big Data Mining*, Philippe Besse, INSA Toulouse <http://www.math.univ-toulouse.fr/~besse/enseignement.html>.

Charpentier A., Flachaire E. et Ly A., *Économétrie et Machine learning*, 2017, en ligne : <http://freakonometrics.hypotheses.org/files/2017/07/econometrics-ML-final-1.pdf>.

Juola, P. & Ramsay S., *Six Septembers : Mathematics for the Humanist*, Zea Books, 2017 doi : <http://dx.doi.org/10.13014/K2D21VHX>.

*Se former avec le numérique*, site du Ministère de l'enseignement supérieur, de la recherche et de l'innovation (cours, exercices, moocs, etc, dans divers domaines, tous niveaux).

Cours *Systèmes dynamiques. Chaos et applications* de Frédéric Faure, UJF Grenoble, en ligne : [https://www-fourier.ujf-grenoble.fr/~7Efaure/enseignement/systemes\\_dynamiques/index.html](https://www-fourier.ujf-grenoble.fr/~7Efaure/enseignement/systemes_dynamiques/index.html). Lire le chapitre d'introduction - et plus si affinités.

#### Articles

T. C. Schelling, *Micromotives and Macrobehavior*, Norton & Co, 1978. Traduction française : *La tyrannie des petites décisions*, PUF, 1980.

Voir sur ce livre le compte-rendu de lecture de E. Reynaud, paru dans la *Revue française de sociologie* en 1982.

Thomas C. Schelling, *The Strategy of Conflict*, Boston, Harvard University Press, 1960. Traduction française *Stratégie du conflit*, Paris, Presses universitaires de France, 1986.

Sur ce thème, voir l'interview de Schelling par Erhard Friedberg et Martha Zuber, 27 oct. 2000 (traduite en français par J. Bouyssou).

J.-P. Nadal et M. B. Gordon, «Choix sous influence sociale : heurs et malheurs de la coordination», dans *Qu'appelle-t-on aujourd'hui les sciences de la complexité?*, sous la direction de G. Weisbuch & H. Zwirn, Vuibert, 2010, en ligne : <http://freakonometrics.hypotheses.org/files/2017/07/econometrics-ML-final-1.pdf>.

Voir aussi les articles donnés comme supports de cours.

### Langue vivante (au choix)

- Volume horaire : 10h
- ECTS : 1

### Séminaire de recherche du master

- **Coordinateurs** : J.B. Camps, Marie Puren, et les enseignants du master.
- **Mode d'évaluation** : participation et discussion (M1) ; présentation de 15 minutes (M2)
- **Volume horaire** : 5h
- **ECTS** : 1

Ce séminaire, qui s'étend sur les quatre semestres de la formation, permet aux étudiants de première année de découvrir les recherches les plus actuelles dans le domaine des humanités numériques, notamment celles menées par des étudiants plus avancés. Aux étudiants de deuxième année, elle fournit également l'occasion de confronter leurs expériences et leurs idées, et de s'initier à la présentation de leurs travaux.

Le séminaire est évalué de la sorte :

- pour les étudiants de M2, ils doivent une fois dans l'année réaliser une présentation de leurs recherches de 15 minutes;

- pour les étudiants de M1, ils doivent endosser une fois dans l'année le rôle de discutant d'une des présentations, et animer les débats autour de celle-ci.

### Éléments bibliographiques

Les compte-rendus des années précédentes sont disponibles sur le carnet de recherche du séminaire :

- Master HN, *Données et sciences humaines : Carnet de recherche du master Humanités numériques de l'Université Paris, Sciences & Lettres*, dir. Jean-Baptiste Camps & Marie Puren, <https://humanodon.hypotheses.org/>

## Semestre 3 – Enseignements optionnels – 12 ECTS

### Spécialisation de SHS Computationnelles (au choix) – 5 ECTS, au choix parmi les deux suivants

#### Philologie computationnelle – 24h, 4 ECTS (ENC 2323)

- **Enseignant** : J.B. Camps (ENC), F. Cafiero (Univ. Sorbonne Paris Cité).
- **Mode d'évaluation** : dossier d'analyse de données textuelles à rendre en janvier.
- **Volume horaire** : 25h
- **ECTS** : 5

Ce cours aborde à la fois les méthodes de **production et d'analyse de corpus textuels** (stylo-métrie notamment).

Ce cours s'adresse à tous ceux qui souhaitent acquérir des notions fondamentales de philologie computationnelle, et veulent découvrir les tenants et aboutissants pratiques et théoriques de la constitution et de l'analyse de corpus. Une très large place est laissée aux travaux pratiques, effectués sur plusieurs corpora. Ce cours envisagera ainsi les méthodes computationnelles en philologie, tant dans la constitution de corpus ou l'édition de textes que dans leur étude et l'établissement de connaissances les concernant. Il suivra la distinction traditionnelle entre haute et basse critique, c'est-à-dire entre problématiques liées à l'attribution des textes, leur datation ou leur localisation, et démarche ecdotique, de la recension des manuscrits à l'établissement de la généalogie des témoins. Le cours envisagera à chaque fois les questions de production ou collecte et formalisation des données, ainsi que leur analyse.

Sur un plan technique, les langages utilisés seront :

- R : le langage d'analyse statistique, utilisé via l'environnement de développement RStudio;
- Python : qui servira notamment pour ses bibliothèques d'apprentissage machine;
- XML : langage descriptif incontournable pour la structuration des données textuelles

Des outils d'administration de corpus et de textométrie seront également présentés, en particulier :

- TXM : plateforme dédiée à la création et l'interrogation de corpus.

### Programme

1. Introduction : fondamentaux de philologie computationnelle;
2. Ecdotique 1 : manuscrits et données matérielles;

3. Ecdotique 2 : transcription, reconnaissance des écritures et système graphique;
4. Ecdotique 3 : alignement des textes et collation;
5. Stylométrie 1 : première approche à la stylométrie : méthodes exploratoires;
6. Stylométrie 2 : méthodes de partitionnement;
7. Stylométrie 3 : méthodes supervisées (KNN, SVM, NN);
8. Séance d'approfondissements et de pratique selon les problématiques des étudiants.

### Éléments bibliographiques

#### Généralités

- Tara Andrews, «The third way : philology and critical edition for a digital age», *Variants : the Journal of the European Society for Textual Scholarship*, 10 (2012), <http://boris.unibe.ch/43071/>.
- Jean-Baptiste Camps, «Où va la philologie numérique?», *LHT Fabula*, 20 (2018), <http://www.fabula.org/lht/20/camps.html>.

#### Stylométrie

- E. Stamatatos, «A Survey of modern authorship attribution methods», *Journal of the American Society for information Science and Technology*, 60–3 (2009), p. 538–556, DOI : <https://dx.doi.org/10.1002/asi.21001>.
- Mike Kestemont, «Function words in authorship attribution : From black magic to theory?», *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 2014, p. 59-66, en ligne : <https://www.aclweb.org/anthology/W14-0908>.
- Moshe Koppel, Jonathan Schler et Shlomo Argamon, « Computational methods in authorship attribution », *Journal of the American Society for Information Science & Technology*, 60–1 (2009), p. 9–26, DOI : <https://dx.doi.org/10.1002/asi.20961>.
- \_Digital Humanities for Literary Studies : Theories, Methods, and Practices, et notamment l'article de M. Eder, J. Rybicki, M. Kestemont, « Measuring Style ».
- Stefan Evert, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch et Thorsten Vitt, «Understanding and explaining Delta measures for authorship attribution», *Digital Scholarship in the Humanities*, 32 (suppl\_2[ 2017]), p. ii4–ii16.

### Traitement et analyse de l'information spatialisée (EHES 103)

- **Enseignant** : Carmen Brando (EHES, CRH)
- **Mode d'évaluation** : dossier à rendre.
- **Volume horaire** : 25h
- **ECTS** : 5

Pour les inscriptions et plus d'informations, voir le site de l'atelier : <https://psigehess.hypotheses.org/1879>. Les Systèmes d'Information Géographique (SIG) et leur développement en ligne (webmapping) apparaissent comme un moyen de faire travailler concrètement les chercheurs ensemble. Le lieu géographique fait le lien pour des chercheurs de différentes disciplines (géographes, archéologues, historiens, sociologues, anthropologues,

informaticiens, etc.). Les SIG induisent la mise en place de nouvelles logiques de recherche. Pour être pleinement opérationnels, ils ne peuvent être dissociés des acteurs, qui dialoguent, négocient et adaptent des cultures et solutions techniques qui leur sont propres. L'examen des modalités de mise en œuvre de référentiels géo-historiques permet de mettre au jour des lignes directrices collectives et de préciser les éléments qui sont à la base de ces nouvelles pratiques de recherche et qui sont des expressions concrètes de ces nouvelles formes d'interdisciplinarités. Ce séminaire se propose d'être un lieu d'échange sur différentes expériences menées en géomatique à l'EHESS et dans d'autres institutions. Il sera associé à une formation initiale sur les SIG permettant d'acquérir les bases du maniement des outils géomatiques et de l'analyse des données cartographiques et de l'imagerie aérienne.

### **Éléments bibliographiques**

- Taylor Arnold, Lauren Tilton, *Humanities Data in R Exploring Networks, Geospatial Data, Images, and Text*, 2015 (Quantitative Methods in the Humanities and Social Sciences).
- Patricia Bordin, *SIG : Concepts, outils et données*, Paris, 2002.
- Philippe Rigaux, Michel Scholl, Agnès Voisard, *Spatial Databases : With Application to GIS*, San Francisco, etc., 2002.

### **Séminaire du directeur de recherche**

- **Volume horaire** : 25h
- **ECTS** : 5

À choisir parmi les séminaires proposés par l'École pratique des hautes études, l'École nationale des chartes ou l'École normale supérieure.

### **Enseignement d'ouverture**

- **Volume horaire** : 10h
- **ECTS** : 2

À choisir parmi les cours et séminaires proposés par l'École pratique des hautes études, l'École des hautes études en sciences sociales, l'École nationale des chartes ou l'École normale supérieure, ainsi que les PSL weeks, l'UE sport, etc. (cf. les possibilités décrites pour les options de M1).

### **Mémoire – bilan d'étape – 5 ECTS**

- **ECTS** : 5

Évaluation de l'avancement du mémoire à l'issue du premier semestre 2, par les directeurs du mémoire. Validation du plan, des recherches bibliographiques et des approches expérimentales. Note sur 20.

## Semestre 4 : Enseignements fondamentaux – 7 ECTS

Le quatrième semestre propose des approfondissements de la spécialisation engagée par l'étudiants, au travers d'options de spécialisation et de séminaires de recherche. Il est **avant tout consacré à l'écriture du mémoire de recherche.**

### Apprentissage machine (ENC 2321)

- **Enseignant** : Daniel Stockholm (EPHE)
- **Mode d'évaluation** : devoir sur table.
- **Volume horaire** : 10h
- **ECTS** : 2

Ce module propose une introduction aux concepts et aux outils d'apprentissage machine.

#### Éléments bibliographiques

- Ian Goodfellow, Yoshua Bengio et Aaron Courville, *Deep Learning*, MIT, 2016, en ligne : <http://www.deeplearningbook.org/>.

### Algorithmique et programmation pour les humanités : Python avancé (ENC 2320)

- **Enseignant** : Chahan Vidal-Gorène
- **Mode d'évaluation** : contrôle continu (1/2) et devoir maison final (1/2)
- **Volume horaire** : 20h
- **ECTS** : 4

Ce module a pour objet de donner des connaissances avancées en Python pour la science des données, traitant en particulier de méthodes et bibliothèques utiles pour l'apprentissage machine (réseaux de neurones...) aux diverses applications (traitement automatique des langues, traitement d'image, etc.). Ce cours est commun avec les M1. Le programme alterne une année sur deux.

#### Détails sur le mode d'évaluation

$$\text{Note finale} = (A + B) / 2$$

où A (note sur 20) = Somme des 4 meilleures notes des devoirs à rendre au cours du semestre (6 à rendre). Chaque devoir produit une note sur 5, chaque devoir non rendu équivaut à une note de 0.

#### Éléments bibliographiques

- MÉLÉARD, S., *Aléatoire, introduction à la théorie et au calcul des probabilités*, Palaiseau, École Polytechnique, 2010.
- HASTIE Trevor, TIBSHIRANI Robert, FRIEDMAN Jérôme, *The Elements of Statistical Learning*, Springer, 2009.

### Séminaire de recherche du master

- **Coordinateurs** : J.B. Camps, Marie Puren, et les enseignants du master.
- **Mode d'évaluation** : participation et discussion (M1) ; présentation de 15 minutes (M2)
- **Volume horaire** : 5h
- **ECTS** : 1

Ce séminaire, qui s'étend sur les quatre semestres de la formation, permet aux étudiants de première année de découvrir les recherches les plus actuelles dans le domaine des humanités numériques, notamment celles menées par des étudiants plus avancés. Aux étudiants de deuxième année, elle fournit également l'occasion de confronter leurs expériences et leurs idées, et de s'initier à la présentation de leurs travaux.

Le séminaire est évalué de la sorte :

- pour les étudiants de M2, ils doivent une fois dans l'année réaliser une présentation de leurs recherches de 15 minutes;
- pour les étudiants de M1, ils doivent endosser une fois dans l'année le rôle de discutant d'une des présentations, et animer les débats autour de celle-ci.

### Éléments bibliographiques

Les compte-rendus des années précédentes sont disponibles sur le carnet de recherche du séminaire :

- Master HN, *Données et sciences humaines : Carnet de recherche du master Humanités numériques de l'Université Paris, Sciences & Lettres*, dir. Jean-Baptiste Camps & Marie Puren, <https://humanodon.hypotheses.org/>

## Semestre 4 – Enseignements optionnels – 10 ECTS

### Spécialisation de SHS Computationnelles (au choix) – 5 ECTS, au choix parmi les trois suivants

#### Traitement automatique de la langue et analyse sémantique (ENS N104)

- **Enseignants** : Thierry Poibeau, Frédérique Mélanie, Clément Plancq (ENS, LATTICE); Carmen Brando (EHESS).
- **Mode d'évaluation** : devoir à rendre en rapport avec le contenu du cours (compte-rendu critique d'un article, petit programme centré sur une tâche de TAL comme la reconnaissance des entités nommées, etc.). La nature exacte du travail à rendre devra faire l'objet de l'accord préalable du responsable du cours.
- **Volume horaire** : 25h
- **ECTS** : 5

Le but de ce cours est de montrer en quoi le traitement automatique des langues peut être utile dans le cadre des humanités numériques. Les séances seront essentiellement pratiques : une large part sera consacrée à la présentation d'outils et à des manipulations simples par les participants, même si les bases théoriques du domaine ne seront pas oubliées. Le cours portera en premier lieu sur la manipulation de corpus (analyse, annotation, extraction d'information) à des fins d'analyse linguistique, littéraire ou de stylistique. Les données et les corpus considérés seront déterminés en fonction des objectifs de la séance et de l'intérêt des participants.

### Éléments bibliographiques

- Dan Jurafsky et James H. Martin. *Speech and Language Processing*. 3e édition en cours, disponible en téléchargement en ligne : <https://web.stanford.edu/~jurafsky/slp3/>. (C'est sans conteste la meilleure introduction au domaine, constamment mis à jour et très pédagogique, en accès libre de surcroît).

### Traitement automatique de l'image (EPHE P102)

- **Enseignant** : Daniel Stockholm (EPHE)
- **Mode d'évaluation** : à définir.
- **Volume horaire** : 25h
- **ECTS** : 5

Ce module présentera des outils et notions utiles au traitement automatique de l'image, depuis des opérations élémentaires (échantillonnage et quantification, transformation en niveau de gris, binarisation), jusqu'à la détection de contours ou régions ou les opérateurs morphologiques. Chaque séance donnera lieu à un travail pratique : un exercice de programmation appliqué à l'image.

### Éléments bibliographiques

- Diane Lingrand, *Introduction au traitement d'images*, 2e édition, Paris, 2008.

### Analyse de réseaux (ENC 2324)

- **Enseignant** : NN (EPHE)
- **Mode d'évaluation** : à définir.
- **Volume horaire** : 25h
- **ECTS** : 5

Ce module présente les concepts et outils nécessaires à l'analyse de données relationnelles et de réseaux sociaux, dans des perspectives historiques, sociologiques, ethnologiques, etc.

### Séminaire du directeur de recherche

- **Volume horaire** : 26h
- **ECTS** : 5

À choisir parmi les séminaires proposés par l'École pratique des hautes études, l'École nationale des chartes ou l'École normale supérieure.

### Mémoire et soutenance – 15 ECTS

- **ECTS** : 15

Soutenance (voir détail du mémoire p. 19)