

INTITULE DE LA MATIERE : **STRUCTURATION ET INTEROPERABILITE DES DONNEES**

CODE : UE-53

SEMESTRE : 5

NOM DE L'ENSEIGNANT RESPONSABLE : P. BELLOT

NOMBRE D'HEURES ENCADREES CM : 12 TD : 12 TP : 16

NOMBRE D'HEURES DE TRAVAIL PERSONNEL DE L'ELEVE – estimation : 25

OBJECTIF GENERAL (pas plus de 5 lignes) :

Ce cours s'intéresse à la question de la modélisation, de la représentation, de la description et de la préparation des données brutes avant traitement et/ou stockage dans une base de données, dans un contexte de données hétérogènes, multimodales ou issues d'objets variés. L'objectif est de comprendre que pour être (ré)utilisables, interopérables, faciles à trouver et accessibles, les données doivent être décrites, respecter des normes et être indépendantes des applications qui les utilisent. La question du partage et de l'ouverture des données (données FAIR, OpenData) est centrale.

CONTENU – PLAN DU COURS (pas plus de 15 lignes) :

- Des données brutes aux bases de données : introduction à la modélisation des Big Data et aux différents modèles de données (statiques ou dynamiques, liées ou non, en flux comme les données météo ou les tweets)
- La question de l'identification des données et de leur description par des méta-données
- Présentation des langages de structuration, d'annotation et des modèles d'accès aux données brutes (JSON, écosystème XML avec DOM, Schémas et DTD)
- Présentation des principaux standards et protocoles internationaux dans le monde de l'internet des objets (ex: JSON-LD, Web Thing Model, Sensor ML...) et des données GPS (GPX)
- Des licences pour comprendre la disponibilité des données (ex. Creative Commons)
- Le plan de gestion des données (but de la collecte, modes d'utilisation, processus de mise à jour, de distribution ou de référencement...)
- Les entrepôts de données ouvertes (exemple de Dataverse)
- Les modèles de processus orientés données
- Expérimentation avec des données issues d'objets connectés (capteurs, unités GPS...) et les standards SensorML, GPX (projet OpenStreetMap)
- Travaux pratiques avec l'environnement logiciel Oxygen

FORME DE L'EVALUATION :

Contrôle continu Contrôle terminal Mémoire/rapport Soutenance
préciser si nécessaire, nombre d'épreuves : et type oral / écrit

ACQUIS DE LA FORMATION ATTENDUS, ET QUI SONT EVALUES (5 à 10 items)

A l'issue de cet enseignement, l'élève est capable de :

Utilisez les termes comprendre, analyser, exploiter, interpréter, calculer, utiliser, mettre en œuvre, choisir, communiquer, rédiger,...

- Choisir les standards internationaux qui représentent le mieux les données manipulées et les objectifs à atteindre
- Rédiger un guide d'annotation des données manipulées (par exemple dans une perspective d'apprentissage automatique ou de fouille de données)
- Comprendre les principes des données ouvertes
- Préparer un plan de gestion des données

PREREQUIS (pas plus de 5 lignes): Aucun

REFERENCES, BIBLIOGRAPHIE (pas plus de 5 lignes) :

- XML: Cours et exercices. Modélisation, Schémas et DTD..., A. Brillant, Eyrolles, 2010
- Open Data - Consommation, traitement, analyse et visualisation de la donnée publique, J.P. Goulgoux, Editions ENI, 2016
- SIG - La dimension géographique du système d'information, H. Pornon, Dunod, 2015
- Introduction to JavaScript Object Notation: A To-the-Point Guide to JSON, L. Bassett, O'Reilly, 2015