Institut de Recherche
en Informatique de Toulouse
CNRS UMR 5505 - Université de Toulouse

Offre de Stage Chaine d'acquisition et alignement de données multimodales pour l'observation des territoires

Mots clefs Séries temporelles, Alignement de données, Intégration, Fusion, multimodalité.

Finalité Ce stage a pour objectif la création de la chaîne de valorisation des données d'un observatoire environnemental multimodal. Il s'intègre dans le projet AUTAN, qui consiste à construire un observatoire environnemental constitué d'un réseau de capteurs environnementaux et intégrant des données de multiples sources (données ouvertes, campagnes participatives, etc.).
À terme, cet observatoire ambitionne de pouvoir contribuer à un **jumeaux numériques de territoire**, apporter des outils pour la surveillance et la modélisation de phénomènes environnementaux comme le *retrait-gonflement des argiles* (RGA), l'impact du vent d'Autan sur la *propagation du radon* produit par le massif granitique du Sidobre, ou encore la *recommandation agricole* pour l'optimisation des cultures.

Sujet du stage

Le territoire de Castres présente plusieurs caractéristiques qui permettraient la constitution d'un observatoire environnemental multimodal. C'est un territoire qui est observé par des études participatives régulières, réalisées notamment par des acteurs associatifs. De plus, il dispose d'un réseau sans fil à longue portée (Lorawan) permettant la construction d'un réseau de capteurs environnementaux (ex: température, pression, hydrométrie, direction et vitesse du vent). Enfin, le bassin de Castres, où cohabitent des milieux urbanisés et ruraux, rend possible des co-observations soumises aux mêmes influences environnementales.
Un tel observatoire est ainsi multimodal sur 3 dimensions : les données sont récoltées longitudinalement (séries temporelles) et sont de différents types (température, pression, hydrométrie, direction et vitesse du vent, radioactivité, évolution géologique et naturaliste…) ; de plus, elles sont d'origine et de sémantique différentes (réseau de capteurs local, récolte citoyenne, données ouvertes, administratives ou satellitaires) ; enfin, ces données portent sur différents milieux observés (urbain et rural) d'un même territoire.

La mise en place d'une chaîne de valorisation des données (récolte, nettoyage, intégration) est essentielle pour donner du sens aux différentes données récoltées et aux analyses qui s'y rapportent. L'objectif de ce stage est de proposer une méthodologie d'intégration et d'alignement temporel/géographique des données pour pouvoir ensuite les publier dans des entrepôts de données ouvertes, et produire des outils de consultation analytique pour les acteurs du territoire.
L'intégration de données multimodales, de granularités temporelle et géographique différentes est une tâche difficile. La fragmentation des données, intrinsèque aux collections de données d'origine différente, est un problème ancien [1] qui repose sur des approches de *data fusion*. Les approches expérimentées mettent en avant les difficultés de modélisation des traitements sans avoir initié l'intégration des données [2] qui doivent au préalable être alignées/appriés. De plus, la fusion de données multimodales est peu explorée quand le nombre de modalités est important [3], et les techniques investiguées s'appliquent mal à la diversité des données multimodales [4]. Malgré les évolutions des approches d'apprentissage profond, celles-ci restent limitées sur leur performance et leur généralisation [5].

Le stage se découpera selon les différentes tâches suivantes :
- Mise en place du réseau de capteurs environnementaux et de la chaîne de valorisation des données.
- Exploration d'approches d'alignement par recherche de variations causées par un même événement (via analyse des données fonctionnelles ou modèle de processus latent).
- Publication des données récoltées et éditorialisées dans des entrepôts de données ouvertes.

Institut de Recherche
en Informatique de Toulouse
CNRS UMR 5505 - Université de Toulouse

Profil des candidat·e·s Étudiant·e en cycle master ou ingénieur, en Informatique ou mention similaires.

Capacités attendues : En complément du niveau de formation défini dans le profil, les éléments suivants sont souhaitables dans le profil des candidats :

- Capacité de réflexion et d'analyse d'un problème, définition d'un protocole d'expérimentations ;
- Rigueur en programmation (Python)
- Capacité à générer de la documentation, avec usage des outils standard (Git, Doctest etc.) ;
- Maîtrise des outils et techniques de l'apprentissage automatique ;
- Capacités en administration de système (VM, déploiement)
- Capacités de synthèse d'information et à rendre compte de son travail régulièrement ;
- Capacité à produire efficacement des documents de communication scientifique.
- Des capacités en réseau et telecom, ou avec les outil de visualisation (grafana) sont bienvenues.

Langues : Maîtrise du Français (Niveau A2), conversationnel et écrit ; Maîtrise de l'anglais scientifique, écrit. **Des certifications attestant votre niveau en langues sont vivement encouragées.**

Conditions de réalisation du stage Le stage sera réalisé au sein de l'**Institut de recherche en Informatique de Toulouse** (IRIT) sur son site de **Castres** (Tarn).

Equipe de recherche : SIG - Systèmes d'Informations Généralisés
Dates & Durée : 6 mois à partir de Février 2027 - Rémunération au tarif légal en vigueur
Lieu du réalisation : IUT de Castres, 5 allée du Martinet 81100 CASTRES
Modalité : Présentiel sur le site de Castres **exclusivement**.

Important

L'IRIT est soumis au régime des **zones à régime restrictif (ZRR)** dans le cadre de la *protection du potentiel scientifique et technique national*. A ce titre, le processus de recrutement peut-être allongé le temps que votre profil soit validé par les autorités compétentes.

Processus de recrutement Envoi par courrier électronique, aux coordonnées ci-dessous, d'un dossier de candidature comprenant :

- un *Curiculum Vitae* (sans période manquante) comportant **absolument** votre adresse postale, votre date et lieu de naissance, et une photo ;
- une copie d'une pièce d'identité valide ;
- une **lettre de motivation**[1] décrivant votre parcours, votre projet professionnel, et les raisons de votre candidature ;
- un **portfolio** (*github* ou équivalent).

A l'issue de votre candidature un ou plusieurs entretiens en visioconférence pourront vous être proposés. Ces entretiens peuvent impliquer des tests techniques. **Seules les candidatures sérieuses seront considérées.**

Date limite de candidature : 31 Décembre 2026

Contact

Dr Hugo Boisaubert (hugo.boisaubert@irit.fr)
IRIT - Equipe SIG (Systèmes d'Informations Généralisés)}
IUT de Castres - 5 allée du Martinet - 81100 CASTRES

[1] Les lettres générées par un outil d'Inférence Automatique (IA) sont à proscrire.

Bibliographie

[1] H. G. Miller and P. Mork, "From data to decisions: a value chain for big data," *It Professional*, vol. 15, no. 1, pp. 57–59, 2013.

[2] S. K. Shankar, M. J. Rodríguez-Triana, A. Ruiz-Calleja, L. P. Prieto, P. Chejara, and A. Martínez-Monés, "Multimodal Data Value Chain (M-DVC): A Conceptual Tool to Support the Development of Multimodal Learning Analytics Solutions," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 15, pp. 113–122, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:218597530

[3] R. Bokade *et al.*, "A cross-disciplinary comparison of multimodal data fusion approaches and applications: Accelerating learning through trans-disciplinary information sharing," *Expert Systems with Applications*, vol. 165, p. 113885, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113885.

[4] D. Lahat, T. Adali, and C. Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015, doi: 10.1109/JPROC.2015.2460697.

[5] L. Zhao, S. Huang, and J. Li, "Multimodal Data Fusion Research Review: From Traditional Methods to Unsupervised Learning Methods," in *Proceedings of the 2025 2nd Symposium on Big Data, Neural Networks, and Deep Learning*, in BDNNDL '25. Jinan, China: Association for Computing Machinery, 2026, pp. 249–254. doi: 10.1145/3784013.3784052.