

**DEMANDE D'ALLOCATION DOCTORALE DE RECHERCHE DE L'ED SIE**  
**Année universitaire 2022/2023**  
**SUJET DE THESE**

<p><b>1. LABORATOIRE</b></p> <p>Nom ou sigle : LISTIC Statut :EA</p>	<p><b>2. DIRECTION DE THÈSE</b></p> <p>Directeur de thèse (HDR) : Flavien VERNIER Codirecteur éventuel : Hervé VERJUS</p>
<p><b>Laboratoire partenaire ou collaborations éventuels :</b></p>	
<p><b>3. SUJET DE THÈSE</b></p>	
<p><b>Titre : Conception et déploiement d'une architecture distribuée pour la mise en œuvre de chaînes de traitements scientifiques autoadaptatives : application aux chaînes de traitement EFIDIR</b></p>	
<p><b>4. RESUME</b> (Français et Anglais)</p>	
<p>Les workflows sont présents dans tous les domaines traitant des données. Ils s'apparentent à une succession d'étapes, chacune pouvant être une unité (un service) de traitement sur les données. On retrouve ainsi des chaînes de traitements scientifiques aux chaînes d'IA en passant par les chaînes de process industriels. Ces chaînes peuvent être complexes, chronophages, sensibles à leur configuration et nécessitent par conséquent une connaissance experte généralement acquise par l'expérience.</p> <p>Dans le cadre de chaînes de traitements scientifiques, la place de l'IA est de plus en plus importante: les chaînes de traitement d'images, de vidéos incorporent de lourds traitements (de reconnaissance par exemple). Ces chaînes de traitement reposent sur l'utilisation d'infrastructures lourdes et complexes (plateformes/centres de calculs, plateformes/centres de données). Il y a de nombreux domaines et de nombreuses applications pour lesquelles les chaînes de traitement sont une solution et tout particulièrement, toutes les applications qui traitent de gros volumes de données. Ces chaînes de traitement sont conçues de manière ad hoc : elles sont peu réutilisables d'une application à une autre, peu adaptables et sensibles (dépendantes) à la plateforme de leur déploiement.</p> <p>L'objectif de la thèse consiste à introduire la notion d'objet sage comme dispositif conceptuel et technologique support au développement et au déploiement des chaînes de traitement. Les objets sages intègrent des mécanismes d'auto-introspection, d'auto-analyse basée sur de l'IA, permettant d'envisager des chaînes de traitement "adaptables" (smart)</p> <p>Le contexte d'expérimentation de ce travail concerne les chaînes de traitement EFIDIR utilisées pour le traitement d'images (satellites ou terrestre, SAR ou optiques).</p> <p><i>The workflows are present in all domains dealing with data. They are defined as a succession of tasks, each one being a unit (a service) of data processing, they can be scientific processing chains, AI chains, industrial process chains... These chains can be complex, time-consuming, sensitive to their configuration and therefore require expert knowledge acquired through experience.</i></p>	

*In the context of scientific processing chains, the place of AI is increasingly important: images and video processing chains incorporate heavy processing (recognition for example). These processing chains rely on the use of heavy and complex infrastructure (platforms/computing centers, platforms/data centers). There are many domains and applications for which processing chains are a solution, and especially all applications that process large volumes of data. These processing chains are designed in an ad hoc manner: they are not very reusable from one application to another, not very adaptable and sensitive (dependent) to the platform of their deployment.*

*The objective of this thesis is to introduce the notion of wise object as a conceptual and technological device supporting the development and deployment of processing chains. The wise objects integrate mechanisms of self-learning, self-analysis based on AI, allowing to consider "adaptable" (smart) processing chains.*

*The experimental context of this work concerns EFIDIR processing chains used for SAR and optical image processing.*

## 5. PROJET DE RECHERCHE DÉTAILLÉ

(2 pages environ)

La notion de workflow apparaît dans la première moitié des années 1990 avec l'arrivée des technologies de l'information et de la communication. Il s'agissait alors d'un enchaînement ordonné d'étapes traitant de données de nature administrative (ex. des documents numériques) qui bien souvent impliquent la participation d'utilisateurs; les workflows ont été expérimentés puis mis en œuvre comme support à des processus administratifs standardisés (aujourd'hui ils sont connus comme permettant la plupart des procédures et processus dématérialisés). Un environnement de workflow permet de modéliser le processus et de l'exécuter de façon automatique ou semi-automatique. Il est courant d'utiliser une plateforme de type GED (gestion électronique de documents) pour stocker de façon centralisée, les documents qui transitent et qui sont traités par les différentes étapes du workflow.

Les workflows sont désormais présents dans tous les domaines traitant des données. Ils s'imposent aussi, et de plus en plus, dans les domaines scientifiques et domaines pour lesquels de gros volumes de données autres que des documents administratifs sont soumis à des traitements (algorithmes) lourds nécessitant de la puissance de calcul. On les retrouve ainsi dans le cas des chaînes de traitements scientifiques, dans le cas des chaînes d'IA en passant par les chaînes de processus industriels. Ces chaînes peuvent être complexes, chronophages, sensibles à leur configuration et nécessitent par conséquent une connaissance experte généralement acquise par l'expérience.

Dans le cadre de chaînes de traitements scientifiques, la place de l'IA est de plus en plus importante : les chaînes de traitement d'images et de vidéos incorporent de lourds traitements (de reconnaissance par exemple). Ces chaînes de traitement reposent sur l'utilisation d'infrastructures lourdes et complexes (plateformes/centres de calculs, plateformes/centres de données). Il y a de nombreuses applications qui utilisent et qui implémentent ces chaînes de traitement (industrie, astrophysique, santé, biologie météorologie, surveillance des phénomènes naturels...). Dans de nombreux cas, ces chaînes de traitement sont conçues de manière ad hoc et répondent à un besoin particulier d'une application particulière. Elles sont assez peu réutilisables et transférables d'une application à une autre, d'un centre de calcul à un autre, alors que les efforts et ressources pour mettre en place de telles chaînes sont lourds et coûteux. Elles sont aussi peu évolutives, peu "plastiques".

L'objectif du sujet de thèse que nous proposons est de s'attacher à l'étude de ces chaînes de traitement par la construction de méthodes et de technologies pour faciliter leur mise en œuvre, le suivi de leurs comportements (on peut parler d'introspection) afin de les rendre plus flexibles et plus adaptées à leurs usages. L'idée de départ de cette thèse est d'incorporer de l'IA au sein de l'architecture même de la chaîne de traitement afin que la chaîne de traitement ait une certaine capacité d'autoraisonnement, d'autoévolution/adaptation. La proposition que nous faisons est d'utiliser les objets sages (wise object) et l'intergiciel Java « Wise Object Framework » (WOF) comme technologies de support à la mise en œuvre des services de traitement et des mécanismes d'orchestration de ces services au sein des chaînes de traitement. L'intérêt des objets sages est leur capacité de raisonnement et d'autoapprentissage, leurs mécanismes d'introspection qu'il serait intéressant de coupler aux services de traitement. Ainsi, chaque service de traitement serait en soi une entité (service) sage disposant de capacités d'auto-analyse et de raisonnement sur ses usages et son environnement.

Cette thèse a pour vocation de faciliter le travail de l'expert concepteur et utilisateur de chaînes de traitement en proposant des workflows capables d'apprendre par eux-mêmes sur leurs caractéristiques et comportements internes : complexités, tâches chronophages, dérives d'exécution, sous-utilisation, surconsommation...

Dans le cadre de cette thèse, nous nous focaliserons sur les workflows scientifiques et plus particulièrement sur ceux décrits via le formalisme EFIDIR. EFIDIR étant un ensemble d'outils de traitements d'images qui est utilisé dans de nombreux projets (ex. analyse d'images radar satellitaires pour l'étude des déplacements de glaciers, terrains – analyse d'images optiques in situ...). L'ensemble EFIDIR propose un système générique de

description de workflows. L'avantage principal de ce dernier est qu'il est simple et complètement ouvert.

La chaîne de traitement dans son ensemble a vocation à constituer un système distribué (les unités de traitement peuvent être distribuées, ainsi que les données/fichiers qui sont traités). Les questions liées à la distribution de l'architecture et ses caractéristiques pourront être des éléments qu'il serait intéressant de prendre en compte dans les analyses, le raisonnement des "services sages" et la prédiction de leurs comportements. Le déploiement de ces chaînes de traitement EFIDIR pourrait s'envisager sur le meso centre MUST, chaque service de traitement serait alors déployé sur un nœud de calcul de l'infrastructure MUST. L'intérêt d'utiliser cette notion de "sagesse" (capacité de raisonnement de chaque service) devient alors évident : après avoir appris sur lui-même, chaque service pourrait donc affiner sa configuration de déploiement pour optimiser la consommation de ressources (CPU, GPU, mémoire) qu'il utilise ainsi que la topologie de déploiement (proximité des nœuds sur lesquels sont déployés les autres services avec lesquels le service initial a de fortes accointances).

Ce sujet est en lien étroit avec une problématique d'actualité qui est celle de l'apprentissage distribué ou fédéré (*federated learning*) qui un l'un des thèmes principaux du laboratoire: en effet, une chaîne de traitement scientifique où l'intelligence se déploie sur chaque entité constitutive de la chaîne devient par nature un système d'apprentissage distribué/fédéré.

Dans un premier temps, le travail de thèse va consister à étudier la manière de concevoir un service sage de traitement, c'est-à-dire à coupler les objets sages aux services de traitement. La deuxième phase envisagée pour ce travail est de proposer des analyses sur le comportement de ces services sages et leur environnement afin de pouvoir suggérer des adaptations de l'orchestration de ces services de traitement, des adaptations dans l'architecture de déploiement. La troisième phase aura pour objectif d'expérimenter plusieurs configurations de déploiement (dont le déploiement distribué) et de mettre en œuvre les analyses précédemment proposées ainsi que les propositions d'adaptation qui en découleront.

Un travail de revue de littérature sera conduit lors de la première année de thèse essentiellement pour appréhender les architectures existantes des chaînes de traitement et les approches et méthodes qui s'intéressent à leur comportement et aux questions de l (auto)adaptation et du déploiement.

[Marsy et al. 2021] Guilhem Marsy, Flavien Vernier, Emmanuel Trouvé, Xavier Bodin, William Castaings, Andrea Walpersdorf, Emmanuel Malet, et al., "Temporal Consolidation Strategy for Ground-Based Image Displacement Time Series: Application to Glacier Monitoring," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 10069-10078, 2021, doi: 10.1109/JSTARS.2021.3115231.

[Alloui et al., 2019] Alloui I., Benoit E., Perrin S., **Vernier F.** Wise Objects for IoT (WIoT): Software Framework and Experimentation. Communications in Computer and Information Science, Springer Verlag, 2019, pp. 349-371.

[Alloui et Vernier 2018] Ilham Alloui, **Flavien Vernier.** WOF: Towards Behavior Analysis and Representation of Emotions in Adaptive Systems. Communications in Computer and Information Science, Springer Verlag, 2018, Software Technologies 12th International Joint Conference, ICISOFT 2017, Revised Selected Papers, 868, pp.244-267. (10.1007/978-3-319-93641-3 12). (hal-01871009)

[Volat et al., 2014] Matthieu Volat, **Flavien Vernier**, Marie-Pierre Doin, Cécile Lasserre, Emmanuel Trouvé, et al.. Improving the execution of workflows for SAR image analysis. Conference on Big Data from Space, Nov 2014, Frascati, Italy. (hal-01487781)

**6. CANDIDAT RECHERCHE** : Détailler en quelques lignes vos besoins et les qualités du candidat recherché...

La/le candidat-e sera issu-e d'une formation de niveau master ou équivalent dans le domaine de l'informatique ou d'un domaine scientifique avec de solides bases en statiques, sciences des données, méthodes et algorithmes d'apprentissage machine (*machine learning*) et en programmation (orientée objet Java, Python). La/le candidat-e aura un intérêt particulier pour les technologies permettant le déploiement de systèmes logiciels distribués (cloud, conteneurisation).

La/le candidat-e saura rédiger de façon rigoureuse des documents structurés et aura une bonne maîtrise (écrite et orale) soit de l'anglais soit du français.

**7. FINANCEMENT DE LA THÈSE** : Le contrat doctoral fixe une rémunération principale, indexée sur l'évolution des rémunérations de la fonction publique. L'arrêté du 11 octobre 2021 modifiant l'arrêté du 29 août 2016 fixant le montant de la rémunération du doctorant contractuel - Article 1 - L'article 1er de l'arrêté du 29 août 2016 susvisé est remplacé par un article ainsi rédigé : « Art. 1.-La rémunération mensuelle minimale des doctorants contractuels est fixée ainsi qu'il suit, sous réserve des dispositions prévues à l'article 2 du présent arrêté «-contrats conclus avant le 1er septembre 2021 : 1 758 euros brut ; «-contrats conclus à compter du 1er septembre 2021 : 1 866 euros brut ; «-contrats conclus à compter du 1er septembre 2022 : 1 975 euros brut. ».

Des heures d'enseignements peuvent être effectuées dans la limite de 64 heures équivalent TD par année universitaire **après autorisation du président de l'université** et rémunérées au taux fixé pour les travaux dirigés en vigueur. D'autres activités complémentaires au contrat doctoral sont prévues par l'article 5 du décret n° 2009-464 du 23 avril 2009 modifié. La durée totale des activités complémentaires aux activités de recherche confiées au doctorant dans le cadre du contrat doctoral ne peut excéder un sixième du temps de travail annuel.

**8. CONTACT :**

Nom prénom : Vernier Flavien et Verjus Hervé

Tél. : +33(0) 4 50 09 65 90, +33 4 50 09 65 94

Email : [flavien.vernier@univ-smb.fr](mailto:flavien.vernier@univ-smb.fr) / [herve.verjus@univ-smb.fr](mailto:herve.verjus@univ-smb.fr)