

LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU
NUMÉRIQUE (LISN) - UMR9015

Projet

Johanne Cohen, Sophie Rosset

2021-2024

Table des matières

1	Avis des conseils de laboratoire	7
2	LISN : processus de construction	9
2.1	Accompagnement Ad’Hoc	9
2.2	La mise en place de la future UMR	10
3	LISN : projet scientifique	13
3.1	Introduction	13
3.2	Département <i>Algorithmes, Apprentissage et Calcul</i>	14
3.2.1	Thèmes	14
3.2.2	Positionnement	15
3.2.3	Composition	15
3.2.4	Projet	16
3.3	Département Interaction avec l’Humain	18
3.3.1	Thèmes	18
3.3.2	Positionnement	18
3.3.3	Composition	19
3.3.4	Projet	20
3.4	Département Mécanique des Fluides - Énergétique	22
3.4.1	Thèmes	22
3.4.2	Positionnement	22
3.4.3	Composition	22
3.4.4	Projet	23
3.5	Département Sciences des Données	24
3.5.1	Thèmes	24
3.5.2	Positionnement	25
3.5.3	Composition	25
3.5.4	Projet	26
3.6	Département Sciences et Technologies des Langues	28
3.6.1	Thèmes	28
3.6.2	Positionnement	28
3.6.3	Composition	29
3.6.4	Projet	30
3.7	Actions Transverses	32
3.8	Insertion dans l’écosystème de l’Université Paris-Saclay	35
4	LISN : organisation, structuration, fonctionnement	39
4.1	Instances	39
4.2	Organigramme	39
4.3	Structuration scientifique	40
4.3.1	Les départements	40
4.3.2	Les équipes	41

4.4	Structuration des services	41
4.4.1	Les cellules	41
4.4.2	Les services support	41
4.4.3	Les services soutien	42
Annexes		45
A Cellule Science Responsable		45

Préambule

Ce document regroupe l'état actuel du projet pour la future unité LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE. En l'état ce document présente tout d'abord le processus de construction et les différentes étapes (voir chapitre 2).

Le projet scientifique est présenté au chapitre 3. Ce projet a été élaboré de façon collective et est encore susceptible d'évoluer.

Le projet de gouvernance, la structuration du laboratoire et l'organigramme cible, brièvement présenté dans le chapitre 4, doivent encore être approfondie mais les grandes lignes du projet et du fonctionnement sont déjà bien dessinées.

La structuration et l'organisation des services est en cours. L'ensemble des responsables de service ont été nommés. Il est à noter qu'actuellement deux responsables de service (Pascal Desroches, infrastructure bâtiment et logistique, et Monique Granon, gestion administrative et financière) ont une double affectation LIMSI/LRI. Nous les remercions chaleureusement pour avoir accepté cette charge de travail supplémentaire. Nous remercions également les membres des quatre services (2 LIMSI et 2 LRI) concernés pour leur soutien.

Les conseils de laboratoire devaient commencer à travailler ensemble au deuxième trimestre 2020, notamment pour entamer un travail sur une première proposition de règlement intérieur. La première réunion en configuration LISN des conseils de laboratoire a eu lieu le 10 juillet. Lors de cette première réunion, les deux conseils de laboratoire ont rendu un avis sur le projet et notre candidature en tant que DU et DUA (voir section 1).

Une première réunion des directoires en mode LISN a eu lieu le 9 juillet 2020. Elle réunissait l'ensemble des responsables des équipes, services et départements. L'objet de cette première réunion était d'amorcer les discussions et réflexions concernant les demandes de moyens et de postes et de préciser les modes de fonctionnement pour l'avenir.

Une AG a été organisée le 30 juin en mode hybride. L'ordre du jour a consisté en une restitution par la cellule Ad'Hoc du travail réalisé et des entretiens individuels, une présentation de l'avancement du projet et une présentation des processus RH à venir avec un bilan du premier CoPil GT RH inter-tutelles qui s'est tenu le 19 juin.

Enfin, une journée des doctorants en mode LISN a été organisée et a été particulièrement apprécié. Nous remercions le comité d'organisation (composé de doctorants LIMSI et LRI) pour leur travail et l'ensemble des personnels pour leur participation à cette journée. Les deux chargées de communication (Bénédicte Daly et Gladys Balayoko) travaillent ensemble et ont préparé la première *newsletter* LISN.

Nous tenons ici à remercier l'ensemble des membres des deux laboratoires qui ont fortement collaboré non seulement à rédaction du document mais aussi à l'élaboration de ce projet.

Chapitre 1

Avis des conseils de laboratoire

Les deux conseils de laboratoire se sont réunis en mode LISN, c'est-à-dire sans les deux représentants de l'équipe VALS. A l'ordre du jour de ce premier CdL commun figurait l'avis à donner sur le projet et la candidature de Johanne Cohen et Sophie Rosset comme DUA et DUA du future laboratoire.

Extrait du CR du conseil de laboratoire en configuration LISN du 10 juillet 2020

Note : le conseil de laboratoire s'est tenu en mode hybride.

Présents dans la salle : Sophie Rosset, Johanne Cohen, Anne Vilnat, François Lusseyran, Valérie Séné, Marie-Christine Duluc, Anthony Pense, Paul Lerner, Nicolas Ferrey

Présents en visio : Nicolas Sabouret, Emmanuelle Frenoux, Nicolas Grenier, Nacera Seghouani, Laurence Rostaing, François Landes, Sonia Vanot, Alain Denise, Monique Granon, Philippe Chatalic, Ioana Vasilescu, Laurent Darré, Nicola Zema, Cédric Fleury, Cyril Grouin, Michael Filhol, Berengère Podvin, Paola Tubaro, Diana Baltean, Thomas Nowak, Monique Granon, Romain Di Vozzo, Philippe Chatalic, Nicolas Thiéry (deuxième partie de séance)

Excusés : Elisabeth Piotelat

Modalité : Sophie Rosset et Johanne Cohen ont quitté la salle durant les débats.

Avis du conseil de laboratoire

Les conseils de laboratoire du LIMSI et du LRI (moins l'équipe VALS), réunis en session commune le 10 juillet 2020, émettent un avis très favorable à l'unanimité sur la direction du projet LISN par Sophie Rosset et Johanne Cohen, et font pleinement confiance à cette équipe pour assurer la direction de cette nouvelle unité. Elles ont montré depuis le 1er janvier 2020 qu'elles ont toutes les compétences requises pour faire vivre ce nouveau laboratoire.

Chapitre 2

LISN : processus de construction

Le point de départ de ce projet correspond à l'annonce de la création d'une nouvelle unité par les différentes tutelles (Université Paris Sud et CNRS) du LRI¹ et du LIMSI lors d'une assemblée générale du 4/12/18. À partir de ce point de départ, un groupe de travail a été créé (proposition par les laboratoires, validation par les tutelles). Ce groupe était constitué de six personnes (G6), trois du LIMSI et trois du LRI. Nous en faisons partie. Durant les 6 premiers mois de l'année 2019, le G6 a mis en place et animé plusieurs groupes de travail afin de construire un projet impliquant les membres des deux laboratoires (IT, EC, chercheurs, non permanents). Cette animation a permis de fournir un projet scientifique construit collectivement et des documents faisant les synthèses des pratiques des deux laboratoires accompagnées de premières propositions (structuration des services, vie quotidienne du laboratoire, gouvernance). En juillet 2019, plusieurs présentations de ces différents travaux ont été réalisées devant les personnels et les différentes tutelles. Cela a permis d'avoir une validation à la fois par les personnels (sous forme de votes) et par les tutelles (sous forme de lettre de validation du projet des tutelles adressée aux personnels des deux laboratoires le 30/07/19). À partir de septembre 2019, un travail de préparation de l'année de transition (2020) a eu lieu (nomination par intérim des équipes de direction du LIMSI et du LRI, élection des conseils de laboratoire, ...). Cette préparation s'est faite en deux parties : accompagnement de la future équipe de direction, à savoir nous-même, par la cellule Ad'Hoc de l'OMES (RH du CNRS) et préparation de la mise en place du LISN.

La frise dans la figure 2.1 récapitule les différentes étapes.

2.1 Accompagnement Ad'Hoc

La cellule ad'hoc de l'OMES (Observatoire des Métiers et de l'Emploi Scientifique) du CNRS a débuté en septembre. Lors d'une assemblée générale, le 24 septembre 2019, une présentation de l'accompagnement proposé par la cellule a été réalisée. Nous avons rencontré deux membres de cette cellule de façon régulière (Isabelle Fraichou, Muriel Perier), entre octobre 2019 et février 2020. Les échanges ont été larges et fructueux. Les discussions ont porté tout d'abord sur le lien entre services et projet scientifique, puis ont successivement abordé le projet scientifique via la présentation des différents départements scientifiques, la vision du projet pour le LISN, et la préparation de l'organigramme cible. Nous profitons de l'occasion qui nous est donnée ici pour les remercier de leur accompagnement et de leur aide efficace dans la préparation de ce projet. Lors de l'assemblée générale du 7 février 2020, en présence des tutelles (CNRS et université) et services RH de la délégation et de l'université, l'organigramme cible a été présenté. De février à avril (malgré le contexte difficile), Mme Fraichou et Mme Perier ont rencontré les collègues sur la

1. En règle générale, dans ce document, lorsque nous employons LRI il s'agit du LRI moins l'équipe VALS qui rejoint le futur LMF qui sera créé à en janvier 2021 à partir de VALS et du LSV

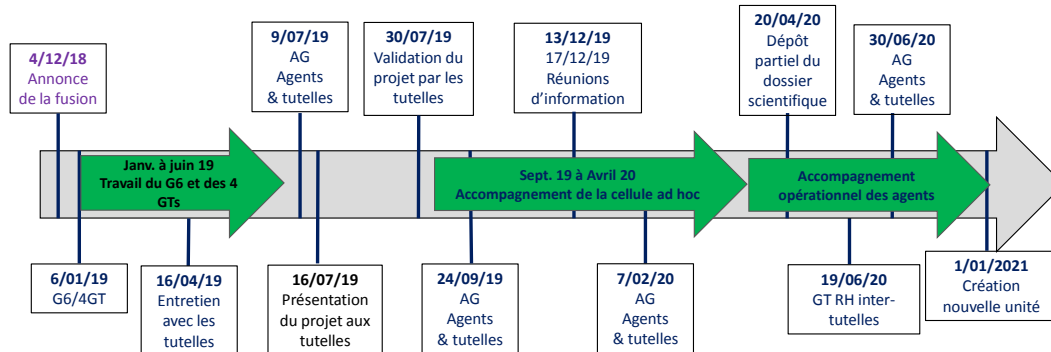


FIGURE 2.1 – Frise chronologique récapitulant les différentes étapes de la création du LISN avec un planning prévisionnel.

base du volontariat et en complet anonymat. Ces rencontres étaient ouvertes à tous les permanents des deux laboratoires afin de discuter de leur projection dans le LISN.

2.2 La mise en place de la future UMR

Dès septembre 2019, nous avons voulu initier une collaboration entre les différents services et les différentes équipes du LIMSI et du LRI. Pour cela, deux commissions ont été créées : la commission “*SI/WEB*” et la commission “*Nom*”. La commission “*SI/WEB*” travaille actuellement à la mise en place d’un système d’information tirant partie du meilleur existant dans chacun des laboratoires. Cette commission est organisée en deux sous-groupes, l’un consacré à la définition d’un cahier des charges pour le site web et l’autre aux aspects techniques pour la mise en place du système d’information.

Sur le plan scientifique, nous avons initié dès novembre 2019 la proposition et le développement d’actions transversales de recherches impliquant fortement des équipes des différents sites. Nous avons profité de l’appel unique du CNRS de 2020 et des financements internes d’appel à projet pour impulser ces différents rapprochements en favorisant les projets de recherche inter-site. Nous avons également favorisé les liens scientifiques avec des membres de la future unité LMF (qui sera composé de l’équipe VALS de l’actuel LRI et du LSV) et des membres du LABORATOIRE INTER-DISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE. Du côté du LIMSI, il y a eu la mise en place des trois départements (qui par ailleurs avait été prévue dans le projet hors fusion du LIMSI), *Mécanique des Fluides-Énergétique*, *Interaction avec l’Humain* et *Sciences et Technologies des Langues*. Du côté du LRI, il y a eu la mise en place d’une structuration en départements ce qui a représenté un très gros travail pour le conseil scientifique, le LRI n’ayant jamais fonctionné avec des départements.

Chaque étape de ce processus s’est effectué dans un esprit de transparence. Par exemple, pendant les six premiers mois de l’année 2019, la participation aux différents groupes de travail s’est faite sur la base du volontariat. Les comptes rendus de réunions étaient publics. Afin de permettre

une communication aussi fluide que possible sur les avancées du travail, un point d'information bimensuel était organisé par le G6. Le G6 a par ailleurs rencontré durant ce premier semestre 2019 chaque service et chaque équipe de recherche impliquée dans la future unité.

De septembre à décembre, nous avons organisé deux réunions d'informations afin d'expliquer le travail réalisé avec la cellule ad'hoc, les actions transversales et l'ensemble du processus en cours. En janvier, nous avons pris les fonctions de directrices d'unités par interim. Afin de pouvoir créer l'organigramme cible, nous avons rencontré les différents responsables de services et avons participé à la commission SI/WEB. Chaque réunion de conseil de laboratoire, tant au LIMSI qu'au LRI, ainsi que chaque réunion du conseil scientifique (LRI) et du directoire (LIMSI) sont également l'occasion de présenter un point d'avancement des travaux.

Le document scientifique a été rédigé, sur la base du document scientifique rédigé par le GT sciences de 2019, par les deux conseils scientifiques. Notons que les deux conseils scientifiques se sont réunis le 7 avril 2020 (par visio) pour discuter de l'organisation en département et de la gouvernance du LISN, que la première réunion des directoires et des conseils de laboratoire en mode LISN ont eu lieu les 9 et 10 juillet 2020.

La commission SI/Web a pour objectif de définir un cahier des charges pour le SI de la future unité en s'appuyant sur un état des lieux de chacun des laboratoire. Participent à cette commission les utilisateurs (le responsable RH de chaque laboratoire, le responsable administratif de chaque laboratoire, le chargé de communication de chaque laboratoire) et les développeurs (les responsables des services "moyens informatique communs et infrastructure informatique" des deux laboratoires, et les ingénieurs spécialistes des aspects SI de chaque laboratoire). Le travail est toujours en cours et s'appuie sur le recrutement de deux ingénieurs en CDD financé par l'institut.

Créée en décembre 2019 sur sollicitation des directions des laboratoires LIMSI et LRI, la commission "Nom" est composée de dix membres motivés par le choix du nom de la future unité de recherche. Afin d'inclure toute la communauté et de créer l'appropriation, la commission a mis à disposition une adresse email dédiée, un fichier partagé (framacalc ayant recueilli 153 propositions), 3 murs de créativité, et organisé trois réunions de type "Café nom" pour informer, écouter et recueillir les avis et idées, tenir compte des inquiétudes, voire les lever. Progressivement, au cours de réunions internes à la commission, des short-lists émergeaient, communiquées par mail aux personnels pour favoriser leur familiarisation et inviter aux réactions, jusqu'à ce qu'une dernière liste de 11 noms soit soumise à consultation, via un sondage LimeSurvey en avril 2020 (8-12 avril 2020). L'analyse des résultats s'est basée sur le jugement majoritaire. Plusieurs membres ayant souhaité qu'un deuxième tour ait lieu, un vote final (15-19 avril 2020) a été organisé pour départager les deux noms arrivés largement en tête. Le nom finalement choisi, et validé par nos tutelles, est arrivé largement en tête (68% de participation sur 347 inscrits, 78.9% pour LISN).

Chapitre 3

LISN : projet scientifique

3.1 Introduction

Le futur LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE (LISN), unissant les membres des laboratoires LRI, UMR 8623 (moins l'équipe Vals) et LIMSI, UPR 3251, a la vocation de devenir un centre de recherche pluridisciplinaire d'envergure mondiale, couvrant un large spectre de recherches fondamentales et appliquées en sciences du numérique. Ce centre relève au premier chef des sections 6-7-10-27-36 du CNRS (rattachement primaire INS2I ; rattachements secondaires INSIS, INSHS et INSB), et 16-27-60-61-62 du CNU. Il participe aux Graduate Schools *Informatique et Sciences du Numérique*, *Ingénierie et Systèmes* et *Sport, Mouvement et Facteurs Humains* de l'Université Paris-Saclay.

Ses membres sont des chercheurs et enseignants-chercheurs de cultures et d'horizons divers, appartenant aux sciences du numérique, aux sciences pour l'ingénieur (section 10 du CNRS), ainsi qu'aux sciences humaines et sociales, sciences cognitives et sciences du mouvement, ce qui inclut les sciences du langage (section 34 du CNRS), la sociologie et les sciences du droit (section 36 du CNRS) et la psychologie (section 16 du CNU). Les membres du LISN interagissent fortement avec les autres disciplines de l'Université Paris-Saclay, essentiellement les mathématiques, la physique, la biologie et la santé.

Les forces de recherche du LISN couvrent d'une part des thématiques cœur des sciences du numérique et des sciences de l'ingénieur, et d'autre part des thématiques interdisciplinaires par nature : intelligence artificielle et science des données, interaction humain-machine, traitement automatique des langues et de la parole, et bio-informatique. Son rayonnement lui permettra d'agréger, en temps voulu, les équipes ou unités de son périmètre scientifique dans le cadre de rapprochements d'intérêt mutuel.

Le projet présenté est construit par les équipes ou groupes des deux laboratoires cités, LRI-sans-Vals et LIMSI. Les projets scientifiques de ces équipes, tels que décrits dans leurs rapports HCERES (2018), restent d'actualité. Par ailleurs, les rapprochements des équipes dans les derniers 18 mois ont permis – au terme d'un travail important de discussion et de prospective scientifique – de faire émerger cinq départements qui positionnent le LISN de façon unique :

- Les départements *Sciences et Technologies des Langues* et *Interaction avec l'Humain* constituent d'ores et déjà des pôles de recherche et d'enseignement de visibilité internationale dans leurs domaines interdisciplinaires respectifs ;
- Les départements *Science des Données et Algorithmes*, *Apprentissage et Calcul* se constituent à l'occasion de la création de ce laboratoire et attaquent des problématiques complémentaires couvrant un large spectre de l'informatique. Ils permettent ainsi de mettre en valeur le potentiel recherche du site dans les domaines de la science de données, des calculs et des algorithmes ;
- Enfin, le département *Mécanique des Fluides-Énergétique* développe une large gamme d'approches méthodologiques innovantes pour la simulation, l'analyse, l'estimation et le contrôle

des écoulements en combinant théorie et expérimentation dans le contexte des sciences de l'ingénieur.

Le LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE a pour ambition, non seulement de contribuer à l'avancement des connaissances, mais aussi de répondre à des défis sociétaux multiples de la Stratégie Nationale de Recherche (SNR). Les domaines applicatifs principaux du département *Mécanique des Fluides-Énergétique* concernent l'efficacité énergétique (Défi 2, *Une énergie propre, sûre et efficace*) et la conception de moyens de transports plus sûrs et moins gourmands en ressources (Défi 6, *Transports et systèmes urbains durables*). Les autres départements s'inscrivent principalement dans le Défi 7 (*Société de l'information et de la Communication*), dont les quatre orientations sont couvertes : les départements *Science des Données*, *Sciences et Technologies des Langues*, *Algorithmes*, *Apprentissage et Calcul* abordent tous le traitement automatisé de grandes masses de données ; le département *Science des Données* traite également des réseaux de cinquième génération et des objets connectés ; le département *Interaction avec l'Humain* étudie notamment la conception d'interfaces collaboratives entre humains et machines. Les questions de traitement de masses de données sont également présentes dans le Défi 4 *Santé et bien-être* pour la bio informatique (*Science des Données*) et le TAL bio-médical (*Sciences et Technologies des Langues*), et dans le Défi 8 *Sociétés innovantes, intégratives et adaptatives* pour l'extraction de connaissances, ainsi que dans le programme d'actions *Big Data* (*Science des Données*). Au-delà de la Stratégie Nationale de Recherche, le LISN aborde d'autres thématiques répondant à des enjeux de société importants, tels que l'aide à l'apprentissage et à l'acquisition de connaissances, l'assistance aux personnes âgées ou diminuées, la protection de la vie privée, ou la souveraineté des infrastructures de calcul et de données.

La pluridisciplinarité du LISN, mettant en jeu des objets et des pratiques scientifiques différents, il requiert de fait des besoins matériels et humains différents (e.g., expérimentations et moyens de calcul). En interaction avec les tutelles, le LISN veillera donc à maintenir un bon équilibre de moyens et de personnels entre les différents départements, en fonction des besoins et en cherchant à mutualiser des équipements de recherche lorsque c'est possible. En particulier, le LISN aura le souci de publier des postes sur l'ensemble des (sous)-disciplines présentes. Enfin il aura à cœur de construire un collectif scientifique porteur d'un projet à long terme, alimentant une riche vie commune de laboratoire, et dont la gouvernance implique l'ensemble de ses membres.

3.2 Département *Algorithmes, Apprentissage et Calcul*

3.2.1 Thèmes

L'algorithmique et le calcul touchent à tous les aspects théoriques et pratiques, logiciels et matériels de l'informatique. Depuis une décennie, l'intelligence artificielle et l'apprentissage s'intéressent à la conception automatique d'algorithmes et de processus de calcul, guidée par les données, l'expert, l'utilisateur et/ou l'environnement. Le département *Algorithmes, Apprentissage et Calcul* a pour objet la conception, la mise en œuvre et l'analyse d'algorithmes en prêtant attention à quatre types de propriétés : i) leur complexité théorique, en particulier en lien avec les paramètres spécifiques des problèmes considérés ; ii) leur convergence, leur passage à l'échelle et leur robustesse par rapport aux divers critères de qualité ; iii) les caractéristiques des algorithmes, des données disponibles *a priori* et générées *a posteriori*, et des processus de calcul en termes de théorie de l'information ; iv) la performance des calculs et leur implantation optimale sur les architectures parallèles les plus récentes.

Les principaux axes de recherche du département concernent les modèles de calcul et leur robustesse (du calcul haute performance au calcul quantique en passant par les réseaux neuronaux et les algorithmes répartis), les architectures de traitement (graphes, traitement distribué, synchrone ou asynchrone), et les méthodes (e.g., optimisation continue, combinatoire, stochastique ; apprentissage statistique et théorie de l'information). Par construction, ces axes de recherche font l'objet de collaborations avec les autres départements, en particulier *Science des Données* et *Mécanique des Fluides-Énergétique*. L'analyse et la conception des modèles et des processus font une

large part aux approches mathématiques (du discret et du continu, en passant par les probabilités, les statistiques et la combinatoire) et de la physique statistique (en particulier sur les phénomènes de transition de phase des systèmes complexes), en lien avec les équipes du LMO et du CMAP (Maths et Maths. Appli), du LAL (Physique), du L2S (Traitement de Signal), ainsi qu’avec le LIX et le futur LMF (Méthodes formelles).

Les domaines d’applications comprennent le calcul scientifique (e.g., algèbre linéaire, calcul tensoriel, optimisation numérique, systèmes dynamiques, simulation d’algorithmes quantiques, physique computationnelle, systèmes d’équations différentielles), le calcul distribué (e.g., cloud, ordonnancement, monnaie virtuelle, informatique ubiquitaire, robots autonomes, circuits micro-biologiques), et l’analyse des données.

3.2.2 Positionnement

Le département comprend une quarantaine de permanents, s’intéressant aux divers axes de l’étude des algorithmes, allant de leur comportement théorique et empirique, à leur robustesse et leur convergence, leur consommation en ressources et en énergie, et enfin, à leurs structures computationnelles. Les tâches considérées vont du calcul scientifique, combinatoire ou quantique, du calcul distribué, de l’algorithmique des graphes, à l’apprentissage statistique.

Parmi les domaines dans lesquels les membres du département ont joué un rôle pionnier, se trouvent l’algorithmique de graphes coloriés, les réseaux sans fil (approche par théorie des jeux, médaille de bronze du CNRS), les réseaux neuronaux (critères d’a-dimensionalité et invariance).

Une activité très visible au niveau international concerne le développement de bibliothèques logicielles du domaine public, notamment dans les domaines de l’algèbre linéaire (numérique et exacte), tenseurs, apprentissage automatique, et calcul combinatoire (SAGE, projet européen DREAM). Ces bibliothèques utilisent le parallélisme massif permis par les processeurs récents et les accélérateurs de type GPU (contributions à la bibliothèque MAGMA).

Une direction de recherche stratégique du département concerne la conception de portefeuilles d’algorithmes et d’implantations, répondant à l’hétérogénéité des instances de problèmes considérés (e.g. en optimisation combinatoire). De cette approche dérive une nouvelle question scientifique, le choix *a priori* de l’algorithme, des ses hyper-paramètres et/ou de l’architecture de calcul la plus adaptée pour une instance donnée. Cette problématique, à notre connaissance originale en France (organisation de challenges internationaux depuis les années 2010s), est étudiée sous le nom de *Sélection d’algorithmes* ou *Ajustement offline ou online d’hyper-paramètres* par quelques centres au niveau international (Polytech Montreal, Univ. Freiburg, Univ. Leiden, CMAP X), et fait l’objet de nombreuses collaborations industrielles (3 thèses CIFRE avec Thalès, dont une en collaboration GALAC et A&O ; des thèses CIFRE avec Renault, ATOS-Bull, ONERA, Orange, ADEME, RTE, CEA ; Chaire industrielle Lusi ; contrats EdF, Total). La vitalité du département est aussi attestée par la création de deux start-ups (Numscale en 2016 et Nukkai en 2018).

3.2.3 Composition

Équipe A&O (Apprentissage et Optimisation, équipe transverse avec le Dpt SDD).

L’équipe A&O – équipe projet commune Paris-Saclay, CNRS et Inria Saclay – s’intéresse à l’apprentissage de modèles à partir des données, en mettant l’accent sur quatre axes fondamentaux. Le premier concerne l’apprentissage adversarial qui repose sur l’interaction de deux ou plusieurs agents apprenants, remplaçant la fonction objectif inconnue par une approche de type min-max (théorie des jeux) ; cet axe est crucial également pour la validation et la certification des réseaux neuronaux. Le second porte sur la sélection et la configuration d’algorithmes *a priori*, en fonction des données disponibles, aussi appelée AutoML ; il s’agit non seulement d’une condition nécessaire pour la démocratisation de l’IA, mais aussi d’un défi ouvert depuis 40 ans, lié à la définition des paramètres d’ordre des données. Le troisième s’attaque aux problèmes d’apprentissage de modèles complexes et traite de l’identification des régularités permettant l’augmentation bien fondée des données dans les nombreux domaines d’application où les données sont petites (*small data*) par rapport à la complexité des modèles cherchés. Enfin, loin de remplacer la connaissance par des

modèles purement issus des données, un objectif consiste à dialoguer avec les connaissances du domaine, par exemple exprimées par des équations aux dérivées partielles ; l'enjeu est ici de faire le pont entre Apprentissage et Numerical Engineering, en lien avec le département *Mécanique des Fluides-Énergétique*.

Équipe GALaC (Graphes, Algorithmes et Combinatoires). Les membres de l'équipe GALaC s'intéressent aux modèles mathématiques discrets pertinents pour l'analyse et la conception d'algorithmes pour les systèmes en réseaux, les problèmes combinatoires et la théorie des graphes, avec des applications issues notamment de la bio-informatique, des réseaux sociaux et des réseaux de senseurs.

En théorie des graphes, la plupart de problèmes considérés (e.g. coloration, domination, Hamiltoniens, connectivité) sont connus pour être NP-difficiles. La question devient alors d'étudier ces problèmes pour des familles spécifiques de graphes ou de proposer des algorithmes approchés avec garanties (en vitesse et/ou en qualité d'approximation ou de non-approximabilité) et d'identifier les paramètres d'ordre contrôlant la croissance exponentielle de la complexité d'algorithmes exacts (algorithmes FPT). Il s'agit également de démontrer des conditions suffisantes garantissant la propriété désirée. L'algorithmique en réseau s'intéresse à la conception, l'analyse et l'optimisation d'algorithmes distribués, répondant à des objectifs appliqués et/ou à des questions théoriques. Le thème combinatoire s'intéresse d'une part à des questions algébriques et énumératives (algèbres de Hopf, théorie des représentations, partitions) et d'autre part aux systèmes symboliques (mots, pavages, automates cellulaires), en relation avec les systèmes dynamiques, les systèmes de numération et l'analyse de complexité calculatoire.

Équipe ParSys (Systèmes Parallèles). L'équipe ParSys est spécialisée dans le calcul haute performance et les algorithmes distribués (aspects théoriques et implémentation). Ceci inclut la conception et l'analyse de complexité d'algorithmes pour les systèmes distribués et le calcul quantique ainsi que le développement de solveurs de calcul haute-performance (HPC). L'activité de l'équipe implique également l'utilisation des architectures parallèles à la pointe de l'état de l'art pour obtenir une performance optimale des codes développés en vue des futures plate-formes exascale. De plus, en s'inspirant des phénomènes naturels, efficaces et robustes, l'équipe étudie des algorithmes distribués naturels pour concevoir des solutions performantes pour les réseaux émergents, et même pour développer des circuits distribués robustes dans des consortiums de cellules bactériennes, à des fins informatiques ou médicales (ordinateurs biologiques, médicaments intelligents, etc.). Les principaux défis concernant les algorithmes étudiés dans l'équipe ParSys sont : l'optimisation des performances, le passage à l'échelle et l'équilibrage de charge, la tolérance aux défaillances et l'économie d'énergie. Les applications (allant du calcul scientifique, quantique et de l'analyse de données aux protocoles réseaux en passant par les circuits micro-biologiques) répondent à des besoins industriels ou scientifiques essentiels et font l'objet de coopérations industrielles avec notamment ATOS-Bull, Total, et EdF.

3.2.4 Projet

Les discussions qui ont abouti à la création du département *Algorithmes, Apprentissage et Calcul* ont conduit à l'identification de plusieurs axes de synergie entre les équipes, qui sont en cours d'émergence. Nous décrirons ensuite les projets des équipes, et leurs liens éventuels avec d'autres départements du LISN.

Complexité paramétrée par apprentissage. La complexité algorithmique (au pire cas ou en moyenne) peut être modélisée finement en identifiant : i) les paramètres d'ordre des instances de problème (e.g. la taille maximum des cliques d'un graphe), et ii) comment la complexité dépend de ces paramètres. Le couplage avec les méthodes d'apprentissage et d'optimisation constitue une voie prometteuse pour identifier de tels paramètres d'ordre, la structure du modèle de complexité, et/ou pour construire des instances de complexité maximale.

Systèmes distribués, Hiérarchie de modèles. L'étude des systèmes distribués est abordée par GALAC, ParSys et A&O sous différents angles, allant de la structure du graphe et des aspects temporels, aux contraintes de ressources et à la puissance des modèles, en passant par l'apprentissage des fonctions de transfert aux nœuds ou graphes neuronaux. L'intérêt de cette recherche jointe est d'établir des liens entre les paramètres de graphe, la complexité locale (e.g. protocoles et mécanismes de diffusion), et les fonctionnalités requises (issues d'applications telles que le dimensionnement de réseaux EdF, réseaux de capteurs ou réseaux bactériens, voir section ci-dessous), et de caractériser une hiérarchie des systèmes distribués sous contraintes de ressources.

Calcul haute performance pour la combinatoire algébrique. La combinatoire algébrique a connu de nombreuses réussites grâce aux calculs explicites de conjectures théoriques. Ceci a été possible par la puissance de calcul offerte par les processeurs parallèles récents (multicœurs, GPU). Néanmoins, en raison de la structure très récursive de certains problèmes, il est nécessaire de repenser les algorithmes afin d'exploiter les possibilités de parallélisme actuelles puis de proposer des implémentations efficaces dans des bibliothèques logicielles du domaine public. Les membres des équipes Galac et ParSys collaborent pour réaliser de tels algorithmes et logiciels afin de permettre des avancées importantes en terme de performance de calcul et de mouvements de données (projet interne soutenu au LRI entre les équipes).

En lien avec le projet du département, l'équipe **GALaC** continuera de développer ses activités en théorie des graphes, algorithmique et combinatoire, structurées selon cinq directions de recherche. La première concerne la **combinatoire algébrique**, les logiciels de mathématiques computationnelles et les systèmes symboliques. La seconde concerne l'étude des **dynamiques, modèles de calcul** incluant toute la problématique des transitions de phases. La **conception d'algorithmes concurrents** dans divers modèles d'environnement (faisant appel à la notamment à la théorie des jeux) forme le troisième axe. Le quatrième et cinquième axes sont relatifs respectivement aux notions de **complexité paramétrée** (incluant optimisation et contrôle des réseaux et des systèmes distribués), et aux **problèmes structurels et algorithmiques pour les graphes** (considérant ici à la fois les graphes classiques et graphes sommet/arête coloriés).

Les directions de recherche de **ParSys** en lien avec les principaux défis de l'équipe et du département incluent six thématiques de recherche. Le premier thème concerne la **conception d'algorithmes distribués fondamentaux tolérants aux pannes** pour des réseaux d'agents mobiles (e.g., protocoles de population auto-stabilisants ou tolérants des pannes byzantines). Le second vise à développer des modèles et des **algorithmes distribués performants et robustes inspirés par la nature** (e.g., imitant le comportement des fourmis, des lucioles). Une direction complémentaire porte sur la **conception et l'implantation de circuits distribués robustes pour les systèmes bactériens**, en exploitant des méthodes de la théorie de calcul distribué, la conception de circuits tolérants aux pannes et la biologie synthétique. Le quatrième thème traite de l'algorithmique parallèle et le **calcul haute performance pour l'algèbre linéaire**, le calcul tensoriel et le traitement des données massives et des enjeux d'ordonnancement et d'équilibrage de charge pour le passage à l'échelle dans les applications parallèles. Le cinquième porte sur le **calcul quantique**, notamment la compilation et la simulation d'algorithmes quantiques avec application à des problèmes réels issus de la physique, de la chimie quantique et de la finance, et en lien avec le calcul haute performance et les méthodes tensorielles. Enfin, l'équipe poursuivra ses travaux sur les **techniques de génération** de codes et langages dédiés en C++ avancé pour la parallélisation automatique des applications.

En collaboration avec le département *Mécanique des Fluides-Énergétique* ainsi qu'avec l'IFPEN, une perspective de recherche majeure de l'équipe **A&O** porte sur la jonction entre **réseaux profonds et équations différentielles**, visant à jeter un pont entre l'identification, la résolution et l'optimisation d'équations différentielles d'une part et les réseaux neuronaux d'autre part. Le

défi est double. Un premier objectif consiste à passer à l'échelle en terme de discrétisation spatio-temporelle. Un second objectif consiste à tirer parti des connaissances du domaine, concernant par exemple les invariances des solutions cherchées pour tout à la fois optimiser la résolution du problème et garantir la validité physique des solutions obtenues. Une troisième dimension stratégique consiste à inspecter et interpréter les architectures de calcul optimisées par réseaux neuronaux (voir section 3.7).

Les autres priorités d'A&O en lien avec le département *Algorithmes, Apprentissage et Calcul* concernent la **certification de modèles neuronaux** (deux thèses co-encadrées avec le CEA) ; le meta-learning, i.e. la **sélection automatique des algorithmes, hyper-paramètres et/ou architectures les plus appropriés** en fonction du contexte (co-organisation d'un challenge avec Google Zurich) ; iii) les **applications** dans les domaines du management de l'énergie (thèse avec RTE ; thèse avec ADEME) et de la physique (en collaboration avec le CERN).

3.3 Département Interaction avec l'Humain

3.3.1 Thèmes

Le département *Interaction avec l'Humain* rassemble des chercheurs et enseignants-chercheurs actuellement répartis au LIMSI (groupes AMI, CPU et VENISE) et au LRI (équipes-projets communes UPSUD-CNRS-Inria EX-SITU, ILDA et AVIZ). Les équipes du département s'intéressent à l'interaction entre l'humain et la machine avec un regard pluridisciplinaire alliant informatique, traitement du signal, psychologie cognitive, psychologie sociale et ergonomie cognitive. Cette thématique englobe non seulement les multiples moyens matériels et logiciels qui peuvent être mobilisés pour établir un lien entre l'humain et l'ordinateur, mais aussi la compréhension de l'humain pour la conception de nouvelles interfaces et la compréhension des interactions sociales entre humains pour assister leur collaboration grâce à l'ordinateur. Le département constitue l'un des principaux pôles internationaux sur cette thématique.

Les équipes de ce département explorent de multiples formes d'interaction : interaction visuelle, tangible, haptique, auditive, gestuelle, ... ; interaction multimodale ; visualisation interactive d'information ; réalité virtuelle ou augmentée ; interactions affectives et sociales ; interactions collaboratives en situation co-localisée ou distante. Elles développent des méthodes de conception (comme la conception participative), des méthodes d'analyse (comme l'analyse d'image pour l'interaction) et elles abordent des domaines d'application variés (comme la prise en compte du handicap). L'éventail de ces recherches permet au département de développer des travaux qui sont à la fois très fondamentaux, en cherchant à modéliser les phénomènes en cours lors de ces interactions, et tournés vers des applications pour évaluer la pertinence de ces modèles et de certains dispositifs dans leur usage réel.

3.3.2 Positionnement

Avec près de 40 chercheurs et enseignants-chercheurs, le département *Interaction avec l'Humain* constitue l'un des plus importants pôles d'enseignement et de recherche dans ce domaine au niveau national, européen et international. Au niveau national, les autres équipes importantes se situent à Grenoble, avec un spectre assez large mais centré sur le génie logiciel des systèmes interactifs, et à Toulouse, avec une spécialisation sur l'interaction pour les systèmes critiques et le handicap. Au niveau international, les plus gros centres sont surtout aux États-Unis (CMU, Georgia Tech, U. Washington, MIT, ...). En dehors de ces centres, peu d'universités ont un spectre thématique aussi large et un nombre de chercheurs aussi important. Le spectre large du département *Interaction avec l'Humain* lui confère un rôle structurant dans le domaine, en s'appuyant sur des chercheurs très reconnus dans leurs thématiques au niveau national (projets ANR, collaborations industrielles) et international (projets européens, ERC, collaborations internationales majeures). Deux masters d'audience internationale sont également portés par des enseignants et chercheurs très impliqués, ce qui contribue à l'aura des équipes du département.

Le département s'appuie sur un réseau d'équipements exceptionnels, développé notamment au travers de l'Equipex Digiscope, dont il assure la direction scientifique et dont il héberge trois des 10 plateformes (EVE, WILD et WILDER), ainsi que le FabLab. Le caractère pluridisciplinaire des recherches du département *Interaction avec l'Humain* est visible à travers différents niveaux de collaboration : à l'intérieur d'une équipe entre plusieurs disciplines (par exemple informaticiens et psychologues dans CPU), entre différentes équipes du département (par exemple VENISE et ExSitu avec les dispositifs Digiscope) ou encore avec des équipes des quatre autres départements du laboratoire. De fortes collaborations ont également été construites au sein de l'Université Paris-Saclay, comme avec la Fédev pour les travaux sur le mouvement ou au travers des actions « Arts et Sciences » dans lesquelles les équipes du département ont toujours été fortement impliquées. Au niveau national, le département est impliqué dans l'animation de groupes de travail thématiques (la journée commune de l'AFIHM et de l'AFIA en 2019 a ainsi été organisée par des chercheurs et enseignants-chercheurs du département) et de conférences scientifiques (par exemple la conférence IHM). Le département a aussi construit de nombreuses collaborations avec des acteurs industriels qui financent régulièrement des thèses CIFRE et des travaux de recherche fondamentale ou appliquée (ex. deux chaires industrielles, participation à l'Institut Carnot « Cognition », collaborations avec EDF R&D). Enfin, au niveau international, les équipes du département collaborent avec les plus grandes universités dans le domaine de l'Interaction Humain-Machine dans le monde, animent la communauté de la RV&A à l'échelon européen (EuroVR) et participent à l'organisation des conférences phares du domaine (CHI 2013 à Paris par exemple avec 3500 participants, et IEEE VIS 2014 avec 1200 participants). Le domaine de l'interaction est en constante mutation du fait des évolutions technologiques qui continuent à modifier le regard que l'on peut avoir sur l'interaction avec une machine. L'introduction dans la vie de tous les jours de ce qui n'apparaît pas toujours comme étant une machine mais avec laquelle il faut malgré tout interagir nécessite de définir de nouveaux paradigmes. Le département a tous les atouts pour être l'un des acteurs majeurs de cette transformation.

3.3.3 Composition

Équipe AMI (Architecture et Modèles pour l'Interaction). L'équipe AMI regroupe des chercheurs et enseignants-chercheurs spécialisés dans différents modes d'interaction tels que la vision, le mouvement, l'interaction tactile ainsi que dans les interfaces multimodales qui combinent différentes modalités pour parvenir à une interaction plus « naturelle ». Les recherches de l'équipe s'articulent autour de trois axes : le traitement d'image et la visualisation d'information appliqués à différents domaines tels que la réalité augmentée vidéo-projetée, l'analyse d'activité, l'aide au diagnostic ; les environnements ambiants et connectés avec des applications dans le domaine des services à la personne et la santé ; l'étude de la multimodalité dans le contexte du handicap sensoriel, moteur et cognitif.

Équipe CPU (Cognition, Perception et Usages). À la frontière entre informatique, psychologie et ergonomie, l'équipe CPU est composé d'enseignants-chercheurs de ces trois disciplines. L'équipe conduit des recherches pluridisciplinaires sur la cognition humaine et artificielle et la conception, l'évaluation et les usages des technologies. Ces recherches ont un double objectif : 1) étudier l'humain pour concevoir des systèmes adaptés et adaptatifs aux caractéristiques humaines dans leurs situations d'usage ; 2) concevoir des interactions humain-machine pour étudier et mieux comprendre l'humain. Ces recherches sont appliquées à l'analyse et à l'apprentissage de compétences sociales et professionnelles, à la simulation de comportements et aux technologies motivationnelles ou persuasives.

Équipe VENISE (Virtual and augmented ENvironments for Simulation and Experiments). L'équipe VENISE a pour objectif scientifique de comprendre et faire progresser le domaine de la Réalité Virtuelle et Augmentée (RV&A). Les recherches de l'équipe visent d'une part des travaux génériques concernant la conception et l'évaluation de nouveaux modèles d'interaction en RV&A, et d'autre part l'élaboration des modèles de données ou de simulation pour

accroître l'utilisabilité de la RV&A dans différentes applications phares. Transversalement à ces deux thèmes, l'équipe étudie les problématiques perceptives et cognitives qui sont propres à ce domaine, telles que les incohérences sensorimotrices, la « cybersickness », les sensations d'immersion ou de présence, et elle travaille sur les évolutions de leurs formalisations théoriques.

Équipe ILDA (Interacting with Large Data). Dans un nombre croissant de domaines, les utilisateurs de systèmes informatiques ont à faire face à de grands jeux de données. Ces jeux de données sont de plus en plus interconnectés et organisés sous forme de structures complexes, grâce notamment aux nouveaux modèles de données que nous voyons apparaître, par exemple, avec le Web des données. Les travaux de recherche de l'équipe ILDA ont pour but de tirer parti de la richesse de ces modèles de données pour concevoir de nouveaux systèmes interactifs qui assistent au mieux les utilisateurs dans leurs tâches de compréhension et de traitement de ces grandes quantités de données.

Équipe EX-SITU (Extreme Situated Computing). L'équipe ExSitu explore les limites de l'interaction en étudiant des utilisateurs « extrêmes », tels que les professionnels créatifs, artistes et designers qui repoussent les limites de la technologie pour atteindre leurs objectifs créatifs, ou les scientifiques qui essaient de comprendre les phénomènes complexes par l'exploration créative de données complexes. L'étude de ces utilisateurs permet non seulement d'anticiper les besoins des usages routiniers de demain, mais aussi de mieux comprendre le phénomène de l'interaction lui-même et de créer de nouvelles formes d'interaction. Le but de l'équipe est de créer des partenariats humains-machines permettant aux utilisateurs de maîtriser la technologie, fondés sur des principes issus de l'approche théorique de l'interaction développée dans l'équipe.

Équipe AVIZ (Visual Analytics). L'équipe Aviz a pour objectif d'améliorer les méthodes, techniques et outils basés sur la visualisation interactive. Traditionnellement, la visualisation a été segmentée en questions spécialisées : visualisation scientifique, techniques d'interaction, ou nouvelles représentations. Aviz, au contraire, étudie le continuum d'activités entre la modélisation, l'exploration, la présentation, la prise de décision et l'éducation. De plus, lorsque la quantité de données augmente, elle dépasse les capacités de transfert des seuls canaux visuels. Aviz travaille aussi sur le « Visual Analytics » : un ensemble de méthodes pour combiner l'analyse de données avec la visualisation afin de permettre l'exploration de données massives. Un des axes d'Aviz est le passage à l'échelle dans l'exploration des données à l'aide d'un nouveau paradigme de calcul : l'analyse progressive de données, qui produit des estimations de résultats en temps borné et améliore ces estimations itérativement. Ce nouveau paradigme a un lien fort avec plusieurs autres départements du laboratoire, comme *Science des Données* et *Algorithmes, Apprentissage et Calcul*.

3.3.4 Projet

Le projet de recherche du département s'inscrit dans la continuité des travaux menés jusqu'à présent en Interaction Humain-Machine, compris dans un sens très large, allant du matériel au logiciel, de la visualisation de données à la réalité virtuelle et augmentée, des modèles informatiques aux études expérimentales avec ou sur les humains. Ils peuvent être regroupés en quatre axes qui sont transverses aux six équipes du département et aux deux sites du laboratoire décrits ci-après.

Étendre les fondements théoriques de l'interaction humain-machine : modèles d'interaction multimodale, modèles d'intelligence artificielle pour l'interaction (apprentissage statistique, modèles symboliques, ...), modèles de perception et d'interaction 3D en situation immersive virtuelle ou augmentée, modèles d'interaction collaborative colocalisée ou distante, modèles de visualisation de données, modèles de traitement d'images, modèles psychologiques liés par exemple à la cognition, aux interactions sociales et aux émotions humain-humain et humain-machine, à la personnalité et aux troubles cognitifs. Nous nous appuyons pour cela sur des études empiriques pour comprendre les aspects moteurs, perceptifs, cognitifs et émotionnels qui sont impliqués dans le phénomène d'interaction avec des machines, qu'il s'agisse d'applications graphiques, d'agents virtuels

animés, de supports d'information exploitant d'autres modalités (par exemple des représentations tangibles) ou d'environnements de Réalité Virtuelle et Augmentée (RV&A) ou ambiants.

Comprendre l'utilisateur dans la spécificité de son activité et dans ses interactions avec son environnement et avec d'autres utilisateurs : qu'il s'agisse de professionnels créatifs (designer graphique, chorégraphe, musicien . . .), de scientifiques ou d'analystes travaillant avec des grandes masses de données potentiellement hétérogènes, de personnes en situation de handicap, de personnes travaillant plus ou moins collaborativement dans différents processus industriels, tous ont des besoins spécifiques en terme d'interaction. En nous appuyant sur des études expérimentales, des observations de terrain, ou encore sur des prototypes de techniques ou de systèmes, nous concevrons avec et pour les utilisateurs des supports interactifs adaptés aux activités humaines en articulant informatique, psychologie et ergonomie.

Concevoir des techniques avancées d'interaction, de visualisation de données et de traitement d'images et de données, de perception et d'interaction immersive multi-sensorimotrice : les environnements multi-surfaces (salles interactives, salle de réalité virtuelle), la variété des dispositifs (de la montre connectée au mur d'écran en passant par les casques de RV&A), et des équipements multi-sensorimoteurs (haptique, audio 3D), d'immersion collaborative co-localisée (multi-stéréoscopie) ou distante, d'augmentation visuelle, audio ou pseudo-haptique, de données issues de l'informatique ubiquitaire, enfouie et ambiante posent de nombreux défis en termes d'interaction, de traitement et de perception de l'information. Par ailleurs, les quantités massives de données, l'hétérogénéité des sources de données et la multitude des tâches utilisateurs constituent des leviers d'innovation sans cesse renouvelés. Ces défis seront abordés en tirant parti aussi bien de modalités d'interaction classiques que de modalités plus exploratoires : gestuelles, incarnées (*embodied* en anglais), tangibles, immersives, spatiales, etc. Nous proposons d'appliquer des principes de conception de l'interaction et de représentation de l'information issus, d'une part des fondements théoriques et des études empiriques du premier axe, et d'autre part des nouvelles capacités matérielles et logicielles, des modèles de données émergents, et des méthodes de calcul issues de l'apprentissage automatique et du calcul haute performance.

Développer des outils, des méthodologies et des modèles conceptuels qui facilitent la conception et la réalisation de systèmes interactifs efficaces : Les travaux porteront à la fois sur les entrées, pour diminuer la difficulté de programmer avec des canaux d'entrée tels que les gestes ou les objets physiques, sur les sorties visuelles, pour assurer des systèmes réactifs même avec de grands volumes de données (visualisation progressive), sur des combinaisons ou substitutions de rendus sensori-moteurs (visuel, audio, tactilo-kinesthésique, olfactif . . .), pour optimiser la communication d'information avec l'humain, sur les nouveaux modèles de conception en réalité virtuelle et augmentée, sur la communication entre utilisateurs, pour améliorer la collaboration locale et à distance, sur les modèles d'interactions ludiques, et sur l'interaction avec des systèmes "intelligents", des agents conversationnels expressifs ou des interactions dynamiques sur objets connectés ou dans des environnements ambiants. Nous travaillerons également à faire évoluer les méthodes de validation des travaux en Interaction Humain-Machine afin de maximiser la validité des évaluations menées, en respectant les contraintes d'une recherche éthique et reproductible. Pour cela nous prendrons en compte les nouveaux contextes d'usages, les nouvelles méthodes statistiques, et revisiterons aussi bien les protocoles expérimentaux que les questionnaires et les études de terrain permettant de collecter des mesures quantitatives et subjectives.

Ces activités de recherche visent à apporter conjointement des solutions pour relever les grands défis actuels de l'interaction avec des systèmes numériques dans les différents domaines liés par exemple à l'éducation, à la santé, au handicap, aux transports, à l'urbanisme, aux processus industriels, au travail, à l'emploi, aux arts et à la culture.

3.4 Département Mécanique des Fluides - Énergétique

3.4.1 Thèmes

Les activités du département se regroupent autour de la Mécanique des Fluides, des Transferts de masse et de chaleur et de l'Énergétique. Les équipes du département mènent des recherches à caractère généralement fondamental, dont les applications se trouvent dans les secteurs de l'énergie, des transports, de la santé et de l'environnement. Elles approfondissent les théories existantes et en développent de nouvelles avec l'objectif d'améliorer la compréhension des phénomènes physiques fondamentaux de ces domaines, d'améliorer leur modélisation et leur simulation avec notamment une visée d'optimisation et de contrôle des processus. Les deux axes principaux pour atteindre ces objectifs sont le développement de méthodes numériques avancées permettant d'accroître les performances des outils de simulation, ainsi que la mise au point de techniques d'analyse efficaces pour la caractérisation de la dynamique des écoulements et des transferts et leur manipulation. Des réalisations expérimentales viennent en support de cette double approche méthodologique et phénoménologique.

3.4.2 Positionnement

Le département de *Mécanique des Fluides-Énergétique* comprend une trentaine de chercheurs, enseignants-chercheurs, et ingénieurs de recherche. Les activités de recherche du département recouvrent un spectre large, allant du numérique à l'expérimental, des échelles nanométriques aux échelles géophysiques, des analyses fondamentales aux études applicatives, et comprennent de larges interfaces avec la physique, les mathématiques appliquées et l'informatique. Elles s'inscrivent dans de nombreuses collaborations nationales et internationales. La simulation numérique représente une part importante des activités de recherche du département, et a conduit au développement d'outils numériques partagés et de supports informatiques mutualisés. La mise au point de solveurs parallèles performants a donné lieu à des collaborations avec l'équipe PARSYS (département *Algorithmes, Apprentissage et Calcul*). La mise en œuvre de méthodes d'apprentissage pour le contrôle a conduit à une ANR avec l'équipe TLP (département *Sciences et Technologies des Langues*), qui se poursuit par plusieurs collaborations avec cette équipe et l'équipe A&O (département *Algorithmes, Apprentissage et Calcul*). Les approches statistiques pour le traitement des données de Mécanique des Fluides ont bénéficié du soutien de l'IRS *Center for Data Science*. La dynamique de recherche globale du département s'inscrit résolument dans le contexte des Sciences de l'Ingénieur, en particulier pour l'aérodynamique et l'énergétique. Elle est soutenue sur le plateau de Saclay par diverses actions, telles que le Labex LaSIPs, ainsi que par des collaborations avec le CEA et le laboratoire EM2C de CentraleSupélec. Le département entretient des liens forts avec le laboratoire FAST, à travers notamment l'organisation d'un séminaire commun, ainsi que plusieurs collaborations exploitant la complémentarité des recherches menées dans les deux laboratoires. Des collaborations, concrétisées dans des ANR, existent également avec le laboratoire GeePS. Le département est organisé en deux équipes qui partagent des approches méthodologiques, notamment des méthodes numériques communes, mais qui sont complémentaires par leurs champs d'intérêt. L'une s'intéresse plus spécifiquement à la dynamique des Fluides, tandis que l'autre vise à caractériser les couplages multiphysiques qui y prennent place.

3.4.3 Composition

Équipe DATAFLOT (Dynamique et Contrôle des écoulements). L'équipe DATAFLOT s'intéresse à la **dynamique des fluides et en particulier à la maîtrise des instabilités et de la turbulence**, qui caractérisent un grand nombre d'applications dans le domaine des sciences de l'ingénieur. Les travaux de l'équipe s'articulent autour de deux thèmes : d'une part, l'étude fondamentale des instabilités et de la turbulence développée, d'autre part l'analyse, la manipulation et le contrôle des écoulements. Pour le premier thème (ITAC - Instability and Transition Characterization), on s'intéresse au comportement des écoulements cisailés, en présence ou non

de rotation et d'entraînement par des parois. Les objectifs d'étude sont la caractérisation de la dynamique linéaire et non-linéaire, pour des régimes allant de l'instabilité primaire à la turbulence pleinement développée. Ces travaux s'appuient sur le développement de méthodes numériques, adaptées aux nouvelles générations de calculateurs, et capitalisées dans des logiciels numériques déposés et ouverts à la communauté. Le second thème (DECIPHER - Data-Enhanced modelling for computational mechanics) s'intéresse à la compréhension, la prédiction et éventuellement le contrôle de systèmes mécaniques complexes, en s'appuyant sur une modélisation augmentée par les données. Il s'agit de développer des méthodes permettant d'enrichir les apports de la simulation numérique et notamment de remédier à ses limites dans des cas réels complexes. La réduction de la dimension, la modélisation stochastique et la quantification d'incertitudes constituent les pierres angulaires de notre approche. Des développements méthodologiques sur plusieurs aspects (réduction de modèles et approximation en grande dimension, assimilation de données et inférence, apprentissage automatique et théorie de l'information) sont indispensables pour surmonter les difficultés liées à nos objectifs, comme notamment la prédiction et le contrôle des écoulements turbulents.

Équipe COMET (Couplages Multiphysiques et Transferts). L'équipe COMET étudie différents phénomènes physiques résultant d'inhomogénéités de température, de densité, de courant électrique ou de champ magnétique, et leurs couplages en volume ou aux interfaces, et aux échelles pertinentes. Les travaux cherchent à améliorer la connaissance de phénomènes physiques isolés, à comprendre le couplage entre différents phénomènes, et à augmenter l'efficacité énergétique d'un système ou d'un ensemble de systèmes. Pour cela, l'équipe met en œuvre des simulations numériques et des expériences de laboratoire, approches complétées par des interprétations analytiques et théoriques. De façon plus détaillée, les thèmes abordés sont :

- la magnétohydrodynamique (MHD), qui décrit le mouvement d'un fluide conducteur de l'électricité dans lequel les champs de vitesse et d'induction magnétique sont couplés par la force de Lorentz et la loi d'Ohm et qui s'intéresse, en particulier, à la compréhension de l'apparition d'un champ magnétique par effet dynamo ;
- la convection, qui apparaît dans les écoulements en présence de gradients de température ou de densité, avec ou sans changement de phase. Ce thème regroupe des études fondamentales (convection naturelle, rayonnement, mélange de fluides miscibles) menées à l'aide de la simulation numérique, ainsi que plusieurs études à contexte applicatif fort sur l'efficacité énergétique (photovoltaïque, récupération de chaleur, réfrigération secondaire) ;
- les transferts aux interfaces, la cryonanothermique et les écoulements diphasiques ; cela va de la résistance thermique de Kapitza existant à l'interface silicium/hélium superfluide à la modélisation numérique des écoulements diphasiques incompressibles ou compressibles à faible nombre de Mach ;
- les phénomènes à l'œuvre dans les machines thermo-acoustiques, notamment au sein d'une pompe à chaleur destinée aux véhicules terrestres, ou dans les systèmes à surface libre (céléfaction ou bien mouillage super-hydrophobe).

3.4.4 Projet

Dans un contexte de développement durable et de maîtrise de l'énergie dans tous les secteurs de l'industrie, les études de Mécanique des Fluides et d'Énergétique apparaissent indispensables afin de conduire à des ruptures technologiques pour la réduction de traînée dans les transports ou la production d'une énergie sûre et efficace. Le succès de ces approches nécessite une modélisation pertinente et une compréhension fine des différents phénomènes physiques en jeu, qui généralement requièrent mais ne peuvent se limiter à un traitement purement numérique.

Une expertise de la situation physique réelle est donc nécessaire, et nous avons la chance dans le Département de pouvoir nous appuyer sur plusieurs expérimentateurs, dont certains sont également spécialistes des méthodes numériques. Une des forces du département de Mécanique réside dans la complémentarité entre les réalisations expérimentales, qui peuvent notamment doubler

ou augmenter les simulations numériques menées au laboratoire, et les approches théoriques, qui peuvent être confrontées aux résultats expérimentaux et numériques.

Les compétences du département *Mécanique des Fluides-Énergétique* en simulation numérique sont larges, et sont notamment matérialisées par l'existence de plusieurs codes de simulation propres au laboratoire, permettant de traiter une variété de géométries et de configurations physiques (turbulence, convection, écoulements multiphasiques). Ces codes sont utilisés pour des applications de calcul intensif (plusieurs projets au laboratoire utilisant ces codes sont associés à des demandes supérieures à un million d'heures sur les calculateurs de l'IDRIS). Ils sont aussi largement utilisés dans le cadre de collaborations nationales et internationales. Le partage par plusieurs utilisateurs d'un même code permet d'accélérer et de rendre robustes les nouveaux développements logiciels. Il favorise également les échanges scientifiques entre différents sujets de recherche.

Outre ses forces en simulation numérique et ses capacités expérimentales, le département *Mécanique des Fluides-Énergétique* a développé une variété d'approches pour la caractérisation, la prédiction et le contrôle des écoulements. Plusieurs de ces approches bénéficient d'une reconnaissance nationale et internationale, concernant notamment la transition à la turbulence ou le contrôle des écoulements par apprentissage. La mise au point de ces outils s'inscrit dans une approche de modélisation augmentée par les données, qui s'appuie sur des techniques comme la réduction de modèle, la quantification d'incertitude et les approches statistiques.

Un premier enjeu pour le département *Mécanique des Fluides-Énergétique* est de préserver et de renforcer la synergie entre les approches expérimentales, numériques et méthodologiques permettant la modélisation, l'analyse et la prédiction de phénomènes fondamentaux dans des situations réelles, en particulier celles mettant en jeu des couplages multiphysiques.

Un deuxième enjeu pour le département *Mécanique des Fluides-Énergétique* est d'intégrer et de développer les techniques d'apprentissage pour le traitement des données de Mécanique des Fluides, dont la complexité est inexorablement croissante, qu'il s'agisse de données issues de l'expérience, de la simulation numérique ou du monde réel. Les collaborations déjà existantes avec les départements *Sciences et Technologies des Langues* ou *Science des Données* ont vocation à se renforcer et à s'enrichir (par exemple par le biais de projets ANR en cours de soumission).

Compte tenu des compétences existantes au sein du département *Mécanique des Fluides-Énergétique*, une dynamique interactive forte est envisageable afin de mener à bien l'ensemble des travaux sous-tendant ces deux enjeux : des bases de données expérimentales ou numériques sont produites en interne et peuvent être analysées avec des méthodes d'apprentissage où l'expertise physique permet de discriminer les plus adaptées ; ces techniques d'apprentissage, contraintes par les lois physiques, peuvent permettre d'identifier des variables pertinentes ou des comportements saillants et de définir des modèles d'ordre réduit basés sur ces variables. Une fertilisation mutuelle peut ainsi nourrir l'interdisciplinarité.

3.5 Département Sciences des Données

3.5.1 Thèmes

Le département *Science des Données* rassemble quatre équipes d'expertises complémentaires, couvrant la modélisation, la collection, la gestion, l'analyse et la construction de données et de connaissances. Des traces numériques de l'ensemble des activités humaines sont aujourd'hui disponibles dans tous les domaines, de la santé à l'éducation, de la finance et des assurances à la production industrielle, des sciences expérimentales aux sciences humaines. Les données disponibles sont souvent massives, hétérogènes, dynamiques et de qualité variable (les 4 V : Volume, Variété, Vitesse, Véracité). Leur exploitation conduit à la définition d'un quatrième paradigme scientifique : la conception et la validation d'hypothèses, de modèles théorique et d'algorithmes, guidées par les données et en interaction forte avec les experts du domaine.

Le département *Science des Données* s'intéresse donc à répondre de façon robuste aux défis des 4V : en termes de passage à l'échelle face au volume et à la vitesse des données, et de ré-

sistance aux biais de diversité et de qualité. Ces objectifs définissent de nouvelles problématiques informatiques dans le domaine du stockage, de la communication, de l'analyse et de l'optimisation des traitements, de l'interrogation et de l'enrichissement des données, de la découverte des connaissances et de l'apprentissage de modèles.

3.5.2 Positionnement

L'expertise des membres du département *Science des Données* couvre un spectre large : bases de données, fouille de données, Web sémantique, représentation des connaissances, algorithmique, combinatoire, optimisation stochastique et distribuée, apprentissage statistique et réseaux neuronaux (supervisés, non supervisés, structurés), réseaux de communication, simulation, validation et transfert vers l'industrie.

Le département, comprenant près de 40 chercheurs et enseignants-chercheurs, dispose ainsi d'une expertise approfondie de la recherche interdisciplinaire et du dialogue avec les experts des domaines d'applications (particulièrement biologie, médecine, sciences humaines et sociales, physique expérimentale), permettant un accès privilégié aux données d'intérêt et à l'évaluation des modèles et des algorithmes.

Les membres du département sont très impliqués dans l'animation de l'Université Paris-Saclay : dans le Labex DigiCosme, dans les Initiatives de Recherche Stratégiques *Center for Data Science* et *Nutriperso*, dans l'Institut de Convergence DataIA, ainsi que dans l'IRT SystemX. Le département a également des partenariats industriels de longue date avec Airbus, IBM, Orange, Safran, EdF, Peugeot, Renault, Thalès, RTE, IFPEN, Facebook. Pour l'année 2019, 14 thèses CIFRE sont en cours.

Sur le plan national, les équipes du département sciences des données sont fortement impliquées dans les GDR MaDICS¹, RO² et BiM³, dans les sociétés savantes en IA (AFIA), en apprentissage automatique (SSFAM), en informatique (SIF), en recherche opérationnelle (ROADEF, vice présidence communication), ainsi que dans les IPL (recherches transversales majeures) d'Inria.

Sur le plan international, les membres du département *Science des Données* sont extrêmement visibles (General chair de NeurIPS 2018; Steering committee de ECML-PKDD) et ont des liens forts avec les centres Alan Turing (Londres), Ivado (Montréal), JST (Tokyo), CDS (New York) ainsi qu'avec les départements de science des données de Berkeley, Manchester, UC San Diego, et plus généralement avec de nombreuses universités et partenaires étrangers. Les équipes font partie de plusieurs réseaux internationaux : NEWCOM, réseau européen d'excellence dans la communication sans fil; EIT Digital, institut Européen pour l'éducation, la recherche et l'innovation; ELIXIR, réseau Européen de coordination des ressources bioinformatiques (bases de données, logiciels, clouds, etc); ELLIS (European Lab for Learning & Intelligent Systems); CLAIRE (Confederation of Laboratories for Artificial Intelligence Research in Europe).

3.5.3 Composition

Équipe LaHDAK (Large-scale Heterogeneous Data and Knowledge). L'équipe LaHDAK s'attaque aux défis liés à la disponibilité croissante des données et des connaissances, impactant tous les domaines d'innovation, dans l'ère du BigData. Elle propose des solutions pour gérer des données massives, complexes ou interconnectées, sémantiquement hétérogènes, incertaines, incomplètes et évolutives, provenant de sources variées, telles que le Web social, le Linked Open Data ou les données scientifiques. L'équipe développe des solutions visant à résoudre les problèmes d'efficacité et de passage à l'échelle, de pertinence sémantique et de robustesse dans les divers scénarios d'accès à l'information. Elle rassemble des expertises complémentaires en bases de données, systèmes de recommandation, fouille de graphes, intelligence artificielle, représentation

1. GdR Science des données : direction du GDR, pilotage d'une action de reproductibilité d'expériences bioinformatiques, co-portage de l'action RoD de raisonnement sur les données.

2. GdR Recherche Opérationnelle, porteur du groupe de travail programmation mathématique.

3. GdR Bioinformatique Moléculaire, membre du conseil scientifique.

des connaissances, raisonnement et Web Sémantique. Ses travaux de recherche sont organisés autour de trois axes : gestion des données, gestion de connaissances, fouille de données et découverte de connaissances.

Équipe ROCS (Réseaux et Optimisation Combinatoire Stochastique). L'équipe ROCS se décline en trois activités : réseaux, optimisation combinatoire et stochastique, optimisation dans les réseaux. L'objectif commun à ces trois activités est de traiter des problèmes issus du large domaine de l'Internet des Objets (IoT) avec la programmation mathématique par exemple en proposant des techniques nouvelles de linéarisation pour les problèmes d'efficacité énergétique. Nous avons ainsi développé des techniques d'optimisation spécifiques pour le problème de réseau à traiter. Plus généralement l'équipe ROCS propose des méthodes de résolution exactes qui serviront de point de référence lors de l'élaboration d'heuristiques visant le passage à l'échelle et/ou la résolution distribuée. La méthodologie générale de l'équipe ROCS consiste dans un premier temps, à avoir une approche théorique, puis les performances sont étudiées par simulation, domaine dans lequel ROCS a développé des compétences fortes en mettant en œuvre des outils logiciels efficaces (en particulier sous NS et OPNET, CPLEX). Les solutions sont également implantées dans des systèmes réels pour conduire des expérimentations. Dans ce cadre, plusieurs plateformes ont été développées (comme la plateforme sur les drones).

Équipe BioInformatique L'objectif de l'équipe Bioinfo est de répondre à des questions biologiques et bioinformatiques à la pointe de l'état de l'art, par la conception et le développement de nouvelles approches informatiques. Le principal atout de cette équipe est non seulement d'être capable de comprendre les problèmes posés par l'étude des systèmes biologiques depuis l'échelle moléculaire jusqu'à celle des cellules et des populations, mais aussi de concevoir des solutions impliquant des compétences issues de domaines très variés (algorithmique, combinatoire, bases de données, représentation des connaissances, extraction de données, analyse statistique, apprentissage automatique, méthodes formelles et simulation). Notre programme scientifique est structuré en trois thèmes, fortement interdépendants : Analyse de séquences et de structures biologiques ; Biologie des systèmes et biologie de synthèse ; Intégration et analyse de données biologiques.

Équipe A&O (Apprentissage et Optimisation, équipe transverse avec le Dpt AAC). L'équipe A&O – équipe projet commune Paris-Saclay, CNRS et Inria Saclay – s'intéresse à l'apprentissage de modèles à partir des données, en mettant l'accent sur trois types de propriétés. La première concerne la robustesse statistique, qui vise la stabilité et la qualité prédictive des modèles face au bruit et à la variabilité des données. La seconde porte sur l'équité des modèles, qui s'intéresse aux impacts des décisions fondées sur les modèles et à l'éthique (recherche des modèles causaux ; analyse des biais des données et des modèles ; traçabilité des décisions). La troisième s'intéresse à l'interprétation des représentations internes de l'apprentissage, en lien avec les structures et les invariances des données, au cœur de l'apprentissage profond.

3.5.4 Projet

Le département *Science des Données* permettra d'explorer les synergies entre expertises liées aux données, à l'apprentissage et à l'optimisation, notamment en lien avec les domaines de la bioinformatique, de l'IoT et des graphes de données, en approfondissant les collaborations existantes ou émergentes.

Un premier axe de collaborations couvre **l'analyse et la modélisation de données bioinformatiques**. Le premier sujet de cet axe concerne l'inférence de réseaux biologiques à partir des données de la littérature et des données -omiques par programmation logique, en collaboration entre les équipes Bioinfo et LaHDAK (co-encadrement d'une thèse, projet ANR ABLISS (2018-2022)). Un second sujet concerne l'apprentissage et la résolution de problèmes inverses pour la génétique des populations à partir des génomes anciens, en collaboration entre les équipes Bioinfo et A&O (co-encadrement d'une thèse). Un troisième sujet démarre sur le calcul de classements

consensus par programmation linéaire pour des données biomédicales, en collaboration entre les équipes Bioinfo et ROCS (projet interne LRI 2020, co-encadrement de stagiaires). Un quatrième sujet en cours d'étude concerne le croisement de concepts et méthodes de provenance utilisées pour la gestion des données et connaissances (expertise équipe LaHDAK) et pour la gestion de workflows biologiques (expertise équipe Bioinfo).

Un second axe de collaborations émergentes intra-département est centré sur la **l'exploitation conjointe d'expertise en gestion de données, analyse de données et modélisation mathématique**. Les domaines de la validation des connaissances (e.g., identification des doublons) ou des politiques de diffusions (e.g., maximisation de l'influence, apprentissage par renforcement) sont attaqués en hybridant les approches d'IA symbolique et de programmation mathématique (projet interne LRI LaHDAK - ROCS, démarré en 2020). D'autres sujets impliquant plusieurs équipes du département comprennent la détection de fraude à la carte bancaire ; l'optimisation des applications en IoT, et l'extraction d'information à partir de documents hétérogènes.

Au niveau des collaborations effectives avec d'autres départements, on peut citer le projet AIDA (financement Banque Publique d'Investissement, démarrage en 2019) qui s'intéresse à la fouille et la recommandation de processus métiers (collaboration entre les départements *Science des Données* et *Sciences et Technologies des Langues*).

Les équipes approfondissent par ailleurs leurs sujets de recherche, dans la continuité des perspectives décrites dans le rapport HCERES 2018.

Dans la continuité du projet du département, l'équipe **LaHDAK** se focalise sur la gestion des données et connaissances massives et hétérogènes suivant trois axes. L'axe **gestion des données** concerne l'étude des polystores, la diffusion d'information dans les réseaux sociaux et le Web, l'estimation et maximisation de l'influence, la recommandation et la fouille de graphes, les données incertaines, la provenance, l'évolution, la recommandation et le profil utilisateur. L'axe **gestion de connaissances** vise à contribuer à la gestion de l'identité dans le Web des données, l'enrichissement des données, la génération et l'évaluation de requêtes sur des ontologies, la sélection de connaissances pertinentes. Le troisième axe concerne la **fouille et découverte de connaissances** avec pour objets d'intérêt l'acquisition de règles et d'axiomes, les clés et clés conditionnelles, les expressions référentielles et les relations causales.

L'équipe **ROCS** travaille sur plusieurs axes importants en IoT en lien étroit avec le projet du département, destiné à devenir dans les prochaines décennies une des sources principales de données à stocker et analyser, en vue d'optimiser notre environnement et ses ressources physiques. Le premier axe porte sur l'**amélioration de la connectivité** des objets et la réduction de leur consommation énergétique dans la **5G et les réseaux locaux et ad hoc**. Le second vise le **développement de l'algorithmique distribuée pour la synchronisation des données** sur des Edge distribués et mobiles. Le troisième concerne le développement des mécanismes de **sécurité distribués tel que la blockchain** pour le stockage des données. Le cinquième porte sur l'**optimisation des processus pour de nouvelles applications** liées à l'IoT, e.g. la charge des véhicules électriques, la logistique, la santé, et la finance.

En lien avec le projet du département, l'équipe **Bioinfo** continue à développer des approches fondées sur les données qui conduisent à de nouveaux concepts et de nouvelles méthodes en informatique et en mathématiques pour résoudre des problèmes en biologie. Nous travaillerons principalement sur les trois grandes problématiques. La première concerne l'**analyse de séquences et structures biologiques**, incluant les statistiques et l'apprentissage pour la génétique des populations, les algorithmes de graphes pour l'étude des liens entre séquence et structure des macromolécules. La seconde, **Biologie des systèmes et biologie synthétique** s'attaque à la conception de méthodes formelles pour l'analyse des réseaux biologiques, de modèles logiques pour l'analyse de flux dans les réseaux métaboliques, de nano-ordinateurs biologiques pour la détection de marqueurs de pathologies. La troisième, l'**intégration et analyse de données biologiques**, a pour objets d'étude principaux les workflows scientifiques, la provenance des données et reproductibilité des résultats, les approches mêlant combinatoire et bases de données pour le classement consensuel de données.

Les projets de recherche de l'équipe **A&O** en lien avec le département *Science des Données*

concernent : (i) la **modélisation causale** (identification de graphes et de mécanismes causaux, conditions d'identifiabilité, estimation de confondeurs dans les problèmes à grand nombre de variables), (ii) l'**identification de biais** dans les données et dans les modèles appris ; (iii) les applications dans le domaine des ressources humaines (appariement et recommandations entre offres d'emploi et CVs ; nutrition et santé) et l'**exploration interactive de grandes masses de documents** (en lien avec les départements *Interaction avec l'Humain* et *Sciences et Technologies des Langues*, e.g. concernant Le Grand Débat ou le Challenge Kaggle des articles portant sur le Covid-19).

3.6 Département Sciences et Technologies des Langues

3.6.1 Thèmes

Le département *Sciences et Technologies des Langues* étudie des questions fondamentales relatives aux systèmes linguistiques, en exploitant de larges corpus collectés, annotés et enrichis de manière non-supervisée ou semi-supervisée. Des modèles d'apprentissage statistique, adaptés au matériau linguistique, sont utilisés pour exploiter ces corpus. En retour, ils permettent d'étudier le fonctionnement des langues, leurs variations (phonétiques-phonologiques, morphologiques-lexicales, syntaxiques et sémantiques) synchroniques et diachroniques, diaphasiques et diatopiques ainsi que de soulever des questions relatives à leur acquisition en tant que langues maternelles ou secondes. Enfin, le département développe les grandes applications du traitement des langues : reconnaissance vocale, traduction automatique, recherche d'information, agents conversationnels, ... qui portent des enjeux sociétaux (sauvegarde des langues en danger, outillage de personnes en situation de handicap, aide au traitement de l'information et de la connaissance médicale) et éthiques de plus en plus importants. Cette approche de la Langue et des langues couvre ainsi un large spectre, des recherches les plus fondamentales aux plus appliquées, dans une grande variété de médias (journaux, médias sociaux, vidéo, téléphone, ...) et toutes les modalités (écrit, parole et signe). Ces recherches sont fortement pluridisciplinaires et elles rassemblent des communautés diverses relevant de l'informatique, des sciences de l'ingénieur et des sciences humaines et sociales, qui se retrouvent autour de l'étude de la langue par l'exploitation de corpus et la culture de l'évaluation quantitative, deux marqueurs forts de leur activité de recherche.

3.6.2 Positionnement

Les travaux menés au sein du département s'attaquent à des données complexes, naturelles, hétérogènes et multidimensionnelles, de grande dimension et d'une grande variété. Le traitement des langues se trouve au cœur de l'intelligence artificielle, qui est abordée en utilisant l'ensemble des modèles qui fondent actuellement cette thématique. Ces travaux nécessitent des adaptations spécifiques des techniques d'apprentissage statistique, en particulier des modèles neuronaux et d'apprentissage profond, mais sans jamais perdre de vue que la finalité est l'étude de la langue et non les techniques d'apprentissage elles-mêmes. Notre apport spécifique concernant l'intelligence artificielle se porte vers une vision éthique, que ce soit dans sa dimension technique (explicabilité des décisions issues du *deep learning*), que ce qui concerne les problèmes potentiels apportés par les systèmes actuels, par exemple agents conversationnels (*bad nudging*) ou robotique émotionnelle, mais également impact environnemental. Cette responsabilité a débouché sur la présence de nombreux membres du département dans différents comités d'éthique chargés en particulier de réguler l'intelligence artificielle (Poletis Paris-Saclay, CERNA, CCNE, Comité d'éthique du numérique).

Le département *Sciences et Technologies des Langues* a pour ambition de contribuer au progrès des sciences et technologies de la langue. Fort de 80 personnes dont 30 CEC permanents, il constitue l'un des pôles majeurs de ce domaine en France et plus largement en Europe et est un acteur reconnu à l'international : projets internationaux (DARPA, NSF), projets européens (T4Me, QT21, MiRoR, ...), fellows ISCA et AMIA, rôle dans les associations savantes nationales et internationales (ISCA, ACL, ELRA, ATALA, AFCP, ...), etc. Le département se caractérise

par une approche globale de la langue dans tous ses états : langue parlée, écrite et signée. Les thématiques sont abordées de manière pluridisciplinaire, du traitement du signal à la sémantique : chercheurs en sciences de l'information (CNRS 07) et en sciences du langage (CNRS 34), collaborations avec les sciences du mouvement (CNRS 26), etc. Le département est fortement impliqué dans la mise en place et l'animation de deux GDR du CNRS, le GDR Traitement Automatique des langues (TAL, INS2I) qui couvre les différentes thématiques du département, et le GDR Linguistique Informatique, Formelle et de Terrain (LIFT, INSHS) plus orienté vers l'interface entre le traitement automatique des langues et la linguistique, et la documentation des langues. Elles impliquent également des collaborations très larges, aussi bien au sein du périmètre Paris-Saclay (on mentionnera en particulier la chaire (hors 3 IA) HUMAINE, CNRS – Institut DATAIA) où le thème « traitement des langues » occupe une place unique, que plus largement en France, en Europe et dans le monde. Ces thématiques sont extrêmement importantes pour le tissu industriel, avec lequel les équipes entretiennent des collaborations suivies. Capitalisant sur les résultats de ses recherches et sur les compétences de ses chercheurs, le département est également engagé dans une démarche d'innovation et valorisation (projets partenariaux, maturation, participation à la création d'entreprise, Institut Carnot).

3.6.3 Composition

Équipe TLP (Traitement du Langage Parlé). L'équipe TLP mène des recherches visant à comprendre les processus de communication humaine par la parole et à développer des modèles pour le traitement automatique de la parole. Ses travaux de recherche sont interdisciplinaires par nature, allant du traitement du signal à la sémantique. L'un des principaux objectifs est l'analyse automatique de la parole visant une description complète du signal audio, qu'il s'agisse d'analyser la scène audio, de détecter la présence de parole, d'identifier la langue parlée, de caractériser le ou les locuteurs (leur identité et détecter leur état émotionnel), de transcrire la parole en texte, d'identifier des entités spécifiques dans la transcription, ou encore de traduire la parole. Les domaines applicatifs ciblés sont par exemple l'audiovisuel (sous-titrage, indexation, et traduction), l'interaction dialogique (robotique) et la sécurité (identification et analyse). Les travaux menés concernent également les corrélats linguistiques des mesures observées sur des grands corpus, qu'il s'agisse de la variation phonétique synchronique, des marqueurs dialectaux, des changements diachroniques ou encore des facteurs extra-linguistiques (par exemple, les émotions) actualisés par des traits spécifiques.

Les recherches de l'équipe sont actuellement structurées en six thèmes interdépendants : Caractérisation du locuteur dans un contexte multimédia ; Dimensions affectives et sociales des interactions parlées avec des (ro)bots et enjeux éthiques ; Perception et traitement automatique de la variation de la parole ; Traduction et apprentissage automatiques ; Reconnaissance de la parole ; Ressources langagières.

Équipe ILES (Information Langue Écrite et Signée). L'équipe ILES se consacre au traitement automatique des langues écrites et signées. Elle étudie la sémantique et les informations véhiculées par des énoncés langagiers, comment les représenter, les détecter, ou encore les générer. Les méthodes employées vont de la modélisation à l'apprentissage automatique. L'équipe travaille sur la modélisation des représentations associées aux traitements, qui s'instancie dans la construction de corpus annotés. En langue des signes, cela repose sur une modélisation en linguistique et en sciences du mouvement. Les axes clés sont la compréhension et la transformation des productions langagières. Ils s'instancient plus spécifiquement dans l'analyse automatique d'énoncés pour extraire, représenter ou accéder aux informations qu'ils contiennent ou dialoguer avec un utilisateur, ainsi que dans le traitement de la similarité textuelle (synonymes, paraphrases, simplification, implication). En langue des signes, ces axes sont complétés par une problématique de génération d'énoncés signés. Ces travaux s'appuient sur l'apprentissage de représentations continues, notamment en exploitant conjointement des textes et des bases de connaissances. Ils se placent souvent dans un contexte où les données sont en quantité insuffisante, ce qui motive l'étude et la conception

de méthodes pour pallier ce manque. Le domaine biomédical est un champ applicatif important pour l'équipe.

Les activités de recherche de l'équipe ILES s'organisent autour des quatre thèmes suivants : Corpus et Représentations ; Multilinguisme et Paraphrase ; Extraction et reconnaissance d'informations précises ; Modélisation et Traitement Automatique des Langues des Signes.

3.6.4 Projet

Le domaine du traitement des langues a beaucoup évolué lors des vingt dernières années, aussi bien dans ses aspects technologiques que ses aspects plus fondamentaux. Cela est dû à de nombreux facteurs, dont deux principaux. Le premier, endogène, est le succès de nos recherches (en tant que communauté) où le traitement du langage quelque soit sa modalité est devenu incontournable des applications de tous les jours, sur son ordinateur ou sur son smartphone, au point d'être un paradigme du succès de l'intelligence artificielle ; cela a conduit les grands acteurs industriels, et en particulier les GAFAM⁴, à investir ces champs de recherche, avec des moyens sans commune mesure avec ce que la recherche publique peut faire. Cela a également conduit le traitement des langues à investir de plus en plus de domaines, comme moyen d'accéder au sens de données massives multimodales.

Le deuxième facteur exogène est l'évolution de la société qui, du fait de la diffusion des outils du numérique (ordinateur, téléphone), a rendu l'usage de certaines applications quasiment obligatoire de fait, en mettant sous le boisseau le fait que le traitement de beaucoup de langues (langues des signes, langues en danger, langues minoritaires) ou de variétés de langues (accents, parole non native) n'est pas existant ou opératoire.

Il n'est plus possible, dans le choix des champs de recherche, qu'elle soit fondamentale ou appliquée, d'oblitérer cette situation. Il est nécessaire, en utilisant le savoir-faire et le savoir-penser acquis au sein du laboratoire, dans tous les champs du traitement des langages, de se repositionner sur des domaines en accord avec l'état du monde et de la société, en conservant notre spécificité, de reposer les recherches sur un matériau réel (le corpus) avec une approche pragmatique, et sur l'évaluation objective des résultats. Pour cela, trois objectifs globaux de recherche ont été définis.

Le **premier objectif**, méthodologique, correspond à l'évidence que notre champ de recherche se trouve au coeur de l'**intelligence artificielle**, qui a bénéficié des progrès spectaculaires liés à l'apprentissage machine comme l'apprentissage profond et les modèles neuronaux, avec une faune complexe et ardue de modèles, qui nécessitent des « zoologistes » compétents à même d'accompagner et de maîtriser les avancées les plus récentes. Il faut conserver cette compétence, en l'augmentant afin de couvrir les nécessités d'éthique et de déontologie scientifique (biais, explicabilité,...) qu'impose notre statut de scientifiques.

L'un des problèmes posés par les modèles statistiques mis en œuvre dans tous les systèmes de traitement statistiques et en particulier les modèles neuronaux est la qualité et le coût des données d'apprentissage. Le département, qui a été un précurseur sur les méthodes d'apprentissage peu supervisé, en particulier pour le traitement de la parole, étend l'usage de ces méthodes pour pallier l'insuffisance des données d'apprentissage et des annotations associées. Afin de rendre possible l'usage de très grands corpus non annotés, hétérogènes et multimodaux, il s'agit de réduire encore le degré de supervision de l'apprentissage complété avec un apprentissage actif visant à sélectionner les données les plus utiles, afin d'élaborer des modèles multidomaines ou multilingues.

Les corpus annotés sont un matériau indispensable pour l'étude et la compréhension des phénomènes en jeu et pour la conception de représentations pertinentes, ainsi que pour la mise au point et l'évaluation de méthodes de traitement automatique. Les ressources linguistiques s'ajoutent aux méthodes dirigées par les données pour les compléter et les renforcer. Ces thématiques, de même que les mesures d'évaluation et les campagnes d'évaluation resteront des activités fortes du département. Ce premier objectif, qui se fonde sur des activités et des compétences largement partagées avec les départements *Science des Données* et *Algorithmes, Apprentissage et Calcul*, sera mené en mutualisant notre savoir-faire sur ces thématiques.

4. Google, Apple, Facebook, Amazon et Microsoft

Le **deuxième objectif** est amené par l’ouverture du champ des possibles dans le domaine des langues et variétés de langues traitées, ce que l’on peut résumer par la formule « **la langue dans tous ses états** ».

L’exploration de la variation dans la parole doit se poursuivre à travers une approche résolument interdisciplinaire de linguistique de corpus, en matière de changements phonétiques, de prosodie expressive, de styles de parole, d’accents, et de dialectes. Elle doit prendre en compte des questions fondamentales relatives à la structure et à l’évolution des langues, et à l’interaction des différents niveaux linguistiques, du signal de parole au discours et à l’expression des affects. En particulier les travaux de cartographie linguistique qui ont débuté sur certaines langues romanes, doivent s’élargir à différentes familles de langues, en combinant enquêtes de terrain, techniques de « crowdsourcing » (sujet commun avec le département *Science des Données*), et expériences de perception et analyse de données au moyen du traitement automatique.

Dans nos mondes numériques, une langue sans outil est une langue en danger. Face à l’hégémonie de l’anglais, il importe de rappeler que le multilinguisme est partout la norme. Il importe donc d’accélérer le développement d’outils pour le français mais également pour toutes les langues avec lesquelles il coexiste, les langues régionales comme les autres langues de l’espace européen et mondial. Il convient également d’accélérer la documentation des nombreuses langues en danger.

Les travaux en langue des signes française (LSF) se poursuivront selon les axes de la modélisation (linguistique et en sciences du mouvement), de l’élaboration de ressources et d’outils pour les manipuler (par exemple l’aide à l’annotation par traitement d’images) et sur les plans de la reconnaissance et de la génération.

La variété de matériau linguistique soulève des problèmes récurrents en traitement automatique des langues et en apprentissage : ceux de la généralité et de la robustesse des représentations et des traitements. Les méthodes d’adaptation au domaine ainsi que les approches non-supervisées ou économes en annotation manuelle présentent de ce fait un grand intérêt.

Enfin, pragmatiquement, il faut regarder stratégiquement quels peuvent être les principaux **apports scientifiques et technologiques du département Sciences et Technologies des Langues** ; c’est le **troisième objectif**. En particulier il convient de positionner le département soit sur des domaines ou des applications où il est possible de jouer à armes égales avec les GAFAM, soit sur des domaines qu’ils ne ciblent pas ou peu. On s’intéressera par exemple à des applications de sécurité et défense, ou aux activités sur la linguistique de corpus. Dans le domaine médical, hautement stratégique, le département bénéficie d’une compétence particulière, et (pour l’instant) d’une volonté étatique de souveraineté dans le domaine de la santé et du soin. Les travaux sur les méthodes permettant d’assurer une meilleure robustesse lors des changements d’échelle, d’adaptation au domaine seront poursuivis.

Pour les travaux sur le dialogue humain-machine, thématique commune entre les départements *Sciences et Technologies des Langues* et *Interaction avec l’Humain*, où les données disponibles sont rares et très coûteuses à obtenir, il s’agit de proposer et d’évaluer des méthodes de génération automatique de ces données, ainsi que de mettre au point des architectures hybrides permettant, lorsque cela est possible, de s’appuyer sur des méthodes robustes mais gourmandes en données (approches à base d’apprentissage profond) tout en injectant des modèles a priori fondés sur des connaissances. Cela implique également le renforcement significatif des interactions entre l’humain et la machine, notamment pour permettre l’apprentissage de connaissances auprès des utilisateurs dans le cadre des systèmes de dialogue.

L’interaction affective et sociale avec des machines en liaison avec l’éthique (portée par l’action transverse Recherche responsable) est un sujet nouveau que va porter le département *Sciences et Technologies des Langues*. Cette activité s’appuiera en particulier sur la chaire HUMAINE⁵. Il s’agit d’auditer l’influence potentielle des systèmes de dialogues affectifs sur les humains et leurs pouvoirs de manipulation afin d’aller vers une conception de “systèmes éthiques par conception”. Les travaux scientifiques interdisciplinaires de la chaire vont se concentrer sur la détection des émotions sociales dans la voix humaine, et sur l’étude des manipulations langagières (les “nudges”). Cette thématique se développera en direction de personnes fragiles (enfants, personnes âgées,

5. CNRS – Institut DATAIA : HUman MAchine Affective INteraction & Ethics

personnes ayant des troubles psychiatriques ou du développement) pour qui ces applications sont d'un intérêt sociétal majeur.

Enfin, les travaux sur l'analyse de scènes audio dans le but de générer une description complète du signal audio incluant l'identification d'événements et d'environnements acoustiques en complément de la transcription de la parole et de la caractérisation des locuteurs seront renforcés.

3.7 Actions Transverses

Les actions transverses mettent en évidence des thèmes de recherche à la croisée de plusieurs départements. Elles se fondent sur les coopérations déjà existantes ou en projet entre les membres de différents départements. Elles feront l'objet d'une attention particulière dans le LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE (e.g. financements incitatifs), visant l'émergence de directions nouvelles et permettant de tirer parti de l'originalité et de la diversité des expertises des membres du LISN. Cinq actions transverses sont décrites ci-après, reflétant l'état actuel des coopérations, projets en cours et échanges. Elles ne sont pas toutes au même stade de développement, mais nous en donnons ici une première description. Elles n'ont pas vocation à être pérennes, et elles seront régulièrement évaluées par le conseil scientifique du LISN. Leur nombre et leur périmètre peuvent donc être amenés à évoluer au cours du quinquennat.

Apprentissage Profond pour la Physique et Physique pour l'Apprentissage. L'apprentissage automatique (Machine Learning, ML) fournit un ensemble de méthodes de traitement de l'information ouvrant des perspectives importantes pour les sciences expérimentales en général, et pour la recherche en mécanique des fluides en particulier. Un certain nombre d'objectifs (tels que la réduction de modèle, le contrôle ou l'optimisation de forme) peuvent en effet être accélérés par l'usage de modèles non-linéaires approchés, appris à partir des données de grande dimension disponibles. Les retombées du ML en mécanique des fluides peuvent opérer à plusieurs niveaux, tels que la modélisation et/ou l'assimilation de données, le couplage simulation-modélisation, les modèles de fermeture, l'accélération des simulations, l'identification des instabilités numériques.

Inversement, la mécanique des fluides, et la Physique d'une manière générale, peut contribuer à plusieurs thématiques de recherche en ML. Une première thématique concerne la définition de contraintes *a priori* (comme l'invariance du modèle vis-à-vis de certains groupes de symétrie), ou la garantie de certaines propriétés ou bornes sur les prédictions (certification, robustesse). Par ailleurs, les données issues de simulations physiques étant abondantes, sans bruit et contrôlables, elles peuvent définir des *benchmarks* de référence, inspirant le développement de nouvelles architectures en ML, et permettant d'explorer les conditions nécessaires pour pouvoir apprendre à partir de *petites données*.

Ce thème transverse vient de démarrer; il réunit l'équipe A&O (départements *Science des Données* et *Algorithmes, Apprentissage et Calcul*) et l'équipe DATAFLOT (département *Mécanique des Fluides-Énergétique*), et fait également l'objet d'une collaboration avec l'IFPEN ainsi qu'avec l'IRT SystemX (thèse co-encadrée).

Sciences du numérique et SHS. Les domaines de recherche des sciences humaines et sociales (SHS) couvrent un large spectre de disciplines incluant, entre autres, la psychologie, l'économie, la sociologie et la linguistique, qui sont des disciplines représentées dans les activités du LISN. Les SHS ont pour but d'étudier divers aspects de la réalité humaine sur le plan de l'individu et sur le plan collectif. Elles cherchent à expliquer des phénomènes au confluent du biologique, du comportement, du cognitif et du social.

Les SHS sont partie prenante des nouvelles technologies de l'information pour déployer de nouvelles orientations théoriques et pratiques et des méthodes innovantes. Comprendre les métamorphoses qu'impose l'introduction généralisée du numérique à nos sociétés et à notre relation au monde représente un enjeu sociétal majeur. Du fait de sa culture interdisciplinaire, le LISN s'avère particulièrement bien pourvu pour relever ce défi scientifique. Plusieurs facettes peuvent

être considérées. Celles qui sont proposées ici ne sont que les plus avancées parmi l'ensemble des possibilités sur lesquelles nous travaillons au LISN.

La première concerne l'apport croisé **de la linguistique et du traitement des données** massives. Il s'agit ici de mutualiser des compétences pour d'une part produire des avancées méthodologiques et pratiques quant à la production de connaissances en linguistique en utilisant des données massives, d'autre part en s'intéressant à des données pour lesquelles nous disposons de connaissances *a priori*, plus ou moins structurées, dont il faut apprendre à tenir compte.

Une autre facette concerne plus spécifiquement l'utilisation du numérique pour aider à mieux comprendre l'humain ou à l'aider dans ses interactions, que celles-ci soient sociales (avec d'autres humains) ou avec des outils numériques. Il s'agit ici de renforcer les liens entre **Sciences cognitives, psychologie, ergonomie et sciences du numérique**.

Une troisième facette porte sur **l'économie et la sociologie du numérique**. Il s'agit ici de porter un nouveau regard sur des grands problèmes sociaux, comme le fonctionnement des marchés du travail mais aussi l'usage des technologies numériques, touchant à des questions comme la vie privée, les modes de consommation, l'accès à l'information et les conditions de travail, ou plus récemment sur les effets socio-économiques de la crise sanitaire due au Covid-19.

Cette action regroupe des recherches menées dans plusieurs départements (*Sciences et Technologies des Langues, Science des Données et Interaction avec l'Humain*), et permet de mettre en lumière l'ancrage du LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE dans les SHS, avec plusieurs thèses co-encadrées par des chercheurs en informatique et en SHS, et des projets de recherche transverses. Ces travaux sont aussi l'occasion de liens forts avec la Maison des Sciences de l'Homme de Paris-Saclay au niveau local, ainsi que dans le TGIR (Très Grande Infrastructure de Recherche) Huma-Num au niveau national.

Arts et Sciences. Cette action a pour objectif de favoriser les collaborations entre la recherche scientifique et les domaines du design, de la création artistique contemporaine et de la médiation. Au-delà de la simple valorisation d'applications artistiques de sujets scientifiques, le but est de développer une approche en recherche et création pouvant servir comme catalyseur de travaux interdisciplinaires et comme incubateur de projets sur des thèmes émergents ou exploratoires. Il s'agit d'une part de faciliter et promouvoir des recherches collaboratives entre les sciences du numérique, les sciences de l'ingénieur et les domaines de la création artistique contemporaine (arts visuels, danse, musique, théâtre et performances) incluant aussi le design et l'architecture. D'autre part, l'action vise à organiser des événements sur les thématiques arts-sciences et d'alimenter la réflexion autour de ces sujets. Les actions arts-sciences peuvent comporter une dimension médiation/vulgarisation de la science, qui pourrait se développer également dans le cadre de cette action.

Cette action s'appuie sur une dynamique déjà existante aussi bien au LIMSI qu'au LRI, au sein desquels de nombreux projets ont été menés ces dernières années. Au LIMSI, l'action transverse VIDA a recensé, sur la période 2013-2018, 23 projets arts-sciences donnant lieu à 50 publications et plusieurs dizaines d'événements à destination du grand public. De nombreux projets arts-sciences et de médiation ont également été menés au LRI, notamment en collaboration avec le LIMSI.

Il s'agit donc de poursuivre et d'étendre cette action dans le cadre du LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE, afin de mieux faire apparaître la diversité des problématiques des projets et les approches conceptuelles développées. Cette action vise à renforcer les aspects recherche et création suivants : (a) l'étude des processus créatifs en Arts & Sciences (maturation de projets, émergence et co-construction de nouvelles problématiques) ; (b) l'échange des pratiques entre artistes et scientifiques (pour l'innovation technologique, le développement d'approches interdisciplinaires en sciences) ; (c) la démarche de formation auprès des étudiants en sciences et en arts (mise en place d'écoles d'été ou d'ateliers Arts & Sciences) et de médiation dans le dialogue sciences et société (festivals, conférences) ; et enfin (d) l'approche réflexive en lien avec les sciences humaines dans le questionnement éthique sur le rôle et l'impact de l'outil numérique dans la co-évolution humain-machine, en lien avec la recherche responsable.

L'action s'appuiera notamment sur la forte dynamique en place sur le plateau de Saclay, en

particulier en coordination avec La Diagonale Paris-Saclay (qui vise à doter le campus de Paris-Saclay d'un grand centre de dialogue entre arts, science et société), le LadHyX (co-porteur depuis 2017 avec l'ENSAD de la chaire art-science de la Fondation Carasso), l'ENS-Paris Saclay dans le parcours Design, le SAS (Science-Art-Société) et le neticLab à l'Université Paris-Saclay.

Visualisation et exploration de grosses masses de données. Cette action transverse vise à mettre en commun des corpus de données, tels que manipulés dans les départements *Science des Données*, *Mécanique des Fluides-Énergétique* ou *Sciences et Technologies des Langues*, des algorithmes, tels que développés dans le département *Algorithmes, Apprentissage et Calcul*, et des outils de visualisation de données développés dans le département *Interaction avec l'Humain*. Elle fédère ainsi l'ensemble de l'activité du LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE autour de la manipulation par les utilisateurs, à travers les outils numériques, des grandes masses données produites par la société (grand débat, épidémie) et les laboratoires (simulation numérique par exemple). L'objectif de cette action est ainsi de mettre en avant l'excellence, au niveau international, de notre unité dans ce domaine de l'exploration et de la visualisation de données. Elle s'appuie pour cela sur le potentiel unique de mise relation des fournisseurs de données (données complexes, données volumineuses et/ou données critiques) avec des experts en visualisation qui peuvent proposer plusieurs niveaux de collaboration (du simple conseil jusqu'à la co-écriture de projet de recherche), et sur les moyens de visualisation interactive et collaborative de l'Equipex DIGISCOPE.

L'action transverse a une double mission au sein du LISN : 1) renforcer les collaborations entre les équipes qui produisent et manipulent des grandes masses données et les problématiques d'interaction humain-machine pour leur visualisation, et 2) constituer un inventaire illustré de systèmes opérationnels, en mettant à contribution l'ensemble du laboratoire pour les évaluations contrôlées de systèmes. La méthodologie dominante du domaine consiste en effet à exposer un large panel d'utilisateurs à des systèmes de visualisation et de récolter performances et préférences dans une démarche comparative. En proposant la participation à nos expérimentations aux membres du LISN, nous augmentons nos cohortes d'utilisateurs tout en faisant la promotion de nouveaux systèmes qui peuvent susciter des questionnements et des collaborations.

L'action est déjà concrétisée par plusieurs collaborations déjà effectives, entre le département *Interaction avec l'Humain* et les départements *Mécanique des Fluides-Énergétique* *Sciences et Technologies des Langues* et *Science des Données* respectivement, comme par exemple le projet Cartolabe (<https://cartolabe.fr>), ou le projet d'atlas des langues (<https://atlas.limsi.fr/>). D'autres collaborations sont à venir dans un futur proche.

Recherche responsable Comme indiqué dans le document de description des services du LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE, l'unité est dotée d'une cellule "Science Responsable" de soutien à la recherche avec trois piliers : développement durable, science ouverte et éthique. La cellule a pour objectif d'établir des bilans (état des lieux et recommandations) concernant les pratiques du laboratoire, de fournir des informations sur les bonnes pratiques (formations, actions de sensibilisation...) et de proposer un guichet d'orientation pour guider les chercheurs lors du montage d'un projet et les adresser à des structures dédiées existantes capables de répondre à leurs interrogations. Ces trois thématiques possèdent un versant recherche de plus en plus important qui est traité dans l'action "Recherche responsable" décrite ici. Cette action regroupe l'évaluation de l'éthique des sciences du numérique, l'impact des sciences du numérique sur l'environnement et les approches et outils pour la science ouverte.

La puissance de l'IA et notamment de l'apprentissage machine et son utilisation massive dans beaucoup d'objets numériques posent des enjeux inédits, et pourraient conduire à des dérives, que les différents acteurs et plus largement la société doivent anticiper. L'**éthique de l'IA** couvre en fait plusieurs aspects que les chercheurs doivent intégrer dans leurs études, créations de solutions et sur leur communication. Le développement des outils numériques (agents conversationnels, robots sociaux et tous les systèmes de décisions fondés sur des approche neuronales) et leur déploiement doivent se faire dans un contexte qui respecte nos valeurs éthiques. Ces valeurs telles que l'équité,

la non-discrimination, la loyauté, le nudge (incitation) pour les systèmes d'apprentissage à base de réseaux de neurones et les corpus utilisés sont un sujet de recherche qui intéresse les départements *Sciences et Technologies des Langues* et *Science des Données*. Deux chaires ont été obtenues sur ces sujets : HUMAINE⁶ et HUMANIA⁷. D'autres projets sont en cours de dépôt ou de montage suite au COVID-19 comme le projet KANOPEE⁸ sur l'évaluation éthique des chatbots développés pour surveiller la dépression, l'addiction de personnes fragiles dans l'après confinement ou un projet de recherche pédagogique (en cours de montage entre *Sciences et Technologies des Langues* et *Science des Données*) sur l'évaluation des risques liés au confinement pour les étudiants isolés à partir d'entretiens oraux et écrits en évitant les biais générés par les données. Dans cette recherche sur l'éthique, l'objectif de recherche est également de démystifier et démocratiser ces sujets auprès du plus grand nombre.

Compte tenu de l'omniprésence du numérique, il est plus que jamais indispensable de comprendre comment nos usages des outils numériques impactent l'environnement. Il s'agit ici de l'étude des **impacts directs du numérique sur l'environnement**, en considérant les impacts énergétiques mais également les problèmes de pollution, d'épuisement des ressources, etc. Des méthodologies d'évaluation d'impact environnemental existent, mais ne sont pas toujours adaptées au numérique, et les données sont difficiles à obtenir dans ce domaine. En outre, les effets indirects, notamment les effets rebonds, ne sont généralement pas pris en compte alors qu'ils sont très importants pour le numérique. Le pendant de cette étude concerne la réduction des impacts environnementaux du numérique : développement de systèmes sobres, c'est-à-dire ayant un impact minimal sur l'environnement, et résilients, c'est-à-dire capables de fonctionner en cas de difficulté d'accès à certaines ressources (énergie, réseau mondial...). Cette action concerne des chercheurs issus des différents départements, et sa structuration est en cours.

Le troisième pan de cette action concerne les recherches associées à la **science ouverte** qui vise à construire un écosystème dans lequel la science est plus reproductible, donc plus cumulative, plus fortement étayée par des données, plus transparente, et d'accès plus universel. Les défis à relever sont nombreux depuis la conception de plate-formes d'échanges de données capables de supporter des volumes de données très importants dans un contexte hautement distribué jusqu'à la conception de solutions pour indexer, comparer, annoter de très larges jeux de données. Les défis se situent également dans la conception de pipelines d'analyse reproductibles qu'il convient de pouvoir échanger, re-exécuter, indexer et comparer. Ces problématiques sont au cœur des activités du département *Science des Données* mais nécessitent des collaborations étroites avec les départements *Sciences et Technologies des Langues* et *Algorithmes, Apprentissage et Calcul*. Construire des pipelines reproductibles d'annotation de données par extraction d'informations dans les textes pose des défis à la croisée du TAL et de la science des données. De même les données de sciences ouvertes, comme par exemple les données LOD (Linked Open Data) forment de très grands graphes de données dont la structure est particulièrement complexe. Naviguer dans ces graphes ou les interroger efficacement pose des problèmes d'algorithmique des graphes particulièrement importants à traiter en collaboration avec le département *Algorithmes, Apprentissage et Calcul* et des problèmes de visualisation interactive à traiter en lien avec l'action Visualisation décrite ci-dessus. Un projet interne au LRI entre *Science des Données* et *Algorithmes, Apprentissage et Calcul* dans ce contexte porte sur la recherche de clés de données dans les grands graphes RDF du LOD.

3.8 Insertion dans l'écosystème de l'Université Paris-Saclay

Le LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE est par nature membre du Département STIC de l'Université Paris-Saclay et de l'Ecole Doctorale STIC, tous deux ac-

6. associant informaticiens et linguistes avec des économistes comportementalistes du RITM (Paris-Saclay), se concentre sur l'étude des manipulations par les "nudges" jouant sur les biais cognitifs des sujets

7. chaire Université Paris-Saclay qui propose d'automatiser des solutions de Machine Learning en s'attachant notamment à éviter les problématiques de biais générés dans les données

8. déposé à l'Agence de l'innovation de défense, avec le CHU de Bordeaux en partenariat *Sciences et Technologies des Langues*

tuellement dirigés par des Professeurs du LISN. Il est également membre du Département MEP de l'Université Paris-Saclay et de l'Ecole Doctorale SMEMAG.

Ainsi que dit dans l'introduction, le LISN participe aux Graduate Schools *Informatique et Sciences du Numérique, Ingénierie et Systèmes et Sport, Mouvement et Facteurs Humains* de l'Université Paris-Saclay.

L'implication pluridisciplinaire du LISN à la frontière des sciences du numérique, des sciences pour l'ingénieur et des sciences humaines et sociales s'inscrit également dans nos participations à plusieurs initiatives et infrastructures de Paris-Saclay, mentionnées dans la présentation des divers départements. Pour récapituler :

- Le Labex *Digicosme* est un laboratoire d'excellence en Sciences du Numérique qui couvre trois grandes thématiques au cœur de la conception des systèmes d'information et de communication de demain : la fiabilité et la sécurité, la gestion des réseaux et la science des données. La plupart des équipes du LISN participent aux activités de Digicosme, qui finance prioritairement des projets collaboratifs entre équipes et laboratoires du Labex.
- L'Institut de Convergence *DataIA* est un institut national s'intéressant aux disciplines fondamentales au cœur de l'intelligence artificielle (IA), ainsi qu'à ses impacts et interactions avec les sciences expérimentales (e.g., physique, biologie, chimie) et les sciences humaines et sociales (e.g., économie, droit, sociologie) ;
- Le Labex *LaSIPS* a pour objectif de favoriser le développement d'activités de coopération scientifique innovantes et transversales dans le domaine des sciences de l'ingénierie et des systèmes, à travers trois pôles d'ingénierie : mécanique, électrique et biologique. Ses actions visent l'avancée des connaissances par synergie entre disciplines, à travers des collaborations scientifiques entre les laboratoires membres du LaSIPS, ainsi qu'avec les autres acteurs du plateau de Saclay.
- Le *Center for Data Science* de Saclay est une initiative de recherche stratégique (2016-2020) dont l'objectif est de mettre l'état de l'art en IA et en apprentissage (ML) à la disposition des laboratoires de sciences expérimentales et/ou humaines et sociales de Paris-Saclay.
- La *Maison des Sciences de l'Homme* de Paris-Saclay réunit les laboratoires de sciences humaines et sociales du Campus, avec le but de développer l'interdisciplinarité entre les SHS et les autres sciences, pour relever les deux défis majeurs que sont la transition numérique et la transition écologique. Un troisième axe transverse concerne les dimensions de l'innovation.
- La *Fédération Demenÿ-Vaucanson* (FéDeV) est une structure fédérative de recherche qui a pour objectif de stimuler les collaborations et la promotion de projets communs dans le champ général des sciences du mouvement du périmètre de Paris-Saclay.
- L'Equipex *Digiscope* est un réseau de 10 plateformes de visualisation et d'immersion interconnectées par un réseau de téléprésence et installées sur le campus Paris-Saclay, ainsi qu'un laboratoire de fabrication numérique (FabLab). Trois de ces plateformes (EVE, WILD, WILDER) et le FabLab sont installés dans les locaux du LISN, et le responsable scientifique du projet est un professeur du laboratoire. Ces plateformes sont utilisées par plusieurs équipes du département *Interaction avec l'Humain*.
- L'équipement de recherche *SaclayIA* apporte à la communauté scientifique et l'écosystème industriel d'Ile-de-France de moyens de calcul indispensables dans la compétition internationale en Intelligence Artificielle. Il inclut la plateforme LabIA, installée dans les locaux de l'IDRIS, qui est utilisée par plusieurs équipes des départements *Sciences et Technologies des Langues, Science des Données et Mécanique des Fluides-Énergétique*.

Les liens avec les autres laboratoires liés aux sciences du numérique de Paris-Saclay, également mentionnés dans les sections précédentes, sont forts et diversifiés. Citons de manière non exhaustive les liens avec le futur LMF (Université Paris-Saclay, ENS Paris-Saclay, CNRS) sur la certification de modèles IA par exemple, avec le L2S (Université Paris-Saclay, Centrale-Supélec, CNRS) sur le traitement du signal, avec le laboratoire MaIAGE (INRAE) et le CEA-LIST dans le domaine du traitement du langage naturel et de la recherche d'information, ou encore avec le laboratoire de

mathématiques d'Orsay (LMO), ainsi qu'à l'Institut Polytechnique de Paris avec le LIX, le CMAP, le LTCI et l'ENSTA en algorithmique, en optimisation, en systèmes et réseaux et en interaction humain-machine.

Le laboratoire a également développé des collaborations avec des laboratoires des Sciences de l'Ingénieur autour de la modélisation physique des systèmes, avec des applications pour la mécanique des fluides et l'énergétique. Des partenaires privilégiés sont le FAST (Université Paris-Saclay), le laboratoire EM2C (Centrale-Supélec), le GEEPS (Université Paris-Saclay, Centrale-Supélec, CNRS, Sorbonne Universités), ainsi que le LADHYX, l'IMSIA et le LMD (Institut Polytechnique de Paris), l'ONERA (DAFE) et différents laboratoires du CEA.

Les collaborations à l'interface des Sciences de la Vie sont nombreuses, notamment avec des laboratoires du campus en sciences du numérique (certains déjà cités plus haut) mais aussi avec des équipes de Génomique et des unités de recherche en Sciences de la Vie comme l'I2BC (Université Paris-Saclay, CEA, CNRS).

Enfin, dans le domaine des SHS, nous entretenons des collaborations de longue date avec le CIAMS (GS sciences du sport et du mouvement humain), et le RITM en économie du numérique.

Chapitre 4

LISN : organisation, structuration, fonctionnement

4.1 Instances

L'organisation du LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE est structurée par différentes instances qui sont présentées ci-dessous : la direction, le conseil de laboratoire, le directoire, le conseil scientifique.

- **La direction** est composée du directeur, du directeur adjoint et du responsable administratif. Elle se réunit toutes les semaines.
- **Le conseil de laboratoire** est une instance officielle de l'unité dont les missions sont précisées dans la Décision numéro 920368 OSI du 28 octobre 1992¹. Il se réunit tous les mois. Les membres de la direction sont membres de droit. Il s'agit d'une instance consultative très importante dans la vie du laboratoire. Il émet un avis sur toutes les questions relatives à la politique scientifique, à la gestion des ressources humaines et financières, aux moyens matériels, à l'organisation et au fonctionnement du laboratoire.
- **Le directoire** est composé de la direction, des responsables de département et des responsables de services. Il peut être élargi à des membres invités pour des questions particulières. Il assiste la direction dans l'impulsion et la conduite de la politique scientifique et des affaires administratives et financières et techniques du laboratoire. Il se réunit chaque mois.
- **Le conseil scientifique** est l'instance de discussion privilégiée concernant la vie scientifique et les grandes orientations scientifiques du laboratoire. C'est là que les décisions concernant les priorités scientifiques du laboratoire peuvent être discutées. Il est composé de la direction, de tous les responsables d'équipes ou de groupes, et des responsables de département. Il se réunit au moins quatre fois par an. Il peut être élargi à des membres invités pour des questions particulières.

4.2 Organigramme

L'organigramme cible de la future unité est présenté dans la figure 4.1. Outre la direction, qui se compose du directeur, du directeur adjoint et du responsable administratif, le laboratoire est composé de cellules, services et départements. Des actions transversales sont également prévues. Les sections suivantes décrivent la structure du laboratoire.

1. http://www.dgdr.cnrs.fr/elections/conseil_unite/textes/DEC920368S0SI_%20Consolidee0316.pdf

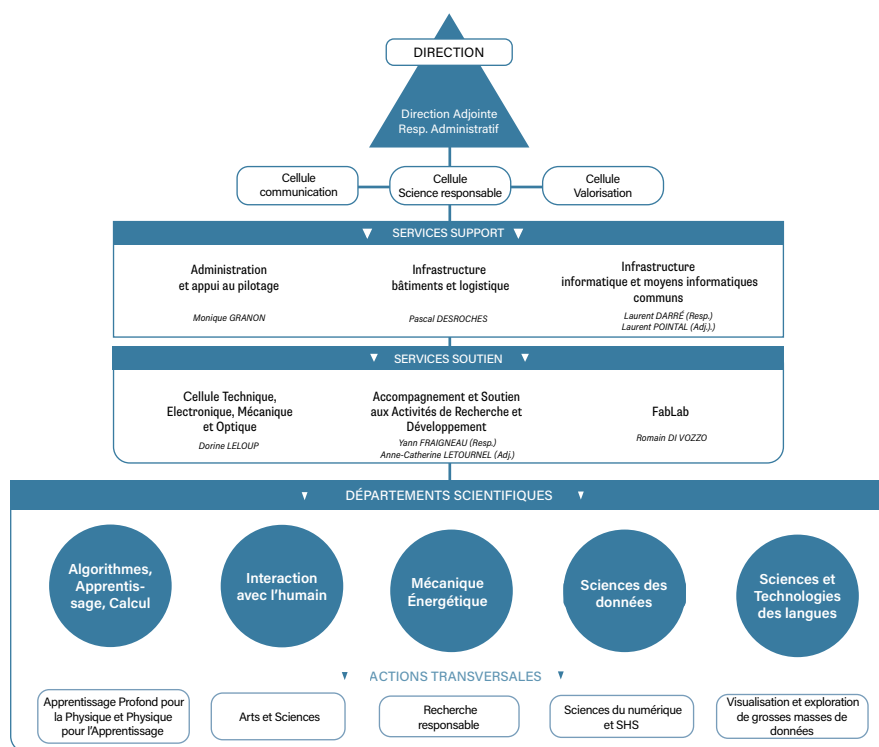


FIGURE 4.1 – Organigramme cible

4.3 Structuration scientifique

Le laboratoire est structuré en départements, eux-mêmes constitués d'équipes de recherche. La recherche se fait principalement au sein des équipes. Les départements permettent de structurer le laboratoire, en directions scientifiques, avec possiblement des actions transversales.

4.3.1 Les départements

Chaque département s'organise autour d'un bureau de département (comprenant au moins le responsable de département et les responsables d'équipe) et d'un conseil scientifique. Les membres du bureau du département ont la capacité de se faire représenter si besoin.

L'organisation de chaque département reflète son autonomie en termes de directions scientifiques et de gestion du budget.

Un membre du laboratoire est nécessairement rattaché à une équipe mais dans une situation transitoire il peut être rattaché à un département.

Le responsable de département coordonne la politique et la vie scientifique au sein de son département (séminaires, actions transversales entre équipes, etc.). Il organise au besoin la tenue d'un conseil scientifique de département.

Il fait le lien entre les membres du département et la direction quand cela est nécessaire. Il anime les discussions sur la stratégie et la planification des besoins à moyen terme, comme les postes par exemple, et fait remonter une vision pluriannuelle des besoins à la direction, et des synthèses des activités du département. Il représente le laboratoire dans le périmètre de son département là où c'est nécessaire.

4.3.2 Les équipes

Le responsable d'équipe anime et organise la vie quotidienne de l'équipe en terme de séminaires, réunions, etc. Il participe aux discussions de départements, et fait remonter, après discussion en interne, les besoins de l'équipe au département. Il gère le budget de l'équipe et veille aux conditions de travail (humaines et/ou matérielles) de façon à permettre à chacun de mener à bien ses activités de recherche.

4.4 Structuration des services

Outre les cellules, nous distinguons dans ce document les services soutiens et les services supports. Ces services sont tous indispensables au bon fonctionnement du laboratoire dans tous ses aspects².

4.4.1 Les cellules

Les cellules sont au nombre de trois. Elles sont toutes placées sous l'autorité directe de la direction du laboratoire. A noter que des discussions sont en cours pour déterminer si deux d'entre elles ne devraient pas être intégrées dans des services. En l'occurrence il s'agit de la cellule *valorisation* qui dans un premier temps pourrait aller dans le service *Accompagnement et Soutien aux Activités de Recherche et Développement* et de la cellule communication qui pourrait rester dans le *Service administratif et d'appui au pilotage*. Ces points sont encore en cours de réflexion.

Cellule communication

Le rôle de la cellule communication est d'assurer une cohérence d'ensemble de toutes les actions de communication interne et externe du laboratoire.

Cellule valorisation

La cellule valorisation³ a pour mission de proposer et de réaliser des opérations de partenariat et de valorisation.

Cellule science responsable

La cellule science responsable permet de rassembler et synthétiser les activités liées à l'éthique et au développement durable. Ses activités sont présentées plus précisément en Annexe A.

4.4.2 Les services support

Les services support sont au nombre de trois. Ils sont chacun placés sous la responsabilité d'un responsable de service.

Service administratif et d'appui au pilotage

Responsable : Monique Granon

Le périmètre des activités du service administratif et soutien au pilotage regroupe les activités du pôle gestion contractuelle et financière, du pôle RH ainsi que du pôle documentation et bibliothèque.

2. Notons ici que le nom des services n'est pas définitif et dépendra du travail actuellement en cours sur leur structuration.

3. À noter qu'une réflexion est en cours concernant le statut de cette cellule qui pourrait dans un premier temps être intégrée dans le service développement logiciel.

Service infrastructure bâtiments et logistique**Responsable :** Pascal Desroches

Le service est chargé de l'entretien, de la maintenance et de l'aménagement des infrastructures liés aux bâtiments du laboratoire et de la logistique.

Services informatiques communs**Responsable :** Laurent Darré, **Co-responsable :** Laurent Pointal

Les missions de ce service concernent la gestion des réseaux et du parc informatique. Ses activités recouvrent également l'administration des services, la mise en place d'applications locales et l'assistance aux utilisateurs.

4.4.3 Les services soutien

Les services soutien sont au nombre de trois. Ils sont chacun placés sous la responsabilité d'un responsable de service.

Cellule Technique, Électronique, Mécanique et Optique (resp. Dorine Leloup)**Responsable :** Dorine Leloup

Les missions de cette cellule sont de deux ordres. Elles se rapportent d'une part à la fabrication, de la conception à l'usinage et au montage et d'autre part à la métrologie, c'est-à-dire les systèmes d'acquisition, les campagnes de mesures et les contrôleurs.

Accompagnement et Soutien aux Activités de Recherche et Développement (ASARD)**Responsable :** Yann Fraigneau, **Co-responsable :** Anne-Catherine Letournel

Ce service prend en charge le support informatique lié aux activités de recherche : service d'appui au développement logiciel, maintenance des plate-formes numériques expérimentales et des moyens de calcul, soutien aux bonnes pratiques de développement logiciel scientifique.

FabLab**Responsable :** Romain Di Vozzo

Fablab Digiscope est un service de prototypage rapide, ouvert à tous. Il a vocation à former aux nouvelles techniques (formation en Design Appliqué à la Fabrication Numérique (DAFN) et plus largement en Design dit "sociétal").

Annexes

Annexe A

Cellule Science Responsable

La responsabilité est le fait de devoir répondre de ses actes ou de ceux d'autrui, individuellement et collectivement, d'en mesurer les conséquences présentes et d'en anticiper les conséquences futures, et d'agir dans la mesure du possible de manière à les rendre compatibles avec les valeurs et les normes d'une société donnée et à en minimiser les effets négatifs. Elle est nécessaire pour l'innovation que les chercheurs développent, mais également en tant que réflexion collective sur la pratique de la recherche, et de ses implications sur la société et les citoyens.

Aujourd'hui de nombreuses initiatives ont été mises en place par nos tutelles : le CNPEN¹ (Comité Pilote d'Éthique du Numérique) au niveau national, Polethis² dans Paris-Saclay pour ce qui concerne l'éthique et l'intégrité scientifique, le groupe de travail Développement durable de Paris-Saclay, le plan national pour la science ouverte du Ministère de la recherche, la feuille de route du CNRS pour la science ouverte, la création de la Task Force « CNRS Agenda 2030 » pour les objectifs du développement durable... Les personnels du LISN, chercheurs, doctorants et IT, se sont impliqués dans une démarche de Science responsable depuis de nombreuses années, au-delà des injonctions des financeurs qui ont tendance à faire rimer uniquement responsabilité avec acceptabilité de l'innovation.

Dans ce contexte, la cellule "Science Responsable" du LISN a pour objectif de proposer un support à la recherche relativement à trois aspects : **éthique, sciences ouvertes, développement durable**.

Sur chacun de ces aspects, la cellule aura en charge de proposer au laboratoire (personnels, direction) :

- i des bilans (état des lieux et recommandations) concernant les pratiques du laboratoire : bilan carbone, enquêtes de déplacement...
- ii des informations sur les enjeux et bonnes pratiques : formations, actions de sensibilisation, documents dédiés...
- iii un guichet d'orientation pour guider les chercheurs en montage de projet et les adresser à des structures dédiées existantes capables de répondre à leurs interrogations.

Les objectifs de la cellule s'articulent avec l'action transversale "Recherche responsable" au sein de laquelle seront menées les opérations de recherche fondamentale et appliquée sur ces trois aspects éthique, sciences ouvertes et développement durable, et vont se nourrir des résultats de celle-ci.

Pour les aspects **éthiques**, la cellule aura pour mission de sensibiliser les personnels aux enjeux éthiques de leur recherche, en particulier lors de l'élaboration de nouvelles directions de recherche, et le montage de projets. Cela se fera :

- par un audit des projets, afin de savoir s'il est nécessaire de demander l'avis d'un Comité d'éthique de la Recherche (CER), et le conseil pour le montage du dossier de soumission au

1. <https://www.ccne-ethique.fr/fr/actualites/creation-du-comite-pilote-dethique-du-numerique>

2. Conseil pour l'éthique de la recherche et l'Intégrité scientifique de l'Université Paris-Saclay <https://www.universite-paris-saclay.fr/recherche/polethis-ethique-et-integrite>

CER si nécessaire et pour l'analyse des implications éthiques des activités pour les projets européens. Pour cela, nous profiterons de l'expérience des membres du LISN qui font ou ont fait partie d'un CER, et en particulier du CER-PS (Comité d'éthique de la Recherche de Paris-Saclay), plusieurs membres du LISN ayant été à l'origine de sa création sous l'égide de POLETHIS ;

- par une diffusion des textes et avis pertinents pour les recherches menées dans le LISN publiés par les différentes instances et comités d'éthique. Pour cela, nous profiterons de la présence de plusieurs membres du LISN au sein du CCNE³, du CNPEN, du COMETS⁴, de POLETHIS, du GFAIH⁵, de l'institut DATAIA⁶.

Pour les aspects **sciences ouvertes**, deux rôles importants seront joués par la cellule.

- D'abord, elle proposera des formations et actions de sensibilisation à l'utilisation d'outils logiciels et à la mise en place de bonnes pratiques notamment au niveau (i) du recensement/de l'annotation de jeux de données produits par les expériences in-silico menées par les chercheurs du laboratoire mais aussi (ii) de la reproductibilité de ces expériences.
- Ensuite, elle aura pour tâche le recensement des publications des chercheurs du laboratoire dans HAL, leur bonne indexation et annotation dans ce système (au niveau des équipes, des départements et du laboratoire) en lien étroit avec les données et informations stockées dans le système d'information du laboratoire.

Pour les aspects **développement durable**, la cellule intégrera la commission développement durable actuelle du LIMSI, qui deviendra une sous-instance de la cellule, et dont les objectifs sont les suivants :

- évaluer les impacts environnementaux des activités de recherche du laboratoire ;
- évaluer les impacts environnementaux liés à l'exploitation des bâtiments, de nos modes de transports ;
- proposer à la direction des pistes d'action pour réduire ces impacts, pour conduire des actions de sensibilisation etc ;
- valoriser et encourager les actions transverses ou non autour de l'écologie.

3. Comité consultatif national d'éthique pour les sciences de la vie et de la santé <https://www.ccne-ethique.fr/>

4. Comité d'éthique du CNRS <https://comite-ethique.cnrs.fr/>

5. Forum mondial sur l'IA pour l'humanité <https://gfaii.org/>

6. Institut de convergence DATAIA <https://dataia.eu/linstitut-dataia>