



Introduction.....	2
Compétences, besoins et projets dans les laboratoires AMU	2
Méthodologies de traitement et fouille (« data mining »)	2
Mise à disposition et préservation.....	3
Usages, droit et éthique	3
Infrastructure et technologies	4
Actions 2015/2016	5
ANNEXES.....	6
Annexe A: Comité de pilotage de l'animation "Big Data" au sein des PR2I AMU	7
Annexe B: Contacts dans les laboratoires.....	8
Annexe C: Compétences dans les laboratoires AMU (en bref).....	10
Annexe D: Projets dans les laboratoires AMU (en bref)	14
Annexe E : Fiches complètes des laboratoires	19
TAGC.....	19
CEREGE	21
CPT.....	22
CRET-LOG.....	23
ESPACE.....	25
Institut Fresnel.....	27
GREQAM.....	28
I2M	30
IBDM.....	31
INT	32
IRSIC.....	33
iSm2	35
LAM.....	36
LBA.....	37
LCB	39
LIF	41
IMBE	43
LPL	45
LSIS.....	46
MESOCENTRE	49
SESSTIM	50
INSERM UMR S910	52
CRCM	53
Eccorev	55
LID2MS	56
CIML.....	58
CRCM	59
NORT.....	61

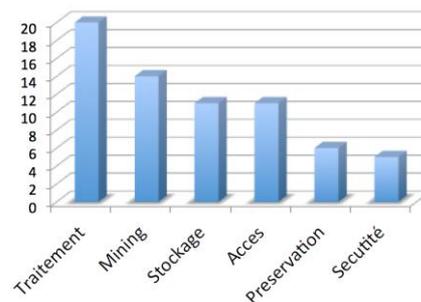
Introduction

Dans le cadre des pôles de recherche interdisciplinaires et intersectoriels de l'AMU, une animation sur les grandes masses de données a vu le jour dans le courant de l'année 2014 (le comité de pilotage en Annexe A).

Les objectifs de cette animations sont :

- Identifier les acteurs AMU dans ce domaine
- Identifier les besoins
- Suivre les appels à projets, documenter et diffuser
- Stimuler les échanges et la construction des consortia
- Faire émerger des nouvelles collaborations
- Connexions au niveau national et international
- Connexion avec les PR2I, Amidex, AMU

Une première prise de contact a été initiée à travers un formulaire distribué dans tous les laboratoire AMU ; 24 laboratoires ont nommé une personne de contact (Annexe B) et envoyé un formulaire rempli (Annexe C). Les laboratoires ont présenté brièvement leurs activités dans le domaine pendant une journée d'animation organisé le 24 Novembre 2015. Cette première réunion a permis une mise en perspective des compétences et des besoins des laboratoires de l'AMU dans le domaine des « Big Data ». Une distribution des mots-clefs mentionnés dans les réponses est présenté plus bas et illustre bien l'envergure et le potentiel très importants des laboratoires AMU dans ce domaine.



Compétences, besoins et projets dans les laboratoires AMU

Méthodologies de traitement et fouille (« data mining »)

Compétences et points forts

- Compétences dans les bases théoriques dans les méthodes de fouille, algorithmes, détection de corrélations, visualisation, classification de grande taille représentation, parallélisations.
- Position de leadership au niveau national et international dans plusieurs domaines liés aux grandes masses de données (astrophysique, physique des particules, génomique)
- Développement d'outils et programmes (ex. bioinformatique), analyse de grandes masses de données.
- Panel de champs disciplinaires très complémentaire et favorisant les coopérations au niveau d'AMU avec un excellent potentiel au niveau international.

Besoins

- Collaborations souhaitées (recherche) entre fournisseurs de Big Data (biologistes, physiciens, chercheurs en sciences sociales, etc.), curateurs de données, statisticiens

- Formation et enseignement : master/licence, et aussi des formation pointues pour des chercheurs avancées.
- Plateformes d'échange données et logiciels, accès à des ressources communes.

Opportunités et projets

- Collaborations sur les méthodes et algorithmes : dimension spatiale des données (mobilités), représentations des données (sémiologie graphique), textes/ontologies.
- Montage des projets thématiques sur le Big Data incluant des domaines des sciences sociales, l'histoire, l'art, les biotechnologies, l'astronomie etc. Exemples de thématiques possibles/engagées : influence de l'e-publicité, traitement du langage, indexation des données sur la biodiversité méditerranéenne, biodiversité des habitats coralligènes, risque en finances et marchés, bio-diversité acoustique, données santé et gestion hospitalière, réseaux sociaux, géosciences, urbain analytics, renforcer nos relations avec des acteurs gestionnaires des territoires.

Mise à disposition et préservation

Compétences et point forts

- Projets d'envergure internationale dans la manipulation et l'accès aux grandes masses de données (Physique des particules, astrophysique, biologie).
- Participation à des projets de bases de données en biologie, pathologie, ressources pour la linguistique, microscopie, documents/Web, réseaux sociaux, données patrimoniales, géosciences ; bases de données relationnelles.
- Traitement systématique des données pluri-disciplinaires, compétences en stockage massif et pérenne. Participation à des projets dédiés à la préservation des données scientifiques au niveau national et international.

Besoins

- Plateformes de mise en commune des données des différentes disciplines qui pourrait stimuler l'échange et l'émergence de projets pluri- et inter-disciplinaires.
- Pour cela, une infrastructure matérielle importante (grappes de calcul, serveurs spécialisés, stockage) permettant de mettre en œuvre ces activités est nécessaire.

Opportunités et projets

- Traitement du langage et de ses bases cérébrales (LPL)
- Banques de données et des bases de données généralistes ou spécialisées, ainsi que des plates-formes logicielles donnant accès à des outils de bioinformatique et d'analyses d'images, des plus génériques aux plus pointus. Préservation pérenne de données scientifiques.
- Intégration de données hétérogènes, la mise en place de chaînes de traitement automatisées, la définition d'ontologies, la fouille de texte.

Usages, droit et éthique

Compétences et points forts

- Enjeux sociétaux du Big Data et de l'Open Data et les dimensions épistémologiques et éthiques de leurs usages: collecte, du traitement et d'usages de grandes masses de

données dans différents domaines tels que la santé, les médias et le journalisme, la publicité et le e-marketing, la recherche scientifique, etc. .

- Communication engageante numérique, économie numérique, intelligence économique et veille informationnelle.
- Usages psychologiques et sociaux des medias sociaux en tant que « conscience collective virtuelle ».

Besoins

- La mise en commun des compétences dans les SHS dans ce domaine.
- Collaboration avec des laboratoires en sciences (médecine, neuro-sciences, physiques, etc) qui disposent des données massives pour étudier les modes d'usages sur le terrain.
- Formation et enseignement en éthique des données et préservation/archivage à long terme.
- Enjeux sur la sécurité, compétences sur le psycho-social et l'impact sur les usages

Opportunités et projets

- Etudes des transformations organisationnelles liées au Big Data.
- Enjeux éthiques et sociétaux de la génomique personnelle
- Enjeux psycho-social et impact sur les usages
- Propriété intellectuelle sur le code, droit et médecine (protection des données médicales), sécurité informatique, philosophie

Infrastructure et technologies

Compétences

- Infrastructures de calcul HPC (méso-centre) et Grille (Tier2 LHC, CPPM)
- Expertise dans l'utilisation massive de grand centres de calcul grille et HPC au niveau national et international.
- Accès à des technologies de pointe ; outils d'accès aux ressources, virtualisation, techniques de calcul sur le « cloud ».

Besoins

- Inter-connectivité, accès aux ressources distribuées, amélioration des bases d'infrastructure informatique dans certains laboratoires.
- Formation pour l'utilisation massive des grandes ressources informatiques.
- Besoins génériques en experts « data management », systèmes/infrastructures (réseau, Cloud, langages), expertises technologiques concernant les outils permettant de traiter le Big Data et d'intégrer des données issues de différents canaux et sources

Opportunités et Projets

- Projet de rapprochement entre le méso-centre et la grille du CPPM est en cours de développement. Lorsqu'elles sont disponibles, les ressources du méso-centre pourront être intégrées à Dirac, le logiciel de gestion de la grille du CPPM et participer ponctuellement au traitement très efficace de données massives.

Actions 2015/2016

-Elargissement du comité de pilotage suivant les premiers contacts dans les laboratoires (en cours)

-Réunions du comité de pilotage (suivi des activités, collaborations, connexions avec les PR2I, AAP etc.).

-Réunions ciblées sur des aspects spécifiques, afin de favoriser l'émergence de projets communs :

ANNEXES

Annexe A: Comité de pilotage de l'animation "Big Data" au sein des PR2I AMU

- **Sciences et Technologies**
 - Cristinel Diaconu (CPPM) : diaconu@cppm.in2p3.fr (coordonnateur)
 - Christian Surace (LAM): christian.surace@lam.fr
 - Thierry Artières (LIF) : thierry.artieres@lif.univ-mrs.fr
 - Nicolas Ferré (Mesocentre) nicolas.ferre@univ-amu.fr
 - Sebastien Fournier (LSIS) sebastien.fournier@lsis.org
- **Environnement**
 - Joel Guiot (ECCOREV) : guiot@eccorev.fr
 - Jean-Pierre Bracco (LAMPEA): jean-pierre.bracco@univ-amu.fr
- **Energie**
 - Mohamed Quafafou (LSIS) (mohamed.QUAFAFOU@univ-amu.fr)
- **Humanités**
 - Patrice Bellot (LSIS) (patrice.bellot@lsis.org)
 - Bernard Bel (LPL) (bernard.bel@lpl-aix.fr)
 - Dominique Augey (Faculté de Droit, AMU) dominique.augey@univ-amu.fr
 - Fidelia Ibekwe-SanJuan IRSIC fidelia.ibekwe-sanjuan@univ-amu.fr
- **Santé**
 - Frédéric Barras (LCB): barras@imm.cnrs.fr
 - Jean-Charles Dufour (SESSTIM) : jean-charles.dufour@univ-amu.fr
 - Christophe Bérout (Genetique) christophe.beroud@inserm.fr
- **AMU Recherche**
 - Mossadek Talby (CPPM): talby@cppm.in2p3.fr

Annexe B: Contacts dans les laboratoires

Nom	Prénom	Institut	e-mail
Fossati	Caroline	Institut Fresnel	caroline.fossati@fresnel.fr
Novelli	Noël	LIF (info)	noel.novelli@lif.univ-mrs.fr
Libes	Maurice	OSU Pytheas	maurice.libes@univ-amu.fr
Humbel	Stéphane	iSm2	stephane.humbel@univ-amu.fr
Takerkart	Sylvain	INT	sylvain.takerkart@univ-amu.fr
Torresani	Bruno	i2M	bruno.torresani@univ-amu.fr
Gonçalves	Bruno	CPT	bgoncalves@gmail.com
Ferré	Nicolas	Mesocentre	nicolas.ferre@univ-amu.fr
Kadoch	Benjamin	IUSTI	benjamin.kadoch@univ-amu.fr
Vicente	Jérôme	IUSTI	jerome.vicente@univ-amu.fr
Ibekwe-Sansuan	Fidelia	IRSIC	fidelia.ibekwe-sanjuan@univ-amu.fr
Ghattas	Badih	I2M	badih.ghattas@univ-amu.fr
Lancini	Agnès	CRET-Log	lanciniagnes@hotmail.com
Blanpain	Cyril	OSU Pytheas	blanpain@osupytheas.fr
David	Romain	IMBE	romain.david@imbe.fr
Guiot	Joël	CEREGE	guiot@cerege.fr
Surace	Christian	LAM	christian.surace@lam.fr
Allard	Paul	ESPACE	paul.allard@univ-amu.fr
Talla	Emmanuel	LCB	talla@imm.cnrs.fr
Lugiez	Denis	LIF	denis.lugiez@univ-amu.fr
Llari	Maxime	LBA	maxime.llari@ifsttar.fr
Beroud	Christophe	INSERM	christophe.beroud@inserm.fr
Laporte	Cathy	AMU/Cellule Europe	cathy.laporte@univ-amu.fr
Laurent	Sébastien	GREQAM	sebastien.laurent@univ-amu.fr
Gingold	Arnaud	LPL	arnaud.gingold@lpl-aix.fr
Bel	Bernard	LPL	bernard.bel@lpl-aix.fr
Spinelli	Lionel	CIML/TAGC	spinelli@ciml.univ-mrs.fr
Chetrit	Bernard	CRCM	bernard.chetrit@inserm.fr
Jaeger	Sébastien	CIML	jaeger@ciml.univ-mrs.fr
Augey	Dominique	LID2MS/GRECAM	dominique.augey@univ-amu.fr
Tsaregorodstev	Andrei	CPPM	atsareg@in2p3.fr
Talby	Mossadek	CPPM	talby@cppm.in2p3.fr
Tadrist	Lounes	IUSTI	lounes.tadrist@univ-amu.fr
Quafafou	Mohamed	LSIS	mohamed.quafafou@univ-amu.fr
Bellot	Patrice	LSIS	patrice.bellot@univ-amu.fr
Dufour	Jean-Charles	SESSTIM	jean-charles.dufour@univ-amu.fr
Diaconu	Cristinel	CPPM	diaconu@cppm.in2p3.fr
van Helden	Jacques	TAGC	Jacques.van-Helden@univ-amu.fr
Martin	Jean-Charles		jean-charles.martin@univ-amu.fr
Damon	Céline	ProtisValor	celine.damon@univ-amu.fr
Bartoli	Jonathan	ProtisValor	jonathan.bartoli@univ-amu.fr

Annexe C: Compétences dans les laboratoires AMU (en bref)

IRSIC

Enjeux sociétaux du Big Data et de l'Open Data et les dimensions épistémologiques et éthiques de leurs usages: collecte, du traitement et d'usages de grandes masses de données dans différents domaines tels que la santé, les médias et le journalisme, la publicité et le e-marketing, la recherche scientifique, etc

LPL

Partenaire du réseau européen CLARIN dédié aux données numériques de la linguistique. Centre de ressources SLDR (<http://sldr.org>) pour la linguistique. Archivage pérenne des données à valeur patrimoniale. Bases de données de parole pathologique très spécifiques (et uniques en France).

LAM

Datamining dans les données astrophysiques, (cosmologie, planètes extra solaires). Visualisation de grands échantillons de données, visualisation 3D, Base de données astrophysiques, Développement d'application de traitement de données.

IMBE Données en écologie et sur la biodiversité, développement de graphes à partir de ces données, fouille de donnée, systèmes d'observations et d'informations répartis
Données génomiques, métagénomiques, transcriptomiques, RNASeq, RADSeq
Analyses de polymorphisme, d'expression de gènes, de phylogénie Après publication les données génétiques sont soumises à des bases de données publiques (Genbank par exemple) mais il y a un intérêt pour les stocker localement afin de poursuivre les analyses.

GREQAM

Estimation et prévision du risque sur les marchés financiers. Utilisation et développement d'outils statistiques et économétriques pour extraire l'information de données à très haute fréquence sur de très grands portefeuilles d'actifs.

Méthodes économétriques/statistiques permettant d'extraire des mesures de risque (volatilités, corrélations) très fines, robustes (à la présence de bruit dans les données observées, à des sauts abruptes dans les prix, ...) et qui sont rapides à obtenir malgré le grand très nombre d'observations.

IBDM

Gestion de grandes masses de données de microscopie (images, films) pour l'exploration quantitative et intégrée de systèmes vivants à différentes échelles spatiales : moléculaires, cellulaires, tissulaires, organismes entiers.

LSIS

Recherche et extraction d'information dans de grandes masses de documents (Web, bibliothèques numériques)

Suivi d'opinion, contextualisation et détection de nouveauté sur les réseaux sociaux
Fouille et Intégration de Données

Gestion de données : élaboration de systèmes et langages, intégration (interopérabilité), qualité des données, données spatiales et biologiques
Cloud computing : Provenance, cloud/web services
Big data analytics : Algorithmes probabilistes et hypergraphes
Architecture orientée services : conception d'architectures offrant différents services distribués pour l'acquisition, le stockage et l'analyse de données massives (Big Data).
Structuration non supervisée et supervisée de masses de données multimodales, vidéo (les collections d'images du web), audiovisuelles (ex : les challenges NIST TREC VIDEO) ou audio (parole et bioacoustique).
Modèles neuronaux profonds, les plus performants en indexation de masses de données d'images et sons
Analyse de données pour la surveillance, pilotage et aide à la décision pour les systèmes Complexes

ESPACE

Traitement systématique des données pluridisciplinaires (sociologie, histoire, écologie...) accumulées depuis 20 ans sur les inondations, les territoires fluviaux (en particulier la Camargue) et les zones humides dans le cadre de programmes de recherche, de thèses, de mémoires terminés ou en cours.

ESPACE

La recherche effectuée par les membres l'UMR ESPACE trouve son unité dans l'attention particulière accordée à l'information géographique. Cette réflexion repose sur des avancées théoriques récentes (théorie de la complexité par exemple), une forte dimension épistémologique, et bien entendu par la pratique quotidienne de cette information et des outils nécessaires à son traitement (Systèmes multi-agents, SIG, Géostatistique, etc.).

CPPM

Traitement de grande masses de données
Infrastructure Grille (LHC)
Système de calcul distribué DIRAC
Projet de préservation de données

CEREGE

Géosciences de l'Environnement : Physique et chimie de la Terre, Biologie et écologie : passé/présent. Ressources et Risques en Environnement.

SESSTIM

Utilisation de données complexes dans les domaines: Médecine personnalisée : cancer, autres pathologies, Veille Sanitaire, Méthodes quantitatives, traitement de l'information médicales
Méthodologies d'enquêtes et recherches axées sur l'apport des média sociaux et technologies du web 3.0 (objets connectés) pour la pratique médicale et la meilleure connaissance des problèmes de Santé en population (veille sanitaire notamment).

LCB

Génomique Comparative / Analyse des génomes
Développement de bases de données relationnelles

Développement d'outils et de programmes bioinformatiques

Analyse des données de séquençage à haut débit

Le laboratoire dispose d'un service de criblage phénotypique des bactéries basé sur la microscopie. Ce service produit de dizaines de milliers d'images lors d'une campagne de criblage. Actuellement, nous travaillons sur les dispositifs nécessaires à l'exploitation de ces grandes masses de données.

AMU mesocentre

Le mésocentre d'AMU est une structure dépendant de la DRV et basée sur un financement Equipex obtenu en 2010. Il donne accès à des ressources de calcul hautes performances (HPC), susceptibles de générer de grandes masses de données et de les post-traiter (visualisation). Ses utilisateurs sont représentatifs de tous les champs disciplinaires d'AMU présentant un besoin en HPC

TAGC

Différentes thématiques liées à la santé (sepsis, malaria, cancer, développement cardiaque) ;
Approches bioinformatiques pour le traitement des données biologique à haut débit (génomomes, transcriptomes, interactomes) ;
Plate-forme technologique spécialisée dans la production de données massives (« Next Generation Sequencing », biopuces).

INT

Acquisition et traitement de grande bases de données issues d'expériences de neurosciences fondamentales et cliniques

LIF

Bases de Données Avancées (BDA) et équipe Apprentissage et Multimédia (QARMA)
Fouille de données, détection de corrélations et de dépendances, approximation avec modèle coût (contrôle des ressources), visualisation de données massives, problème de classification de grande taille (nombres d'exemples, nombre de classes, dimensionnalité), apprentissage budgétisé, apprentissage de représentations pour le texte et l'image, parallélisation."

CPT

Mining et l'analyse à grande échelle des données de réseau social pour l'étude du comportement social et humain.

INSERM_UMRS910

Etude génétique des maladies rares.

Gestion des grands volumes de données générées par le séquençage des exomes et génomes.

Compétences dans la création de bases de données génétiques et cliniques, dans le développements d'algorithmes permettant l'analyse de ces données, développement de

systèmes de prédiction du caractère pathogène des mutations et de validation d'hypothèses thérapeutiques.

CRET-LOG

"Pas de compétences techniques en matière de traitement des big data
Intérêt au traitement des big data en vue d'aide à la décision dans plusieurs domaines
- décision marketing (compréhension du comportement du consommateur)
- décision logistique (liée aux données de commandes clients, à la traçabilité totale des supply chains, à la maîtrise des risques dans les supply chains)
Intérêt pour les transformations organisationnelles et technologiques de l'entreprise nécessaires pour absorber, traiter et analyser les Big Data
- notamment les big data client, multi-canal, logistique
Intérêt porté au lien entre Big Data et Connaissance
"

iSm2

(pas d'info spécifique fournie)
Calculs en chimie quantique, conservation de données.

IFSTTAR

(pas d'info spécifique fournie)

IMM

Traitement, mining, stockage, signal, génomique
(pas d'info spécifique fournie)

FRESNEL

Traitement tensoriel du signal pour les « Big-Data », Acquisition de données
(pas d'info spécifique fournie, à compléter)

CIML

Utilisation de diverses techniques expérimentales générant de grandes quantités de données dans le domaine de l'immunologie et s'articulant essentiellement autour de 2 axes :
(i) Traitement de données d'expériences à haut débit (Biopuces, « Deep Sequencing »...) y compris au niveau unicellulaire par association avec des techniques de microfluidique., (ii) Analyses de données d'imagerie (« Cell Tracking, imagerie bi-photonique, cinétique 3D...).

CRCM

- Le développement d'outils et de programmes dans les domaines de la bioinformatique et de la chemoinformatique.
- L'analyse des données de séquençage à haut débit.
- Le développement d'applications de traitement de données massive.
- Une infrastructure de calcul et de développement à haut débit
- La génération de données de masse dans le domaine de la cancérologie (omics: NGS, protéomique, cytométrie de flux,...) et particulièrement dans le champ de la médecine personnalisée.

Annexe D: Projets dans les laboratoires AMU (en bref)

IRSIC

- Réception et Influence de l'e-publicité (display) : expérimentation en collaboration avec l'Université de Cologne (contexte psycho -socio-cognitif)
- Expérimentation pour mettre au point un dispositif numérique sur le web (projet MOVIDA (Contrat de recherche avec le Ministère de l'écologie, du développement durable et de l'énergie 190 – 0190-THUR-BSAF) : ""Apports de la communication engageante dans le cadre du co-voiturage et de l'autopartage"", en collaboration avec l'entreprise Auxilia (Paris)
- Mises en œuvre de nouvelles techniques de communication par e-mailing pour les associations humanitaires (dans le cadre des collectes de fond ex : collaboration avec l'UNICEF) et pour les entreprises commerciales (convention avec une entreprise d'e-commerce en PACA) : recherche action dans le cadre théorique de la « communication engageante numérique »
- Economie numérique, intelligence économique et veille informationnelle : Etude des connaissances et de la complexité ; Usages du web dans un contexte d' « intelligence économique » (ex : études des dépôts de brevets ; Les enjeux économiques de la géolocalisation pour les réseaux sociaux numériques""
- Réflexion sur une problématique des usages psychologiques et sociaux des medias sociaux en tant que « conscience collective virtuelle » et leurs liens avec les émotions collectives.
- La relation entre communication et engagement en actes des acteurs, nos travaux enrichissent les théories et pratiques consacrées au changement en étudiant comment les acteurs individuels et collectifs entrent dans un cours d'action et adoptent de nouvelles signification et valeurs. Cette question concerne les mutations technologiques, culturelles, organisationnelles, institutionnelles, environnementales, économiques et sociétales. De nombreux projets de recherche financés sur ce thème et de nombreuses publications.

LPL

"Le LPL est porteur du LabEx BLRI (<http://blri.fr>) consacré à l'étude du traitement du langage et de ses bases cérébrales.

Il aussi activement associé au programme Erasmus Mundus (<https://www.em-multi.eu>).

"

LAM

Grandes masses de données en astrophysique: mining, preservation, accès

IMBE

data mining concernant les données sur la biodiversité des habitats coralligènes
indexation des données sur la biodiversité méditerranéenne

GREQAM

méthodes pour la mesure et la prévision du risque en finance
modélisation du risque de marché dans les séries financières

IBDM "– ERC Advanced Grants BIOMECAMORPH (T. Lecuit)

- ERC Consolidator Grants SPOT.COM (B. Prud'homme)
- LabEx INFORM (T. Lecuit)
- Infrastructure nationale biologie - santé FBI (PF. Lenne)
- AAP Emergence & innovation : SpongEX (A. Le Bivic)
- ANR ATMIR (H. Cremer)
- ANR EVOLNET (B. Prud'homme)
- FRM « Mesoscopic origins of cell behaviours during tissue morphogenesis: mechanics and biochemical signaling » (PF. Lenne)
- FdF « Traitement des informations thalamiques au sein des ganglions de la base en situation normale et parkinsonienne. » (L. Kerkerian)

"

LSIS

Scaled Acoustic Biodiversity (SABIOD)

Recherche d'information et bases de données (Fouille de videos, e-santé, semantic keywords)

Gros Volumes de Données Spatiales

Cloud computing

Environnement pour le développement d'applications NFC sécurisées

l'exploitation d'un entrepôt de données pour la gestion hospitalière

l'extraction d'information de textes à base d'ontologie

Réseaux sociaux et Ciblage, Comportement d'achat

Observation sociale des territoires fluviaux

ESPACE

Il n'y a pas actuellement de projet spécifique au sein du laboratoire sur ces questions

CPPM

Distributed Infrastructure with Remote Agent

Control Interware l'interface permettant à une communauté

scientifique de mettre en place l'accès à des

ressources distribuées GRID, HPC ou Cloud

Tier2 infrastructure de calcul locale, accepte plusieurs Virtual Organizations (VO) :

Project interdisciplinaire pour la préservation de données scientifiques

CRET-LOG

"Pas de projet en cours

Des publications « prospectives »

- sur l'intérêt des big data et leurs usages

- sur la traçabilité totale et l'usage de ces données

"

CEREGE

"projets ANR , projets CCE : Planétologie, Champ magnétique terrestre ; Propriétés physiques des roches , Géodynamique ; Variations du climat et de la biodiversité ; Risques ; Géochronologie

Collaborations avec les entreprises (Energie, Risques sismique, Risques environnementaux (pollution , nanoparticules...)) ;

"

FRESNEL Collaboration avec un industriel (bourse CIFRE-Défense)

SESSTIM

Accès aux grandes bases de données françaises en santé (caisse d'assurance maladie, hôpitaux, interopérabilité et développement d'algorithmes

Surveillance sanitaire en lien avec l'épidémiologie et les sciences humaines et sociales, notamment

LCB

Analyse des données issues de séquençage à haut débit (RNA-seq, ..)

Analyse à grande échelle de données génomiques procaryotes

Développement d'outils pour l'identification et l'analyse des séquences répétées au sein de genomes

iSm2

IFSTTAR

AMU mesocentre Actuellement, un projet de rapprochement entre le mésocentre et la grille du CPPM est en cours de développement. Lorsqu'elles sont disponibles, les ressources du mésocentre pourront être intégrées à Dirac, le logiciel de gestion de la grille du CPPM et participer ponctuellement au traitement très efficace de données massives.

IMM ANR

TAGC

"ANR « ORICHOICE » (2014-2018), collaboration avec Marcel Mechali et Naomi Taylor (Institut de Génétique Humaine de Montpellier).

-Collaboration nationale sur le projet sepsis : Didier Payen pour un projet centré sur le contrôle génétique de l'apparition de formes graves et de la mortalité due au sepsis

-Collaboration internationale sur le paludisme : Serga Sawadogo (Université de Ouagadougou,, Burkina Faso).

-« Projet Européen « Blueprint » : Analyse des données épigénomique du système hématopoïétique humain «

« Projet Amidex Etoiles Montantes : Identification et caractérisation des ARNs longs non-codant dérégulés dans les leucémies »"

INT

Acquisition de données en continue

Installation d'un nouvel imageur IRM 3T (financé par FLI)

Création d'une plateforme d'imagerie photonique

LIF

"Dans le cadre des appels à projets MASTODONS du CNRS depuis 2012 :

- Projet AMADEUS 2012 à 2014 (LaBRI, LIF, LAM, LIRMM, CEREGE, Espace DEV, Université de TRENTE Canada) ;

- Projet PetaSKY depuis 2014 (LaBRI, LIF, LAM, LIRMM, Université de TRENTE Canada, LIRIS, LIMOS, Observatoire de Paris, PRISM).

Projet ANR :

- Blanc ANR Class-Y (fin 02/2015) ;

- Blanc ANR GRETA (fin septembre 2015).

Projet européen BioASQ (fin 12/2014).

Collaborations avec UJF (Grenoble), Demokritos Athènes (Grèce), S. McConnell Université de TRENTE (Canada).

Projet eBob 2.0, 2014 à 2016 (investissement d'avenir) :

Gestion de grandes masses de données avec différents partenaires industriels sur le thème du e-achat et exploration des technologies liées au Big Data."

CPT

Collaboration avec MIT, Northeastern University en Boston, USA, l'IFISIC en Espagne et la Fondazione ISI en Turin, Italie sur l'épidémiologie et l'étude du comportement humain et social.

INSERM_UMRS910

"Projet Européen RD-CONNECT, responsable du WP "Clinical Bioinformatics" et participation à la création de la plateforme permettant le lien entre les données -omics, les registres et les biobanques;

Projet National de création de la banque de donnée nationale de variations issues des expériences de NGS dans le cadre des maladies rares;

Projet national de création de la banque nationale des CNV;

TREAT-NMD Alliance : développement et hébergement des banques de données internationales des gènes DMD et SMN (50 pays impliqués);

Développement, hébergement, curation, de nombreuses LSDB de référence au niveau international;

Développement, hébergement de registres de patients nationaux et internationaux."

CIML

L'utilisation du séquençage à haut débit se généralisant progressivement dans le domaine de l'immunologie, à l'heure actuelle plus de la moitié des projets développés au centre produisent des données de NGS (~15 projets ANR, 4 projets ERC, 5 Amidex, ainsi que divers financements FRM, ARC, InCA...).

Par ailleurs, si la quasi-totalité des projets ont recours à des données d'imagerie standard ou de cytométrie de flux (confocale, FCS, FACS...), on estime à 25% le nombre de projets générant massivement des données d'imagerie (3D, biphoton, etc.) incluant 2 projets ERC et 4 projets ANR.

CRCM

SIRIC Marseille (Site de Recherche Intégrée en Cancérologie) (Label Institut National du Cancer – Inserm – Direction Générale des Soins)

Institut Carnot : CALYM (Le Consortium pour l'Accélération de l'innovation et de son transfert dans le domaine du Lymphome)

Investissement d'Avenir : Démonstrateur : MIMAbs : Antibodies for Medicine

Projet Européen TRANSCAN : New technologies for early detection of HLA loss leukemia relapse after allogeneic hematopoietic stem cell transplantation, and identification of relevant risk factors

A*MIDEX interdisciplinaire « Méthodes optiques innovantes pour l'analyse des interactions et signalisations moléculaires en biologie », avec l'Institut Fresnel, "BIOCLOAK: Bio-Cloaking with Graphene and Metamaterials" avec Institut Fresnel, A*MIDEX Emergence «PDZ targeting hits cancer communication»,

A*MIDEX interdisciplinaire "NOvel UltraSonic BIOMarkers of Tumor response to systemic anticancer therapies and their integration on mathematical predictive model", projet le Laboratoire de Mécanique et d'Acoustique (LMA)

A*MIDEX Chaire d'Excellence: Pedro BALLESTER (cancérologie/bio informatique)

Projets avec Institut Fresnel (molécule unique, vectorisation de médicaments, imagerie endoscopique), CINAM (interaction cellules-substrats, vectorisation de médicaments), Ecole des Mines de Gardanne (cellules tumorales circulantes)

Annexe E : Fiches complètes des laboratoires

TAGC

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Lab. Technologies Avancées pour la Génomique et la Clinique (TAGC)
 Unité INSERM U1090, 163, Avenue de Luminy, 13288 MARSEILLE cedex 09. France
 Bureau: Bâtiment INSERM, bloc 6, niveau 2
 Directrice : Catherine Nguyen

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Jacques van Helden, Jacques.van-Helden@univ-amu.fr
 Rarement joignable par téléphone.

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Le laboratoire TAGC rassemble

- des chercheurs en biologie travaillant sur différentes thématiques liées à la santé (sepsis, malaria, cancer, développement cardiaque) ;
- des chercheurs qui développent des approches bioinformatiques pour le traitement des données biologique à haut débit (génomés, transcriptomes, interactomes) ;
- une plate-forme technologique spécialisée dans la production de données massives (« Next Generation Sequencing », biopuces).

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
 Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

1. Traitement des données
2. Données à caractère personnel
3. Enjeux éthiques de la génomique personnelle

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

ANR « ORICHOICE » (2014-2018), collaboration avec Marcel Mechali et Naomi Taylor (Institut de Génétique Humaine de Montpellier).

Collaboration nationale sur le projet sepsis : Didier Payen pour un projet centré sur le contrôle génétique de l'apparition de formes graves et de la mortalité due au sepsis

Collaboration internationale sur le paludisme : Serga Sawadogo (Université de Ouagadougou,, Burkina Faso).

« Projet Européen « Blueprint » : Analyse des données épigénomique du système hématopoïétique humain »

« Projet Amidex Etoiles Montantes : Identification et caractérisation des ARNs longs non-codant dérégulés dans les leucémies »

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Collaborations internationales sur le paludisme : Fuad Iraqi (Tel Aviv University), Francine Ntoumi (Fédération de la recherche médicale au Congo, Brazzaville), Alioune Dleye (Institut Pasteur Dakar).

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Au-delà des défis techniques posés par la génomique et autres applications liées au séquençage à haut débit, Jacques van Helden est particulièrement intéressé par les enjeux éthiques et sociétaux de la génomique personnelle, et aimerait trouver dans le projet interdisciplinaire des compétences complémentaires pour développer cette thématique.

Ces compétences incluent :

- Informatique (stockage, transfert, protection et traitement de données massives).
- Propriété intellectuelle sur le code.
- Droit et médecine (protection des données médicales).
- Sécurité informatique.
- Philosophie.

CEREGE

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
CEREGE, UM 34 , Pôle Environnement Arbois Méditerranée
Nicolas THOUVENY PR Cle exc AMU
2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone :: BLANPAIN Cyrille IR CNRS , blanpain@cerge.fr
Affecté à l'OSU Pythéas
3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Géosciences de l'Environnement : Physique et chimie de la Terre, Biologie et écologie :
passé/présent. Ressources et Risques en Environnement.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection,
sécurité informatique, sécurité juridique
Tous ces mots clés plus « modélisation »
4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

projets ANR , projets CCE : Planétologie, Champ magnétique terrestre ; Propriétés physiques des
roches , Géodynamique ; Variations du climat et de la biodiversité ; Risques ; Géochronologie

Collaborations avec les entreprises (Energie, Risques sismique, Risques environnementaux
(pollution , nanoparticules...) ;
5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
Mêmes thématiques qu'actuellement jusqu'en 2018.
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations
recherchées ou souhaités au sein de l'AMU
Signalement de l'appartenance à un OSU : Institut Pythéas et collaborations étroites avec les
UMR et l'UMS de l'OSU

CPT

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Centre de Physique Theorique (CPT), UMR 7332, Avenue de Luminy, Case 907, Marseille, Thierry Martin

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Bruno Gonçalves, bgoncalves@gmail.com, 07 50 27 86 52

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

Mining et l'analyse à grande échelle des données de réseau social pour l'étude du comportement social et humain. Pour plusieurs informations : www.bgoncalves.com

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, Mining, Stockage

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Collaboration avec MIT, Northeastern University en Boston, USA, l'IFISIC en Espagne et la Fondazione ISI en Turin, Italie sur l'épidémiologie et l'étude du comportement humain et social.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Mobilité humaine et le tourisme.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

CRET-LOG

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

CRET-LOG – EA 881
 413 Avenue Gaston Berger
 13 625 AIX Cedex 1
 dir : Pr. N. Fabbe-Costes

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Agnès Lancini, agnes.lancini@univ-amu.fr, 06.18.40.49.10
 Nathalie Fabbe-Costes nathalie.fabbe-costes@univ-amu.fr 04 42 26 46 33

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Pas de compétences techniques en matière de traitement des big data

Intérêt au traitement des big data en vue d'aide à la décision dans plusieurs domaines

décision marketing (compréhension du comportement du consommateur)
 décision logistique (liée aux données de commandes clients, à la traçabilité totale des supply chains, à la maîtrise des risques dans les supply chains)
 Intérêt pour les transformations organisationnelles et technologiques de l'entreprise nécessaires pour absorber, traiter et analyser les Big Data
 notamment les big data client, multi-canal, logistique
 Intérêt porté au lien entre Big Data et Connaissance

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
 Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Préalables organisationnels et technologiques pour un traitement performant du Big Data
 Impacts organisationnels et technologiques du traitement du Big Data
 Traitement du big Data pour une meilleure gestion de la connaissance Client.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Pas de projet en cours
 Des publications « prospectives »

sur l'intérêt des big data et leurs usages
 sur la traçabilité totale et l'usage de ces données

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Projet sur le passage du Big Data à la connaissance client, ou comment les nouveaux usages du Web 2.0 transforment la Gestion de la Relation Client. (A. Lancini et O. Chanut)

Volonté d'identifier les évolutions organisationnelles et technologiques nécessaires (projet avec le groupe Allianz et la cantine numérique-Toulouse)

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Besoin en expertises technologiques concernant les outils permettant de traiter le Big Data et d'intégrer des données issues de différents canaux et sources

Collaboration avec tous chercheurs intéressés par l'étude des transformations organisationnelles liées à l'émergence du Digital.

ESPACE**Reponse Voiron**

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

UMR 7300 ESPACE, directrice Christine Voiron PR de géographie– chritine.voiron@unice.fr - 04 93 37 54 64
 Site de l'AMU : UMR 7300 ESPACE (CNRS)
 Technopôle de l'environnement Arbois - Méditerranée
 BP 80
 Bâtiment Laennec
 Avenue Louis Philibert
 13545 Aix-en-Provence Cedex 04
 Tél. 04 42 90 85 25
 e-mail : espace-arbois@univ-amu.fr

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

ALLARD Paul, PR d'histoire - Paul.allard@univ-amu.fr - 06 22 70 64 94
 LABEUR Christine, sociologue – labeur.christine@orange.fr - 0676297825

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Traitement systématique des données pluridisciplinaires (sociologie, histoire, écologie...) accumulées depuis 20 ans sur les inondations, les territoires fluviaux (en particulier la Camargue) et les zones humides dans le cadre de programmes de recherche, de thèses, de mémoires terminés ou en cours.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
 Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Data mining, statistiques textuelles sur de grands corpus, , stockage, identification et accès aux banques de données

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Projets de recherches en cours (enquêtes) dans le cadre de l'axe « Observation sociale des territoires fluviaux » de la Zone Atelier du Bassin du Rhône (ZABR) regroupant 21 institutions de recherche et de gestion riveraines du Rhône dont l'Agence de l'Eau, EDF, la CNR.
 Projets en cours dans le cadre de l'Observatoire Homme Milieu (OHM) de la Vallée du Rhône

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Constitution d'une base de données pluridisciplinaire sur le Rhône et la Camargue en collaboration avec le site de l'UMR ESPACE d'Avignon, la ZABR, l'OHM Vallée du Rhône
 Recherches sur les inondations dans le bassin du Rhône, comparaisons internationales

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

-Les grands corpus de données en sciences sociales sont de plus en plus nombreux et souvent dispersés, ils constituent un champ d'investigations incontournable dans les années à venir. La mise en commun des expériences est souhaitable, quelles que soient les disciplines. Mais tout cela ne se fera pas sans moyens pérennes en postes et en enseignement.

Reponse Oliveau

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

ESPACE, UMR 7300, 29 avenue Schuman, Sébastien Oliveau (directeur adjoint)

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Sébastien Oliveau, sebastien.olievau@univ-amu.fr, 06 28 04 39 37

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

La recherche effectuée par les membres l'UMR ESPACE trouve son unité dans l'attention particulière accordée à l'information géographique. Cette réflexion repose sur des avancées théoriques récentes (théorie de la complexité par exemple), une forte dimension épistémologique, et bien entendu par la pratique quotidienne de cette information et des outils nécessaires à son traitement (Systèmes multi-agents, SIG, Géostatistique, etc.).

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, mining, représentation graphique (dont cartographie), réflexion épistémologique

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Il n'y a pas actuellement de projet spécifique au sein du laboratoire sur ces questions

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

La question des Big Data est une question supplémentaire dans nos réflexions sur le traitement de l'information géographique mais ne constitue pas, pour l'instant une voie particulière d'exploration.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Les membres du laboratoire se proposent pour des collaborations qui nécessiteraient une expertise sur la dimension spatiale des données (par exemple sur les questions relatives aux mobilités) ainsi que sur les questions de représentations (sémiologie graphique).

Institut Fresnel

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
Institut Fresnel, CNRS-UMR 7249, Avenue Escadrille Normandie Niemen, 13013 Marseille,
Directeur : Stefan ENOCH

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
a. Nom, mail, téléphone

Salah BOURENNANE, salah.bourennane@fresnel.fr, 04 91 28 80 38

3. Domaines d'études relatifs aux grandes masses de données
a. Courte description des intérêts et compétences du laboratoire
- Traitement tensoriel du signal pour les « Big-Data », Acquisition de données.

b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection,
sécurité informatique, sécurité juridique (ajoutez si besoin)

- Traitement, compression, stockage.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Collaboration avec un industriel (bourse CIFRE-Défense)

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations
recherchées ou souhaités au sein de l'AMU

GREQAM

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Groupement de Recherche en Économie Quantitative d'Aix-Marseille (GREQAM), UMR 7316, Centre de la Vieille-Charité, 2, rue de la Charité, 13236 Marseille cedex 02. Directeur : Nicolas Gravel

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

LAURENT Sébastien, sebastien.laurent@univ-amu.fr, 0788101258

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Estimation et prévision du risque sur les marchés financiers. Utilisation et développement d'outils statistiques et économétriques pour extraire l'information de données à très haute fréquence sur de très grands portefeuilles d'actifs. Dans ce contexte, il y a une double dimension Big Data, c.-à-d.

- un nombre d'observations par actif qui tend vers l'infini lorsque le temps entre deux prix observés tend vers 0 ;

- un nombre d'actifs très grand, e.g. les 500 actifs constitutifs de l'indice SP500 ;

ce qui donne lieu à des bases de données de plusieurs centaines de millions d'observations par année.

Nous étudions des méthodes économétriques/statistiques permettant d'extraire des mesures de risque (volatilités, corrélations) très fines, robustes (à la présence de bruit dans les données observées, à des sauts abruptes dans les prix, ...) et qui sont rapides à obtenir malgré le grand nombre d'observations.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

- Projet IUF de Sébastien Laurent (débutant en Octobre 2014) sur le thème 'Nouvelles méthodes pour la mesure et la prévision du risque en finance' qui traite explicitement du problème Big Data et risque financiers.

- Projet PACA 'MODRISK' sur la modélisation du risque de marché dans les séries financières
- Divers projets de recherche avec des collaborations internationales (Belgique, Pays-Bas, Danemark, Angleterre).

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Projet d'ANR en discussion sur le thème des risques extrêmes en finance.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Manque de structure informatique sur AMU permettant du calcul (estimations et simulations) sur des problématiques Big Data. Par structure informatique nous entendons :

- Serveurs puissants ;
- Capacités de stockage importantes ;
- Personnel de maintenance qualifié.

I2M

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Institut de Mathématiques de Marseille (I2M), UMR 7373, CMI , 31 rue Joliot-Curie, 13453
Marseille Cedex 13, directeur : B. Torrèsani

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

Mr Badih GHATTAS,
badih.ghattas@univ-amu.fr,
tel: 04.91.82.90.89

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

L'I2M est un laboratoire de Mathématiques Fondamentales et Appliquées. Son spectre thématique recouvre presque l'ensemble des grandes thématiques des mathématiques, ainsi qu'un certain nombre de domaines d'applications en Physique, Mécanique, Chimie, Sciences de la vie et la santé et sciences pour l'ingénieur.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, signal, génomique

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

- L'équipe Signal-Image est partie prenante dans plusieurs projets à l'interface avec l'apprentissage statistique, en particulier dans le cadre de l'ANR GRETA (2011-15) et l'ANR MAD (2014-18). Les principaux thèmes sont l'optimisation (algorithmes gloutons, algorithmes variationnels) et des applications en traitement des signaux audio. Ces projets se font en interaction avec le LIF.
- L'équipe Statistique de l'I2M s'intéresse à la théorie de l'apprentissage statistique et développe des approches originales exploitant des techniques d'arbre de régression, boosting,... Les domaines d'application principaux se situent dans le domaine de l'environnement et de la génomique.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Les compétences présentes au sein de l'I2M se situent principalement à un niveau assez amont par rapport aux applications: modélisation probabiliste et statistique, optimisation numérique, analyse d'algorithmes... L'I2M est ouvert à des collaborations, notamment dans des domaines d'application actuellement abordés (génomique, traitement des images,...)

IBDM

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Nom : Institut de Biologie du Développement de Marseille

Sigle : IBDM

Numéro : UMR 7288

Adresse : IBDM - Case 907 - Parc Scientifique de Luminy - 13288 Marseille Cedex 9 France

Directeur : André LE BIVIC

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Nom : Magali CONTENSIN

Mail : magali.contensin@univ-amu.fr

Tél : 04 91 26 93 24

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Gestion de grandes masses de données de microscopie (images, films) pour l'exploration quantitative et intégrée de systèmes vivants à différentes échelles spatiales : moléculaires, cellulaires, tissulaires, organismes entiers.

b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
 Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, stockage, accès, préservation, protection, sécurité informatique, sécurité juridique

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)
 - ERC Advanced Grants BIOMECAMORPH (T. Lecuit)
 - ERC Consolidator Grants SPOT.COM (B. Prud'homme)
 - LabEx INFORM (T. Lecuit)
 - Infrastructure nationale biologie - santé FBI (PF. Lenne)
 - AAP Emergence & innovation : SpongEX (A. Le Bivic)
 - ANR ATMIR (H. Cremer)
 - ANR EVOLNET (B. Prud'homme)
 - FRM « Mesoscopic origins of cell behaviours during tissue morphogenesis: mechanics and biochemical signaling » (PF. Lenne)
 - FdF « Traitement des informations thalamiques au sein des ganglions de la base en situation normale et parkinsonienne. » (L. Kerkerian)
5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
 - Projet Emergence PhagoHox (Y. Graba)
 - Projet Emergence SUZIDEX (B. Prud'homme)
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

INT

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Institut de Neurosciences de la Timone (INT), UMR7289 CNRS-AMU
Directeur : Guillaume Masson

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

Sylvain Takerkart, Sylvain.Takerkart@univ-amu.fr, 04 91 324 007

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

Acquisition et traitement de grande bases de données issues d'expériences de neurosciences fondamentales et cliniques

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, extraction d'information

Stockage, bases de données

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Acquisition de données en continue

Installation d'un nouvel imageur IRM 3T (financé par FLI)

Création d'une plateforme d'imagerie photonique

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

IRSIC

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Institut de Recherche en Sciences de l'Information et de la Communication (IRSIC EA-4262), 21, rue Virgile Marron - 13392 Marseille Cedex 05

Directeur : Françoise Bernard

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Ibekwe-SanJuan, fidelia.ibekwe-sanjuan@univ-amu.fr, 0684372708

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

Le laboratoire IRSIC s'intéresse aux enjeux sociétaux du Big Data et de l'Open Data et aux dimensions épistémologiques et éthiques de leurs usages.

Plus précisément, des chercheurs de ce laboratoire s'intéressent aux enjeux de la collecte, du traitement et d'usages de grandes masses de données dans différents domaines tels que la santé, les médias et le journalisme, la publicité et le e-marketing, la recherche scientifique, etc

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Usages des TICs, Traitement, mining, droit d'accès, perspectives socio-cognitives, éthique, épistémologie, humanités numériques, Big Data, Open Data, web 2.0, observation participante (ajoutez si besoin)

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

- Réception et Influence de l'e-publicité (display) : expérimentation en collaboration avec l'Université de Cologne (contexte psycho -socio-cognitif)
- Expérimentation pour mettre au point un dispositif numérique sur le web (projet MOVIDA (Contrat de recherche avec le Ministère de l'écologie, du développement durable et de l'énergie 190 – 0190-THUR-BSAF) : "Apports de la communication engageante dans le cadre du co-voiturage et de l'autopartage", en collaboration avec l'entreprise Auxilia (Paris)
- Mises en œuvre de nouvelles techniques de communication par e-mailing pour les associations humanitaires (dans le cadre des collectes de fond ex : collaboration avec l'UNICEF) et pour les entreprises commerciales (convention avec une entreprise d'e-commerce en PACA) : recherche action dans le cadre théorique de la « communication engageante numérique »
- Economie numérique, intelligence économique et veille informationnelle : Etude des connaissances et de la complexité ; Usages du web dans un contexte d'« intelligence économique » (ex : études des dépôts de brevets ; Les enjeux économiques de la géolocalisation pour les réseaux sociaux numériques)
- Réflexion sur une problématique des usages psychologiques et sociaux des medias sociaux en tant que « conscience collective virtuelle » et leurs liens avec les émotions collectives.
- La relation entre communication et engagement en actes des acteurs, nos travaux enrichissent les théories et pratiques consacrées au changement en étudiant comment les acteurs individuels et collectifs entrent dans un cours d'action et adoptent de nouvelles signification et valeurs. Cette question concerne les mutations technologiques, culturelles,

organisationnelles, institutionnelles, environnementales, économiques et sociétales. De nombreux projets de recherche financés sur ce thème et de nombreuses publications.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU
 - Accroître les occasions de rencontres interdisciplinaires
 - Identifier les domaines d'expertises des différents laboratoires sur la thématique de Big Data et de l'Open Data
 - Organiser des colloques / journées d'étude à l'AMU autour de cette thématique
 - Mettre en place un master recherche sur les « Humanités Numériques » fédérant plusieurs disciplines (SHS et Sciences)
 - Favoriser l'émergence des projets AMU sur ce thème pour ensuite se préparer à la réponse à d'autres projets d'envergure

iSm2

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
Institut des sciences moléculaire (iSm2) UMR 7313 Campus St Jerome

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
Stéphane Humbel stephane.humbel@univ-amu.fr 04 91 28 86 67

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire
 - partie expérimentale (au spectropole de la fédération de chimie) acquisition + stockage de spectres d'analyse
 - théorique : calculs sur le cluster de la fédération de chimie ou sur le cluster d'équipe (84 CPU – 512 Go Mémoire – 5 To Disque)

 - b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)
- Calculs en chimie quantique, conservation de données.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

LAM

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
Laboratoire d'Astrophysique de Marseille (LAM)
38 rue Joliot Curie, 13388 Marseille cedex 13
directeur : Jean-Gabriel Cuby
2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone
SURACE Christian : Christian.surace@lam.fr
3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire
 - i. Intérêt dans l'étude de datamining dans les données astrophysiques, (cosmologie, planètes extra solaires).
 - ii. Collaboration : Etudes de data mining commencées avec S. MacConnell (U Trent Canada), N. Novelli (LIF).
 - iii. Compétences : visualisation de grands échantillons de données, visualisation 3D, Base de données astrophysiques, Développement d'application de traitement de données.
 - b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
 - i. Traitement
 - ii. data mining
 - iii. stockage
 - iv. accès
 - v. préservation
 - vi. visualisation
4. Projets en cours dans le laboratoire (collaborations nationales, internationales)
 - a. BigData
 - i. Projet MASTODONS : PETASKY (coll nationale)
 - ii. Projet MASTODONS : PREDON (coll nationale)
 - b. Autres projets astrophysiques (internationaux)
 - i. EUCLID. Mission ESA
 - ii. PFS. Projet télescope Subaru (Japon). Collaboration internationale
 - iii. SVOM. Mission CNES et agence spatiale chinoise.
 - iv. FIREBALL. Projet Ballon CNES
 - v. PLATO. Mission ESA
 - vi. CHEOPS. Mission ESA
 - vii. ...
5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
 - a. Projets astrophysiques
 - i. E-ELT. European Extremely Large Telescope
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU
 - a. Recherche collaboration pour
 - i. Data mining dans les données

LBA

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
Laboratoire de Biomécanique Appliquée LBA UMR_T24
Faculté de Médecine – secteur Nord – Boulevard Pierre Dramard – 13916 Marseille Cedex 20
Directeur : Pr S. BERDAH
Directeur Adjoint : P.J. ARNOUX

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

a. Nom, mail, téléphone

LLARI Maxime

maxime.llari@ifsttar.fr

Tel : 0491658009

3. Domaines d'études relatifs aux grandes masses de données

a. Courte description des intérêts et compétences du laboratoire

Le LBA est une unité mixte de recherche IFSTTAR/Aix Marseille Université composée de 39 collaborateurs. Implanté au cœur de la Faculté de Médecine, sur le Campus Hospitalo-Universitaire Nord, la singularité du Laboratoire de Biomécanique Appliquée provient de l'approche pluridisciplinaire et transversale entre Sciences pour l'Ingénieur et Médecine. Ceci a permis de développer de manière synergique [des thématiques de recherches basées sur l'étude biomécanique du corps humain depuis le traumatisme jusqu'à sa réparation](#). Ses recherches conjuguent connaissances cliniques et anatomiques, approches expérimentales et théoriques du comportement des tissus et structures du corps humain, modélisation et simulation numérique.

[Nos thématiques de recherche suivent un objectif unique : l'Homme Virtuel](#). Pour cela, le laboratoire a cultivé une originalité scientifique en développant différentes familles de modèles numériques du corps humain et privilégiant une analyse biomécanique multi-échelle (du tissu isolé jusqu'au corps humain).

[Nos perspectives de recherche sont centrées sur l'Homme Virtuel pour comprendre les traumatismes, les prévenir et les réparer](#).

Le LBA de par ses activités de simulation numérique (plusieurs clusters de calcul, des modèles éléments finis lourds), ses implications cliniques et ses activités expérimentales génère et manipule de gros volumes de données, et doit avoir une politique de sécurité robuste (confidentialité des données, coopération avec des industriels sur des sujets sensibles avec des problématiques de brevets, expertises...).

Enfin, le LBA que ce soit avec ses tutelles (AMU et Ifsttar) ou encore dans le cadre de partenariat doit concilier des besoins prégnants d'espaces de partage de l'information (en particulier dans le cadre d'un Laboratoire International AMIDEX BSIP avec le Québec).

b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

[Stockage, accès, sécurité informatique, simulation numérique, expérimental](#).

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Laboratoire International Associé **LIA AMIDEX BSIP** avec les partenaires l'Ifsttar LBA, l'APHM, le CNRS, L'ETS Montréal, Polytechnique Montréal, L'hôpital du Sacré Coeur de Montréal et l'hôpital Sainte Justine de Montréal. Implanté sur deux villes avec 6 sites géographiques : Montréal et Marseille.

Coopérations internes AMU et Ifsttar (département transport santé sécurité)

Projets de recherche du LBA, ...

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Projet SECU2RM (en négociation sur financement Fondation en sécurité routière)

Projet de DHU (AMIDEX) simulation en santé

Plus largement l'ensemble des coopérations industrielles du Laboratoire.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Actuellement les moyens disponibles coté AMU sont peu connus et mal identifier. Par défaut, le Laboratoire a jusqu'ici privilégié l'utilisation du réseau Ifsttar.

Il serait donc très intéressant d'avoir un échange approfondi sur notre environnement informatique et les outils et dispositifs développés coté AMU en vue d'une meilleure utilisation (par le Laboratoire) et d'un meilleur appui informatique de l'AMU.

LCB

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Laboratoire de Chimie Bactérienne
 UMR7283 CNRS AMU
 31 Chemin Joseph Aiguier
 13009 Marseille
 Directeur: Prof. BARRAS Frédéric

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

TALLA Emmanuel
talla@imm.cnrs.fr
 0491164508

3. Domaines d'études relatifs aux grandes masses de données

- a. *Courte description des intérêts et compétences du laboratoire*

Les thématiques de recherche relatives aux grandes masses de données sont les suivantes :

- Génomique Comparative / Analyse des génomes
- Développement de bases de données relationnelles
- Développement d'outils et de programmes bioinformatiques
- Analyse des données de séquençage à haut débit

Le laboratoire dispose d'un service de criblage phénotypique des bactéries basé sur la microscopie. Ce service produit de dizaines de milliers d'images lors d'une campagne de criblage. Actuellement, nous travaillons sur les dispositifs nécessaires à l'exploitation de ces grandes masses de données.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

- Traitement
- Data mining
- Stockage
- Accès

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

- Analyse des données issues de séquençage à haut débit (RNA-seq, ..)
- Analyse à grande échelle de données génomiques procaryotes
- Développement d'outils pour l'identification et l'analyse des séquences répétées au sein de génomes

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

- Deux appels à projets ANR 2015 ont été soumis en lien avec l'analyse des données de séquençage à haut débit.
- Un projet de collaboration avec une équipe du laboratoire CREATIS à Lyon (CNRS UMR 5220 – INSERM U1044 – Université Lyon 1). Ce projet concerne l'annotation automatique d'images de bactéries en fluorescence.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Les banques de données biologiques, l'acquisition des images ainsi que les données issus des analyses devenant de plus en plus importantes, nous aimerons à l'avenir avoir à disposition des

banques de données et des bases de données généralistes ou spécialisées, ainsi que des plateformes logicielles donnant accès à des outils de bioinformatique et d'analyses d'images, des plus génériques aux plus pointus. Ces outils et collections de données peuvent être développés en interne ou déployés à partir de ressources publiques. Cela permettra également l'intégration de données hétérogènes, la mise en place de chaînes de traitement automatisées, la définition d'ontologies, la fouille de texte. Pour cela, une infrastructure matérielle importante (grappes de calcul, serveurs spécialisés, stockage) permettant de mettre en œuvre ces activités est nécessaire.

LIF

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Laboratoire d'Informatique Fondamentale (LIF), UMR CNRS 7279 ;

LIF

Parc Scientifique et Technologique de Luminy

163, avenue de Luminy - Case 901

F-13288 Marseille Cedex 9 ;

Directeur : Jean-Marc TALBOT (jean-marc.talbot@lif.univ-mrs.fr)

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Noël NOVELLI ; noel.novelli@lif.univ-mrs.fr ; 04 91 82 93 70

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

2 équipes de recherche *Bases de Données Avancées (BDA)* et *équipe Apprentissage et Multimedia (QARMA)* sont directement concernés par le thème *Big Data*.

Les intérêts sont les suivants :

- fouille de données, détection de corrélations et de dépendances, approximation avec modèle coût (contrôle des ressources), visualisation de données massives, problème de classification de grande taille (nombres d'exemples, nombre de classes, dimensionnalité), apprentissage budgétisé, apprentissage de représentations pour le texte et l'image, parallélisation.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement et gestion des données, Mining, Accès et optimisation, Visualisation, Apprentissage statistique pour des grandes collections de données plates et structurées.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Dans le cadre des appels à projets MASTODONS du CNRS depuis 2012 :

- Projet AMADEUS 2012 à 2014 (LaBRI, LIF, LAM, LIRMM, CEREGE, Espace DEV, Université de TRENT Canada) ;

- Projet PetaSKY depuis 2014 (LaBRI, LIF, LAM, LIRMM, Université de TRENT Canada, LIRIS, LIMOS, Observatoire de Paris, PRISM).

Projet ANR :

- Blanc ANR Class-Y (fin 02/2015) ;
- Blanc ANR GRETA (fin septembre 2015).

Projet européen BioASQ (fin 12/2014).

Collaborations avec UJF (Grenoble), Demokritos Athènes (Grèce), S. McConnell Université de TRENTO (Canada).

Projet eBob 2.0, 2014 à 2016 (investissement d'avenir) :

Gestion de grandes masses de données avec différents partenaires industriels sur le thème du e-achat et exploration des technologies liées au *Big Data*.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Projets en cours de montage sur le diagnostic médical et l'annotation de documents médicaux / dossiers médicaux.

Dépôt d'ANR Défi de tous les savoirs avec le laboratoire d'anthropologie ADES, Timone.

Poursuite des travaux cités au paragraphe 4.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Collaboration déjà en cours avec le LAM (BDA et QARMA) à poursuivre.

Autre laboratoire identifié : CPPM.

Besoins :

- Une plateforme *Big Data* pour les tests opérationnels dans différents contextes (environnement logiciels, solutions techniques, paramétrages...).
- Des données réelles ainsi que des experts des données.
- RH (IE/IR) pour la gestion de la plateforme et des tests.

IMBE

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

**UMR 7263 Institut Méditerranéen de Biodiversité et d'Ecologie
marine et continentale (IMBE)**

Station marine d'Endoume
rue de la Batterie des Lions
13007 Marseille
France

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

DAVID Romain romain.david@imbe.fr 0491041651 et 0614743054

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

Données en écologie et sur la biodiversité, développement de graphes à partir de ces données, fouille de donnée, systèmes d'observations et d'informations répartis
Données génomiques, métagénomiques, transcriptomiques, RNASeq, RADSeq
Analyses de polymorphisme, d'expression de gènes, de phylogénie

- a. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, mining, normalisation, interopérabilité, traçabilité, accessibilité des données, ontologie et vocabulaire contrôlé, droit d'accès, analyses statistiques

Après publication les données génétiques sont soumises à des bases de données publiques (Genbank par exemple) mais il y a un intérêt pour les stocker localement afin de poursuivre les analyses.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

<http://www.cigesmed.eu> WP6 (data mining concernant les données sur la biodiversité des habitats coralligènes)

<http://www.indexmed.eu> (indexation des données sur la biodiversité méditerranéenne)

ANR ADACNI (<http://adacni.imbe.fr>)

Projets internes à l'unité IMBE

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Optimiz (Optimization of uses and capabilities of ecological services provided by the hard substrates of biogenic origin with high added value in the context of the integrated coastal zone management (CZIM): Comprehensive approach (ecological, economic and sociological) and use of large sets of heterogeneous data. Acronym: OPTIMIZ)

Divers projets liés à la génomique

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Bien identifier les compétences en rapport, leur disponibilité et leur intérêt pour monter des projets en commun et les réaliser efficacement

Des collaborations sont déjà en cours avec le service informatique de l'OSU Pythéas pour stocker ces données. Ceci pourrait être mutualisé plus largement à l'échelle d'AMU

Il y a un intérêt fort pour l'organisation de sessions de formation en bioinformatique pour la manipulation de ces données: les compétences existent au sein d'AMU pour organiser cela.

LPL

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
Laboratoire Parole et Langage (LPL), UMR 7309
5 rue Pasteur
13100 Aix-en-Provence
Directeur : Noël Nguyen

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
a. Nom, mail, téléphone
Bernard Bel, bernard.bel@lpl-aix.fr, 06 70 08 19 90

3. Domaines d'études relatifs aux grandes masses de données
a. Courte description des intérêts et compétences du laboratoire

1) Le LPL est partenaire de l'EquipEx ORTOLANG (<http://ortolang.fr>) qui vise à instaurer en France un sous-ensemble du réseau européen CLARIN (<http://clarin.eu>) dédié aux données numériques de la linguistique. Le LPL est à l'initiative du centre de ressources SLDR (<http://sldr.org>) qui est appelé à fusionner avec le CNRTL de Nancy (<http://www.cnrtl.fr>) sur la plateforme commune d'ORTOLANG. ORTOLANG est complémentaire pour la linguistique de l'offre générale proposée par Huma-Num. Il s'appuie sur le dispositif Huma-Num/CINES pour l'archivage pérenne des données à valeur patrimoniale.

2) Le LPL est dépositaire de bases de données de parole pathologique très spécifiques (et uniques en France).

Voir <https://speedi-db.lpl-aix.fr/physio/>

b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, stockage, accès, identifiants pérennes, stockage sécurisé, archivage pérenne, droit d'accès, sécurité juridique, Code du patrimoine

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)
Le LPL est porteur du LabEx BLRI (<http://blri.fr>) consacré à l'étude du traitement du langage et de ses bases cérébrales.

Il aussi activement associé au programme Erasmus Mundus (<https://www.em-multi.eu>).

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
Voir <http://www.lpl-aix.fr/index.php?id=38>

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

LSIS

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

LSIS UMR 7296, Domaine universitaire de Saint Jérôme, Avenue Escadrille Normandie Niemen, 13397 MARSEILLE Cedex 20,
Directeur : Mustapha Ouladsine

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Patrice Bellot, patrice.bellot@lsis.org, 04 91 05 60 79
- b. Mohamed QUAFAROU, mohamed.quafafou@univ-amu.fr, 04.91.05.60.58

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

- Recherche et extraction d'information dans de grandes masses de documents (Web, bibliothèques numériques)
- Suivi d'opinion, contextualisation et détection de nouveauté sur les réseaux sociaux
- Fouille et Intégration de Données
- Gestion de données : élaboration de systèmes et langages, intégration (interopérabilité), qualité des données, données spatiales et biologiques
- Cloud computing : Provenance, cloud/web services
- Big data analytics : Algorithmes probabilistes et hypergraphes
- Architecture orientée services : conception d'architectures offrant différents services distribués pour l'acquisition, le stockage et l'analyse de données massives (Big Data).
- Structuration non supervisée et supervisée de masses de données multimodales, vidéo (les collections d'images du web), audiovisuelles (ex : les challenges NIST TREC VIDEO) ou audio (parole et bioacoustique).
- Modèles neuronaux profonds, les plus performants en indexation de masses de données d'images et sons
- Analyse de données pour la surveillance, pilotage et aide à la décision pour les systèmes complexes.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

- Text Mining
- Traitement, mining, stockage, accès, préservation, droit d'accès
- protection, sécurité informatique
- clustering non supervisé
- Data Mining
- Accès
- Digital Libraries
- Infrastructure, stockage, traitement, accès, droit d'accès, sécurité informatique, qualité, provenance, workflows scientifiques, services, cloud.
- Apprentissage, classification

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

--- Equipex OpenEdition.org (DILOH — Digital Library for Open Humanities)

- Investissement d'Avenir – Fonds Société Numérique : projet Intertextes
- Investissement d'Avenir – Fonds Société Numérique : projet MAGE
- pilotage du projet Scaled Acoustic Biodiversity (SABIOD) pour la Mission Interdisciplinaire du CNRS MASTODONS, Masses de DONNEES SCIENTIFIQUES, regroupant IN2SI, INSU, INSB, INEE, financé depuis 2012, sur appel à projet national. SABIOD regroupe plusieurs UMR et laboratoires internationaux, voir détails sur <http://sabiod.org>
- Projet européen COST « Keystone » : Recherche d'information et bases de données
- Projet AMIDEX catégorie « Transfert » (LSIS/LIF) : Fouille de vidéos
- Projet AMIDEX e-santé
- Action Prospective « Gros Volumes de Données Spatiales », GDR MAGIS (Méthodes et Applications pour la Géomatique et l'Information Spatiale),
- ICT COST Action IC1302 KEYSTONE « semantic KEYword-based Search on sTructured data sOURcEs », membre du comité de pilotage (pour la France)
- « Gestion de la Provenance dans le Cloud » : thèse en cours et participation à l'action transversale "Cloud Computing" du GDR ASR (Architecture et Systèmes Répartis)
- Environnement pour le développement d'applications NFC sécurisées : CIFRE Gemalto
- Analyser les données concernant plus de 1 200 000 associations et leurs adhérents afin de mettre en place des synergies entre associations, citoyens et sponsors locaux (entreprises, commerçants) grâce à des services de financement innovants du secteur associatif.
- collaboration avec l'Assistance Publique Hospitalière de Marseille et dans le cadre de la thèse de L. El Sarraj (soutenue en juillet 2014), nous avons développé un prototype de système d'aide à l'exploitation d'un entrepôt de données pour la gestion hospitalière
- collaboration avec l'Université du Pernambuco (Recife, Brésil) nous travaillons sur l'extraction d'information de textes à base d'ontologie
- Réseaux sociaux et Ciblage (Partenariat LSIS, Galoo)
- Comportement d'achat (Partenariat LSIS, Marketshot)
- collaborons depuis plusieurs années avec l'UFPE (Université Fédérale du Pernambuco (Recife- Brésil),
- Projet Européen Avec ST Rousset sur l'analyse des données pour la qualité des Wafers.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

- Thème Recherche d'information sémantique en domaine médical (fouille de dossiers patients, de documents médicaux) a ANR (partenaire = CHU Rouen ; contacts en cours avec AP-HM)
- Thème Recherche d'information dans des bases de brevets a ANR (partenaire = Office des Brevets, Genève)
- Autres thèmes : Gestion des risques,
- « Gestion de Big data biologiques (ADN) » : thèse en collaboration avec l'NSERM (fin 2014)
- ANR EnVGI (appel 2014), « Une nouvelle approche d'intégration, d'analyse et de visualisation de l'information géographique volontaire basée sur la qualité pour l'observation de l'environnement ».
- Autres projets « Big Data » : discussions en cours avec Atos, GFI, ...
- Big Data et géolocalisation (Université de la Californie): Acquérir des données en continue et les analyser afin de mettre en évidence des phénomènes locaux

géolocalisés et modéliser leur évolution dans le temps.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

--- Besoins génériques: au vu de ma connaissance du paysage AMU, experts « data management », systèmes/infrastructures (réseau, Cloud, langages), statisticiens (à un degré moindre).

--- Collaborations souhaitées (recherche): fournisseurs de Big Data (biologistes, physiciens, chercheurs en sciences sociales, etc.), curateurs de données, statisticiens,

...

--- Collaborations pour la formation : interaction avec les responsables des cursus en sciences exactes et SHS.

--- Montage des projets thématiques sur le Big Data dans des domaines tels que l'histoire, l'art, les biotechnologies, l'astronomie, etc.

--- Faire le lien entre le Big Data et la société, par exemple Urbain Analytics. Cela nous amène à renforcer nos relations avec des chercheurs des sciences sociales et des acteurs gestionnaires des territoires.

--- Mise en commun d'infrastructure exploitée par une recherche pluridisciplinaire.

MESOCENTRE

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Mésocentre AMU, Site Etoile/St Jérôme Case 521, Avenue Escadrille Normandie Niemen, 13397 Marseille Cedex 20

Responsable scientifique : Pr. Nicolas Ferré

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Nicolas Ferré, nicolas.ferre@univ-amu.fr, 0413550532

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

Le mésocentre d'AMU est une structure dépendant de la DRV et basée sur un financement Equipex obtenu en 2010. Il donne accès à des ressources de calcul hautes performances (HPC), susceptibles de générer de grandes masses de données et de les post-traiter (visualisation). Ses utilisateurs sont représentatifs de tous les champs disciplinaires d'AMU présentant un besoin en HPC.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, stockage, accès, droit d'accès, protection.

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Actuellement, un projet de rapprochement entre le mésocentre et la grille du CPPM est en cours de développement. Lorsqu'elles sont disponibles, les ressources du mésocentre pourront être intégrées à Dirac, le logiciel de gestion de la grille du CPPM et participer ponctuellement au traitement très efficace de données massives.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

La question de la résidence des données est essentielle pour répondre efficacement à la problématique Big Data. En collaboration avec la grille du CPPM, nous avons proposé et demandé le financement d'une approche à 2 niveaux, un stockage sécurisé non pérenne à proximité des moyens de calcul, géré par le mésocentre et/ou la grille, et un stockage sécurisé pérenne assuré notamment par la DOSI. Ce modèle peut être discuté et amélioré.

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Le découpage support/soutien à l'activité de recherche mis en place par AMU doit être rationalisé de façon à optimiser les solutions de stockage Big Data, leurs interconnexions, les migrations entre les solutions etc. La mutualisation des ressources de stockage et leur intégration dans un plan à l'échelle d'AMU doit permettre de limiter les coûts supportés par les unités.

SESSTIM

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur
SESSTIM – UMR912, Jean-Paul MOATTI
2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone
DUFOUR Jean-Charles, jean-charles.dufour@univ-amu.fr, 04 91 32 46 00
3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

L'UMR 912- SESSTIM est une unité interdisciplinaire composée de chercheurs et d'enseignant-chercheurs ayants différentes formations (Economie, Médecine de différentes spécialités, Sciences Humaines, Biostatistiques, Sciences de l'information, Informatique médicale,...) et travaillant au sein de différentes institutions (Hôpitaux, Observatoire Régional de la Santé, Institut Hospitalo Universitaire Méditerranée infection, Institut Cancérologie,...).

Le SESSTIM est composé de 3 équipes dont les thématiques générales sont :

- Équipe 1 : Cancer, Biomédecine & Société (CAN-BIOS)
- Équipe 2 : Maladies transmissibles, systèmes de santé, sociétés (ESSEM)
- Équipe 3 : Économie de la santé et vieillissement (CRISSPOP)

Ces équipes travaillent au sein de 9 groupes projets transversaux dont certains ont un rapport direct avec les Big Data parce qu'ils font déjà usage de ce type de données et/ou développent des méthodes d'analyses les concernant. Il s'agit des groupes projets :

- Médecine personnalisée : cancer, autres pathologies
- Veille Sanitaire
- Méthodes quantitatives, traitement de l'information médicales

De plus, le SESSTIM porte un intérêt tout particulier aux méthodologies d'enquêtes et recherches axées sur l'apport des média sociaux et technologies du web 3.0 (objets connecté) pour la pratique médicale et la meilleure connaissance des problèmes de Santé en population (veille sanitaire notamment).

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Interopérabilité des données de santé
Nomenclatures et référentiels médicaux
Analyses spatiales appliquée à la Santé
Traitements et méthodologies Biostatistiques
Epidémiologie

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)
 - Projet INTERFACE dont l'objectif est la mise en place d'une plateforme nationale d'accès aux grandes bases de données françaises en santé (caisse d'assurance maladie, hôpitaux, ...). Le SESSTIM participe au workpackage « interopérabilité et développement

d'algorithmes » et se positionne sur l'étude et la mise en œuvre des métadonnées du projet INTERFACE.

- Projet Medihandtrace : Analyse des parcours des soignants et de l'hygiène des mains en utilisant le recueil en continu des signaux émis par des puces RFID implantées dans les sabots du personnel du service MIT (hôpital Nord, Pr Brouqui)
 - D'autres projets dans le domaine de la surveillance sanitaire en lien avec l'épidémiologie et les sciences humaines et sociales, notamment, sont en cours
5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
- Participation à la préparation du projet européen PHRIMA (Design study à l'initiative de l'ITMO Santé Publique) : Proposal for a European Public Health Research Infrastructure for Sharing of health and Medical Administrative data - Ce projet se fait en lien avec des partenaires européens (Italie [POLIMI], Grande-Bretagne [UOM], Espagne [ISCI], Danemark [SSI], Allemagne [UDUS+FAU], Scandinavie [Nordforsk]) ainsi que le CNRS en France.
 - Projet PHASE (Public Health Awareness from Social Exchanges) ;FUI19 : Le projet a pour but la mise en place d'une surveillance qui permettra l'exploitation dans le domaine de la santé et des épidémies d'informations provenant des médias sociaux. Une fois captée et analysée, cette information sera mise à la disposition des citoyens, organismes publiques ou privés, laboratoires de recherche, ministères et organismes étatiques... sous la forme d'une plateforme internet dédiée à la surveillance et l'analyse de l'activité d'échange social sur internet en lien avec la santé publique. Partenaires industriels Inbox, Invoxis, Sindup.
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Nous sommes particulièrement intéressés par des collaborations dans lesquelles nous pourrions d'une part apporter notre **expertise SHS et Santé** pour la **valorisation de Big Data** dans ce domaine et d'autre part bénéficier de soutien pour l'extraction et la mise à disposition de ces Big Data.

INSERM UMR S910

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

INSERM UMR_S910, Faculté de Médecine La Timone, 4ème étage, 27 boulevard Jean Moulin 13385 Marseille Cedex 05

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Christophe Bérout, christophe.beroud@inserm.fr

3. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

Laboratoire impliqué dans l'étude génétique des maladies rares. Gestion des grands volumes de données générées par le séquençage des exomes et génomes. Compétences dans la création de bases de données génétiques et cliniques, dans le développements d'algorithmes permettant l'analyse de ces données, développement de systèmes de prédiction du caractère pathogène des mutations et de validation d'hypothèses thérapeutiques.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Stockage, traitement, extraction de savoir, alalgorithmes, outils de prédiction, génomique, phénotype, mutations

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Projet Européen RD-CONNECT, responsable du WP "Clinical Bioinformatics" et participation à la création de la plateforme permettant le lien entre les données -omics, les registres et les biobanques;

Porjet National de création de la banque de donnée nationale de variations issues des expériences de NGS dans le cadre des maladies rares;

Projet nationla de création de la banque nationale des CNV;

TREAT-NMD Alliance : développement et hébergement des banques de données internationales des gènes DMD et SMN (50 pays impliqués);

Développement, hébergement, curation, de nombreuses LSDB de référence au niveau international;

Développement, hébergement de registres de patients nationaux et internationaux.

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Création d'une infrastructure européenne dédiée à la "genotype-based medicine"

Développement de partenariats public/privés pour favoriser le développement de bases de données de référence dans le domaine de la génétique humaine dans le cadre d'activités de diagnostic

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

CRCM

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Centre de recherche en cancérologie de Marseille (CRCM)
Inserm (UMR1068), CNRS (UMR7258) , Université Aix-Marseille (UM105)
27 Boulevard Leï Roure BP30059
13273 Marseille Cedex 9
Directeur : Jean-Paul Borg

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

Bernard Chetrit Bernard.chetrit@inserm.fr 04 86 97 72 09

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

La marque de fabrique du CRCM est de faire progresser la biologie et la médecine des cancers en travaillant à l'interface du physiologique et du pathologique, du fondamental et de l'appliqué, de la recherche et de la clinique.

L'inclusion plus large de patients dans des essais cliniques, et l'utilisation de nouvelles technique à haut débit, permet de générer de plus en plus de données biologiques en lien avec des formes cliniques précises, en augmentant la complexité et les demandes en analyse et traitement de données, et avec l'obligation de traiter des volumes de données considérables. Ce changement d'échelle permet d'affiner le diagnostic, le pronostic et les traitements en cancérologie et d'aboutir à ce qu'on appelle la médecine personnalisée.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, Data mining, stockage, accès, préservation, droit d'accès, cancérologie, médecine personnalisée

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)
5. Projets futurs en préparation au laboratoire: thématiques, appel à projet
6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

L'information scientifique générée par les laboratoires de recherche en biologie, que ce soit en immunologie, ou cancérologie ou autres possède une immense valeur pour les patients et les chercheurs. Cependant, seule une partie limitée de cette information est utilisée par les biologistes et cliniciens.

Parmi les raisons de ces limitations sont la difficulté des laboratoires à gérer rationnellement l'information qu'ils génèrent, la complexité de l'extraction de connaissances biomédicale à partir de bases de données existantes, et les difficultés d'intégration des données et de leur confrontation à

des données tierces. De plus, les données biomédicales disponibles ne sont souvent que partiellement annotées par les équipes qui les ont générées, que ce soit sur des aspects clinique ou phénotypique. Aux problématiques de l'exploitation des données existantes, s'ajoute la problématique de l'évolution très rapide des technologies en biologie. Ces technologies demandent de nouveaux types de traitement informatique et apportent de l'information à résolution beaucoup plus élevée, par exemple en biologie moléculaire, notamment en séquençage ou en protéomique, ce qui augmente considérablement la complexité des traitements et le volume des données.

Eccorev

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

ECCOREV, FR3098, Europôle de l'Arbois, 13545 Aix-en-Provence cedex 4

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

- a. Nom, mail, téléphone

Rotereau, Armand, 06 75 18 07 66

3. Domaines d'études relatifs aux grandes masses de données

- a. Courte description des intérêts et compétences du laboratoire

ECCOREV est une fédération de recherche qui regroupe une quarantaine de laboratoires en sciences de l'environnement ; à ce titre, elle fédère un certain nombre de bases de données pluridisciplinaires sur le climat, la biodiversité, le fonctionnement des forêts, la pollution de l'air, de l'eau et du sol.

- b. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, stockage, accès

4. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Labex OT-Med : catalogue des données sur le climat de la Méditerranée

European Pollen Database (réseau européen)

Coopérate pour la sauvegarde et l'accès aux données de fonctionnement de la forêt (Labex OT-Med, OSU Pytheas)

Observatoire du Chêne Blanc à l'OHP (Labex OT-Med, OSU Pytheas)

Observatoire Homme-Milieu Bassin Minier de Provence (réseau Labex DRIIHM)

5. Projets futurs en préparation au laboratoire: thématiques, appel à projet

AMEDEES : Interaction climat et sociétés en région Méditerranéenne (projet soumis à l'ANR)

Extension de COOPERATE : Fédération de l'accès aux données des stations de mesures en forêt en région Méditerranéenne

Big City : interaction des bases de données sur l'environnement de la Métropole d'Aix-Marseille (climat, pollution, biodiversité, urbanisme, sociologie, santé)

6. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Faciliter l'accès à un centre de stockage et de traitement des données de l'AMU avec une aide informatique

Etablir un réseau de collaborations concernant l'interopérabilité des bases de données et la fouille de donnée

Formation au niveau master : data mining, aspects juridiques, gestion des bases de données

LID2MS

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Laboratoire interdisciplinaire de droit des médias et des mutations sociales
EA 4328
3 Avenue Robert Schuman
Espace René Cassin
13628 Aix-en-Provence cedex 1
France
Directeur : Hervé ISAR

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :

Dominique Augey
dominique.augey@univ-amu.fr
Tel : 06 31 40 23 03

3. Domaines d'études relatifs aux grandes masses de données
 - Droit et économie des données personnelles
 - Droit des bases de données
 - Protection de la vie privée et des libertés fondamentales
 - Contrôle du traitement des données à caractère personnel et de la libre circulation de ces données
4. Courte description des intérêts et compétences du laboratoire

Le LID2MS compte aujourd'hui 17 enseignants/chercheurs permanents, dont 14 HDR, auxquels s'associent 17 autres enseignants, chercheurs et professionnels partenaires. Il abrite 20 doctorants et a fait soutenir, en moins de trois ans, plus d'une dizaine de thèses.

Scientifiquement la démarche pluridisciplinaire constitue la posture centrale du Laboratoire même s'il revendique néanmoins un attachement fort à la recherche en science juridique et en science économique. Par ailleurs, la recherche du laboratoire est organisée autour des trois triptyques suivants :

- Médias, communication et information
- Art, culture et identité
- Libertés, mémoires et religions

C'est évidemment dans le cadre du triptyque « Médias, communication et information » que se situe l'intérêt du laboratoire pour le « Big Data ».

En effet, les aspects juridiques des données massives, d'ailleurs ici intimement liés aux aspects économiques de cette nouvelle réalité, sont assez nombreux et particulièrement porteurs en ce qui concerne l'étude de possibles mutations sociales.

Les chercheurs du LID2MS sont ainsi présents sur les thèmes :

- Droit et économie des données personnelles (spécifiquement de la protection des données personnelles et du marketing des données en nombre).
- Droit et économie des réseaux et des services de télécommunications

- Droit et économie des objets connectés
- Droit et économie des bases de données
- Droit et économie des villes intelligentes

5. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Protection des données personnelles, Vie privée, CNIL, marketing des données, sécurité

6. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Un colloque intitulé Médias 014 Is 'Big Data' beautiful a été organisé par le laboratoire le 12 décembre à la Faculté de Droit sur le site aixois de Schuman et sa publication, sous forme électronique, est en cours.

Par ailleurs, une thèse sur les aspects juridiques de l'e-réputation réalisée dans le cadre d'un contrat SISE vient de débiter et elle contiendra des développements importants sur le thème de la désanonymation des données.

7. Projets futurs en préparation au laboratoire : thématiques, appel à projet

Les suites des révélations d'Edward Snowden et les débats autour du droit au déréférencement consacré par la Cour de justice de l'Union Européenne vont nécessairement continuer à alimenter les réflexions des juristes comme des économistes sur le traitement et la conservation des données massives. A ce titre, le projet de règlement européen sur les données personnelles sera au centre des recherches du LID2MS, tout comme les mesures visant à organiser une lutte coordonnée de l'Europe contre le terrorisme (PNR, etc.) ou le projet de loi relatif au renseignement actuellement en discussion au niveau du parlement.

8. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaitées au sein de l'AMU

Concernant l'étude des aspects juridiques et économiques du big data, les chercheurs du LID2MS ont surtout besoin de rencontrer des chercheurs d'autres disciplines abordant les données massives dans une démarche appliquée. Ce seront en effet les aspects tangibles et concrets de leur recherche qui seront potentiellement révélateurs des dimensions juridiques et/ou économiques de leur démarche et qui permettront de mettre en évidence de nouvelles problématiques ou pistes croisées de recherche.

Par ailleurs, les corpus juridiques étant en eux-mêmes susceptibles de constituer des « données massives », une recherche sur ce type particulier de « Big data » pourrait certainement être envisagée.

CIML

1. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Centre d'Immunologie de Marseille – Luminy (CIML)
Aix Marseille Universit_e UM2, Inserm, U1104, CNRS UMR7280,
Parc Scientifique de Luminy, case 9056
13288 Marseille cedex 09, France
Directeur : Eric Vivier

2. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

- Lionel Spienelli, spinelli@ciml.univ-mrs.fr, 04 91 26 91 90.
- Sébastien Jaeger, jaeger@ciml.univ-mrs.fr, 04 91 26 91 90.

3. Domaines d'études relatifs aux grandes masses de données
 - b. Courte description des intérêts et compétences du laboratoire

Le CIML est composé de :

- 17 équipes de recherche utilisant divers organismes modèles et approches complémentaires pour explorer le système immunitaire. Ces équipes sont donc en majorité constituées de biologistes avec une minorité de bioinformaticiens, physiciens ou statisticiens.
- Une petite dizaine de plateformes techniques commune générant une grande quantité de données expérimentales (cytometrie de flux, microscopie, histologie, FCS, etc.)
- Une plateforme bioinformatique prenant en charge une grande partie des traitements de données de type « Big Data » (NGS, biopuces) ainsi qu'une expertise en statistique et modélisation.

- c. Choisissez des mots clés qui vous semblent pertinents pour votre activité
Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Immunologie, génomique, modélisation, traitement de données de séquençage à haut débit, imagerie du vivant.

7. Projets en cours dans le laboratoire (collaborations nationales, internationales)

Plus d'une cinquantaine de projets sont actuellement en cours au CIML dont plus de la moitié génèrent des données de type « Big Data », que ce soit sur le plan génomique (NGS) ou imagerie.

8. Projets futurs en préparation au laboratoire: thématiques, appel à projet

Là encore, la liste est longue...

9. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

Le CIML, de par les compétences qu'il peut apporter et les problématiques qu'il rencontre aujourd'hui, se place certainement d'avantage du côté de l'utilisateur et bénéficiaire des progrès récents et futur en matière de « Big Data » qu'en tant que réel acteur de ces progrès. En effet, l'explosion exponentielle de la quantité de data générées par la plupart des techniques émergentes en Immunologie et plus généralement en biologie nous amène au constat que ni l'infrastructure ni les compétences raisonnablement disponibles ne nous permettront de faire face à tous les problèmes associés à cette explosion. Nous sommes donc particulièrement conscient de l'importance que représente le traitement et la gestion de « Big Data » et sommes tout disposés à apporter tout le soutien et l'aide possible, toute dérisoire soit-elle.

CRCM

7. Identification laboratoire : Nom (acronyme), numéro, adresse, directeur

Centre de recherche en cancérologie de Marseille (CRCM)
 Inserm (UMR1068), CNRS (UMR7258) , Université Aix-Marseille (UM105)
 27 Boulevard Leï Roure BP30059
 13273 Marseille Cedex 9
 Directeur : Jean-Paul Borg

8. Personne(s) de contact pour la thématique « Big Data » PR2I-AMU :
 - a. Nom, mail, téléphone

Bernard Chetrit Bernard.chetrit@inserm.fr 04 86 97 72 09

9. Domaines d'études relatifs aux grandes masses de données
 - a. Courte description des intérêts et compétences du laboratoire

La marque de fabrique du CRCM est de faire progresser la biologie et la médecine des cancers en travaillant à l'interface du physiologique et du pathologique, du fondamental et de l'appliqué, de la recherche et de la clinique.

L'inclusion plus large de patients dans des essais cliniques, et l'utilisation de nouvelles technique à haut débit, permet de générer de plus en plus de données biologiques en lien avec des formes cliniques précises, en augmentant la complexité et les demandes en analyse et traitement de données, et avec l'obligation de traiter des volumes de données considérables. Ce changement d'échelle permet d'affiner le diagnostic, le pronostic et les traitements en cancérologie et d'aboutir à ce qu'on appelle la médecine personnalisée.

Les thématiques de recherche du laboratoire

Biologie : Cancérologie, médecine de précision, médecine personnalisée, (in)stabilité du génome, réparation de l'ADN, immunité antitumorale, signalisation cellulaire, polarité cellulaire, migration cellulaire, cancer du sein, cancer du pancréas, leucémies, hémopathies malignes, bioinformatique, biologie computationnelle, drug design and discovery, chémobiologie, chémoinformatique, stress cellulaire, oncologie moléculaire, hématopoïèse, cellules souches cancéreuses, thérapies ciblées, recherche translationnelle, immunothérapie, anticorps monoclonaux, nanobodies, nanovecteurs,

génomique, épigénétique, protéomique, Next Generation Sequencing, transcriptomique, approches molécule unique, imagerie

b. Choisissez des mots clés qui vous semblent pertinents pour votre activité

Traitement, mining, stockage, accès, préservation, droit d'accès, protection, sécurité informatique, sécurité juridique (ajoutez si besoin)

Traitement, Data mining, stockage, accès, préservation, droit d'accès, cancérologie, médecine personnalisée

10. Projets en cours dans le laboratoire (collaborations nationales, internationales)

SIRIC Marseille (Site de Recherche Intégrée en Cancérologie) (Label Institut National du Cancer – Inserm – Direction Générale des Soins)

Institut Carnot : CALYM (Le Consortium pour l'Accélération de l'innovation et de son transfert dans le domaine du Lymphome)

Investissement d'Avenir : Démonstrateur : MIMAbs : Antibodies for Medicine

Projet Européen TRANSCAN : New technologies for early detection of HLA loss leukemia relapse after allogeneic hematopoietic stem cell transplantation, and identification of relevant risk factors

A*MIDEX interdisciplinaire « Méthodes optiques innovantes pour l'analyse des interactions et signalisations moléculaires en biologie », avec [l'Institut Fresnel](#), "**BIOCLOAK: Bio-Cloaking with Graphene and Metamaterials**" avec [Institut Fresnel](#), A*MIDEX Emergence «PDZ targeting hits cancer communication»,

A*MIDEX interdisciplinaire "**NOVel UltraSonic BIOMarkers of Tumor response to systemic anticancer therapies and their integration on mathematical predictive model**", projet le Laboratoire de Mécanique et d'Acoustique ([LMA](#))

A*MIDEX Chaire d'Excellence: **Pedro BALLESTER** (cancérologie/bio informatique)

Projets avec Institut Fresnel (molécule unique, vectorisation de médicaments, imagerie endoscopique), CINAM (interaction cellules-substrats, vectorisation de médicaments), Ecole des Mines de Gardanne (cellules tumorales circulantes)

11. Projets futurs en préparation au laboratoire: thématiques, appel à projet

- Structuration du NGS (Next Generation Sequencing) à visée diagnostique des cancers.
- Médecine personnalisée dans le traitement du cancer (projet SIRIC Marseille sur le cancer du sein, pancréas, gliome et leucémies)

12. Commentaires et propositions afin d'identifier les besoins, expertises, collaborations recherchées ou souhaités au sein de l'AMU

L'information scientifique générée par les laboratoires de recherche en biologie, que ce soit en immunologie, ou cancérologie ou autres possède une immense valeur pour les patients et les

chercheurs. Cependant, seule une partie limitée de cette information est utilisée par les biologistes et cliniciens.

Parmi les raisons de ces limitations sont la difficulté des laboratoires à gérer rationnellement l'information qu'ils génèrent, la complexité de l'extraction de connaissances biomédicale à partir de bases de données existantes, et les difficultés d'intégration des données et de leur confrontation à des données tierces. De plus, les données biomédicales disponibles ne sont souvent que partiellement annotées par les équipes qui les ont générées, que ce soit sur des aspects clinique ou phénotypique. Aux problématiques de l'exploitation des données existantes, s'ajoute la problématique de l'évolution très rapide des technologies en biologie. Ces technologies demandent de nouveaux types de traitement informatique et apportent de l'information à résolution beaucoup plus élevée, par exemple en biologie moléculaire, notamment en séquençage ou en protéomique, ce qui augmente considérablement la complexité des traitements et le volume des données.

Le laboratoire peut mettre à disposition dans le cadre de collaboration ses compétences de développement d'outils et de programmes dans les domaines de la bioinformatique et de la chemoinformatique, L'analyse des données de séquençage à haut débit. Le développement d'applications de traitement de données massive. Une infrastructure de calcul et de développement à haut débit, La génération de données de masse dans le domaine de la cancérologie (omics: NGS, protéomique, cytométrie de flux,...) et particulièrement dans le champ de la médecine personnalisée

NORT

- **Laboratoire / Directeur / Personnel total (permanent recherche, total, visiteurs/an-estimation).**

Laboratoire NORT « Nutrition Obésité et Risque Thrombotique », DU Marie-Christine Alessi, personnel : 43 permanents, 80 total, visiteurs/an environ 10

- **Thématiques de recherche aux laboratoires (grandes lignes et sous-thématiques ; par exemple : Physique, nanotechnologies, lasers etc.)**

Thème général : science de la vie et de l'environnement (SVE)

Sous thèmes : Physiologie, Physiopathologie, Cardiologie, Technologies Médicales

Domaines applicatifs : santé humaine et nutrition

La thématique du laboratoire NORT est basée sur l'identification de biomarqueurs et de nouvelles voies physiopathologiques dans le domaine de la santé cardiovasculaire.

Les approches de biologie classique, qui ont permis un certain nombre de découvertes majeures, présentent par ailleurs certaines limites, constituant autant de verrous scientifiques, qui peuvent être en partie levés par des approches globales, basées notamment sur la biologie des systèmes, combinant diverses techniques dites « omiques » dont la génomique avec l'essor des nouvelles technologies de séquençage haut débit, la transcriptomique, la métabolomique et les statistiques appliquées. Ces approches agnostiques sont des approches modernes et puissantes particulièrement adaptées à l'identification des déterminants non classiques, nutritionnels et génétiques de la pathologie cardiometabolique. La prise en compte multi-échelle des phénomènes biologiques permet une compréhension fine des phénotypes observés, mais nécessite la mise en place d'infrastructures lourdes, de moyens humains et de compétences techniques et scientifiques spécifiques notamment celles utiles à la gestion et à la modélisation des données (big data). Les moyens tant humains que financiers issus des activités de contractualisation constituent un levier efficace pour lever les verrous technologiques induits par ce nouveau domaine de recherche, mais aussi pour investir:

développement de nouvelles procédures analytiques, contribution à l'achat d'équipements, constitution d'une banque de données de référence, modélisation statistique et développement d'algorithmes en lien notamment avec la start-up de traitement statistique « MS nutrition » du laboratoire.

- **Equipes de recherche liées aux thématiques big data (thèmes, mots clefs).**

Equipe 1 : micronutriments, impact en santé métabolique et vasculaire Directeur P Borel

Equipe 2: biomarqueur du risque vasculaire Directeur MC Alessi

Plateforme CRIBIOM (Criblage Biologique Marseille) gérée par protisvalor issue d'un partenariat NORT avec 6 partenaires privés et académique pour offrir des prestations d'explorations du vivant à haut débit (omiques) : Responsable / JC Martin.

Start'up MS Nutrition : en lien avec CRIBIOM, traitement statistique des données à haut débit ; responsable M Maillot F Vieux

- **Thèmes de recherche liés aux grandes masses de données.**

Equipe 1 L'un des thèmes de recherches est l'identification chez l'humain de polymorphismes génétiques liés à la biodisponibilité de micronutriments d'intérêt santé (vitamines, composés phytochimiques comme les caroténoïdes...). L'objectif à terme est d'optimiser les apports en fonction du fond génétique et de proposer des solutions de nutrition personnalisée. Pour cela, nous menons des travaux cliniques sur des cohortes limitées mais très bien définie d'un point de vue génétique ; nous avons établi des scores génétiques « multiplexes » à partir de plusieurs centaines de milliers de polymorphismes, permettant d'associer des ensembles de polymorphismes capables d'expliquer une part importante des variations interindividuelle en concentration sanguine de certaines molécules d'intérêts après un repas (Desmarchelier C, J Clin Endocrinol Metab, 2014).

Equipe 2 : Nous disposons de plusieurs larges cohortes d'individus et de familles atteintes de pathologie thrombotique veineuse et artérielle. Nous avons acquis ces dix dernières années une expertise internationalement reconnue dans le domaine des études d'association génome entier et sommes leader en France. Grâce à cette approche menée sur 7,507 patients et 52,632 contrôles, nous venons d'identifier deux gènes de susceptibilité de la thrombose (Germain M, Am J Hum Genet. 2015) offrant de nouvelles opportunités prometteuses pour comprendre la physiopathologie et traiter la thrombose veineuse.

Nos projets sont de continuer avec les stratégies « omiques » en utilisant les outils de dernières générations plus discriminants. En collaboration avec la suède (Science for Life Laboratory, J Odeberg, Stockholm) nous développons une approche protéomique originale visant à la découverte de nouveaux biomarqueurs. Grâce à cette approche nous avons déjà identifié une protéine d'intérêt ayant donné lieu à un dépôt de brevet. Cette technologie va être appliquée à 1500 patients. Ce programme intitulé PROVENCE (PROteomics based approach to biomarker discovery for diagnosis of Venous thrombosis and prediction of venous thrombosis recurrENCE program) a pour but d'extrapoler le travail pilote à des échantillons de taille plus importante et de prédire la survenue et la récurrence de thrombose veineuse mais aussi de mettre en place des programmes de screening de molécules pharmacologiques ciblées sur les nouvelles voies identifiées.

Plateforme CRIBIOM (<http://www.protisvalor.com/site/fr/cribiom>) et PF BIOMET de Métabolomique (labellisée AMU)

Cette plateforme gérée par PROTISVALOR (filiale AMU) a été fondée par un partenariat contractuel (2014-2020) avec 5 partenaires industriels et 1 académique afin de fournir des prestations dans le domaine des sciences du vivant appliqué à la santé humaine. Sa spécificité est de s'appuyer sur les technologies de biologie des systèmes d'exploration du vivant développé au niveau académique pour mettre en évidence et caractériser les effets biologiques de molécules bioactives issues de l'alimentation et de contaminants liés à l'industrie nucléaire. Ces technologies à haut débit sont à base de « omiques », comme la transcriptomique, l'épigénétique et la métabolomique dont pour cette dernière nous disposons de ressources matérielles particulièrement importantes (3 spectromètres de masses dont deux à haute résolution). Nous avons développé toute une infrastructure reconnue internationalement pour analyser et traiter les mégadonnées issues de ces technologies (Grison, S, et al, *Metabolomics*, 9, 1168-1180, 2013, Martin, JC et al, *Metabolomics*, 11, 807-821, 2015 et Martin, JC et al, *Am J Physiol Heart Circ*, 309, H935-H945). L'une de atouts pour cette plateforme est également de nous adosser sur la start'up MS Nutrition du laboratoire pour le traitement statistique avancé des mégadonnées générées.

- Programmes interdisciplinaires en cours : description rapide mais axée sur l'excellence scientifique (mentionner les thématiques, les coopérations internationales, rôles de coordination des chercheurs, "milestones", publications, prix etc.) Mentionner ici les programmes d'envergure nationale (Labex, Equipex etc.) et internationale.

Notre unité a acquis au cours de la dernière décennie une reconnaissance nationale et internationale dans le domaine de la nutrition et en particulier sur les déterminants de la biodisponibilité et les effets cardiométaboliques des micronutriments. Nous avons consolidé et développé notre expertise dans le domaine de la biologie des systèmes, incluant la métabolomique et dans l'utilisation et l'exploitation des données issues des techniques de séquençage de nouvelle génération (GWAS, NGS, séquençage d'exons). Notre positionnement santé, nutrition thrombose et biologie des systèmes est unique en France. Notre efficacité et donc la valeur ajoutée de notre prestation repose sur le maintien et le développement de notre savoir-faire, ainsi que sur la disponibilité de ressources humaines suffisantes pour mener à bien des projets. La mise à disposition de cette expertise et de notre infrastructure en contre-partie d'un financement contractuel de ressources humaines et de moyens renforce l'efficacité de nos prestations et notre réactivité pour les partenaires. Nous sommes leader national et international dans le domaine de la thrombose. Nous coordonnons le centre de référence national sur les pathologies plaquettaires, ainsi qu'un workpackage du labex Genmed et avons coordonné la plus large méta-analyse internationale dans le domaine des déterminants génétiques de la thrombose veineuse.

JC Martin a coordonné un programme transnational de validation de l'approche ouverte de métabolomique (16 plateformes, 3 pays et une structure recherche de la commission européenne : « Can we trust untargeted metabolomics : Results of the metaboring initiative, a large-scale multi-instruments inter-laboratory study », *Metabolomics*, 2015). Il est aussi membre actif du CA du réseau francophone de métabolomique et fluxomique.

programmes financés par les « investissements avenir »,.....

Grâce à notre expertise unique en métabolomique, nous sommes partenaires de 2 PIA, lancés en 2014 : le projet PRIODAC (2,5 M€) et le projet POLYGONE (1,2 M€). Le premier projet vise à redéfinir la posologie du traitement à l'iode stable des populations consécutivement à un accident nucléaire, et aboutira à une nouvelle AMM du produit mis

à disposition du public. Outre notre laboratoire, il comprend comme autre partenaire l'IRSN, le CEA et la pharmacie centrale des armées (PCA).

Le projet POLYGONE a pour objectif d'identifier pour le colza (première plante cultivée en Europe) des pratiques culturales durables permettant d'obtenir des plantes à plus value santé pour le consommateur.

Dans ces deux projets, les technologies d'exploration biologique à haut débit sont utilisées afin d'identifier et de caractériser les impacts sur la santé humaine.

- **Liste des enseignements délivrés par le laboratoire dans des thématiques proches des sciences de données. Noms des enseignants, niveau, nombre d'heures, audience (nombre d'étudiants).**

PE Morange : Responsable du module de m : « Master 2 maladies polyfactorielles » dans le cadre du master pathologie humaine Faculté de médecine marseille (1 sem module optionnel). Cet enseignement est totalement dédié aux approches omiques/

MC Alessi : Responsable du module de master 2 intitulé « mécanismes moléculaires et cellulaires de la thrombose » dans le cadre de la spécialité « nutrition, régulations métaboliques et pathophysiologie vasculaire » Enseignement d'une semaine associant cours théorique et analyse d'articles portant en partie sur les maladies rares de l'hémostase et l'intérêt des approches génomiques (NGS) dans l'identification des gènes coupables

JF Landrier (enseignant et responsable du module), Implication dans le Master 2 Physiopathologie Humaine, UE Santé et alimentation, de l'épidémiologie aux mécanismes moléculaires ; cours sur la nutrigenomique, nutrigenétique,, épigénétique (12h) ;

JC Martin, Intervenant dans 2 masters à Aix-Marseille Université et 1 master à l'université de Toulon sur l'application de la métabolomique et des big data dans le domaine de la santé (12h).

- **Nom des personnes et équivalent temps plein qui pourrait justifier d'une participation à des projets de recherche dans un IC.**

Members			Permanent	FTE	Total
Plateforme CRIBIOM/start'up MS Nutrition					
Researcher	DR2	MARTIN Jean-Charles	YES	0.5	
	CR1	LANDRIER Jean-François	YES	0.5	
		TOURNIAIRE Franck	CDI	1	
engineer	IR	SVILAR Ljubica	CDI	1	
Lecturer	MCF	DEFOORT Catherine	YES	0.1	
Technical staff	TR	ANTONA Claudine	YES	1	
	AI	TARDIVEL Catherine	YES	1	
CDI	TR	Recruitment underway	YES	1	
	TR	ASTIER Julien	YES	1	
Data Engineer	IR	MAILLOT Matthieu	YES	1	
	IR	VIEUX Florent	YES	1	
TEAM 3					

Researcher	CR1	PEIRETTI Franck	YES	0.2	
	CR1	CANAULT Mathias	YES	1	
	CR1	GRINO Michel	YES	0.2	
	CR1	RITEAU Beatrice	YES	0.2	
Lecturer	MCF	POGGI Marjorie	YES	1	
Clinical researcher	PUPH	ALESSI Marie-Christine	YES	0.5	
	PUPH	MORANGE Pierre	YES	0.4	
	PUPH	DUTOUR Anne	YES	0.4	
	MCUPH	GABORIT Benedicte	YES	0.4	
	PUPH	BONNET J- Louis	YES	0.1	
	PUPH	CUISSET Thomas	YES	0.2	
	PUPH	CHAMBOST Hervé	YES	0.2	
	PH	BARTHET M-Christine	YES	0.5	
	PH	BRUNET Dominique	YES	0.5	
	MCUPH	BACCINI Veronique	YES	0.1	
AHU	SUCHON Pierre	No	0.5		
Post doc	Post doc	THOMAS Grace	No	1	
Technical staff	TECH	BONARDO Bernadette	YES		
	AI	BASTELICA Delphine	YES	0.8	
	ADT	GUICHARD Philippe	YES	1	
	AI	ANCEL Patricia	No	0.5	
	ING	GOUMIDI Louisa	YES	1	
hospital resource	ING	SAUT Noemie	YES	0.2	
	TEC	GEORGELIN Odile	YES	0.5	
	TEC	ROSTAN Agathe	YES	0.5	
TEAM 1					
Researcher	DR1	BOREL Patrick	YES	0.5	
Post-Doc		DESMARCHELIER Charles	NO	1	
Technical Staff	TR	NOVICKI Marion	YES	0.5	
	TR	HALIMI Charlotte	YES	0.5	

- **Infrastructures liées à cette thématique dans le laboratoire.**

Plateforme Métabolomique par spectrométrie de masse (labellisée AMU)

- **Liaison avec le monde socio-économique (contact entreprises, projets communs)**

plateforme CRIBIOM (Criblage Biologique Marseille) gérée par protisvalor issue d'un partenariat NORT avec 6 partenaires privés et académique pour offrir des prestations d'explorations du vivant à haut débit (omiques),
start'up MS Nutrition collaborant avec CRIBIOM pour l'analyse des méga-données générées.

Screening de drogues (Projet PROVENCE)

- **Projets qui pourraient être mis en œuvre au sein de l'IC :**

Programmes de recherche interdisciplinaires possibles au sein d'une structure locale comme IC

Amélioration de la visualisation des données et de l'extraction

Collaboration santé / mathématiques : développement d'algorithmes de prédiction

Groupement des projets par thème majeurs de santé

Formation / Enseignement (un master interdisciplinaire sur la science des données est en discussion, pour réunir les initiatives plus ciblées et donner plus d'ampleur, tout en profitant de la large assise pluridisciplinaire de l'AMU) ;

Infrastructure interdisciplinaire :

Création de centre de requêtes

Plateforme ouverte sur les big data interconnectées

Création de CRB multidisciplinaire

Projet liés au monde socio-économique

Plateforme CRIBIOM : utilisation des big data pour améliorer la compétitivité et l'innovation de partenaires industriels et académiques. Nouvelle AMM pour un produit destiné à la protection de l'ensemble de la population Française et Européenne (nouvelle posologie pour les tablettes d'iode stable administré en cas d'accident nucléaire, projet PRIODAC). Impact sur le monde agricole (nouvelles pratiques culturales pour le colza dans un contexte de durabilité, projet POLYGONE).

