

Vague A : Campagne d'évaluation 2014 - 2015

Unité de recherche

Dossier d'évaluation

Nom de l'unité : Entrepôts, Représentation et Ingénierie des Connaissances
Acronyme : ERIC
Nom du directeur pour le contrat en cours : D.A. Zighed, J. Darmont (depuis juillet 2012)
Nom du directeur pour le contrat à venir : J. Darmont

Type de demande :

Renouvellement à l'identique

Restructuration

Création ex nihilo

Choix de l'évaluation interdisciplinaire de l'unité de recherche :

Oui

Non

I. DOSSIER D'ÉVALUATION

1. Présentation de l'unité

Le laboratoire ERIC¹ est une unité de recherche (EA 3083) dont les établissements de tutelle sont l'Université Lumière Lyon 2 et l'Université Claude Bernard Lyon 1. Fondé à Lyon 2 en 1995 en tant que jeune équipe, ERIC a obtenu le statut d'équipe d'accueil en 1999. À partir de 2009, un regroupement, officialisé en 2010, a été opéré avec l'équipe MA²D (Méthodes et Algorithmes pour l'Aide à la Décision) de Lyon 1, sur la base de complémentarités thématiques en matière de recherche dans le domaine du décisionnel, d'une sensibilité commune au domaine des Sciences Humaines et Sociales (SHS) et de nombreuses collaborations sur les plans scientifique et pédagogique. Le début du contrat en cours a donc été un moment charnière d'intégration de nouveaux collègues au sein d'ERIC.

Le laboratoire est principalement hébergé à Lyon 2, sur le campus de la Porte des Alpes. Ses locaux, partagés avec le Département Informatique et Statistique (DIS) de l'Institut de la Communication (ICOM), représentent une surface utile de 400 m². Les membres Lyon 1 d'ERIC ne disposent sur le campus de la Doua, hormis leurs bureaux d'enseignants, que d'un bureau recherche partagé avec des collègues extérieurs au laboratoire. Les doctorants basés sur ce campus peuvent bénéficier d'un *open space*. Des espaces partagés ont donc été libérés à Lyon 2 pour les collègues de Lyon 1 de passage. ERIC a actuellement atteint sa capacité maximale en termes d'occupation des bureaux, personnels permanents et non-permanents compris.

¹ <http://eric.univ-lyon2.fr>

ERIC occupe une situation originale dans le paysage informatique lyonnais, d'une part par son positionnement scientifique ciblé sur l'informatique décisionnelle, alors que les trois autres laboratoires d'informatique de la place de Lyon, le LIRIS, le LIP et le DISP, sont positionnés sur des créneaux plus généraliste pour le premier (image, données, connaissances, services) et également très spécialisés (parallélisme et systèmes de production, respectivement) pour les deux autres. Toutefois, des proximités thématiques existent avec le pôle *data science* du LIRIS, l'équipe Modélisation, Intégration, Système d'Information du DISP, ainsi qu'avec l'équipe Systèmes d'Information du Centre Magellan (laboratoire de gestion) de l'Université Jean Moulin Lyon 3. Elles se traduisent actuellement par des participations croisées à des journées scientifiques, séminaires et jurys de thèse, par exemple. D'autre part, la tutelle de l'Université Lyon 2 et l'intégration en 2012 du laboratoire à l'Institut des Sciences de l'Homme (ISH) de Lyon, Unité de Service et de Recherche multitutelée dirigée par un membre d'ERIC, ouvrent des terrains de collaboration privilégiés au laboratoire dans les domaines des lettres, langues et SHS.

Politique scientifique

Le laboratoire s'est donné pour objectif scientifique de valoriser les grandes bases de données complexes, notamment dans les domaines des SHS, mais aussi en lien avec l'industrie à travers des thèses CIFRE et des contrats. Les recherches d'ERIC se situent dans les domaines suivants (Figure 1) :

- les entrepôts de données : intégration intelligente de données complexes, modélisation multidimensionnelle d'objets complexes, analyse en ligne personnalisée, sécurité du processus d'entreposage ;
- la fouille de données et la décision : apprentissage automatique, étude et fouille de graphes, analyse de données complexes, agrégation multicritère, fouille d'opinion, logiciels de fouille de données.

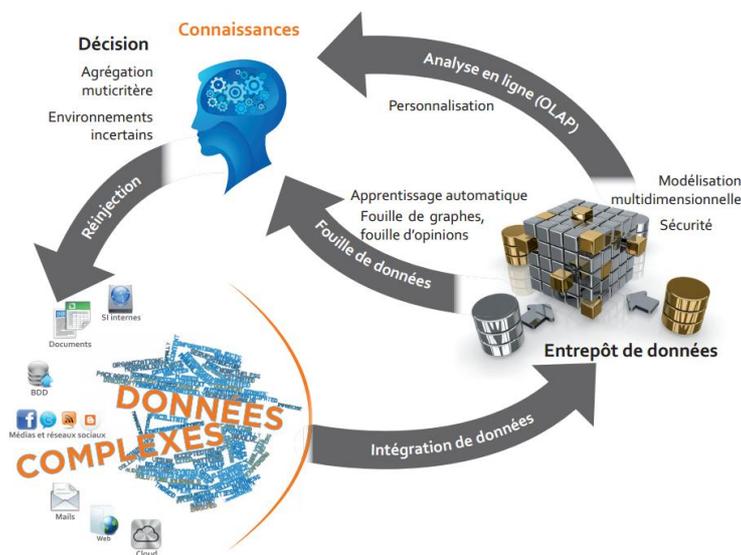


Figure 1 : Thématiques de recherche d'ERIC

Par ailleurs, suite à l'évaluation du laboratoire portant sur la période 2005-2008 (rapport en Annexe 15a²), nous nous sommes donné quatre grands objectifs, déclinés ci-dessous ainsi que les actions que nous avons mises en œuvre pour les atteindre et, le cas échéant, les résultats que nous avons obtenus.

1. *Amélioration de la qualité globale de la production scientifique* : tous les membres du laboratoire ont été incités à publier dans les revues internationales, ainsi que dans les conférences référencées dans les classements CORE/ERA et plus récemment QUALIS. Il a été fait appel à des services de traduction en anglais quand nécessaire. Le nombre de publications a conséquemment augmenté de 67 %³ dans les revues internationales et de 39 % dans les conférences internationales référencées dans les classements CORE/ERA/QUALIS (Section 2.1).

² Depuis sa fondation, ERIC organise statutairement (Annexe 5) une évaluation de son activité à mi-parcours, en sus des évaluations ministérielles ou de l'AERES. En 2012, l'Université Lyon 2 a de plus requis cette évaluation de toutes ses équipes d'accueil. Pour la mettre en œuvre, nous avons contacté la Professeure Maylis Delest, déléguée scientifique à l'AERES, qui a désigné pour nous un jury à qui nous avons demandé d'évaluer le laboratoire « à la façon de l'AERES » pour la période 2009-2012. Ses conclusions sont fournies en Annexe 15b.

³ Ce ratio, ainsi que toutes les évolutions quantitatives indiquées dans ce rapport, est calculé en rapportant la production scientifique de la période 2009-2014 à la même durée que la période 2005-2008, afin que la comparaison soit valide.

2. *Augmentation des ressources du laboratoire et de sa visibilité par des projets sélectifs de type ANR et PCRD* : les membres permanents du laboratoire ont été incités à porter et à s'associer à des projets de ce type. Deux projets européens et un projet ANR ont été obtenus pendant la période de référence, ainsi que la participation à un Projet Investissement d'Avenir démarré en février 2014, pour un montant global d'environ 600 k€.
3. *Clarification du positionnement stratégique du laboratoire dans son environnement immédiat* : le laboratoire s'est clairement engagé sur le terrain d'application des SHS, notamment en intégrant l'ISH, ce qui a permis de nombreuses collaborations avec ses laboratoires membres, en recrutant l'ex-directeur de la MSH de Toulouse et en participant au projet de Labex (non retenu) Humanités et Humanités Numériques.
4. *Adoption d'une structure plus simple à gérer que la matrice projets/axes de recherche avec rattachement principal et secondaire des enseignants-chercheurs proposée lors de l'intégration de l'équipe MA²D* : le laboratoire a été restructuré en 2012 en deux équipes de recherche thématiques (contre trois axes précédemment) avec un seul rattachement par membre du laboratoire.

De plus, nous avons travaillé de manière transversale sur la communication du laboratoire, afin de renforcer sa visibilité nationale et internationale vis-à-vis de partenaires potentiels (universités, entreprises, laboratoires de SHS...), par exemple dans le cadre d'appels à projets. Pour cela, nous avons financé une plaquette (en français et en anglais), fait réaliser un clip vidéo de présentation sous-titré en anglais (grâce au service audiovisuel de l'ISH) et présenté le laboratoire dans le Bulletin de l'AFIA (n° 81, novembre 2013) et la revue *ACM SIGMOD Record* (article en cours de révision). Ces supports de communication sont disponibles en ligne.

Au plan national, ERIC est *leader* des communautés scientifiques qui se sont constituées autour des entrepôts de données (journées EDA) et de la fouille de données (conférence EGC). Enfin, l'activité du laboratoire est reconnue au plan international, comme en attestent la production (revues et conférences internationales classées) et le rayonnement scientifiques de ses membres (participation aux comités de lecture des grandes revues et conférences internationales du domaine, mentionnées en Annexe 6), ainsi que les collaborations et projets internationaux qu'ils mènent (Figure 2).

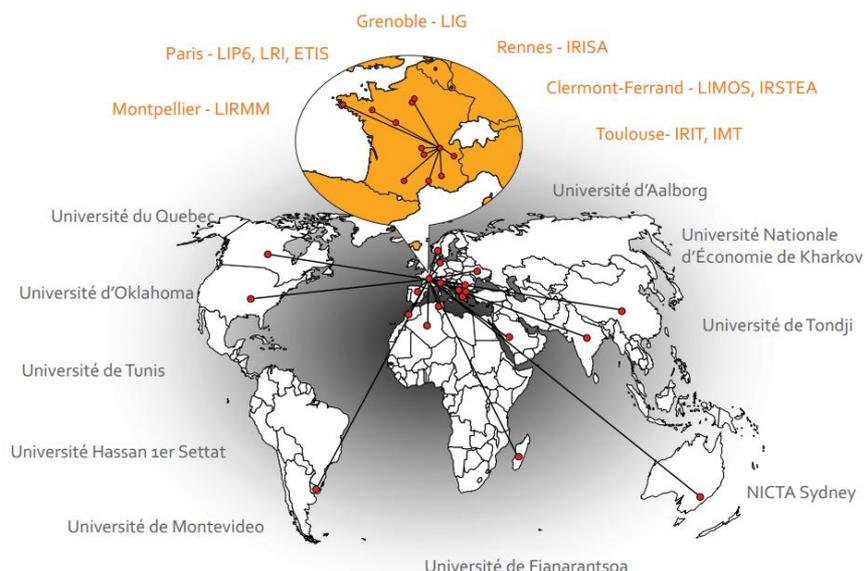


Figure 2 : Principales collaborations

Profil d'activités

| Unité/Équipe | Recherche académique | Interactions avec l'environnement | Appui à la recherche | Formation par la recherche | Total |
|--------------|----------------------|-----------------------------------|----------------------|----------------------------|-------|
| Ensemble | 40 % | 15 % | 15 % | 30 % | 100 % |

La production scientifique d'ERIC inclut de nombreuses copublications avec des collègues français (80) et étrangers (50). Nous contribuons également à la recherche académique par l'organisation de journées thématiques (deux à trois par an depuis 2012, rassemblant chacune 20 à 40 participants). Les interactions avec l'environnement prennent la forme de missions de conseil (Feu Vert, Aéroport de Lyon, CAF, INSERM...) que nous essayons de faire fructifier en contrats de recherche ou en thèses CIFRE. Le laboratoire a également hébergé de jeunes entreprises en lien avec l'incubateur CREALYS et invite régulièrement des professeurs étrangers (Canada, Espagne, Italie, République tchèque, Roumanie, Suisse, Ukraine ; un ou deux par an), ainsi que des doctorants étrangers pour des séjours courts dans le cadre des programmes PROFAS, TASSILI et Erasmus Mundus eLink (8 séjours d'une semaine à six mois). Notre contribution à l'appui à la recherche se fait

par les prototypes logiciels que nous mettons à disposition de la communauté (Annexe 6) et l'animation de groupes de recherche (AS INFORSID Décisionnel dans le nuage, GDR MACS et ROADEF, Pretoplogics) et de conférences aux plans national (EGC, EDA) et international (ateliers QIMIE@PAKDD et Cloud-I@VLDB). Enfin, la formation par la recherche inclut l'encadrement de projets et de stages orientés recherche dès le master 1, un encadrement doctoral actif et l'accueil régulier de doctorants et postdoctorants étrangers (Section 2 et Annexe 11). Les membres d'ERIC sont impliqués dans l'animation de deux masters orientés vers la recherche : Extraction de Connaissance à partir des Données (cohabilité par l'Université de Nantes) et *Data Mining & Knowledge Discovery* (Erasmus Mundus en partenariat avec l'Université Pierre et Marie Curie, Polytech' Nantes, l'Université Polytechnique de Catalogne, l'Université Polytechnique de Bucarest et l'Université du Piémont Oriental). Le directeur du laboratoire est également membre du conseil de l'école doctorale InfoMaths et de son comité des thèses, ainsi que représentant de l'école doctorale à Lyon 2.

Organisation et vie de l'unité

Organisation. Compte-tenu des axes de recherche développés au sein d'ERIC, le laboratoire s'est progressivement structuré d'un point de vue scientifique, depuis le regroupement de 2009, en deux équipes :

- Systèmes d'Information Décisionnels (SID), composée de 8 membres permanents ;
- Data Mining et Décision (DMD), composée de 14, puis de 13 membres permanents suite au détachement de B. Jouve au CNRS.

Chaque équipe est gérée par un responsable et dispose d'un budget propre depuis 2010, pris sur la dotation générale de fonctionnement globale (Lyon 1 et Lyon 2) du laboratoire et établi au prorata des effectifs (membres permanents et non-permanents). Chaque équipe renforce ces ressources par des contrats et des projets. De nombreuses collaborations existent entre les équipes, à travers des coencadrements de thèses et des projets transversaux comme ImagiWeb (ANR) ou REQUEST (PIA).

De plus le laboratoire met à disposition de ses membres des ressources mutualisées, dont des serveurs virtuels (stockage, espace web, calcul), un intranet documentaire, un photocopieur multifonctions (mutualisé avec le DIS) et une petite bibliothèque contenant des ouvrages dont les doctorants peuvent avoir besoin sur de longues périodes, ainsi que des revues auxquelles le laboratoire est abonné en propre.

Enfin, le laboratoire organise des séminaires communs réguliers (en moyenne une douzaine par an ; Annexe 14). Les intervenants sollicités sont à la fois des universitaires et des industriels (EADS, Orange, Xerox, Masa Group, AMI Software...) venant de toute la France et de l'étranger (Canada, Suisse, Royaume Uni, Italie, Roumanie...). Les séminaires sont ouverts à tous, étudiants de master compris. Enfin, chaque année, une journée « hors les murs » constitue une occasion de présenter des travaux (de doctorants ou de permanents) en interne à tout le laboratoire, ou en externe, comme lors de la journée de démonstrations de logiciels qui s'est déroulée à l'ISH en 2013.

Pilotage. D'un point de vue fonctionnel, compte-tenu de la double tutelle d'ERIC, il est apparu indispensable que le laboratoire soit dirigé par un directeur d'un des deux établissements de tutelle et un directeur adjoint de l'autre. L'organigramme du laboratoire est présenté en Annexe 4. Le conseil de laboratoire, qui se réunit tous les mois, est constitué de tous les membres permanents, d'un représentant élu du personnel administratif, d'un représentant élu des doctorants et post-doctorants et d'un représentant élu des membres associés.

Afin de proposer au conseil des éléments d'animation scientifique et de stratégie de laboratoire, nous avons mis en place une instance collégiale intermédiaire : le conseil de direction, constituée du directeur, du directeur adjoint, des responsables des équipes de recherche et du responsable administratif et financier. Le conseil de direction se réunit tous les mois, entre les conseils de laboratoire. Les statuts du laboratoire et son règlement intérieur sont fournis en Annexe 5.

Moyens humains. Au 1^{er} juin 2014, le laboratoire est composé de 52 membres, dont 21 enseignants-chercheurs (dont 9 HDR), 1 personnel BIATSS, 25 doctorants, 2 post-doctorants ou ATER et 3 membres associés. L'évolution des effectifs d'ERIC depuis 2009 est donnée dans la Figure 3. L'augmentation significative des effectifs en 2010 correspond au regroupement avec l'équipe MA²D. La Figure 4 précise l'origine des recrutements d'enseignants-chercheurs effectués au laboratoire depuis 2009. Sont considérés comme internes les recrutements de collègues ayant préparé leur thèse ou leur HDR au sein d'ERIC. L'apport lié au regroupement avec l'équipe MA²D est également dissocié des recrutements effectifs, tous effectués à Lyon 2. L'âge moyen des enseignants-chercheurs d'ERIC est d'environ 44 ans et la tranche d'âge la plus représentée est celle des 31-40 ans.

Depuis 2010, la responsabilité administrative et financière d'ERIC est assurée par un personnel BIATSS affecté à temps plein par l'Université Lyon 2. La direction accorde à sa formation continue une attention particulière (Annexe 13). Occasionnellement, le laboratoire recrute également sur ses fonds propres des personnels en CDD pour mener à bien des missions précises. Par exemple, un programmeur a été embauché (six mois à temps partiel en 2011-2012) pour mettre au point un outil de *reporting* des publications du laboratoire. Enfin, ERIC fait également ponctuellement appel aux services du technicien informatique du DIS, dont il partage les locaux. Ces prestations sont facturées au laboratoire.

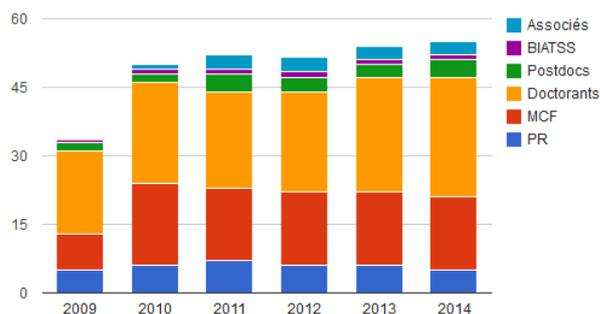


Figure 3 : Évolution des effectifs

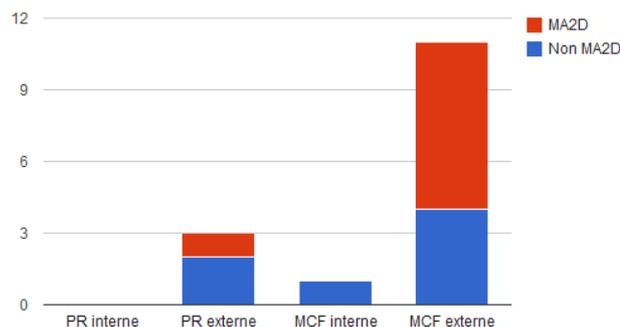


Figure 4 : Origine des recrutements de permanents

Ressources financières. Les recettes perçues par ERIC, dont l'évolution (hors taxe) depuis 2009 est présentée dans la Figure 5, proviennent de différentes sources de financement que nous classons en deux grandes familles :

1. les Dotations Générales de Fonctionnement (DGF) allouées par les universités de tutelle ;
2. les contrats académiques (Europe, état français, région Rhône-Alpes...) ou industriels (dont CIFRE) et autres subventions, sur lesquels est opéré un prélèvement de 10 % (hors ressources fléchées) par le laboratoire.

La Figure 5 montre la croissance importante des recettes du laboratoire (+24 % par an en moyenne), grâce à une augmentation de la DGF en 2010 suite à la fusion avec l'équipe MA²D, puis, la DGF diminuant lentement, mais régulièrement ensuite, surtout grâce à l'effort porté sur l'obtention de contrats de recherche. Les ressources issues des DGF et des prélèvements sur ressources propres sont allouées, d'une part, aux dépenses mutualisées du laboratoire (fonctionnement courant, matériel informatique, personnel le cas échéant) et, d'autre part, aux budgets propres des équipes de recherche, établis au prorata des effectifs (permanents et doctorants). Cette répartition est représentée pour 2014 (à titre d'exemple) dans la Figure 6.

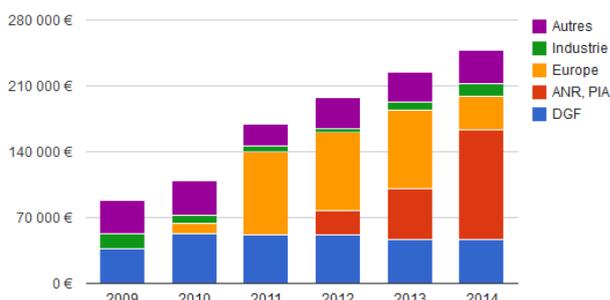


Figure 5 : Évolution des ressources financières

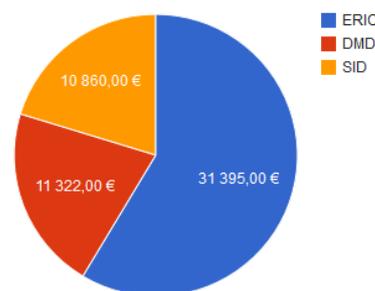


Figure 6 : Répartition 2014 DGF/prélèvements

Hygiène et sécurité. Les activités de recherche d'ERIC et les matériels dont nous disposons ne génèrent pas de risque particulier pour le personnel et ne requièrent donc pas de précaution particulière en termes d'hygiène et de sécurité. En revanche, les locaux Lyon 2 dans lesquels est hébergé le laboratoire présentent des défauts d'isolation et d'étanchéité. De l'amiante est également présent dans la colle des dalles du sol et dans certains plafonds mal identifiés. Ce constat, dressé en 2008, n'a été pris en charge que très partiellement (installation d'une climatisation dans trois salles ou bureaux en 2011, sur fonds propres du laboratoire). Toutefois, aucun accident n'a heureusement été à déplorer. Enfin, certains membres du laboratoire (direction, administration) ont suivi des formations de secourisme et de gestion des alertes incendies.

Faits marquants

- 2010 : Regroupement avec l'équipe MA²D. Cet événement représente une évolution importante du laboratoire, d'une part en lui permettant d'atteindre une taille critique et, d'autre part, en participant à sa stratégie de se renforcer dans le domaine des mathématiques appliquées, notamment en anticipant le départ à la retraite de deux collègues issus des statistiques. Cette stratégie s'est traduite par le recrutement de 3 MCF (dont un a été promu PR dans un autre laboratoire et remplacé) et 1 PR (actuellement en détachement au CNRS) de 26^e section. Les solides compétences en mathématiques appliquées (optimisation, notamment) développées au sein de MA²D par des collègues de 27^e section sont venues compléter ce renforcement.
- 2011-2014 : Coordination scientifique du projet européen FLURESP. Ce projet nous a permis réaliser un transfert de technologie des méthodes d'analyse multicritère vers le domaine de la santé, pour la

description et l'aide au choix de stratégies d'intervention en cas de pandémie humaine au niveau européen. Dans ce cadre, ERIC a développé une application à destination des professionnels de santé et des publications ont été rédigées dans ce domaine et sont en cours de soumission ("Multi-criteria assessment of influenza human pandemic public health responses", "Multidimensional analyses of public health intervention's characteristics targeting influenza human pandemic").

- 2012 : Présidence de l'organisation des conférences internationales jointes *Algorithmic Learning Theory* (ALT, de rang A) et *Discovery Science* (DS). Outre la portée scientifique de l'événement, son organisation en collaboration avec les laboratoires Hubert Curien (Université Jean-Monnet Saint-Etienne) et LIRIS (Université Claude Bernard Lyon 1 et INSA Lyon) a participé à la structuration du « paysage informatique » au sein de l'Université de Lyon (PRES puis COMUE). Enfin, les synergies mises en œuvre sont également nationales, avec l'implication de Telecom Bretagne.
- 2012-2015 : Porteur du projet ANR ImagiWeb, dont l'enjeu est de dévoiler les mécanismes qui procèdent à la production, la diffusion, l'évolution des opinions des internautes relativement à des entités qui peuplent le Web. La principale originalité du projet tient en un travail pluridisciplinaire qui se concrétise en une double contribution : en informatique, avec le développement de nouveaux algorithmes d'analyse des opinions pour prendre en compte l'aspect multi-facettes et temporel ; et en sciences sociales, avec le renouvellement des méthodes d'enquête traditionnelles pour les adapter aux nouveaux médias du Web.
- 2014 : Présidence de l'organisation du 32^e congrès INFORSID. Obtenue sous l'impulsion d'ERIC, l'organisation de cette manifestation scientifique française de référence dans le domaine des systèmes d'information a été pensée d'emblée comme une collaboration avec les laboratoires du site de Lyon, poursuivant ainsi la dynamique engagée avec le LIRIS (Université Claude Bernard Lyon 1 et INSA Lyon) et l'étendant au centre Magellan (Université Jean Moulin Lyon 3).
- 2014-2017 : Participation au Projet Investissements d'Avenir REQUEST. Porté par l'entreprise Thales et regroupant un consortium d'une dizaine d'entreprises (dont Orange et la SNCF), institutions et d'une demi-douzaine de laboratoires de recherche, son objectif est le traitement des problèmes liés au *big data analytics* à travers une démarche innovante et unique sur le plan international, qui allie les domaines *big data*, *big analytics*, *visual analytics* et *cloud computing*. ERIC est coordinateur d'une tâche relative au requêtage intelligent des données, dont l'objectif est de définir des méthodes innovantes de recherche par similarités, dans le cadre d'un processus mêlant recherche d'information et découverte de connaissances, itératif, interactif et tirant profit des technologies *cloud*.

2. Réalisations

2.1. Note sur le classement des publications

En l'absence de référentiel national, nous nous sommes basés sur deux classements internationaux (aucun ne couvrant totalement la production scientifique d'ERIC, notamment en mathématiques appliquées, et chacun étant critiquable à divers égards) pour attribuer un rang aux publications internationales du laboratoire :

- le classement australien CORE 2013/ERA 2010⁴, qui se base sur des critères subjectifs non divulgués ;
- le classement brésilien QUALIS⁵, qui se base sur un calcul de facteur d'impact et est remis à jour annuellement.

La grille de classement que nous avons conçue sur cette base est donnée dans le Tableau 1. « NR » y signifie « publication non référencée » dans le classement en question. De plus, dans la suite de ce rapport, la codification employée pour décrire les publications est la suivante :

- O - Ouvrages et direction d'ouvrages,
- R - Revues,
- C - Conférences avec comité de lecture et actes,

⁴ <http://core.edu.au>

⁵ <http://qualis.capes.gov.br/webqualis/>

- A - Autres publications ;
- I - Portée internationale,
- N - Portée nationale.

Ainsi, un article publié dans une revue internationale est par exemple noté RI.

| Rang CORE/ERA | Rang QUALIS | Rang ERIC |
|---------------|-------------|-----------|
| A* A | A1 A2 B1 NR | A |
| A* A | B2 B3 B4 B5 | B |
| B | A1 A2 | A |
| B | B1 B2 B3 NR | B |
| B | B4 B5 | C |
| C | A1 | A |
| C | A2 B1 B2 | B |
| C | B3 B4 B5 NR | C |
| NR | A1 A2 | A |
| NR | B1 B2 | B |
| NR | B3 B4 B5 | C |
| NR | NR | NR |

Tableau 1 : Classement des publications

Les listes des réalisations et produits de la recherche, des contrats institutionnels sur financements publics, des contrats industriels et sur financement privé, classés par équipe de recherche, figurent en Annexes 6 et 7

2.2. Équipe SID

Production scientifique

Les thèmes de recherche de l'équipe SID se situent dans le domaine de l'informatique décisionnelle (*business intelligence* ou BI). Nos travaux incluent à la fois des aspects méthodologiques et formels, ainsi que des aspects appliqués. Les problématiques étudiées sont concentrées autour des questions suivantes : intégration, modélisation et analyse en ligne (*On-Line Analytical Processing* ou OLAP) de données massives, entrepôts de données NoSQL (*Not Only SQL*) et décisionnel dans le nuage. D'autres travaux de recherche menés au sein de l'équipe se sont intéressés à la conception et à la construction d'entrepôts centrés utilisateurs, tout en assurant la qualité et la sécurité de leurs données.

Pour la période 2009-2014, nos activités de recherche se sont structurées en trois axes principaux : 1) entrepôts de textes, OLAP sémantique et personnalisation ; 2) analyse des réseaux d'information et 3) *big data analytics* et sécurité.

Axe 1 : Entrepôts de textes, OLAP sémantique et personnalisation

1.1. Entrepôts de textes. Les travaux de recherche dans le domaine des entrepôts de données se sont principalement focalisés sur l'analyse de données structurées ou semi-structurées pour lesquelles les dimensions d'analyse sont souvent représentées par des métadonnées préalablement définies. Cependant, ces méthodes d'entreposage ne sont pas appropriées pour analyser des contenus textuels pour lesquels les dimensions d'analyse, représentées par les thèmes abordés dans le texte, ne sont pas explicites mais enfouies dans le contenu du document. Pour aborder ce problème, nous avons proposé une nouvelle approche d'entreposage de données textuelles qui s'appuie sur des techniques de recherche d'information (RI) [Aknouche-12a]. Cette approche inclut l'intégration de données textuelles et un modèle d'entrepôt de textes en constellation étendue, qui prend en compte les dimensions lexicales (termes) produites à partir de la phase d'indexation de documents textes et les dimensions sémantiques (thèmes) extraites à partir d'une ontologie de domaine, ainsi qu'une sélection des faits inspirée des modèles statistiques de langue en RI [Aknouche-13b]. Pour analyser en ligne les données textuelles entreposées, nous avons proposé une mesure textuelle fondée sur un modèle vectoriel adapté [Oukid-13a] ainsi qu'un opérateur d'agrégation, ORANK (*OLAP documents RANKing*), permettant de classer les documents par ordre de pertinence [Oukid-13b]. Ces travaux constituent une avancée importante dans l'entreposage de données textuelles et marquent un début prometteur dans l'analyse en ligne de données textuelles massives.

1.2. *Entrepôts de données évolutifs et personnalisés.* Un entrepôt de données extrait ses données à partir de sources de données, souvent hétérogènes, qui évoluent, et des besoins d'analyse des utilisateurs (décideurs), qui peuvent changer au fil du temps. En relâchant la contrainte de schéma fixe du modèle en étoile, nous avons défini un modèle d'entrepôt qui permet d'ajouter, de façon dynamique, un nouveau niveau de hiérarchie dans une dimension [Bentayeb-09a]. Ces travaux ont été étendus afin de permettre la recommandation aux utilisateurs finaux de nouvelles hiérarchies de dimension via la découverte par classification de nouvelles structures naturelles [Bentayeb-13]. D'autres travaux se sont focalisés sur la modélisation multidimensionnelle de données complexes, avec comme principale contribution la définition d'un modèle multidimensionnel d'objets complexes [Boukraa-13] basé sur le paradigme objet et dans lequel un objet peut jouer le rôle d'un fait ou d'une dimension. Un opérateur de projection cubique a été également proposé [Boukraa-11a]. Lorsque les données sources évoluent constamment, comme c'est le cas des données massives, nous proposons une analyse en ligne à la demande qui s'appuie sur un système de médiation et sur des ontologies du domaine [Maïz-10].

Par ailleurs, les technologies décisionnelles ont longtemps nécessité des investissements lourds, uniquement possibles dans les grands organismes et entreprises. C'est pourquoi nous avons travaillé à les rendre accessibles aux PME/TPE, en proposant dans un premier temps des solutions simples et efficaces (magasin de données en mémoire vive) basées sur des technologies ouvertes (notamment issues du Web avec un navigateur comme client léger) [Grabova-10], mais nécessitant encore une installation locale. Avec l'apparition de l'infonuagique (*cloud computing*) et du paiement de ressources (espace de stockage, temps de calcul) à la demande, il est devenu possible de démarrer un projet décisionnel à coût réduit. Nous proposons donc un système décisionnel en mode *Software as a Service* qui permet une prise en main simplifiée du processus décisionnel, en masquant les phases d'intégration de données et de conception d'un entrepôt et en permettant une navigation graphique dans les données étudiées [Grabova-14]. Nous travaillons également à l'intégration de données situationnelles, c'est-à-dire issues du Web, de portée limitée thématiquement et dans le temps, dans les processus décisionnels et plus particulièrement dans des structures dynamiques nommées *fusion cubes* [Abello-13]. Ces travaux s'inscrivent également dans la mouvance du décisionnel à la demande.

1.3. *OLAP sémantique.* Une des caractéristiques de l'analyse en ligne et qui constitue malheureusement l'une de ses limites, est de se restreindre à des aspects exploratoires et navigationnels. L'équipe s'est donc investie dans cette problématique et a proposé l'enrichissement de l'OLAP en le combinant avec des techniques de fouille et d'analyse de données, qui ont permis de définir de nouveaux opérateurs de structuration par la méthode des *k-means* [Bentayeb-09], de prédiction avec les arbres de régression, d'explication avec les règles d'association et de visualisation avec l'analyse formelle de concepts (AFC) [Loudcher-11a]. L'équipe a également développé des travaux sur l'analyse en ligne des données complexes structurées en format XML. Dans ce contexte, nous avons proposé une plateforme guidée par les événements pour intégrer des données du Web de façon autonome et en temps quasi-réel, qui s'appuie sur la technologie AXML et les services web [Salem-13]. Nous avons également travaillé à une analyse OLAP à partir d'entrepôts XML qui prennent en compte les problèmes spécifiques qu'elle pose, comme l'additivité, qui est un problème d'autant plus aigu que les données sont complexes [Hachicha-12].

Axe 2 : Analyse des réseaux d'informations. L'expansion des réseaux sociaux a drainé de très larges volumes de données sur le Web et induit de nouveaux enjeux. Par exemple, les marques prêtent une attention particulière à Facebook ou Twitter, où des millions d'utilisateurs échangent des informations. Il devient crucial de trouver des solutions adéquates permettant d'identifier, d'analyser et de valoriser cette information. Dans cet axe de recherche, nous avons travaillé à l'élaboration de nouveaux modèles multidimensionnels de données issues des réseaux sociaux. Nous avons notamment proposé un cube OLAP de *tweets* dont les dimensions sont les sujets et leurs domaines, extraits à partir des contenus textuels des *tweets* grâce à l'ontologie ODP (*Open Directory Project*), et dont la mesure est représentée par une distribution des sujets sur l'ensemble des *tweets* [Hannachi-13]. Les résultats obtenus sont encourageants et peuvent aussi trouver un terrain d'application dans la détermination du profil des utilisateurs, par exemple. Un autre type d'analyse auquel nous nous sommes intéressés est la détection des communautés dans les réseaux sociaux, en proposant une approche basée sur l'AFC et les treillis de gallois qui s'appuie sur une fonction de modularité adaptée qui améliore la qualité des communautés obtenues [Selmane-14].

Axe 3 : Big data analytics et sécurité

3.1. *Big data analytics.* Plusieurs travaux sont dédiés à cette thématique, dont l'équipe SID a fait une action prioritaire ces deux dernières années. Nous avons notamment proposé des algorithmes parallèles pour les

opérateurs OLAP [Arres-13a], ainsi que des modèles d'entrepôts de données NoSQL [Dehdouh-14a]. De plus, en l'absence d'opérateurs OLAP dans les modèles NoSQL, nous avons défini un opérateur d'agrégation nommé CN-CUBE (*Columnar NoSQL CUBE*) qui permet de calculer des cubes de données à partir d'entrepôts NoSQL orientés colonnes [Dehdouh-14b]. Nos expériences montrent que CN-CUBE est plus performant que l'opérateur CUBE relationnel. D'autres travaux ont permis d'étudier la performance des requêtes dans le nuage, qui doit optimiser simultanément les temps de réponse et le coût monétaire d'utilisation du nuage. Pour cela, nous avons proposé de nouveaux modèles de coût qui intègrent le paradigme de paiement à la demande en vigueur dans les nuages [Nguyen-12]. Sur la base de ces modèles, nous sélectionnons par optimisation des vues matérialisées minimisant le coût d'interrogation et de maintenance de la base de données, ainsi que le temps de réponse à une charge de requêtes donnée. Nos expériences montrent que la matérialisation de vues dans le nuage est toujours avantageuse.

3.2. Sécurité. Afin de garantir la sécurité des données stockées dans les entrepôts, dont l'exploitation fait de plus en plus appel à des services et des applications internet, nous avons travaillé à la fois sur la détection d'intrusions et sur la cryptographie. Sur le premier plan, nous avons notamment exploité des techniques de fouille de données, notamment la classification non supervisée floue [Nguyen-11a] et les méthodes d'ensemble appliquées à la fois sur les entêtes des paquets réseau et sur le contenu, qui permettent d'améliorer sensiblement la précision de détection des cyber-attaques [Nguyen-11b]. Par ailleurs, des résultats importants de cryptographie ont montré que tout calcul effectué sur des données distribuées peut se faire en garantissant leur confidentialité [Gavin-11]. Nous avons également proposé de nouveaux outils pour développer des cryptosystèmes complètement homomorphiques, qui sont des outils majeurs pour garantir la sécurité du calcul dans le nuage. Toutefois, les preuves de sécurité restent partielles et beaucoup de travail reste à faire pour l'obtention d'une preuve de sécurité complète.

Évolution de la production scientifique. Sur les derniers cinq ans et demi, l'équipe SID a pris en compte les remarques et conseils du comité d'évaluation de la période 2005-2008 pour améliorer sa production scientifique tant en termes quantitatif que qualitatif. La Figure 7 montre clairement l'évolution du nombre de publications de l'équipe dans des revues et conférences internationales, en augmentation de 65 %. La Figure 8 permet de constater l'amélioration de la qualité des publications de l'équipe, avec une augmentation de 56 % et 4 % du nombre de revues et de conférences classées (notamment de rang A et B), respectivement. La Figure 9 présente finalement les publications de l'équipe SID réalisées en collaboration avec d'autres chercheurs, notamment étrangers. Enfin la production scientifique de l'équipe inclut 2 HDR et 6 thèses de doctorat.

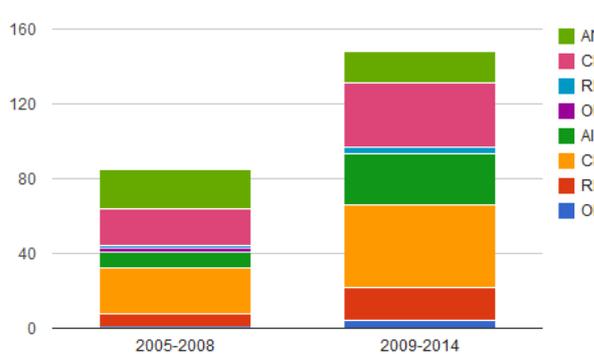


Figure 7 : Évolution de la production scientifique par type de publication (SID)

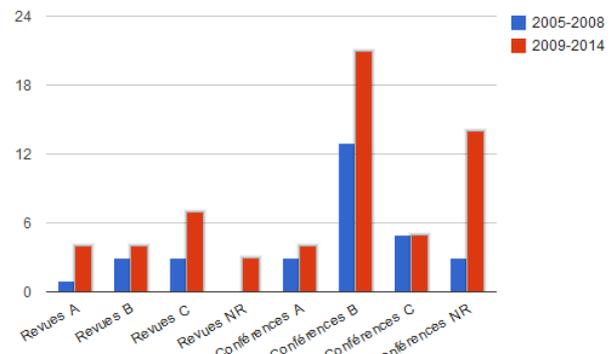


Figure 8 : Évolution de la production scientifique par rang de publication (SID)

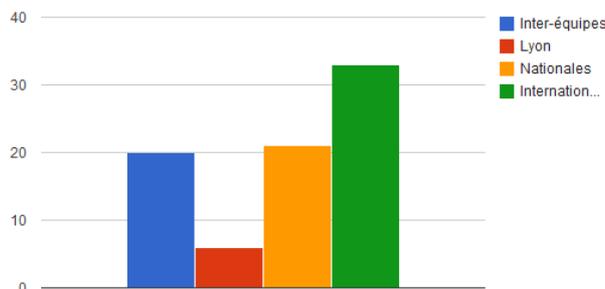


Figure 9 : Publications en collaboration (SID 2009-2014)

Rayonnement et attractivité académiques

Participation à des réseaux scientifiques. L'équipe SID est fondatrice de la conférence francophone EDA (Entrepôt de Données et Analyse en ligne) et cofondatrice de la conférence maghrébine ASD (Avancées des Systèmes Décisionnels), qui structurent la communauté francophone de notre domaine, ainsi que d'ateliers nationaux (Aide@EGC) et internationaux (Cloud-I@VLDB) dans lesquels elle est toujours fortement impliquée. Les membres de SID participent à des comités éditoriaux et de programme aux plans national et international (Annexe 6). Ils sont aussi régulièrement sollicités pour présider des sessions de conférences et ateliers auxquels ils participent, comme par exemple DAWAK, DEXA, DBKDA, EDA, ASD... Par ailleurs, plusieurs membres de SID sont invités pour des séminaires nationaux et internationaux : journée « Big data and cloud analytics », Université de Blida, Algérie (2014) ; école d'été « Les nouvelles technologies appliquées aux pratiques des chercheurs francophones en sciences humaines et sociales », Université de Bucarest, Roumanie (2013) ; séminaire Dagstuhl « Data Warehousing: from Occasional OLAP to Real-time Business Intelligence », Allemagne (2011)... Enfin, l'équipe SID entretient de bonnes relations de travail avec les principaux laboratoires reconnus pour leurs recherches en systèmes d'information d'aide à la décision en France et développe avec eux de nombreuses collaborations. Citons par exemple les laboratoires IRIT (Toulouse), LIRMM (Montpellier), LIMOS (Clermont-Ferrand), ETIS (Cergy-Pontoise), LIAS (Poitiers), LI (Blois) et l'IRSTEA de Clermont-Ferrand.

Implication dans des projets nationaux ou internationaux. SID entretient des collaborations avec de nombreux chercheurs en Europe et hors Europe. En Europe, les collaborations, sous la forme de copublications, d'organisation de manifestations scientifiques ou de jurys de thèses, même si certaines d'entre elles sont informelles, sont significatives. Hors Europe, nos accords de collaborations sont nombreux et prennent des formes différentes. Nous avons par exemple des relations bilatérales étroites avec un certain nombre de laboratoires, qui donnent lieu à des échanges de chercheurs et/ou des cotutelles de thèses, comme avec l'Algérie (Université d'Oran), la Tunisie (Institut Supérieur de Gestion de Tunis), le Canada (Laboratoire de Recherche sur l'Information Multimédia, Université du Québec en Outaouais) ou l'Ukraine (Université Nationale d'Économie de Kharkiv). Nous avons également un programme de coopération franco-algérien (TASSILI) avec l'Université Saad Dahleb (Blida, Algérie). Ces collaborations sont favorisées par les possibilités d'accueil de chercheurs sur des postes invités proposés par l'Université Lyon 2 (Rokia Missaoui, 2012 ; Ivan Sydorenko, 2013 et 2014).

Animation scientifique. L'équipe SID se réunit chaque vendredi après-midi, avec des réunions d'une durée de deux à trois heures. Plusieurs séminaires internes sont programmés dans l'année, à raison de deux séminaires par mois. L'équipe a mené des actions d'animation scientifique au plan local, comme l'organisation de la journée thématique « le décisionnel dans le nuage » en 2013 et au plan national, avec un fort investissement dans l'organisation du congrès INFORSID 2014.

Expertises. Les membres de l'équipe interviennent régulièrement dans les travaux d'expertise de dossiers de recherche auprès d'organismes comme l'ANRT (Association Nationale de la Recherche et de la Technologie), COFECUBE (Comité Français d'Évaluation de la Coopération Scientifique et Universitaire avec le Brésil), l'AUF (Agence Universitaire de la Francophonie), le CNU (Conseil National des Universités) section 27 (mandature 2012-2015)... Les membres de SID participent régulièrement à de nombreux jurys de thèses et de HDR, ainsi qu'à des comités de sélection extérieurs à Lyon pour le recrutement d'enseignants-chercheurs.

Contrats de recherche. Une part non négligeable de l'activité de recherche de l'équipe SID a été consacrée au montage et à la réalisation de projets/contrats de recherche collaboratifs (Annexe 7).

Interactions avec l'environnement social, économique et culturel

Contrats industriels. Les collaborations avec des industriels ou d'autres organismes publics ou privés définissent des contextes applicatifs importants pour l'équipe, afin de mettre en œuvre et tester les méthodologies, modèles et autres heuristiques développés dans l'équipe. L'équipe SID maintient un lien fort avec les entreprises, que ce soit pour des encadrements de thèses CIFRE (AID, AMI Software, CPCAM) ou pour intervenir en tant que conseil (CAF, Buzzinbees ; Annexe 7).

Synthèse

Les cinq dernières années ont permis à l'équipe SID d'obtenir les résultats suivants :

1. une consolidation de ses activités de recherche, attestée par une progression significative en nombre et en qualité de la diffusion scientifique et la soutenance de deux HDR et de six thèses ;

2. une visibilité accrue permettant à l'équipe de s'insérer dans les réseaux de recherche et de prendre une place significative dans l'animation scientifique nationale et internationale ;
3. une attractivité liée à un fort potentiel recherche de notre équipe et attestée par les nombreux doctorants étrangers boursiers d'excellence de leur pays (Algérie, Iran, Madagascar, Thaïlande, Vietnam), ainsi que par les doctorants financés sur contrats CIFRE.

2.3. Équipe DMD

Production scientifique

L'objectif de l'équipe DMD est de concevoir de nouveaux systèmes, modèles et algorithmes pour la fouille de données complexes et l'aide à la décision. Les données complexes sont des données structurées (par exemple sous forme de graphes), hétérogènes (descriptions attributs-valeurs, textes, images, etc.), dynamiques, imprécises, volumineuses. Pour manipuler ces données, l'équipe s'appuie sur des approches principalement statistiques (analyse des données, statistique inférentielle) et inspirées de l'intelligence artificielle (apprentissage automatique, représentations floues, raisonnement dans l'incertain, etc.).

L'activité de l'équipe sur la période 2009-2014 peut être découpée en quatre axes thématiques, même si les chercheurs travaillant dans ces axes sont souvent amenés à collaborer :

1. apprentissage automatique pour la fouille de données ;
2. modélisation, caractérisation, fouille dans les graphes ;
3. modèles d'aide à la décision multicritère ;
4. analyse des données complexes et fouille d'opinions.

Les principales contributions sont détaillées ci-dessous selon ces quatre axes thématiques.

Axe 1 : Apprentissage automatique pour la fouille de données. Une partie des travaux de l'équipe a porté sur le développement de nouvelles techniques d'apprentissage automatique (*machine learning*). L'une des contributions consiste à construire des combinaisons de classifieurs (ensembles) afin d'améliorer les performances des algorithmes. L'équipe a réalisé des contributions théoriques sur des ensembles composés de forêts aléatoires [Pisetta-10]. Dans une autre optique, nous nous sommes aussi intéressés à l'extraction de règles d'association de classe et à leur évaluation à l'aide de mesures d'intérêt. En collaboration avec Telecom Bretagne, nous avons identifié et généralisé un certain nombre de propriétés d'antimonotonie et nous avons ensuite établi des conditions nécessaires et/ou suffisantes pour qu'une mesure d'intérêt possède ces propriétés [LeBras-12a]. L'équipe a également étudié le problème de la classification non supervisée dans son ensemble : critères de classification, mesures de similarités [AhPine-13b], modélisation sous forme de problèmes d'optimisation [AhPine-10b] et algorithmes qui passent à l'échelle [AhPine-09a]. D'autres travaux concernent les modèles d'apprentissage non supervisé dédiés à l'analyse de données textuelles. En particulier, l'équipe a proposé des algorithmes d'étiquetage des catégories thématiques et d'évaluation de la qualité des thématiques obtenues. Réalisé en collaboration avec des chercheurs de l'Ecole Polytechnique de Bucarest, l'algorithme d'évaluation permet d'émuler le jugement humain en se basant sur une recherche de correspondance (*mapping*) entre les thématiques et une base de connaissance lexicale [Musat-11]. Notons enfin des travaux sur l'apprentissage automatique de la structure de réseaux bayésiens lorsque le nombre de variable est très grand, dont l'idée consiste à réaliser l'apprentissage sur des sous-ensembles de variables de taille raisonnable, puis de recombinaison les différents résultats [Thibault-09]. Pour finir, des méthodes numériques basées sur des algorithmes de type réseaux de neurones ont été appliquées pour la modélisation économique et financière [Deperetti-09].

Axe 2 : Modélisation, caractérisation, fouille dans les graphes. Dans ce deuxième axe, les chercheurs de l'équipe ont travaillé sur l'analyse de graphes, avec des contributions sur les voisinages dans les graphes, sur la recherche d'information dans ces graphes et sur la caractérisation de familles de graphes. En particulier, l'équipe a développé une approche d'apprentissage qui utilise des graphes de voisinage pour la comparaison, le regroupement et l'équivalence topologique de mesures de proximité [Zighed-12]. D'autres travaux ont permis d'étudier les propriétés structurelles des graphes. Ainsi, l'équipe a étudié la décomposition de graphes (graphes orientés et 2-structures) en 2-clans, qui sont des modules (ou clans) à deux éléments, ainsi que leur caractérisation par l'utilisation d'homotopies [Jouve-09]. Des travaux récents ont également été menés sur la recherche d'information au sein de graphes, dans une approche de fouille du Web (*web mining*). L'approche

consiste à prendre en compte à la fois la structure du graphe (degrés, composantes connexes, etc.) et le contenu textuel des nœuds. Nous avons montré qu'il était possible d'extraire le réseau social sous-jacent à des discussions en ligne [Forestier-11a], d'en extraire les messages les plus intéressants [Stavrianou-09a] et les acteurs jouant des rôles clefs comme celui de célébrité [Forestier-12]. Un algorithme d'extraction des communautés, basé sur une approche prétopologique et réalisé en collaboration avec le laboratoire Hubert Curien de Saint-Étienne, donne des résultats probants sur des données bibliographiques [Muhlenbach-09]. D'autres approches pour la détection de communautés qui utilisent des mesures d'association développées en statistique ont également été proposées et donnent des résultats encourageants dans le cadre de graphes aléatoires [AhPine-13b].

Axe 3 : Modèles pour l'aide à la décision. Les membres de l'équipe qui travaillent dans cet axe de recherche se sont intéressés aux méthodes et aux modèles d'aide à la prise de décision collective, multicritère et multiobjectif. Des recherches théoriques ont été menées pour étudier les propriétés de nouvelles méthodes dans le cadre de l'analyse multicritère : propriétés des intégrales de Sugeno [Prade-09] ou de Choquet [Lust-13], modèles à base de points de référence [Rolland-13a], modèles à base de bicapacités [Mayag-12, AhPine-13]. D'autres travaux ont été plus particulièrement axés sur l'interprétation de l'intégrale de Sugeno en terme de bases possibilistes [Dubois-14], sur la théorie de l'évidence et des possibilités [Dubois-11], mais aussi sur l'étude des liens et interactions possibles entre l'aide à la décision multicritères et les autres champs de la théorie des processus décisionnels et de l'apprentissage non supervisé [AhPine-10]. On citera aussi des travaux sur l'agrégation de relations binaires par l'utilisation de la programmation linéaire en nombres entiers [AhPine-10b]. Des applications ont été développées en collaboration avec d'autres partenaires : utilisation de l'intégrale de Choquet pour la reconnaissance d'image en collaboration avec le LIRMM [Strauss-09a], aide à la sélection de fruits en collaboration avec l'INRA d'Avignon [Ould-Sidi, à paraître], caractérisation de classements de pays [Rolland-13b], adaptation de méthodes d'analyse multicritère en santé (projet FLURESP). Du point de vue de la prise de décision dans le cadre d'une optimisation multiobjectif, les travaux développés proposent un nouvel algorithme évolutionnaire hybride HEMH combinant différentes techniques (DM-GRASP, *Path-Relinking* et recherche locale) pour résoudre le problème du sac à dos multiobjectif [Kafafy-12a].

Axe 4 : Analyse des données complexes et fouille dans les médias sociaux. Dans cet axe, l'équipe a privilégié l'élaboration de nouvelles techniques d'analyse des données et leur utilisation sur des données complexes : données temporelles/évolutives, textuelles, hétérogènes (variables quantitatives et qualitatives). De nouvelles méthodes ont été proposées : analyse des associations dissymétriques dans le cadre de l'analyse des correspondances et de l'analyse de la variance, traitement des données mixtes (quantitatives et qualitatives) dans le contexte de l'analyse en composantes principales et enfin, des techniques de classement et de prédiction par l'analyse discriminante sur données évolutives, sur données mixtes ou encore à plusieurs variables cibles-groupes [Abdesslam-10]. À partir d'entités décrites dans le temps (pays, hommes politiques, internautes), l'équipe a proposé un nouvel algorithme de *clustering* semi supervisé capable de prendre en compte la dimension temporelle et de construire le graphe des évolutions temporelles des catégories [Rizoïu-14]. L'équipe s'est également attaquée au cas des données textuelles complexes issues des nouveaux médias sociaux (blogs, forums de discussion, réseaux sociaux) afin d'en capturer les thématiques et les événements (de « quoi » on parle) mais également les opinions (« comment » on en parle). Ces informations sont produites par des acteurs eux-mêmes inscrits dans un réseau de relations qui, de plus, évolue au fil du temps. Ces données posent de nombreux problèmes d'analyse descriptive, de modélisation, de visualisation, de validation. Sur ces thèmes, l'équipe a notamment proposé un nouveau modèle de diffusion de l'information sur les médias sociaux [Guille-13a], ainsi qu'un nouvel outil de visualisation pour le suivi de données textuelles évolutives en collaboration avec l'IRISA (Rennes) et l'Université de Zagreb [Petrovic-09]. Dans le cadre du projet ImagiWeb, une plateforme d'annotation a été mise en place afin de pouvoir suivre l'évolution de l'opinion des internautes au sujet de l'image de marque d'entités [Velcin-14]. L'étape suivante du projet consiste à élaborer de nouveaux modèles et algorithmes pour l'analyse automatique de ces données d'opinion.

Évolution de la production scientifique. En termes quantitatifs, la Figure 10 montre clairement que le volume de publications internationales a augmenté de manière significative (28 %) depuis la dernière évaluation AERES. Cette évolution s'explique en partie par l'intégration d'anciens membres de l'équipe MA²D du LIRIS et en partie par les recrutements de plusieurs collègues enseignants-chercheurs ces dernières années. En termes qualitatifs, la Figure 11 permet de constater un saut positif très visible sur les revues et les conférences classées (+89 % et +73 %, respectivement), notamment dans les supports de publication classés A, et qui ne résulte pas uniquement de cet effet lié au volume des publications. Cette évolution résulte essentiellement dans une politique affirmée de publier les résultats des travaux de l'équipe dans les revues et conférences internationales reconnues et sélectives dans notre domaine de recherche. Enfin, la Figure 12 permet de

constater un nombre important de publications résultant de collaborations avec d'autres équipes à tous les niveaux de granularité, mais plus particulièrement au niveau national. Ces collaborations participent au rayonnement scientifique de l'équipe que nous décrivons plus en détail ci-dessous.

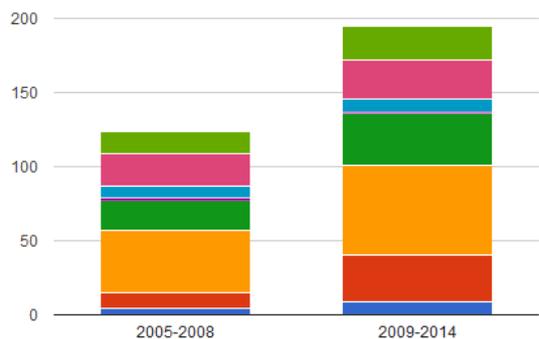


Figure 10 : Évolution de la production scientifique par type de publication (DMD)

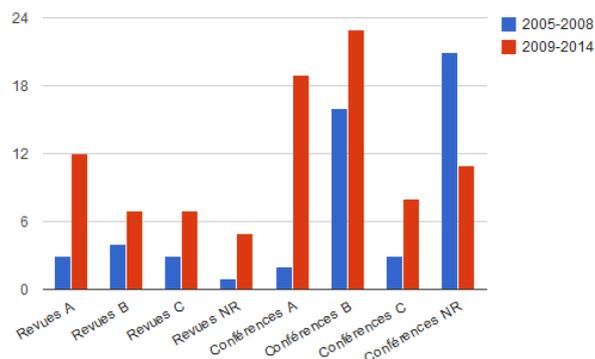


Figure 11 : Évolution de la production scientifique par rang de publication (DMD)

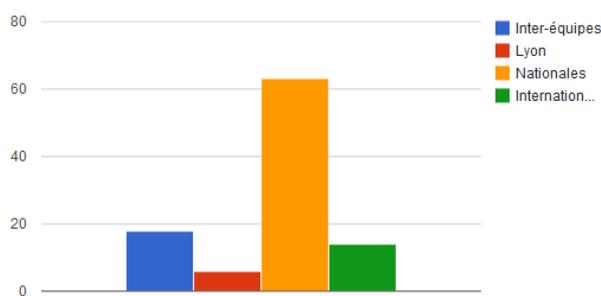


Figure 12 : Publications en collaboration (DMD 2009-2014)

Rayonnement et attractivité académiques

L'équipe DMD est clairement positionnée sur les domaines de la fouille des données et de l'aide à la décision, domaines sur lesquels elle a acquis une solide reconnaissance. Avec plusieurs laboratoires d'informatique (LINA, LRI, LIRMM, LIRIS, TELECOM-ParisTech), elle fait partie des membres fondateurs de l'association Extraction et Gestion des Connaissances (EGC), qui organise régulièrement la conférence du même nom. Les membres de l'équipe continuent à animer la communauté francophone de fouille de données par le biais de l'association. L'un des membres de l'équipe est le codirecteur de la Revue des Nouvelles Technologies de l'Information (RNTI). Dans le domaine de la décision, l'équipe DMD a également animé le groupe SCDD (Systèmes Complexes et Décision Distribuée) associé à la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF) et le GdR MACS (modélisation, analyse et conduite des systèmes dynamiques). Elle a fondé et coanimé l'association PretopologiCS qui a pour but la promotion, la valorisation et la diffusion de la recherche en prétopologie et modélisation des systèmes complexes. Au niveau international, les membres de l'équipe animent des événements scientifiques internationaux (pilotage des workshops QIMIE@PAKDD et MSND@WWW 2012, organisation des conférences ALT et DS en 2012). Ils participent bien sûr à l'évaluation de la recherche par leur implication dans des comités éditoriaux de journaux, des comités de programme de conférences, des jurys de thèses et d'HDR. En matière d'expertises, les membres de l'équipe DMD sont régulièrement sollicités pour évaluer des dossiers de projets de recherche (par exemple, pour l'ANR en France ou pour le NSERC au Canada) et des projets de thèse CIFRE pour l'ANRT. Les membres de l'équipe sont très impliqués dans des projets nationaux (pilotage du projet ANR ImagiWeb, participation au projet d'investissements d'avenir REQUEST) et internationaux (pilotage du projet européen FLURESP). Durant la période concernée, l'équipe a accueilli de nombreux enseignants-chercheurs étrangers (T. Aluja, J. Antoch, M. Becue, L. Saitta, S. Trausan-Matu, I. Zolotaryova) et a recruté deux professeurs et quatre maîtres de conférence (dont un a obtenu sa qualification et obtenu un poste de PR dans un laboratoire voisin). On notera enfin de nombreuses collaborations scientifiques nationales, académiques (D. Dubois à l'IRIT, P. Lenca de Telecom Bretagne, M. El-Bèze du LIA...) et industrielles (EDF, Technicolor, Thales, Xerox...).

Interactions avec l'environnement social, économique et culturel

Depuis mars 2012 sont organisés à Lyon des « cafés de la statistique » sur le modèle de ce qui existe depuis dix ans à Paris. Il s'agit, quatre fois par an, de montrer au grand public comment la statistique peut éclairer le

débat public. Ces cafés sont coorganisés par un membre de l'équipe DMD et une professeure de l'Institut Camille Jordan (UMR 5208).

L'équipe a participé à l'événement BLEND Web Mix 2013, organisé conjointement par la Cuisine du Web et l'Université de Lyon et soutenu par le Grand Lyon. Il s'agissait de deux jours de conférences, ateliers et rencontres qui s'adressaient à des publics variés faisant partie de l'écosystème web de la région lyonnaise. C'est dans ce cadre que des membres de l'équipe ont présenté des projets de recherche à destination du grand public dans une optique de diffusion de la culture scientifique.

Synthèse

Les cinq dernières années ont permis à l'équipe DMD d'obtenir les résultats suivants :

1. une amélioration significative de la qualité des publications réalisées, en particulier dans les meilleures revues et conférences au niveau international ;
2. une réelle implication dans des projets de recherche d'envergure, avec le pilotage d'un projet ANR et d'un projet européen, mais aussi la participation récente à un Projet Investissement d'Avenir ;
3. l'organisation d'événements scientifiques au niveau international (DS-ALT en 2012, ateliers adossés à des conférences reconnues comme WWW ou PAKDD) ;
4. des collaborations nouvelles avec des acteurs reconnus de la recherche, que ce soit dans le milieu académique (D. Dubois et H. Prade de l'IRIT, M. El-bèze au LIA) ou industriel (EDF, Technicolor, Thales, Xerox).

3. Implication de l'unité ou de l'équipe dans la formation par la recherche

Le laboratoire ERIC est rattaché à l'école doctorale Informatique et Mathématiques (InfoMaths, ED 512), qui regroupe une grande partie des laboratoires et des chercheurs en informatique et mathématiques à Lyon (Université Claude Bernard Lyon 1, Université Lumière Lyon 2, INSA Lyon, ENS Lyon, École Centrale de Lyon et, depuis peu, Université Jean Moulin Lyon 3). Le directeur du laboratoire est membre du conseil de l'école doctorale et de son comité des thèses, ainsi que représentant d'InfoMaths à Lyon 2.

ERIC est majoritairement alimenté en doctorants par le biais de son réseau international, grâce à une bonne attractivité de boursiers d'excellence financés par leur gouvernement et à de nombreuses cotutelles (Figure 13). Toutefois, malgré un environnement très concurrentiel, ERIC est aussi un membre actif l'école doctorale InfoMaths et dispose de divers contacts industriels, ce qui permet de fournir un débouché aux étudiants de masters qui souhaitent poursuivre en thèse avec un contrat doctoral (ministériel, régional ou lié à un projet) ou un financement CIFRE. La liste des thèses en cours est fournie en Annexe 10.

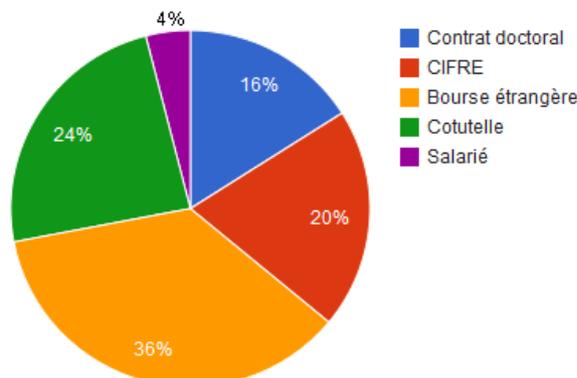


Figure 13 : Mode de financement des doctorants en cours de thèse

Le cas échéant, le classement des projets doctoraux soumis à l'école doctorale est discuté et voté en conseil de laboratoire, dans sa formation restreinte aux enseignants-chercheurs. Enfin, le conseil de direction d'ERIC, dans sa formation restreinte aux enseignants-chercheurs, fait également office de comité des thèses. Il valide les premières inscriptions en thèse et les jurys de soutenance avant transmission à l'école doctorale, et

organise une soutenance à mi-parcours (rapport et présentation orale évalués par deux collègues extérieurs au laboratoire) pour chaque doctorant, en liaison avec le(s) directeur(s) de thèse.

17 thèses et 2 habilitations à diriger des recherches ont été soutenues à ERIC sur la période 2009-2014. De plus, 9 membres d'ERIC ont demandé et obtenu leur qualification (4 PR, 5 MCF). La durée moyenne des thèses est d'environ quatre ans. Elle s'explique par les nombreuses cotutelles gérées par le laboratoire, dont la durée est typiquement supérieure à trois ans, ainsi qu'aux boursiers étrangers, dont certains sont d'emblée financés sur quatre ans (par exemple, les bourses du gouvernement égyptien) afin de permettre l'apprentissage du français et l'adaptation des doctorants. Le taux d'abandon est faible, avec seulement 1 abandon pendant la période écoulée. Pour finir, nous effectuons un suivi systématique du devenir de nos docteurs, qui est indiqué en Annexe 10 avec la liste des thèses soutenues au laboratoire.

Par ailleurs, les membres du laboratoire ERIC sont tous des enseignants-chercheurs relevant des sections 26 (Mathématiques appliquées) et 27 (Informatique) du CNU. Ils sont très impliqués dans la vie des universités Lyon 1 et Lyon 2 et dans leurs enseignements d'informatique (Annexe 12). Ils s'adressent à la fois à des publics de spécialistes (Départements d'Informatique et de Mathématique de la Faculté des Sciences et Technologies et école d'ingénieurs Polytech' Lyon, Département Informatique et Statistique de l'Institut de la Communication et IUT Lumière à Lyon 2) et à des non-spécialistes (Facultés de Sciences Économiques et de Gestion, de Sociologie et de Droit, Institut de la Communication et IUT Lumière à Lyon 2). Compte-tenu du sous-encadrement en informatique à l'Université Lyon 2, la majorité des membres d'ERIC effectue plus que son service statutaire minimum.

Les formations animées par des membres d'ERIC et susceptibles d'offrir la possibilité d'une thèse au laboratoire dépendent essentiellement de la spécialité Fouille de Données et Gestion des Connaissances (FDGC) du master d'Informatique de Lyon 2 :

- parcours Extraction des Connaissances à partir des Données (ECD) en partenariat avec Polytech' Nantes (une vingtaine d'étudiants en M2) ;
- parcours Erasmus Mundus *Data Mining & Knowledge Management* (DMKM) en partenariat avec Polytech' Nantes, l'Université Paris 6, l'Université Polytechnique de Catalogne (Barcelone, Espagne) l'Université du Piémont Oriental (Alessandria, Italie) et l'Université Polytechnique de Bucarest (Roumanie) (une vingtaine d'étudiants en M2).

De plus, les étudiants des formations à vocation générale ou professionnelle dans lesquelles sont impliqués des membres d'ERIC participent également à la recherche au laboratoire, par des projets ou des stages :

- spécialité Informatique Décisionnelle et Statistique (IDS) du master d'Informatique de Lyon 2 (entre 130 et 150 étudiants en M2, formation continue incluse) ;
- spécialité e-MIAGE du master d'Informatique de Lyon 1 ;
- licences Informatique Décisionnelle et Statistique (IDS) et Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales (MIASHS) de Lyon 2 ;
- Licence pro Chargé(e) d'Etudes Statistiques (CESTAT) de l'IUT Lumière.

Les membres d'ERIC sont également très impliqués dans des formations internationales : doubles diplômes de licence et de master en informatique à l'université de Ho Chi Minh (Saigon, Vietnam) à Lyon 1, master Erasmus Mundus DMKM et double-diplôme de master franco-ukrainien Informatique Décisionnelle et Statistique pour le Management (IDSM-Kharkiv) à Lyon 2 ; ce qui contribue à l'attractivité du laboratoire. Un membre d'ERIC et notamment représentant de l'Université Lyon 1 au consortium de l'Agence Universitaire de la Francophonie.

Enfin, les membres d'ERIC s'impliquent régulièrement dans le pilotage et l'organisation d'écoles pour jeunes chercheurs, comme les écoles d'été Web Intelligence (ARC 6 région Rhône-Alpes) 2010 et 2013 ou la WI-IAT 2011 *summer school*.

4. Stratégie et perspectives scientifiques pour le futur contrat

2.1. Projet global

Objectifs scientifiques. L'objectif du laboratoire est de conforter son positionnement *leader* dans le domaine de l'informatique décisionnelle (modélisation, exploitation d'entrepôts et fouille de données massives, processus d'aide à la décision) au plan national, ainsi que de renforcer sa visibilité internationale.

Orientations scientifiques et choix stratégiques. Le laboratoire va poursuivre sa politique de collaboration scientifique avec les SHS, qui lui permet d'occuper un positionnement original au sein de l'informatique lyonnaise, par exemple avec des projets pluridisciplinaires comme l'*Infonomics Resources Facility*⁶, qui pose un cadre de recherche scientifique et méthodologique, ainsi qu'un environnement technologique visant à développer des recherches sur la production, la diffusion et l'utilisation de données, d'informations, de connaissances et de savoir-faire. ERIC est également partie prenante, avec les UMR Environnement Ville et Société et Centre Max Weber, ainsi que le soutien du Labex Intelligence des Mondes Urbains, d'un projet de financement CPER déposé auprès de la Région Rhône-Alpes pour renforcer la plateforme PANELS de l'ISH, qui vise à exploiter des corpus massifs. De plus, les recherches en informatique et en mathématiques appliquées menées à ERIC s'inscrivent naturellement dans les axes de recherche prioritaires 2016-2020 de l'ISH « numérique » et « complexité et décision ». Les synergies avec les laboratoires de SHS de l'établissement permettent également de développer des collaborations autour des humanités numériques, autre axe de recherche commun à l'Université Lyon 2 et à l'ISH. Enfin, ERIC vise une reconnaissance par le CNRS, une piste possible étant, en lien avec des partenaires de la COMUE Lyon-Saint-Étienne, de développer un pôle pluridisciplinaire autour des humanités numériques.

Moyens à mobiliser pour atteindre les objectifs. Afin de renforcer la visibilité internationale du laboratoire, il s'agit de continuer à consolider le niveau de qualité de la production scientifique, qui doit encore être améliorée, notamment en termes de publications dans des journaux internationaux de rang A, ainsi que d'affirmer la présence du laboratoire dans les comités éditoriaux et les manifestations scientifiques internationaux. En plus du dépôt de projets nationaux, la participation à des projets internationaux, européens ou autres, fait également partie de la stratégie du laboratoire. Par ailleurs, la complémentarité continuera d'être favorisée au sein du laboratoire, à la fois entre équipes de recherche et entre sections CNU (26^e et 27^e), pour porter des projets, coencadrer des thèses, etc. Enfin, la collaboration initiée pour la mise en place d'un master humanités numériques à l'échelle de la COMUE Lyon-Saint-Étienne permettra de renforcer les liens avec les équipes de SHS et de mettre en œuvre une synergie enseignement-recherche pluridisciplinaire et innovante dans ce domaine. Compte-tenu de la charge de travail actuelle des membres permanents d'ERIC, tant sur le front de la formation que de celui de la recherche, il sera impératif pour mener à bien toutes ces actions de recruter, hors renouvellements de postes, des enseignants-chercheurs. L'accent ayant été mis sur le renforcement de la 26^e section pendant la période écoulée, nous allons demander au cours du prochain contrat quinquennal la création ou le redéploiement de 4 postes (2 PR et 2 MCF) relevant majoritairement de la 27^e section, sur des profils centrés sur les humanités numériques et l'exploitation des **-data (open data, big data, linked data...)*. Le recrutement d'un ingénieur de recherche susceptible d'assurer de manière pérenne le suivi du développement de prototypes logiciels nous semble également indispensable. Finalement, la pratique dans nos disciplines étant d'effectuer la recherche au laboratoire et les locaux actuels d'ERIC sur le campus Porte des Alpes étant saturés, nous aurons besoin de nouveaux espaces (essentiellement des bureaux) pour accueillir ces nouveaux membres, ainsi que des doctorants.

Renouvellement des partenariats. Sur le plan local, l'appartenance d'ERIC à l'ISH va permettre une poursuite aisée des collaborations avec les laboratoires de SHS de la COMUE. Par ailleurs, le LIRIS, l'équipe Systèmes d'information du Centre Magellan de Lyon 3 et ERIC (noyau susceptible de s'élargir à d'autres laboratoires d'informatique) ont décidé d'actions communes destinées à permettre l'émergence de collaborations inter-établissements (rencontres et échanges scientifiques sur certaines thématiques, cofinancement de stages de master recherche...). ERIC poursuivra également son implication au sein de l'ARC 6 « T.I.C. et usages informatiques innovants » de la Région Rhône-Alpes. Au niveau national, les réseaux centrés sur les conférences EGC et EDA sont bien structurés, mais ERIC participera également aux efforts « transthématiques »

⁶ <http://infonomics.ish-lyon.cnrs.fr/irf/>

mis en œuvre par le GdR I3 (information, interaction, intelligence). Enfin, au niveau international, il s'agit de capitaliser sur les collaborations existantes (consortium du master Erasmus Mundus DMKM, entre autres) pour étendre le réseau du laboratoire.

Nouvelles thématiques scientifiques. Fort de son expérience dans la gestion et l'analyse de données dites complexes, ERIC est idéalement placé pour contribuer aux travaux scientifiques visant la valorisation des données massives, caractérisées par trois ou quatre « V », selon les sources : volume, variété, vélocité et véracité. Le projet REQUEST (Programme d'investissements d'avenir) a d'ailleurs récemment débuté sur le thème *big data analytics*. Par ailleurs, la conception de nouvelles méthodes d'aide à la décision pour l'analyse des médias sociaux nous semble d'une importance stratégique majeure. Ces méthodes pourront être mises en œuvre avec l'aide de spécialistes en sciences de l'information et de la communication, de sociologie, etc. et être appliqués à d'autres domaines des SHS.

La liste des enseignants-chercheurs permanents du laboratoire est fournie en Annexe 9. Pour terminer, nous synthétisons dans la Figure 14 sous forme de matrice SWOT les atouts et les handicaps liés à la stratégie que nous venons de présenter.

| | |
|---|---|
| FORCES | FAIBLESSES |
| Thématiques de recherche sur laquelle ERIC bénéficie d'une reconnaissance nationale et internationale | Bilocalisation du laboratoire |
| Positionnement au sein d'un secteur d'application (SHS) en demande d'expertise dans notre domaine | Moyens matériels (locaux à Lyon 1 et à Lyon 2) et humains (personnel administratif et technique) insuffisants, comme déjà souligné par l'AERES à l'issue du précédent contrat |
| OPPORTUNITÉS | MENACES |
| Forte demande de valorisation des données massives | Pas d'affectation de nouveaux moyens matériels et humains par les établissements de tutelle |
| Afflux d'étudiants dans nos formations de master, notamment internationales | Augmentation des charges administratives des enseignants-chercheurs |
| Développement de collaborations plus étroites avec les laboratoires lyonnais et rhônalpins aux thématiques voisines | |

Figure 14 : Matrice SWOT du projet d'ERIC

4.2. Équipe SID

Objectifs scientifiques. L'équipe SID, sur la base d'une expertise développée depuis près de quinze ans dans le domaine de l'entreposage et de l'analyse en ligne (OLAP) des données complexes, a pour objectif de développer des méthodes innovantes pour une analyse en ligne de qualité, sécurisée et performante des données massives. Cet objectif s'inscrit dans la continuité des travaux développés dans l'équipe depuis deux ans autour de l'informatique décisionnelle dans le nuage (*cloud analytics*), ce qui lui permettra de renforcer sa visibilité nationale et internationale dans ce domaine.

Orientations scientifiques et choix stratégiques. L'objectif scientifique principal de l'équipe est d'aborder différents aspects du *big data analytics* de manière cohérente et globale dans toutes les phases du processus décisionnel : modélisation, intégration et analyse en ligne de données massives. Les quatre « V » soulèvent ou accentuent en effet des verrous scientifiques nombreux. Comment prendre en compte la variété et de l'évolutivité des données, ainsi que l'évolution potentiellement rapide des besoins des utilisateurs, lors des phases d'intégration des données et d'analyse ? Comment gérer la vélocité et le volume des données lors des phases de stockage (stratégies de traitement à la volée vs. stratégies de stockage, d'agrégation ou d'oubli) et de calcul (parallélisation de tâches d'intégration et d'analyse) ? Comment contrôler la véracité (qualité, intégrité) des données en entrée pour une aide à la décision correctement informée ? Par ailleurs, il est nécessaire et stratégique pour l'équipe SID, afin de valider ses propositions de manière expérimentale, de

constituer des corpora de données, de rechercher des terrains d'applications et de développer des collaborations locales, notamment grâce à son encrage SHS à Lyon 2, et nationales grâce à ses contacts industriels (entreprise de e-marketing LiveBanner, par exemple).

Moyens pour atteindre les objectifs. L'équipe SID affiche actuellement une forte attractivité internationale, notamment en termes de flux de doctorants. Elle compte poursuivre le développement de ces échanges internationaux pour recruter de nouveaux doctorants. Par ailleurs, les partenariats industriels (thèses CIFRE) seront aussi recherchés de manière proactive. Enfin, les efforts de l'équipe pour cibler des supports de publications de qualité (notamment des revues et conférences internationales de rang A) seront reconduits pour améliorer sa visibilité internationale. Dans cette optique, nous poursuivrons et amplifierons si nécessaire le recours à des services de traduction/relecture en anglais.

Renouvellement des partenariats. L'équipe SID a un positionnement original dans l'environnement lyonnais, de par ses travaux de recherche en informatique décisionnelle. Toutefois, des proximités thématiques existent avec d'autres équipes de Lyon (LIRIS et Centre Magellan), Saint-Etienne et Grenoble, qui font partie d'un groupe de travail traitant de l'infonuagique⁷. L'engagement de SID dans son projet scientifique *big data analytics* va lui permettre de renforcer les liens avec ces équipes de recherche et d'initier de nouvelles collaborations avec les entreprises. Cela se traduit notamment par la création en projet d'un sous-groupe de travail sur la gestion et l'analyse des données massives et le partage de ressources matérielles et logicielles pour la validation et l'étude de performance. Au plan national, le réseau des journées EDA permet de développer des projets de type ANR en jouant sur la complémentarité des équipes de recherche du domaine des entrepôts de données. Enfin, les partenariats internationaux de l'équipe seront également confortés (réseau de la conférence ASD, contacts avec le Brésil...), que ce soit en termes d'animation scientifique ou de projets de recherche.

Nouvelles thématiques scientifiques. Grâce à son expertise dans l'entreposage et l'analyse en ligne de données complexes, l'équipe SID inscrit ses nouvelles thématiques scientifiques dans une continuité naturelle pour l'analyse en ligne et la valorisation des données massives. Il s'agit de tenir compte des évolutions technologiques actuelles pour permettre l'élaboration de nouveaux modèles d'entrepôts utilisant les modèles NoSQL, tout en permettant un passage à l'échelle aisé en déployant une architecture décisionnelle dans le nuage. L'équipe s'intéresse ainsi à la BI à la demande, qui vise à fournir des outils d'analyse en ligne à des petites structures (TPE, associations, simples citoyens), leur permettant ainsi de s'approprier les outils décisionnels habituellement dévolus aux grosses structures. En association avec ces objectifs, l'équipe prévoit de travailler à l'utilisation d'agents intelligents et de services web pour l'acquisition, l'intégration, l'analyse en ligne et la sécurité des données dans un environnement collaboratif dans le nuage. Plus précisément, l'équipe prévoit de développer trois axes de recherche qui s'inscrivent dans le cadre décrit ci-avant.

Axe A : Analyse en ligne de données massives. Les données massives sont souvent présentées comme un virage comparable à celui du Web en son temps. Il y a de fait actuellement une forte demande sociale pour la valorisation des données massives, mais les défis scientifiques à relever restent nombreux. Il s'agit pour l'équipe de définir des modèles multidimensionnels permettant le stockage et le traitement des données massives dans le nuage. Les modèles de données NoSQL s'avèrent adaptés au stockage, mais il faut alors créer/définir des opérateurs OLAP pour pouvoir proposer une stratégie d'analyse en ligne à la demande.

Axe B : Analyse en ligne collaborative. À l'image du Web, qui est un lieu d'information et d'échanges qualifié de social, participatif et collaboratif, les applications décisionnelles se doivent de fournir une analyse en ligne pouvant être partagée par plusieurs utilisateurs selon leur profil (OLAP collaboratif). Ces derniers ont alors la possibilité d'annoter, de donner leur avis à propos des analyses partagées. Dans ce contexte, nous nous intéressons particulièrement à plusieurs verrous scientifiques qui doivent être levés dans le but de faciliter le processus de personnalisation et de recommandation d'analyses en ligne collaborative : 1) comment enrichir les modèles d'entrepôts avec les profils des utilisateurs, les annotations, etc. ; 2) comment intégrer ces informations dans un processus d'analyse en ligne ; 3) comment évaluer la qualité des analyses personnalisées et comment les sécuriser ?

Axe C : Entrepôts de données ouvertes. Collectivités et entreprises mettent aujourd'hui à disposition du public de plus en plus de données. L'équipe souhaite développer une activité de recherche dans ce contexte, qui nous apparaît socialement important. Notre objectif est de construire des entrepôts de données

⁷ <http://liris.cnrs.fr/cloud/wiki/doku.php>

susceptibles d'être enrichis par des informations et des connaissances issues d'autres applications, ou plus largement provenant du Web. En effet, de nombreuses applications (de santé publique, par exemple) nécessitent, en plus des données issues de sources locales ou de sources externes, des informations complémentaires qui peuvent provenir du Web. Pour reprendre l'exemple de la santé, l'objectif pourrait être de prévenir une pandémie au lieu de la constater *a posteriori*. Pour atteindre ce but, différentes stratégies, qui induisent autant de verrous scientifiques, peuvent être envisagées : 1) intégrer dans le processus d'entreposage une ontologie du domaine étudié afin d'alimenter l'entrepôt avec des informations pertinentes (instances et relations entre instances dans l'ontologie) ; 2) alimenter l'entrepôt avec des données de simulation afin d'affiner des résultats d'analyse ; 3) de manière un peu plus orientée vers les données du Web (réseaux sociaux, blogs, etc.), combiner les technologies du Web sémantique combinée avec l'OLAP.

4.3. Équipe DMD

Objectifs scientifiques. La fouille des données et son utilisation pour l'aide à la décision constituent le cœur des préoccupations des chercheurs de l'équipe DMD. Dans ce contexte, l'objectif de l'équipe est de concevoir des méthodes originales pour réaliser des analyses fines sur les données complexes (relationnelles, temporelles, textuelles, volumineuses), tout en veillant à mener une réflexion sur la qualité des modèles (parcimonieux, réflexifs, contraints) et des algorithmes (de faible complexité, parallélisables). Les méthodes développées sont ensuite testées sur des problèmes issus du monde socioéconomique dans des domaines d'application variés (SHS, médecine...).

Orientations scientifiques et choix stratégiques. Forts de compétences variées reposant sur des profils orientés aussi bien en mathématiques appliquées (section CNU 26) qu'en informatique (section CNU 27), nous souhaitons développer davantage ces synergies pour résoudre les verrous scientifiques qui se posent à nous. Ainsi, nous proposons d'étudier comment l'aide à la décision multicritère peut profiter aux modèles d'apprentissage automatique et de fouille de données, ou au contraire en tirer profit. Nous tenons beaucoup à effectuer des contributions aussi bien théoriques que pratiques, avec des applications dans le monde socio-économique, en particulier sur l'analyse des médias sociaux. Même si la dimension SHS est privilégiée, avec des projets en cours avec des linguistes, des sociologues, des historiens, etc., nous comptons appliquer nos modèles à d'autres domaines, comme celui de la médecine, par exemple.

Moyens à mobiliser pour atteindre les objectifs. L'attractivité de l'équipe, soutenue par un investissement dans les formations orientées recherche de l'Université Lyon 2, permet de recruter régulièrement des étudiants pour des stages de master et surtout des doctorants d'origines variées. Les financements sont à la fois publics (école doctorale, projet européen ou ANR, etc.) et proviennent également d'une activité soutenue de collaboration avec les entreprises (thèses CIFRE). Nous comptons poursuivre cette activité orientée projets qui permet de compléter utilement le financement récurrent de l'équipe. Concernant la valorisation du travail effectué dans l'équipe, nous avons l'objectif de continuer à publier dans les conférences et les journaux internationaux reconnus du domaine, tout en interagissant avec l'environnement régional et national dans lesquels nous nous sommes beaucoup investis. Depuis peu, nous réservons une partie du budget de l'équipe afin de faire relire les articles en anglais pour atteindre les meilleures revues internationales.

Renouvellement des partenariats. Nos partenariats nationaux et internationaux seront bien sûr reconduits, qu'il s'agisse de collaborations académiques (UPC en Espagne, UPB en Roumanie) ou industrielles (EDF, Technicolor, Thales, Xerox). De nouvelles collaborations sont actuellement mises en place (par exemple avec Télécom Bretagne et Centrale Paris dans le cadre du projet *Decision Deck*, ou avec l'INSERM sur des données médicales). Nous cherchons à davantage communiquer sur notre activité de recherche (journées thématiques ouvertes depuis 2012 attirant de nombreux participants extérieurs) et à construire des partenariats durables avec les équipes de recherche de la région Rhône-Alpes (Lyon 1, Grenoble, Saint-Etienne).

Nouvelles thématiques scientifiques. Nous proposons d'organiser l'activité de recherche de l'équipe en trois axes : a) étudier les relations entre l'apprentissage automatique et les méthodes d'optimisation combinatoire, b) étudier les relations entre l'apprentissage automatique et l'agrégation multicritère, c) développer des nouveaux modèles d'apprentissage automatique capables de prendre en compte le contexte (temporel, spatial) de données hétérogènes. Précisons que ces axes ne sont pas disjoints puisque, par exemple, la construction de nouveaux modèles capables de prendre en compte le contexte (axe C) profiteront utilement des travaux sur les liens avec l'optimisation (axe A) ou avec l'agrégation multicritère (axe B), et ainsi de suite. Plusieurs domaines d'applications sont envisagés, en particulier en lien avec les SHS (projets avec des linguistes, des géographes, des sociologues, etc.) mais également sur l'analyse des médias sociaux qui nous fournissent des données

présentant toutes les caractéristiques pour mettre nos techniques à l'épreuve : hétérogènes (blogs qui comportent du texte, des images, des métadonnées), temporelles (*tweets*, fils de discussion, flux RSS), structurées (Web social), mais également liées aux données ouvertes et au Web sémantique. Les différents projets menés ces dernières années nous garantissent d'avoir un accès privilégié à des sources variées de données correspondant à nos problèmes.

Axe A. La recherche en optimisation consiste à développer des processus automatiques de recherche de bonnes solutions, guidée par une ou plusieurs fonctions objectif à optimiser. L'apprentissage a pour but la construction de modèles à partir de données, en se passant de fonction objectif. L'intégration de techniques d'optimisation dans les méthodes d'apprentissage automatique est habituelle (minimisation des fonctions d'erreur). Néanmoins, le rôle de l'optimisation en apprentissage est devenu indispensable pour permettre en particulier le passage à l'échelle. D'un autre côté, l'intégration de techniques d'apprentissage dans la construction de processus de recherche de bonnes solutions est moins habituelle : comment intégrer l'historique des processus d'optimisation pour améliorer le processus lui-même ? Notre équipe travaille au développement de nouvelles combinaisons nécessaires entre ces deux domaines afin de rendre possible à la fois la construction des modèles et l'amélioration de la recherche de bonnes solutions. Notre objectif est de nous attaquer aussi bien aux verrous théoriques qu'algorithmiques, tout en travaillant à la résolution de problèmes académiques et industriels. Par exemple, nous nous intéressons à l'intégration de l'algorithme MCTS (*Monte Carlo Tree Search*) dans l'apprentissage de la structure de réseaux bayésiens, mais aussi à l'intégration de l'apprentissage dans l'algorithme MCTS pour résoudre des problèmes d'optimisation combinatoire. La recherche des caractérisations mathématiques des problèmes et des algorithmes constitue également une partie importante du travail de l'équipe.

Axe B. Il s'agit ici d'approfondir les liens existants entre l'apprentissage automatique et l'agrégation multicritère. La modélisation de prise de décision s'est développée dans plusieurs champs en parallèle. L'apprentissage automatique (ML) est basé sur une approche statistique et vise à découvrir les paramètres de modèles de classifications. L'agrégation multicritère (MCDA) est plus orientée vers l'accompagnement d'un décideur dans un processus décisionnel. Cependant, la communauté multicritère s'intéresse de plus en plus à la problématique d'élicitation des paramètres de modèles à partir de données. Il est alors intéressant de confronter les différentes approches, en prenant en compte les divergences entre domaines : i) grand nombre de données dans le cas ML, peu de variables et d'individus en MCDA ; ii) notions d'erreur dans les données en ML, d'incohérences dans les préférences en MCDA ; iii) méthodes quantitatives en ML, quantitative, mixte ou qualitative en MCDA ; etc. En particulier, le domaine de l'apprentissage des préférences est en plein essor et tend à faire le lien entre les communautés (conférence ADT, par exemple).

Axe C. De nos jours, l'accès aux données est devenu très facile (données ouvertes, par exemple), à tel point que l'on parle de « déluge » de données. Ces données sont directement disponibles sur Internet, d'où elles peuvent être récupérées et analysées par des outils automatiques. Cependant, le caractère fortement hétérogène des données issues de blogs, de médias et réseaux sociaux ou de sites d'actualité rend difficile l'utilisation d'algorithmes standards d'apprentissage automatique, qu'ils soient supervisés, non ou peu supervisés. En effet, le volume des données, la grande dimension de l'espace de description, le faible taux de données étiquetées et le caractère lacunaire des informations recueillies constituent encore aujourd'hui un verrou important. Dans ce contexte, intégrer des informations supplémentaires peut grandement améliorer l'efficacité de ces modèles, que ces informations proviennent du contexte des données (contexte topologique, spatial, temporel) mais également de sources complémentaires (comme celles issues du Web sémantique). Il s'agit par exemple de prendre en compte le contenu textuel de messages courts (comme des *tweets* ou des messages issus de forums) mais également leur environnement immédiat (voisinage des *retweets*, fil de discussion) pour capturer la dynamique temporelle des opinions exprimées dans un grand corpus de textes. Une autre illustration peut-être la correction des erreurs d'étiquetage en utilisant le voisinage des individus. Intégrer ces différentes dimensions hétérogènes (textuelle, temporelle, contextuelle, sémantique) pose de sérieux problèmes théoriques et pratiques car elles sont rarement commensurables. L'un des verrous auquel nous souhaitons nous attaquer dans le cadre de cette prise en compte du contexte est la difficulté d'agréger différents niveaux d'information, différentes granularités. Par exemple, comment peut-on prendre convenablement en compte la dimension temporelle, au-delà d'un découpage arbitraire de la frise chronologique ?

II. ANNEXES

Annexe 1 : Présentation synthétique (*Executive Summary*)

Les activités de recherche du laboratoire ERIC visent à valoriser les grandes bases de données complexes, notamment dans les domaines des sciences humaines et sociales. Les champs d'expertise d'ERIC couvrent les problématiques liées à la modélisation et l'exploitation des entrepôts de données complexes, la fouille de données hétérogènes, massives et peu structurées et les processus d'aide à la décision. Le laboratoire ERIC est composé de 52 membres : 21 enseignants-chercheurs, 1 personnel BIATSS, 25 doctorants, 2 post-doctorants ou ATER et 3 membres associés. Il est structuré en deux équipes de recherche : Systèmes d'Information Décisionnels (SID) et Data Mining et Décision (DMD).

La production scientifique globale sur la période 2009-2014 couvre 46 articles dans des revues internationales (dont 14 de rang A), 105 articles dans des conférences internationales (dont 23 de rang A), 13 ouvrages, directions d'ouvrages ou de revues, 65 articles dans des revues et des conférences nationales et 97 autres publications. 4 prototypes logiciels ont également été conçus. 17 thèses et 2 habilitations à diriger des recherches ont été soutenues.

Les ressources financières du laboratoire sont en constante progression (+24 % par an en moyenne depuis 2009).

ERIC participe à plusieurs programmes internationaux (dont 2 projets européens), invite régulièrement des professeurs et des chercheurs étrangers (20 de 2009 à 2014), a accueilli 9 étudiants en thèse dans le cadre de cotutelles, 3 postdoctorants étrangers, et collabore avec une trentaine d'universités du monde entier.

ERIC participe à plusieurs projets nationaux et locaux (dont 1 projet ANR dont nous sommes porteurs et 1 PIA) et collabore avec une vingtaine d'autres laboratoires français (dont 18 UMR).

Localement, les membres d'ERIC sont impliqués dans plusieurs structures fédératives (Institut des Sciences de l'Homme, Communauté de Recherche Académique régionale ARC6). Ils contribuent également aux formations d'enseignement supérieur des universités Lyon 1 et Lyon 2.

Executive summary. Research activities at ERIC aim at extracting value from very large, complex databases, especially in the domains of humanities. ERIC's fields of expertise include the issues related to modeling and exploiting complex data warehouses, mining heterogeneous, massive and loosely structured data, as well as decision-support processes. The ERIC lab is constituted of 52 members: 21 professors and associate professors, 1 administrative staff member, 25 Ph.D. students, 2 postdoctoral fellows and 3 associate members. It is composed of two research teams: Decision-support Information Systems (DIS) and Data Mining and Decision (DMD).

ERIC's global scientific production from 2009 to 2014 includes 46 papers in international journals (14 of which are ranked A), 105 papers in international conferences (23 of which are ranked A), 13 books, edited books or journal issues, 65 papers in domestic journals and conferences and 97 other publications. 4 software prototypes have also been designed. 17 Ph.D. theses and 2 qualifications for supervising research have been defended.

The lab's budget is continuously growing (+24% per year on an average since 2009).

ERIC is involved in several international programs (including 2 European projects), regularly invites foreign professors and researchers (20 from 2009 to 2014), housed 9 Ph.D. students in collaboration with foreign labs, 3 foreign postdoctoral fellows, and collaborates with about 30 universities worldwide.

ERIC is involved in many national and local projects (including 1 national project we lead) and collaborates with more than 20 research structures in France.

Locally, members of the lab are involved in several federating bodies (Human Sciences Institute and regional Academic Research Community #6). They also contribute to the higher education activity of the universities Lyon 1 and Lyon 2.

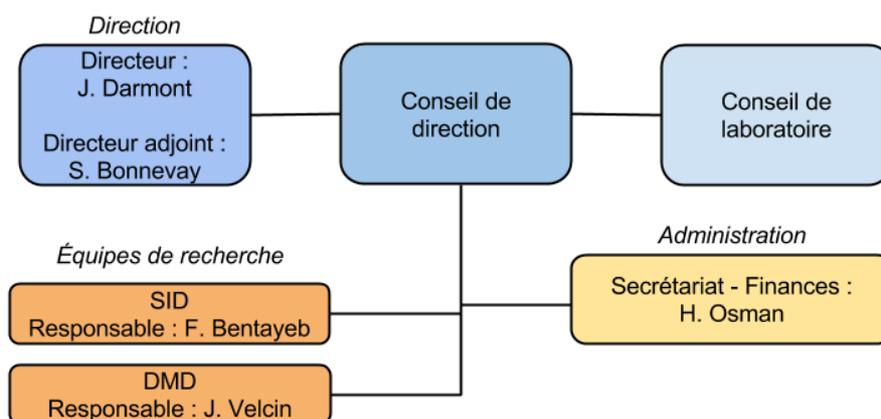
Annexe 2 : Lettre de mission contractuelle

N/A

Annexe 3 : Équipements, plateformes

- Serveur mail / web / intranet (PC Core i5 3,1 GHz - 4 Go de RAM - 2 disques de 500 Go)
- Serveur de calcul dédié au projet Mediamining (2 processeurs Intel(R) Xeon(R) CPU E5-2430, chacun avec 6 cœurs en hyperthreading - 64 Go de RAM - 3 To de disque)
- NAS Synology DS1511+ pour sauvegarde et partage de fichiers (2 disques d'1 To)
- Machines virtuelles louées à l'ISH
 - Serveur mail / web / intranet (migration du serveur actuel) : 1 vCore de CPU sandy bridge 2 GHz - 4 Go de VRAM - 100 Go de vDisk sauvegardés
 - Serveur de fichiers / annuaire / versions : 1 vCore de CPU sandy bridge 2 GHz - 1 Go de VRAM - 400 Go de vDisk sauvegardés
 - Serveur d'applications Linux : 8 vCore de CPU sandy bridge 2 GHz - 8 Go de VRAM - 400 Go de vDisk sauvegardés
 - Serveur d'applications Windows : 4 vCore de CPU sandy bridge 2 GHz - 4 Go de VRAM - 200 Go de vDisk sauvegardés
- Photocopieur / scanner / imprimante connecté Toshiba e-Studio 256se (location)

Annexe 4 : Organigramme fonctionnel



Annexe 5 : Statuts et règlement intérieur du laboratoire

STATUTS (MAI 2013)

Préambule

Les présents statuts et règlements ne doivent en aucun cas se substituer aux textes réglementaires du ministère de tutelle ni à ceux des établissements de rattachement du laboratoire. Ils sont destinés à assurer une transparence dans le fonctionnement et une gestion démocratique du laboratoire. En cas de conflit avec les textes réglementaires, ces derniers s'appliquent en priorité.

Ces statuts sont complétés par un règlement intérieur validé par le conseil de laboratoire.

Article 1 : Dénomination

ERIC (Entrepôts, Représentation et Ingénierie des Connaissances) est un laboratoire labélisé par le Ministère de l'Enseignement Supérieur et de la Recherche en tant qu'Équipe d'Accueil (EA) sous le numéro 3083.

Article 2 : Établissements de tutelle

ERIC est un laboratoire commun aux deux universités Claude Bernard Lyon 1 et Lumière Lyon 2. L'établissement de rattachement principal est l'Université Lumière Lyon 2.

Article 3 : Objet et Mission

Le laboratoire ERIC a pour objectif de développer des travaux de recherche théoriques et appliqués, ainsi que la création de logiciels, dans le domaine de l'informatique décisionnelle. Cela couvre notamment tous les aspects liés à la gestion et à l'exploitation des données et des connaissances (notamment dans les entrepôts de

données), à l'extraction de connaissances à partir des données (en particulier la fouille de données) et aux processus d'aide à la prise de décision.

Les missions du laboratoire couvrent :

- la production de connaissances,
- la diffusion de connaissances et de savoir-faire dans le domaine,
- la formation à la recherche par la recherche,
- participation à des projets scientifiques nationaux et/ou internationaux,
- valorisation de la recherche vers le grand public,
- création de valeur économique par des dépôts de brevets, l'incubation d'entreprises ou la création d'entreprises.

Article 4 : Organisation scientifique

Le laboratoire est organisé en équipes de recherche. Cette organisation vise à structurer le projet scientifique global du laboratoire en thèmes pouvant être traités en parallèle, mais également combinés pour traiter de questions scientifiques plus vastes. Les thèmes sont définis par le conseil de laboratoire au début du contrat quinquennal. Ces thèmes peuvent être redéfinis en cas de nécessité, notamment en cas d'arrivée d'un grand nombre de nouveaux chercheurs.

Une équipe de recherche regroupe au minimum quatre enseignants-chercheurs permanents. Le responsable doit nécessairement être un enseignant-chercheur permanent. Une équipe de recherche est l'unité opérationnelle de recherche. Selon leurs besoins, les membres d'une équipe sont libres de s'organiser de la façon qui leur paraît la plus appropriée.

Article 5 : Composition

Le laboratoire est composé de membres permanents, de membres associés et de membres non permanents.

Membres permanents

La qualité de membre permanent est accordée à toute personne qui en fait la demande et qui remplit les conditions suivantes :

- être enseignant et/ou chercheur sur un poste statutaire permanent dans un établissement de l'enseignement supérieur et/ou de la recherche reconnu par le Ministère de l'Enseignement Supérieur et de la Recherche et relever d'une discipline informatique, mathématiques appliquées ou toute autre discipline scientifique concourant à l'objet de recherche du laboratoire ; ou être personnel administratif sur poste statutaire (BIATSS...) ;
- avoir reçu l'approbation de la majorité des membres du conseil de laboratoire d'ERIC ;
- s'engager à respecter les présents statuts.

Membres associés

La qualité de membre associé est accordée à toute personne qui en fait la demande et qui remplit les conditions suivantes :

- occuper un poste fixe dans un établissement public ou privé ;
- effectuer sa recherche dans un domaine en rapport avec les travaux menés au sein du laboratoire (par exemple, enseignant-chercheur en SHS et conduisant des recherches nécessitant un recours significatif à l'informatique ou à la modélisation mathématique) ;
- s'engager à réaliser des projets de recherche s'inscrivant dans la programmation scientifique du laboratoire ;
- avoir reçu l'approbation de la majorité des membres du conseil de laboratoire ;
- s'engager à respecter les présents statuts.

Membres non permanents

La qualité de membre non permanent est accordée à toute personne de statut transitoire : ATER, PAST ou personnel administratif contractuel d'une part, et doctorant, chercheur postdoctorant ou chercheur contractuel exerçant une activité sous la direction d'un membre permanent et sur la base d'un projet

scientifique précis, d'autre part. Le recrutement d'un membre non permanent doit être approuvé par le conseil de laboratoire.

Perte du statut de membre

La qualité de membre permanent, non permanent ou associé peut être perdue dès que la personne :

- quitte le laboratoire ;
- en fait la demande par courrier adressé au directeur ;
- perd sa qualité d'enseignant et/ou de chercheur titulaire ou de personnel administratif (uniquement pour les membres permanents) ;
- ne respecte pas les statuts après deux rappels notifiés par écrit ;
- est exclue sur délibération du conseil de laboratoire aux deux tiers des votants sans qu'un cinquième ou plus des membres soit contre la décision.

Article 6 : Organisation

Le laboratoire est composé d'un conseil de direction, d'un conseil de laboratoire et est dirigé par un directeur, assisté d'un ou plusieurs directeurs adjoints et d'un responsable administratif et financier.

Directeur

Il est le garant de l'unité et de l'intérêt général. Le directeur est chargé de veiller à l'application de toutes les décisions prises par les deux conseils. Il est le représentant officiel du laboratoire auprès de toutes les instances internes et externes. Il valide tous les actes engageant le laboratoire notamment les contrats, le budget, les ordres de mission relevant de la recherche, l'accueil de nouveaux membres, la réalisation de toutes les recettes, l'engagement de toutes les dépenses.

Le directeur est élu par le conseil de laboratoire pour une durée de cinq ans renouvelable, parmi les personnels habilités à diriger les recherches.

Le directeur préside et anime le conseil de direction ainsi que le conseil de laboratoire. En cas d'absence, il peut confier cette mission à l'un des directeurs adjoints.

Directeurs adjoints

Le directeur est assisté d'un ou plusieurs directeurs adjoints en fonction de la taille du laboratoire et de sa répartition géographique. Les directeurs adjoints sont proposés par le directeur et approuvés par le conseil de laboratoire.

Un directeur adjoint a des missions délimitées par le directeur et approuvées par le conseil de laboratoire.

Responsable administratif et financier

Le directeur est assisté d'un responsable administratif et financier dans les tâches administratives et pour toutes les questions relatives au budget du laboratoire et à son exécution.

Conseil de direction

Il est composé du directeur, des directeurs adjoints, des responsables d'équipes de recherche et du responsable administratif et financier.

Le rôle du conseil de direction est de coordonner les activités d'animation et de valorisation scientifiques du laboratoire, d'en proposer les orientations stratégiques et de les soumettre au conseil de laboratoire. Il veille à l'application des décisions du conseil de laboratoire. Il participe à toutes les actions qui touchent à la vie collective : préparation du budget, organisation de manifestations scientifiques, affectation des doctorants et des personnes recrutées dans les locaux, élaboration des priorités scientifiques, notamment pour l'attribution des bourses de thèse. Il se réunit de façon ordinaire, une fois par mois, et peut-être convoqué à tout moment par le directeur du laboratoire.

Conseil de laboratoire

Le conseil de laboratoire se compose :

- des membres permanents
- de deux membres élus parmi les membres non permanents et associés, dont au moins un doctorant.

Les chercheurs associés sont invités au conseil de laboratoire avec voix consultative.

Le conseil de laboratoire est l'instance de validation des choix stratégiques du laboratoire. Il valide le budget, le classement des candidats aux allocations de recherche, l'attribution des moyens aux enseignants-chercheurs, le choix des projets portés par le laboratoire et toutes les actions relatives à la politique scientifique.

Le conseil se réunit de façon ordinaire une fois par mois et peut-être convoqué par le conseil de direction, le directeur ou le tiers de ses membres à tout moment. L'ordre du jour est établi à l'avance par le directeur et communiqué aux membres avant la réunion.

Sauf mention contraire explicite, les modalités de vote du conseil de laboratoire sont fixées comme suit.

- Seuls les membres présents ou représentés prennent part aux votes.
- Un votant ne peut porter plus de deux procurations de membres permanents absents.
- Les votes à bulletin secret sont organisés à la demande d'au moins deux membres présents ou en cas de décision concernant des positions nominatives.
- Les décisions sont prises à la majorité simple des votes exprimés sauf mention contraire explicitée.
- Un quorum de 25 % des membres du conseil de laboratoire est requis pour valider un vote.

Formation restreinte du conseil de laboratoire

Lorsqu'il procède au classement des allocataires de recherche, à l'élaboration des profils de poste recherche de maître de conférences ou de professeur des universités, le conseil de laboratoire siège en formation réduite aux seuls membres permanents.

Article 7 : Ressources du laboratoire

Le budget global du laboratoire est constitué de la dotation allouée par les deux établissements de tutelle et de prélèvements sur les contrats publics ou privés obtenus par ses membres, qui contribuent aux frais de fonctionnement du laboratoire. Les ressources du budget global sont gérées par le directeur selon le budget proposé par le conseil de direction et voté par le conseil de laboratoire. Les dépenses sur le budget global sont essentiellement dédiées au fonctionnement quotidien du laboratoire et aux actions de d'animation et de valorisation scientifiques.

Chaque équipe de recherche dispose en outre d'un budget propre dont le montant, prélevé du budget global, est décidé par le conseil de laboratoire sur proposition du conseil de direction. La clé de répartition de ce montant global entre les équipes de recherche est précisée dans le règlement intérieur du laboratoire.

Les contrats, publics ou privés, constituent la principale source de financement de l'activité de recherche. Le laboratoire prélève un pourcentage sur toutes les recettes de contrats. Ce pourcentage est précisé dans le règlement intérieur du laboratoire. La gestion et le suivi effectif du contrat sont assurés par le responsable du contrat conformément au budget proposé. Les contrats sont suivis sur le plan comptable par le responsable administratif du laboratoire. Seuls les fonds faisant l'objet d'un contrat avec le laboratoire entrent dans le cadre du bilan quantitatif du laboratoire.

Article 8 : Publications

Les publications produites par tout membre du laboratoire doivent mentionner son nom, ainsi que le nom de l'établissement de rattachement selon les règles proposées par L'Université de Lyon. Le libellé exact est précisé dans le règlement intérieur du laboratoire. Seules ces publications feront partie du bilan du laboratoire.

Article 9 : Modification des statuts

Ces statuts peuvent être modifiés à la demande d'au moins les deux tiers des membres du conseil du laboratoire.

Article 10 : Rattachement à une école doctorale

Le rattachement du laboratoire à une école doctorale est décidé par le conseil de laboratoire après recueil des avis de chaque membre permanent habilité à diriger les recherches. Ce rattachement est précisé dans le règlement intérieur du laboratoire.

Article 11 : Rapport d'activité

À mi-parcours du contrat quinquennal, le conseil de direction rédige un rapport d'activité. Ce rapport est soumis au conseil du laboratoire pour validation. Une évaluation externe pourra être organisée par des experts.

RÈGLEMENT INTERIEUR (MARS 2014)

Préambule

Les établissements de tutelle du laboratoire ERIC (EA 3083) sont les universités Claude Bernard Lyon 1 (UCBL) et Lumière Lyon 2 (ULL2, tutelle principale). ERIC est également unité membre de l'Institut des Sciences de l'Homme de Lyon (ISH).

Par ailleurs, le laboratoire ERIC est rattaché à l'École Doctorale Informatique et Mathématiques (ED 512).

Article 1 : Dispositions générales

Tous les membres d'ERIC tels que définis dans l'article 5 des statuts du laboratoire, ainsi que tout visiteur lié au laboratoire par une convention (chercheurs étrangers, stagiaires...), est assujéti :

- au règlement intérieur de l'ULL2 dans ses locaux du campus de la Porte des Alpes ;
- au règlement intérieur de l'UCBL dans ses locaux du campus de la Doua ;
- au présent règlement.

Article 2 : Fonctionnement financier du laboratoire

Conformément à l'article 7 des statuts du laboratoire, ce dernier effectue des prélèvements sur les contrats publics ou privés obtenus par ses membres pour contribuer à ses frais de fonctionnement. Ce prélèvement forfaitaire se monte à 10 % du montant global du contrat affecté au fonctionnement. En cas d'autres prélèvements (par exemple, par les établissements de tutelle), le prélèvement du laboratoire sera minoré afin que le prélèvement total ne dépasse pas 15 % du montant global du contrat.

Conformément aux articles 4 et 7 des statuts du laboratoire, ce dernier est structuré en équipes de recherche :

- Systèmes d'Information Décisionnels (SID) ;
- Data Mining & Décision (DMD).

Le budget global alloué chaque année aux équipes de recherche est soumis à l'approbation du conseil de laboratoire et réparti au prorata de leurs effectifs (membres permanents et non-permanents).

Article 3 : Utilisation des ressources du laboratoire

Le photocopieur/scanner ainsi que ses consommables (notamment le papier), les postes téléphoniques et les ordinateurs mis à disposition des membres du laboratoire sont à usage strictement professionnel.

Hormis les logiciels libres et/ou gratuits, tout logiciel installé sur les ordinateurs du laboratoire doit faire l'objet d'une licence acquittée par le laboratoire lui-même ou par les établissements de tutelle d'ERIC ou l'ISH et dont ERIC est bénéficiaire.

Tout membre du laboratoire présent dans les locaux du campus de la Porte des Alpes s'engage à respecter la charte informatique de l'ULL2 (Annexe 1).

Article 4 : Éthique scientifique

Tout membre du laboratoire s'engage à respecter l'éthique de la publication scientifique (Annexe 2). De plus, toute publication produite au laboratoire doit porter la mention « Université de Lyon (Laboratoire ERIC) ».

Conformément au règlement de l'École Doctorale Informatique et Mathématiques, tout doctorant, ainsi que son ou ses encadrants, doivent signer la Charte des thèses de l'Université de Lyon (Annexe 3).

Article 5 : Dispositions relatives aux membres non permanents

Ces dispositions s'appliquent également aux visiteurs liés au laboratoire ERIC par une convention (chercheurs étrangers, stagiaires...).

Le laboratoire alloue à chaque membre non permanent un espace de travail et un ordinateur.

Hors périodes de cours, de missions (ces dernières ayant fait l'objet d'un ordre de mission dûment validé) ou de vacances, les membres non permanents doivent être présents au laboratoire pendant ses horaires d'ouverture (du lundi au vendredi de 8h30 à 18h, hormis durant les périodes de fermeture des universités de tutelle) dans la limite de 35 heures hebdomadaires. Les doctorant(e)s en entreprise ne sont assujéti à cette contrainte que durant leurs périodes de présence au laboratoire, stipulées dans la convention liant le laboratoire et l'entreprise d'accueil.

Toute absence doit être signalée au(x) membre(s) permanent(s) en charge de l'encadrement du membre non permanent (par exemple, le directeur de thèse ou le tuteur de stage), ainsi qu'au secrétariat du laboratoire. Tout manquement délibéré et répété à cette obligation pourra aboutir à des sanctions telles que la non autorisation de réinscription en thèse ou la suspension du paiement de gratification de stage.

Les annexes ne figurent pas dans ce document du fait de leur taille.

Annexe 6 : Liste des réalisations et produits de la recherche

PRODUCTION SCIENTIFIQUE

Équipe SID

Ouvrages ou direction d'ouvrages de portée internationale

- [Bentayeb-10] BENTAYEB F., BOUSSAID O., DARMONT J., HARBI N., LOUDCHER S.. «Warehousing and Mining Complex Data: Applications to Biology, Medicine, Behavior, Health and Environment». (2010).
- [Darmont-12a] DARMONT J., Pedersen T. B. «Proceedings of the 1st International Workshop on Cloud Intelligence (Cloud-I 2012)». (2012). Ref HAL: [hal-00727331_v1](https://hal.archives-ouvertes.fr/hal-00727331_v1).
- [Darmont-13a] DARMONT J., Pedersen T. B. «Proceedings of the 2nd International Workshop on Cloud Intelligence (Cloud-I 2013)». (2013). Ref HAL: [hal-00921824_v1](https://hal.archives-ouvertes.fr/hal-00921824_v1).

Reuves internationales

- [Abello-13] Abello A., DARMONT J., Etcheverry L., Golfarelli M., Mazón J.-N., Naumann F., Pedersen T. B., Rizzi S., Trujillo J., Vassiliadis P., Vossen G. «Fusion Cubes: Towards Self-Service Business Intelligence». In *International Journal of Data Warehousing and Mining*, vol. 9/2 (2013) p.66-88. DOI: [10.4018/jdwm.2013040104](https://doi.org/10.4018/jdwm.2013040104). Rang C.
- [Aouiche-09a] Aouiche K., DARMONT J.. «Data Mining-based Materialized View and Index Selection in Data Warehouses». In *Journal of Intelligent Information Systems*, vol. 33/1 (2009) p.65-93. DOI: [10.1007/s10844-009-0080-0](https://doi.org/10.1007/s10844-009-0080-0). Rang B.
- [Aufaure-13] Aufaure M.-A., Cuzzocrea A., FAVRE C., Marcel P., Missaoui R. «An Envisioned Approach for Modeling and Supporting User-Centric Query Activities on Data Warehouses». In *International Journal of Data Warehousing and Mining*, vol. 9/2 (2013) p.89-109. DOI: [10.4018/jdwm.2013040105](https://doi.org/10.4018/jdwm.2013040105). Rang C.
- [Bimonte-12] Bimonte S., Bertolotto M., Gensel J., BOUSSAID O.. «Spatial OLAP and Map Generalization: Model and Algebra». In *International Journal of Data Warehousing and Mining*, vol. 8/1 (2012) p.24-51. DOI: [10.4018/jdwm.2012010102](https://doi.org/10.4018/jdwm.2012010102). Rang C.
- [Boukraa-10] BOUKRAA D., BOUSSAID O., BENTAYEB F., LOUDCHER S.. «OLAP Operators For A Complex Object-Based Multidimensional Model». In *International Journal of Business Intelligence and Data Mining* (2010). Rang C.
- [Boukraa-12] BOUKRAA D., BOUSSAID O., BENTAYEB F.. «Complex Object-Based Multidimensional Modeling and Cube Construction In the International». Accepted in *Fundamenta Informaticae* (2012). Rang B.
- [Cuzzocrea-09] Cuzzocrea A., DARMONT J., MAHBOUBI H.. «Fragmenting Very Large XML Data Warehouses via K-Means Clustering Algorithm». In *International Journal of Business Intelligence and Data Mining*, vol. 4/3/4 (2009) p.301-328. DOI: [10.1504/IJBIDM.2009.029076](https://doi.org/10.1504/IJBIDM.2009.029076). Rang C.
- [Derrar-12] Derrar H., Ahmed-Nacer M., BOUSSAID O.. «Exploiting data access for dynamic fragmentation in data warehouse In the International». In *International Journal of Intelligent Information and Database Systems*, vol. 7/1 (2012) p.34-52. DOI: [10.1504/IJIDS.2013.051736](https://doi.org/10.1504/IJIDS.2013.051736). Rang C.
- [Farid-10] Farid D. M., HARBI N., Rahman M. Z. «Combining Naïve Bayes and Decision Tree for Adaptive Intrusion Detection». In *International journal of Network Security & Its Applications (IJNSA)*, vol. 2/2 (2010) p.12. DOI: [10.5121/ijnsa.2010.2202](https://doi.org/10.5121/ijnsa.2010.2202).
- [Farid-10a] Farid D. M., DARMONT J., Rahman M. Z. «Attribute Weighting with Adaptive NBTree for Reducing False Positives in Intrusion Detection». In *International Journal of Computer Science and Information Security*, vol. 8/1 (2010) p.19-26. Rang C.
- [Hacene-12] Hacène D., Mohamed A.-N., BOUSSAID O.. «Particle swarm optimisation for data warehouse logical design». In *International Journal of Bio-Inspired Computation* (2012). Ref HAL: [hal-00712141_v1](https://hal.archives-ouvertes.fr/hal-00712141_v1). Rang B.
- [Hachicha-13] HACHICHA M., DARMONT J. (Corresp.). «A Survey of XML Tree Patterns». In *IEEE Transactions on Knowledge and Data Engineering*, vol. 25/1 (2013) p.29-46. Ref HAL: [hal-00623652_v1](https://hal.archives-ouvertes.fr/hal-00623652_v1). Rang A.
- [Loudcher-12] LOUDCHER S., BOUSSAID O.. «OLAP on Complex Data: Visualization Operator Based on Correspondance Analysis». In *Lecture Notes in Business Information Processing*, vol. 107 (2012) p.172-185. Ref HAL: [halshs-00677401_v1](https://hal.archives-ouvertes.fr/halshs-00677401_v1).
- [Salem-13] SALEM R., DARMONT J., BOUSSAID O.. «Active XML-based Web Data Integration». *Information Systems Frontiers*, vol. 15/3 (2013) p. 371-398. DOI: [10.1007/s10796-012-9405-6](https://doi.org/10.1007/s10796-012-9405-6). Rang B.

- [Wolf-10] Wolf C. GAVIN G. «Inference and parameter estimation on hierarchical belief networks for image segmentation». In *Neurocomputing*, vol. 73/4-6 (2010) p.563-569. DOI: [10.1016/j.neucom.2009.07.017](https://doi.org/10.1016/j.neucom.2009.07.017). Rang A.

Conférences internationales avec comité de programme et actes

- [Aknouche-12] AKNOUCHE R., ASFARI O., BENTAYEB F., BOUSSAID O. «Integrating Query Context and User Context in an Information Retrieval Model Based on Expanded Language Modeling». In proc. of *International Cross Domain Conference and Workshop (CD-ARES 2012)* p.244-258. Ref HAL: [hal-00911860_v1](https://hal.archives-ouvertes.fr/hal-00911860_v1).
- [Aknouche-13a] AKNOUCHE R. (Corresp.), ASFARI O., BENTAYEB F., BOUSSAID O. «Decisional Architecture for Text Warehousing: ETL-Text Process and Multidimensional Model TWM». In proc. of *International Conference on Management of Data (COMAD 2013)*. Rang B.
- [Arres-13a] ARRES B. (Corresp.), KABACHI N., BOUSSAID O. «Building OLAP cubes on a Cloud Computing environment with MapReduce - Benchmarking Hive». In proc. of *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2013)* p.26. Ref HAL: [hal-00907004_v1](https://hal.archives-ouvertes.fr/hal-00907004_v1). Rang B.
- [Bahri-11] BAHRI E. (Corresp.), HARBI N., NGUYEN H.-H. «How to detect better and faster intrusions in computer networks using ensemble methods». In proc. of *International Conference on Computer Networks and Security (ICCNS 2011)*.
- [Bahri-11a] BAHRI E., HARBI N., NGUYEN H.-H. «Approach Based Ensemble Methods for Better and Faster Intrusion Detection». In proc. of *Computational Intelligence in Security for Information Systems (CISIS 2011)* p.17-24. DOI: [10.1007/978-3-642-21323-6_3](https://doi.org/10.1007/978-3-642-21323-6_3). Rang B.
- [Bahri-12a] BAHRI E., HARBI N. (Corresp.), NGUYEN H.-H. «A Multiple Classifier System Using an Adaptive Strategy for Intrusion Detection». In proc. of *International Conference on Intelligent Computational System (ICICS 2012)* p.124-128.
- [Basileu-11] BASILEU C., KABACHI N., LAMURE M. «Towards an intelligent system for decision making in pandemics case». In proc. of *European Social Simulation Association Conference (ESSA 2011)*.
- [BenHassine-09a] BEN HASSINE S. (Corresp.). «Data quality evaluation in an e-business environment: a survey». In proc. of *International Conference on Information Quality (ICIQ 2009)*. Ref HAL: [hal-00519723_v1](https://hal.archives-ouvertes.fr/hal-00519723_v1). Rang B.
- [Bentayeb-09a] BENTAYEB F. (Corresp.), FAVRE C. «RoK: Roll-Up with the K-Means Clustering Method for Recommending OLAP Queries». In proc. of *International Conference on Database and Expert Systems Applications (DEXA 2009)* p.501-505. DOI: [10.1007/978-3-642-03573-9_43](https://doi.org/10.1007/978-3-642-03573-9_43). Rang B.
- [Bentayeb-13] BENTAYEB F. (Corresp.), KHEMIRI R. «Adapting OLAP Analysis to User's Constraints through Semantic Hierarchies». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2013)* p.193-200. Rang B.
- [Boukraa-10b] BOUKRAA D., BOUSSAID O., BENTAYEB F. «OLAP Operators for Complex Object Data Cubes». In proc. of *East-European Conference on Advances in Databases and Information Systems (ADBIS 2010)* p.103-116. DOI: [10.1007/978-3-642-15576-5_10](https://doi.org/10.1007/978-3-642-15576-5_10). Rang B.
- [Boukraa-11a] BOUKRAA D., BOUSSAID O., BENTAYEB F. «Vertical Fragmentation of XML Data Warehouses using Frequent Path Sets». In proc. of *International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2011)* p.103-116. DOI: [10.1007/978-3-642-23544-3_15](https://doi.org/10.1007/978-3-642-23544-3_15). Rang B.
- [Boukraa-13] BOUKRAA D., BOUSSAID O., BENTAYEB F. (Corresp.), Zegour D.-E. «A Layered Multidimensional Model of Complex Objects». In proc. of *International Conference on Advanced Information Systems Engineering (CAISE 2013)* 498-513. DOI: [10.1007/978-3-642-38709-8_32](https://doi.org/10.1007/978-3-642-38709-8_32). Rang A.
- [Clement-10] Clément D. BEN HASSINE S. (Corresp.), Laboisse B. «Data Quality as a Key Success Factor for Migration Projects». In proc. of *International Conference on Information Quality (ICIQ 2010)*. Ref HAL: [hal-00519703_v1](https://hal.archives-ouvertes.fr/hal-00519703_v1). Rang B.
- [Farid-09] Farid D. DARMONT J., HARBI N., NGUYEN H.-H., Rahman M. Z. «Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification». In proc. of *International Conference on Computer Systems Engineering (ICCSE 2009)*. Ref HAL: [hal-00503951_v1](https://hal.archives-ouvertes.fr/hal-00503951_v1).
- [Farid-10b] Farid D. NGUYEN H.-H., DARMONT J., HARBI N., Rahman M. Z. «Scaling up Detection Rates and Reducing False Positives in Intrusion Detection using NBTree». In proc. of *International Conference on Data Mining and Knowledge Engineering (ICDMKE 2010)*. Ref HAL: [hal-00503961_v1](https://hal.archives-ouvertes.fr/hal-00503961_v1).
- [Farid-10c] Farid D. M. HARBI N., BAHRI E., Rahman M. Z. Rahman C. M. «Attacks Classification in Adaptive Intrusion Detection using Decision Tree». In proc. of *International Conference on Computer Science (ICCS 2010)*.
- [Farid-10d] Farid D. DARMONT J., HARBI N., Rahman C. M. «A NEW SUPERVISED LEARNING ALGORITHM USING NAÏVE BAYESIAN CLASSIFIER». In proc. of *IADIS International Conference on Information Systems (IS 2010)* p.78-84. Ref HAL: [hal-00503956_v1](https://hal.archives-ouvertes.fr/hal-00503956_v1). Rang C.
- [Ferrarons-14] Ferrarons J. Adhana M. Colmenares C. Pietrowska S. BENTAYEB F., DARMONT J. «PRIMEBALL: a Parallel Processing Framework Benchmark for Big Data Applications in the Cloud». In proc. of *TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2014)* p.109-124. Ref Arxiv: [1312.6293](https://arxiv.org/abs/1312.6293).

- [Gavin-09] GAVIN G. Minier M. «Oblivious Multi-variate Polynomial Evaluation». In proc. of *International Conference on Cryptology (INDOCRYPT 2009)* p.430-442. Rang B.
- [Hachicha-10] HACHICHA M. DARMONT J. «Pattern tree-based XOLAP rollup operator for XML complex hierarchies». In proc. of *International Conference on Machine and Web Intelligence (ICMWI 2010)*. Ref Arxiv: [1102.0952](#).
- [Hachicha-12] HACHICHA M. KIT C. DARMONT J. «A Novel Query-Based Approach for Addressing Summarizability Issues in XOLAP». In proc. of *International Conference on Management of Data (COMAD 2012)* p.56-67. Ref Arxiv: [1308.6682](#). Rang B.
- [Hajlaoui-10] HAJLAOUI K. Camarinha-Matos L. M. Afsarmanesh H. Boucher X. BOUSSAID O. «UNICOMP: Identification of Enterprise Competencies to Build Collaborative Networks». In proc. of *IFIP Working Conference on Virtual Enterprises (Pro-VE 2010)*. Ref HAL: [emse-00662022_v1](#). Rang C.
- [Hannachi-12] Hannachi L. ASFARI O. (Corresp.), BENTAYEB F. KABACHI N. BOUSSAID O. «Community Extraction based on Topic-Driven-Model for Clustering Users Tweets». In proc. of *International Conference on Advanced Data Mining and Applications (ADMA 2012)* p.39-51. DOI: [10.1007/978-3-642-35527-1_4](#). Rang B.
- [Hannachi-13] Hannachi L. Benblidia N. BENTAYEB F. BOUSSAID O. «Social Microblogging Cube». In proc. of *ACM International Workshop on Data Warehousing and OLAP (DOLAP 2013)* p.19-26. DOI: [10.1145/2513190.2513200](#). Rang B.
- [Harbi-10] HARBI N. (Corresp.), BAHRI E. «Approach based aggregation for intrusions detection». In proc. of *International Conference on Computer, Electrical, and Systems Science, and Engineering (ICCESSE 2010)*.
- [Khemiri-13] KHEMIRI R. BENTAYEB F. (Corresp.). «FIMIOQR: Frequent Itemsets Mining for Interactive OLAP Query Recommendation». In proc. of *International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2013)* p.9-14. Rang C.
- [Kit-12] KIT C. HACHICHA M. DARMONT J. «Benchmarking Summarizability Processing in XML Warehouses with Complex Hierarchies». In proc. of *ACM International Workshop on Data Warehousing and OLAP (DOLAP 2012)* p.73-80. Ref Arxiv: [1308.6683](#). Rang B.
- [Madani-11] Madani A. BOUSSAID O. Zegour D. «Clust-XPaths: Clustering of XML Paths». In proc. of *International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2011)* p.294-305. DOI: [10.1007/978-3-642-23199-5_22](#). Rang B.
- [Mahboubi-09c] MAHBOUBI H. DARMONT J. «Enhancing XML Data Warehouse Query Performance by Fragmentation». In proc. of *Annual ACM Symposium on Applied Computing (SAC 2009)* p.1555-1562. Ref Arxiv: [0908.3957](#). Rang A.
- [Mahboubi-11] MAHBOUBI H. DARMONT J. «XWeB: the XML Warehouse Benchmark». In proc. of *TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2011)* p.185-203. Ref Arxiv: [1102.0372](#).
- [Maiz-10] MAIZ N. Fahad M. BOUSSAID O. BENTAYEB F. «Automatic Ontology Merging by Hierarchical Clustering and Inference Mechanisms». In proc. of *International Conference on Knowledge Management and Knowledge Technologies (I-KNOW 2010)* p.103-116. Rang C.
- [Nguyen-11] NGUYEN H.-H. HARBI N. DARMONT J. «An Efficient Fuzzy Clustering-Based Approach for Intrusion Detection». In proc. of *East-European Conference on Advances in Databases and Information Systems (ADBIS 2011)* p.117-127. Ref Arxiv: [1110.2704](#). Rang B.
- [Nguyen-11a] NGUYEN H.-H. HARBI N. DARMONT J. «An Efficient Local Region and Clustering-Based Ensemble System for Intrusion Detection». In proc. of *International Database Engineering and Application Symposium (IDEAS 2011)* p.185-191. Ref HAL: [hal-00631503_v1](#). Rang B.
- [Oukid-13] Oukid L. ASFARI O. (Corresp.), BENTAYEB F. Benblidia N. BOUSSAID O. «CXT-Cube: Contextual Text Cube Model and Aggregation Operator for Text OLAP». In proc. of *ACM International Workshop on Data Warehousing and OLAP (DOLAP 2013)* p.27-32. DOI: [10.1145/2513190.2513201](#). Rang B.
- [Pitarch-10] Pitarch Y. FAVRE C. Laurent A. Poncelet P. «Context-Aware Generalization for Cube Measures». In proc. of *ACM International Workshop on Data Warehousing and OLAP (DOLAP 2010)* p.99-104. Ref HAL: [lirmm-00798821_v1](#). Rang B.
- [Pitarch-12] Pitarch Y. FAVRE C. Laurent A. Poncelet P. «Enhancing Flexibility and Expressivity of Contextual Hierarchies». In proc. of *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2012)* p.1-8. DOI: [10.1109/FUZZ-IEEE.2012.6251176](#). Rang A.
- [Rezoug-11] Rezoug N. BOUSSAID O. Nader F. «Integration of Profile in OLAP Systems». In proc. of *International Conference on Knowledge Discovery and Information Retrieval (KDIR 2011)* p.320-332.
- [Salem-10] SALEM R. (Corresp.), BOUSSAID O. DARMONT J. «Conceptual Workflow for Complex Data Integration using AXML». In proc. of *International Conference on Machine and Web Intelligence (ICMWI 2010)* p.6. Ref HAL: [hal-00560970_v1](#).
- [Salem-11] SALEM R. DARMONT J. BOUSSAID O. «Efficient Incremental Breadth-Depth XML Event Mining». In proc. of *International Database Engineering and Application Symposium (IDEAS 2011)* p.197-203. Ref Arxiv: [1110.3158](#). Rang B.

- [Salem-12] SALEM R. BOUSSAID O. DARMONT J. «An Active XML-based Framework for Integrating Complex Data». In proc. of *Annual ACM Symposium on Applied Computing (SAC 2012)* p.888-892. Ref HAL: [hal-00686001_v1](#). Rang A.
- [Selmane-14b] SELMANE S. A. (Corresp.), BENTAYEB F. Missaoui R. BOUSSAID O. «Efficient method for community detection based on formal concept analysis». In proc. of *International Symposium on Methodologies for Intelligent Systems (ISMIS 2014)* p.61-72. Rang B.
- [Triki-10] TRIKI S. Ben-Abdallah H. Feki J. HARBI N. «MODELING CONFLICT OF INTEREST IN THE DESIGN OF SECURE DATA WAREHOUSES». In proc. of *International Conference on Knowledge Engineering and Ontology Development (KEOD 2010)* p.445. Ref HAL: [hal-00625078_v1](#). Rang C.
- [Triki-11] TRIKI S. (Corresp.), Ben-Abdallah H. HARBI N. BOUSSAID O. «Securing Data Warehouses: A Semi-automatic Approach for Inference Prevention at the Design Level». In proc. of *International Conference on Model & Data Engineering (MEDI 2011)* p.71. Ref HAL: [hal-00634705_v1](#).

Autres publications internationales

- [Aknouche-13] AKNOUCHE R. (Corresp.), ASFARI O. BENTAYEB F. BOUSSAID O. «Integration Process for Multidimensional Textual Data Modeling». In proc. of *International Workshop in Software Evolution and Modernization (SEM-ENASE 2013)* p.119-126. Ref HAL: [hal-00911862_v1](#).
- [Aouiche-09] Aouiche K. DARMONT J. «Index and Materialized View Selection in Data Warehouses». *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (2009) p.693-700.
- [BenHassine-10] BEN HASSINE S. (Corresp.), Coppet O. Labosse B. «Statistics and Data Quality Towards more collaboration between these communities». Ref HAL: [hal-00519979_v1](#).
- [Bentayeb-09] BENTAYEB F. FAVRE C. BOUSSAID O. «Dynamic Workload for Schema Evolution in Data Warehouses: a Performance Issue». *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development* (2009) p.28-46.
- [Bentayeb-11a] BENTAYEB F. MAIZ N. MAHBOUBI H. FAVRE C. LOUDCHER S. HARBI N. BOUSSAID O. DARMONT J. «Innovative Approaches for efficiently Warehousing Complex Data from the Web». *Business Intelligence Applications and the Web : Models, Systems and Technologies* (2011) p.20. Ref HAL: [halshs-00606792_v1](#).
- [Boukraa-09] BOUKRAA D. Ben Messaoud R. BOUSSAID O. «Modeling XML warehouses for complex data: the new issues». *Open and Novel Issues in XML Database Applications* (2009) p.287-307.
- [Darmont-09] DARMONT J. «Database Benchmarks». *Encyclopedia of Information Science and Technology, Second Edition* (2009) p.950-954.
- [Darmont-09a] DARMONT J. «Data warehouse benchmarking with DWEB». *Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytic* (2009) p.302-323. DOI: [10.4018/978-1-60566-232-9.ch015](#).
- [Darmont-11] DARMONT J. «Privacy and performance of cloud data warehouses». Ref HAL: [hal-00667301_v1](#).
- [Darmont-12] DARMONT J. Pedersen T. B. Middelfart M. «Cloud Intelligence: What is REALLY New?». DOI: [10.1145/2347673.2347683](#).
- [Darmont-13] DARMONT J. Pedersen T. B. «Cloud intelligence: challenges for research and industry». DOI: [10.1145/2501928.2506094](#).
- [Darmont-14] DARMONT J. (Corresp.). «Data Processing Benchmarks». *Encyclopedia of Information Science and Technology, Third Edition* (2014). Ref HAL: [hal-00978026_v1](#).
- [Darmont-14a] DARMONT J. Pedersen T. B. «Report on the Second International Workshop on Cloud Intelligence (Cloud-I 2013)». Ref HAL: [hal-00978029_v1](#).
- [Favre-09] FAVRE C. BENTAYEB F. BOUSSAID O. «A Survey of Data Warehouse Model Evolution». *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (2009) p.129-136. Ref HAL: [hal-00701445_v1](#).
- [Jakawat-13] JAKAWAT W. FAVRE C. LOUDCHER S. (Corresp.). «OLAP on Information Networks: a new Framework for Dealing with Bibliographic Data». In proc. of *International Workshop on Social Business Intelligence (SoBI-ADBIS 2013)* p.361-370. Ref HAL: [hal-00834929_v1](#).
- [Khemiri-12] KHEMIRI R. BENTAYEB F. (Corresp.). «Interactive Query Recommendation Assistant». In proc. of *Recommender Systems meet Databases (RSmeetDB 2012)* p.93-97.
- [Loudcher-11a] LOUDCHER S. BOUSSAID O. «OLAP Visualization Operator for Complex Data». In proc. of *Forum of International Conference on Advanced Information Systems Engineering (CAISE Forum 2011)* p.139-146. Ref HAL: [halshs-00603841_v1](#).
- [Mahboubi-09] MAHBOUBI H. DARMONT J. «Indices in XML databases». *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (2009) p.674-681. DOI: [10.4018/978-1-60566-242-8.ch072](#).
- [Mahboubi-09a] MAHBOUBI H. RALAIVAO J.-C. LOUDCHER S. BOUSSAID O. BENTAYEB F. DARMONT J. «X-WACoDa: An XML-based approach for Warehousing and Analyzing Complex Data». *Data Warehousing*

Design and Advanced Engineering Applications: Methods for Complex Construction (2009) p.38-54. Ref HAL: [hal-00515436_v1](#).

- [Mahboubi-09b] MAHBOUBI H. HACHICHA M. DARMONT J. «XML Warehousing and OLAP». *Encyclopedia of Data Warehousing and Mining, Second Edition* (2009) p.2109-2116.
- [Mahboubi-10] MAHBOUBI H. DARMONT J. «Query Performance Optimization in XML Data Warehouses». *E-Strategies for Resource Management Systems: Planning and Implementation* (2010) p.22. DOI: [10.4018/978-1-61692-016-6.ch014](#).
- [Nguyen-12] Nguyen T.-V.-A. D'Orazio L. Bimonte S. DARMONT J. «Cost Models for View Materialization in the Cloud». In proc. of *Workshop on Data Analytics in the Cloud (EDBT-ICDT/DanaC 2012)*. Ref HAL: [hal-00686027_v2](#).
- [Selmane-12] SELMANE S. A. (Corresp.), BENTAYEB F. Missaoui R. BOUSSAID O. «Community detection and analysis in social networks». In proc. of *International Workshop on Web Intelligence & Communities (WI&C 2012)*.

Revue nationale

- [Boukraa-11] BOUKRAA D. BOUSSAID O. BENTAYEB F. Zegour D. «Modèle multidimensionnel d'objets complexes : Du modèle d'objets aux cubes d'objets complexes». In *Ingénierie des Systèmes d'Information* (2011) p.41-65. DOI: [10.3166/isi.16.6.41-65](#).
- [Favre-09a] FAVRE C. Rougié M. BENTAYEB F. BOUSSAID O. «Gestion et analyse personnalisées des demandes marketing. Cas de LCL-Le Crédit Lyonnais». In *Ingénierie des Systèmes d'Information*, vol. 14/3 (2009) p.119-139.
- [Pitarch-11] Pitarch Y. FAVRE C. Laurent A. Poncelet P. «Généralisation contextuelle de mesures dans les entrepôts de données - Application aux entrepôts de données médicales». In *Ingénierie des Systèmes d'Information*, vol. 16/6 (2011) p.67-90. Ref HAL: [lirmm-00798664_v1](#).
- [Triki-12] TRIKI S. (Corresp.), Ben-Abdallah H. Feki J. HARBI N. «Sécurisation des entrepôts de données contre les inférences précises et partielles». Accepted in *Ingénierie des Systèmes d'Information* (2012). Ref HAL: [hal-00634704_v1](#).

Conférences nationales avec comité de programme et actes

- [Aknouche-11] AKNOUCHE R. (Corresp.). «Entreposage et analyse des données complexes : cas des données texte». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 2011)* p.419-420. Ref HAL: [hal-00911859_v1](#).
- [Aknouche-12a] AKNOUCHE R. (Corresp.), BOUSSAID O. BENTAYEB F. «RICSH : Recherche d'information contextuelle par segmentation thématique de documents». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2012)* p.327-332. Ref HAL: [hal-00666588_v1](#). Rang C.
- [Aknouche-13b] AKNOUCHE R. (Corresp.), ASFARI O. (Corresp.), BENTAYEB F. BOUSSAID O. «ETL-Text: Extract-Transform-Load Processes for Textual Data Warehousing». In proc. of *Portuguese Conference on Artificial Intelligence (EPIA 2013)* p.308-319. Ref HAL: [hal-00911861_v1](#). Rang B.
- [Arres-13] ARRES B. (Corresp.), KABACHI N. BOUSSAID O. «Vers la construction de cube OLAP avec MapReduce dans un environnement Cloud Computing». In proc. of *Conférence Maghrébine sur les Avancées des Systèmes Décisionnels (ASD 2013)* p.13-24. Ref HAL: [hal-00907013_v1](#).
- [Arres-14] ARRES B. (Corresp.), KABACHI N. BENTAYEB F. BOUSSAID O. «Application du paradigme MapReduce aux données ouvertes. Cas : Accessibilité des personnes à mobilité réduite aux musées». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2014)* p.497-500. Ref HAL: [hal-00941510_v1](#). Rang C.
- [Attasena-13] ATTASENA V. HARBI N. DARMONT J. «Sharing-based Privacy and Availability of Cloud Data Warehouses». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2013)* p.17-32. Ref HAL: [hal-00836924_v1](#).
- [Bentayeb-09b] BENTAYEB F. BOUSSAID O. FAVRE C. Ravat F. Teste O. «Personnalisation dans les entrepôts de données : bilan et perspectives». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2009)* p.7-22.
- [Bouattour-09] BOUATTOUR S. BOUSSAID O. «Les entrepôts de données actifs : modélisation des règles d'analyse». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2009)* p.139-154.
- [Boukraa-10a] BOUKRAA D. BOUSSAID O. BENTAYEB F. «Opérateurs OLAP pour des cubes d'objets complexes : construction, visualisation et analyse». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)*.
- [Choquet-10] Choquet R. Qouiya S. Pasche E. Daniel C. BOUSSAID O. Jaulent M.-C. «Un modèle de connaissances pour mesurer la qualité d'une source d'information». In proc. of *Journées Francophones d'Ingénierie des Connaissances (IC 2010)*. Ref HAL: [hal-00487591_v1](#).
- [Dehdouh-13] DEHDOUH K. (Corresp.), BENTAYEB F. KABACHI N. «Performances de requêtes OLAP dans les bases de données en colonnes». In proc. of *Conférence Maghrébine sur les Avancées des Systèmes Décisionnels (ASD 2013)* p.439-444.

- [Dehdouh-14] DEHDOUH K. (Corresp.), BENTAYEB F. KABACHI N. BOUSSAID O. «Construction de cube OLAP à partir d'un entrepôt de données orienté colonnes». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2014)* p.525-528. Rang C.
- [Favre-11] FAVRE C. Laurent A. Pitarch Y. Poncelet P. «Représentation graphique des hiérarchies contextuelles : modèles avec satellites». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2011)* p.23-38. Ref HAL: [lirmm-00798665_v1](#).
- [Hachicha-10a] HACHICHA M. DARMONT J. «Modèles d'arbre pour XOLAP». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)* p.97-106. Ref HAL: [hal-00503966_v1](#).
- [Hachicha-13a] HACHICHA M. DARMONT J. «Problèmes d'additivité dus à la présence de hiérarchies complexes dans les modèles multidimensionnels : définitions, solutions et travaux futurs». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2013)* p.7-16. Ref HAL: [hal-00836927_v1](#).
- [Mabit-10] Mabit L. LOUDCHER S. BOUSSAID O. «Analyse en ligne d'objets complexes avec l'analyse factorielle». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2010)* p.381-386. Ref HAL: [halshs-00476570_v1](#). Rang C.
- [Madani-10] Madani A. BOUSSAID O. Abed H. «F-CheX : Une approche de fouille dans les documents XML». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)*.
- [Midouni-09] Midouni S. A. D. DARMONT J. BENTAYEB F. «Approche de modélisation multidimensionnelle des données complexes : Application aux données médicales». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2009)* p.155-166. Ref HAL: [hal-00411237_v1](#).
- [Oukid-13a] Oukid L. ASFARI O. (Corresp.), BENTAYEB F. BOUSSAID O. «Cube de textes et opérateur d'agrégation basé sur un modèle vectoriel adapté». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2013)*. Rang C.
- [Perriot-13] Perriot R. Pfeiffer J. D'Orazio L. Bachelet B. Bimonte S. DARMONT J. «Modèles de Coût pour la Sélection de Vues Matérialisées dans le Nuage, Application aux Services Amazon EC2 et S3». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2013)* p.53-68. Ref HAL: [hal-00836947_v1](#).
- [Pitarch-10a] Pitarch Y. FAVRE C. Laurent A. Poncelet P. «Analyse flexible dans les entrepôts de données : quand les contextes s'en mêlent». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)*.
- [Pitarch-12a] Pitarch Y. FAVRE C. Laurent A. Poncelet P. «Vers davantage de flexibilité et d'expressivité dans les hiérarchies contextuelles des entrepôts de données». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2012)* p.12. Ref HAL: [lirmm-00798260_v1](#). Rang C.
- [Salem-10a] SALEM R. DARMONT J. BOUSSAID O. «Toward Active XML Data Warehousing». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)* p.65-80. Ref HAL: [hal-00503963_v1](#).
- [Selmane-14a] SELMANE S. A. (Corresp.), BENTAYEB F. Missaoui R. BOUSSAID O. «GNOM-FCA: Une extension de la méthode de Falzon de détection de communautés.». In proc. of *Conférence en Recherche d'Information et Applications (CORIA 2014)* p.345-349. Rang C.
- [Selmane-14c] SELMANE S. A. (Corresp.), BENTAYEB F. BOUSSAID O. «Règles d'Association Triadiques pour la personnalisation de requêtes décisionnelles». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014)* 16 pages full paper.
- [Triki-10b] TRIKI S. Ben-Abdallah H. Feki J. HARBI N. «Sécurisation des entrepôts de données contre les inférences en utilisant les réseaux Bayésiens». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2010)*.

Autres publications nationales

- [Aknouche-12b] AKNOUCHE R. BENTAYEB F. «Prise en compte du contexte utilisateur pour améliorer la recherche d'information : Application au corpus 20 Newsgroups». In proc. of *Atelier Systèmes Décisionnels (ASD), Maghrebien Conference on Information Technologies (ASD-MCSEAI 2012)* p.71-82. Ref HAL: [hal-00713095_v1](#).
- [Bahri-12] BAHRI E. HARBI N. (Corresp.). «Vers une meilleure détection d'intrusions dans les SI avec les méthodes ensemblistes». In proc. of *séminaire de Veille Stratégique Scientifique et Technologique (VSST 2012)*.
- [Basileu-12] BASILEU C. KABACHI N. LAMURE M. «Structuration prétopologique d'un ensemble à partir d'une famille de graphes». In proc. of *séminaire de Veille Stratégique Scientifique et Technologique (VSST 2012)*. Ref HAL: [hal-00702876_v1](#).
- [BenHassine-09] BEN HASSINE S. (Corresp.), Clément D. «Règles d'association pour la qualité des données». In proc. of *Atelier Qualité des données et des connaissances EGC (QDC-EGC 2009)* p.33-39. Ref HAL: [hal-00519731_v1](#).
- [Bentayeb-11] BENTAYEB F. «Entrepôts et analyse en ligne de données complexes centrés utilisateur : un nouveau défi». Ref HAL: [tel-00752126_v1](#).

- [Darmont-13b] DARMONT J. Pedersen T. B. «Report on the First International Workshop on Cloud Intelligence (Cloud-I 2012)». Ref HAL: [hal-00936489_v1](#).
- [Favre-09b] FAVRE C. BENTAYEB F. BOUSSAID O. «Personnalisation collaborative pour l'enrichissement des analyses dans les entrepôts de données complexes». In proc. of *Atelier Fouille de Données Complexes EGC (FDC-EGC 2009)*.
- [Favre-13] FAVRE C. (Corresp.), BENTAYEB F. BOUSSAID O. DARMONT J. GAVIN G. HARBI N. KABACHI N. LOUDCHER S. «Les entrepôts de données pour les nuls... ou pas !». In proc. of *Atelier aide à la Décision à tous les Etages (AIDE-EGC 2013)*. Ref HAL: [hal-00783638_v1](#).
- [Grabova-14] GRABOVA O. (Corresp.), SOBATI MOGHADAM S. CHAGHERI S. DARMONT J. (Corresp.). «BI4people : le décisionnel pour tous». In Proc. of 10es journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014), démo. Ref HAL: [hal-00980915_v1](#).
- [Harbi-10a] HARBI N. BAHRI E. «Une approche basée sur l'agrégation pour une meilleure détection d'intrusions». In proc. of *Atelier Qualité des données et des connaissances EGC (QDC-EGC 2010)* p.45-46.
- [Karkouda-12] Karkouda K. HARBI N. DARMONT J. GAVIN G. «Confidentialité et disponibilité des données entreposées dans les nuages». Ref HAL: [hal-00667304_v1](#).
- [Karkouda-12a] Karkouda K. HARBI N. DARMONT J. GAVIN G. «Confidentialité et disponibilité des données entreposées dans les nuages». In proc. of *Atelier Fouille de Données Complexes (FDC-EGC 2012)*. Ref HAL: [hal-00667416_v1](#).
- [Loudcher-11] LOUDCHER S. «Vers l'OLAP sémantique pour l'analyse en ligne des données complexes». Ref HAL: [tel-00606847_v1](#).
- [Loudcher-13] LOUDCHER S. (Corresp.), FAVRE C. JAKAWAT W. «Que peut apporter l'OLAP à l'analyse de réseaux d'informations bibliographiques ?». In proc. of *Conférence sur les modèles et l'analyse des réseaux : approches mathématiques et informatiques (MARAMI 2013)*.
- [Selmane-14] SELMANE S. A. (Corresp.), BENTAYEB F. Missaoui R. BOUSSAID O. «Identification des communautés au sein des réseaux sociaux par l'Analyse Formelle de Concept.». In proc. of *Atelier Fouilles de Grands Graphes (FGG-EGC 2014)*.
- [Triki-09a] TRIKI S. Feki J. Ben-Abdallah H. HARBI N. «Sécurisation des entrepôts de données : Etat de l'art et proposition d'une architecture». In proc. of *Atelier Systèmes Décisionnels (ASD), Maghrebien Conference on Information Technologies (ASD-MCSEAI 2009)*.

Équipe DMD

Ouvrages ou direction d'ouvrages de portée internationale

- [Berka-09] Berka P. Rauch J. ZIGHED D. A. «Data Mining and Medical Knowledge Management: Cases and Applications». (2009). DOI: [10.4018/978-1-60566-218-3](#).
- [Briand-09] Briand H. Guillet F. Ritschard G. ZIGHED D. A. «Advances in Knowledge Discovery and Management». (2009). DOI: [10.1007/978-3-642-00580-0](#).
- [Guillet-10] Guillet F. Ritschard G. ZIGHED D. A. Briand H. «Advances in Knowledge Discovery and Management. Studies in Computational Intelligence». (2010). Ref HAL: [hal-00461286_v1](#).
- [Guillet-12] Guillet F. Ritschard G. ZIGHED D. A. «Advances in Knowledge Discovery and Management. Studies in Computational Intelligence - Volume 2, Vol. 398». (2012). DOI: [10.1007/978-3-642-25838-1](#).
- [Guillet-13] Guillet F. Pinaud B. Venturini G. ZIGHED D. A. «Advances in Knowledge Discovery and Management, Vol. 3». (2013). DOI: [10.1007/978-3-642-35855-5](#).
- [Lenca-09] Lenca P. LALLICH S. «First Workshop Quality Issues, Measures of Interestingness and Evaluation of data mining models (PAKDD-QIMIE'09), Bangkok, Thailand». (2009). Ref HAL: [hal-00608489_v1](#).
- [Lenca-09a] Lenca P. LALLICH S. «Quality issues, measures of interestingness and evaluation of data mining models». (2009). Ref HAL: [hal-00459789_v1](#).
- [Zighed-09] ZIGHED D. A. Tsumoto S. Ras Z. HACID H. «Mining Complex Data». (2009).

Revue internationale

- [Bourdes-10] Bourdès V. (Corresp.), BONNEVAY S. Lisboa P. Defrance R. Pérol D. Chabaud S. Bachelot T. Gargi T. Négrier S. «Comparison of Artificial Neural Network with Logistic Regression as Classification Models for Variable Selection for Prediction of Breast Cancer Patient Outcomes». In *Advances in Artificial Neural Systems*, vol. 2010 (2010). DOI: [10.1155/2010/309841](#).
- [Bourdes-11] Bourdès V. Ferrières J. Amar J. Amelineau E. BONNEVAY S. Berlion M. Danchin N. «Prediction of persistence of combined evidence-based cardiovascular medications in patients with acute coronary syndrome after hospital discharge using neural networks». In *Medical and Biological Engineering and Computing*, vol. 49 (2011) p.947-955. DOI: [10.1007/s11517-011-0785-4](#). Rang A.
- [Deperetti-10] Deperetti C. SIANI C. «Graphical methods for investigating the finite-sample properties of confidence regions». In *Computational Statistics & Data Analysis*, vol. 54/2 (2010) p.262-271. DOI: [10.1016/j.csda.2009.09.012](#). Rang A.

- [Dubois-14] Dubois D. Prade H. RICO A. «The logical encoding of Sugeno integrals». In *Fuzzy Sets and Systems*, vol. 241 (2014) p.61-75. DOI: [10.1016/j.fss.2013.12.014](https://doi.org/10.1016/j.fss.2013.12.014). Rang A.
- [Farrusseng-09] Farrusseng D. CLERC F. Mirodatos C. RAKOTOMALALA R. «Virtual screening of materials using neuro-genetic approach: concepts and implementation». In *Computational Materials Science*, vol. 45/1 (2009) p.52-59. DOI: [10.1016/j.commatsci.2008.03.060](https://doi.org/10.1016/j.commatsci.2008.03.060). Rang A.
- [Forestier-12] FORESTIER M. STAVRIANOU A. VELCIN J. ZIGHED D. A. «Roles in Social Networks: Methodologies and Research Issues». In *Web Intelligence and Agent Systems*, vol. 10/1 (2012) p.117-133. Rang C.
- [Lanagan-14] Lanagan J. (Corresp.), Anokhin N. VELCIN J. «Early Stage Conversation Catalysts on Entertainment-Based Web Forums». Accepted in *Social Network Analysis and Mining* (2014).
- [LeBras-10a] Le Bras Y. Lenca P. LALLICH S. «Mining interesting rules without support requirement: a general universal existential upward closure property». In *Annals of Information Systems*, vol. 8/Part 2 (2010) p.75-98. DOI: [10.1007/978-1-4419-1280-0](https://doi.org/10.1007/978-1-4419-1280-0).
- [LeBras-12a] Le Bras Y. LENCA P. LALLICH S. «Optimonotone measures for optimal rule discovery». In *Computational Intelligence*, vol. 28/4 (2012) p.475-504. DOI: [10.1111/j.1467-8640.2012.00422.x](https://doi.org/10.1111/j.1467-8640.2012.00422.x). Rang A.
- [Lenca-10] Lenca P. LALLICH S. Vaillant B. «Construction of an off-centered entropy for the supervised learning of imbalanced classes : some first results». In *Communications in Statistics - Theory and Methods*, vol. 39/3 (2010) p.493-507. Ref HAL: [hal-00445166_v1](https://hal.archives-ouvertes.fr/hal-00445166_v1). Rang B.
- [Lust-13] LUST T. ROLLAND A. (Corresp.). «Choquet optimal set in biobjective combinatorial optimization». In *Computers and Operations Research*, vol. 40/10 (2013) 2260-2269. DOI: [10.1016/j.cor.2013.04.003](https://doi.org/10.1016/j.cor.2013.04.003). Rang A.
- [Marino-10] Marino P. SIANI C. Roché H. Protière C. Fumoleau P. Spielmann M. Martin A.-L. Viens P. Le corroller soriano A.-G. «Cost-effectiveness of adjuvant docetaxel for node-positive breast cancer patients: results of the PACS 01 economic study.». In *Annals of Oncology*, vol. 21/7 (2010) p.1448-54. DOI: [10.1093/annonc/mdp561](https://doi.org/10.1093/annonc/mdp561). Rang A.
- [MendezHernandez-09] Méndez-Hernández P. Flores Y. SIANI C. LAMURE M. Dosamantes-carrasco L. D. Halley-Castillo E. Huitrón G. Talavera J. O. Gallegos-Carrillo K. Salmerón J. «Physical activity and risk of metabolic syndrome in an urban Mexican cohort.». In *BMC public health*, vol. 9 (2009) p.276. DOI: [10.1186/1471-2458-9-276](https://doi.org/10.1186/1471-2458-9-276). Rang B.
- [Metzger-12] Metzger M.-H. (Corresp.), Durand T. LALLICH S. Salamon R. Castets P. «The use of regional platforms for managing electronic health records for the production of regional public health indicators in France.». In *BMC medical informatics and decision making*, vol. 12/1 (2012) p.28. DOI: [10.1186/1472-6947-12-28](https://doi.org/10.1186/1472-6947-12-28). Rang B.
- [Perrier-12] Perrier L. (Corresp.), Buja A. Mastrangelo G. Vecchiato A. Sandonà P. Ducimetièrre F. Blay J.-Y. Gilly F. SIANI C. Biron P. Ranchère-Vince D. Decouvelaere A.-V. Thiesse P. Bergeron C. Dei Tos A. Coindre J.-M. Rossi C. Ray-Coquard I. «Clinicians' adherence versus non adherence to practice guidelines in the management of patients with sarcoma: a cost-effectiveness assessment in two European regions.». In *BMC Health Services Research*, vol. 12/1 (2012) p.82. DOI: [10.1186/1472-6963-12-82](https://doi.org/10.1186/1472-6963-12-82). Rang B.
- [Petrovic-09] Petrovic S. Bašić B. D. Morin A. Zupan B. CHAUCHAT J.-H. «Textual features for corpus visualization using correspondence analysis». In *Intelligent Data Analysis*, vol. 13/5 (2009) p.795-813. DOI: [10.3233/IDA-2009-0393](https://doi.org/10.3233/IDA-2009-0393). Rang C.
- [Rico-10] RICO A. (Corresp.), Strauss O. «Imprecise expectations for imprecise linear filtering». In *International Journal of Approximate Reasoning* (2010) p.16. DOI: [10.1016/j.ijar.2010.06.003](https://doi.org/10.1016/j.ijar.2010.06.003). Rang A.
- [Rizoiu-13a] RIZOIU M.-A. (Corresp.), VELCIN J. LALLICH S. «Unsupervised Feature Construction for Improving Data Representation and Semantics». In *Journal of Intelligent Information Systems*, vol. 40/3 (2013) p.501-527. DOI: [10.1007/s10844-013-0235-x](https://doi.org/10.1007/s10844-013-0235-x). Rang B.
- [Rizoiu-14] RIZOIU M.-A. (Corresp.), VELCIN J. LALLICH S. «How to use Temporal-Driven Constrained Clustering to detect typical evolutions». Accepted in *International Journal on Artificial Intelligence Tools* (2014). Rang C.
- [Rizoiu-14a] RIZOIU M.-A. (Corresp.), VELCIN J. LALLICH S. «Semantic-enriched Visual Vocabulary Construction in a Weakly Supervised Context». Accepted in *Intelligent Data Analysis*, vol. 19/1 (2014). Ref HAL: [hal-00959355_v1](https://hal.archives-ouvertes.fr/hal-00959355_v1). Rang C.
- [Rolland-12b] ROLLAND A. (Corresp.), Kasparian J. «OECD's "Better life index": can any country be well ranked?». In *Journal of Applied Statistics*, vol. 39/10 (2012) p.2223-2230. DOI: [10.1080/02664763.2012.706265](https://doi.org/10.1080/02664763.2012.706265). Rang C.
- [Rolland-13a] ROLLAND A. «Reference-based Preferences Aggregation Procedures in Multicriteria Decision Making». In *European Journal of Operational Research*, vol. 225/3 (2013) p.479-486. DOI: [10.1016/j.ejor.2012.10.013](https://doi.org/10.1016/j.ejor.2012.10.013). Rang A.
- [Šilic-12] Šilic A. (Corresp.), Morin A. CHAUCHAT J.-H. Bašić B. D. «Visualization of temporal text collections based on Correspondence Analysis». In *Expert Systems with Applications*, vol. 39/15 (2012). DOI: [10.1016/j.eswa.2012.04.040](https://doi.org/10.1016/j.eswa.2012.04.040). Rang A.

- [Strauss-12a] Strauss O. RICO A. «Towards interval-based non-additive deconvolution in signal processing». In *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 16/5 (2012) p.809-820. DOI: [10.1007/s00500-011-0771-7](https://doi.org/10.1007/s00500-011-0771-7). Rang C.
- [Wei-09] WEI Z., Miao D. CHAUCHAT J.-H., Zhao R. Li W. «N-grams based feature selection and text representation for Chinese Text Classification». In *International Journal of Computational Intelligence Systems (IJCIS)*, vol. 2/4 (2009) p.365-374. DOI: [10.2991/ijcis.2009.2.4.5](https://doi.org/10.2991/ijcis.2009.2.4.5).

Conférences internationales avec comité de programme et actes

- [Abdesselam-12] ABDESSELAM R. ZIGHED D. A. «Statistical comparisons for topological equivalence of proximity measures». In proc. of *International Conference on Stochastic Modeling Techniques and Data Analysis (SMTDA 2012)*.
- [AhPine-13] AH-PINE J. (Corresp.), Mayag B. ROLLAND A. «Identification of a 2-Additive Bi-Capacity by Using Mathematical Programming». In proc. of *Algorithmic Decision Theory* p.15-29. DOI: [10.1007/978-3-642-41575-3_2](https://doi.org/10.1007/978-3-642-41575-3_2). Rang C.
- [AhPine-13a] AH-PINE J. «A General Framework for Comparing Heterogeneous Binary Relations.». In proc. of *Geometric Science of Information (GSI 2013)* p.188-195. DOI: [10.1007/978-3-642-40020-9_19](https://doi.org/10.1007/978-3-642-40020-9_19).
- [AhPine-13b] AH-PINE J. «Graph Clustering by Maximizing Statistical Association Measures.». In proc. of *International Symposium on Intelligent Data Analysis (IDA 2013)* p.56-67. DOI: [10.1007/978-3-642-41398-8_6](https://doi.org/10.1007/978-3-642-41398-8_6). Rang B.
- [Bahri-09b] BAHRI E. LALLICH S. «FCP-Growth to extract itemsets for class association rules». In proc. of *International Conference Florida Artificial Intelligence Research Society (FLAIRS 2009)*. Rang B.
- [Bahri-09e] BAHRI E. LALLICH S. «Pruning for Extracting Class Association Rules without Candidate». In proc. of *International Conference on Data Mining (DMIN 2009)* p.11-17.
- [Chauchat-11] CHAUCHAT J.-H. (Corresp.), Zolotariova I. «Competence Orientation in Business Informatics International Master Program». In proc. of *International Conference on Perspective in Business Informatics Research (BIR 2011)* p.99-103. DOI: [10.2478/v10143-011-0014-5](https://doi.org/10.2478/v10143-011-0014-5). Rang C.
- [Clinchant-11] Clinchant S. AH-PINE J. Csurka G. «Semantic Combination of Textual and Visual Information in Multimedia Retrieval». In proc. of *International Conference on Multimedia Retrieval (ICMR 2011)* p.44. DOI: [10.1145/1991996.1992040](https://doi.org/10.1145/1991996.1992040). Rang A.
- [Deperetti-09] Deperetti C. SIANI C. Cerrato M. «An Artificial Neural Network Based Heterogeneous Panel Unit Root Test». In proc. of *International Joint Conference on Neural Networks (IJCNN 2009)* p.2487-2493. DOI: [10.1109/IJCNN.2009.5178885](https://doi.org/10.1109/IJCNN.2009.5178885). Rang A.
- [Deperetti-09a] Deperetti C. SIANI C. Cerrato M. «A Bootstrap Artificial Neural Network Based Heterogeneous Panel Unit Root Test in Case of Cross Sectional Independence». In proc. of *International Conference on Neural Information Processing (ICONIP 2009)* p.441-450. DOI: [10.1007/978-3-642-10677-4_50](https://doi.org/10.1007/978-3-642-10677-4_50). Rang A.
- [Dubois-12] Dubois D. Prade H. RICO A. «Qualitative integrals and desintegrals - Towards a logical view». In proc. of *International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2012)* p.127-138. DOI: [10.1007/978-3-642-34620-0_13](https://doi.org/10.1007/978-3-642-34620-0_13). Rang B.
- [Dubois-12c] Dubois D. Prade H. RICO A. «Qualitative Integrals and Desintegrals: How to handle positive and negative scales in evaluation». In proc. of *Information Processing and Management of Uncertainty (IPMU 2012)* p.306-316. DOI: [10.1007/978-3-642-31718-7_32](https://doi.org/10.1007/978-3-642-31718-7_32). Rang B.
- [Dubois-13b] Dubois D. Prade H. RICO A. «Qualitative desintegrals as lower and upper possibilistic expectations». In proc. of *conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2013)*. DOI: [10.2991/eusflat.2013.51](https://doi.org/10.2991/eusflat.2013.51).
- [Dubois-13c] Dubois D. Prade H. RICO A. «Qualitative capacities as imprecise possibilities». In proc. of *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2013)*. DOI: [10.1007/978-3-642-39091-3_15](https://doi.org/10.1007/978-3-642-39091-3_15). Rang B.
- [Forestier-11] FORESTIER M. VELCIN J. ZIGHED D. A. «Extracting Social Networks to Understand Interaction». In proc. of *International Conference on Advances in Social Network Analysis and Mining (ASONAM 2011)* p.213-219. DOI: [10.1109/ASONAM.2011.64](https://doi.org/10.1109/ASONAM.2011.64). Rang C.
- [Forestier-11a] FORESTIER M. VELCIN J. ZIGHED D. A. «Extracting Social Networks Enriched by Using Text». In proc. of *International Symposium on Methodologies for Intelligent Systems (ISMIS 2011)* p.140-145. Rang B.
- [Forestier-12a] FORESTIER M. VELCIN J. ZIGHED D. A. «Analyzing Social Roles using Enriched Social Network on On-Line Sub-Communities.». In proc. of *International Conference on Digital Society (ICDS 2012)* p.17-22. Rang C.
- [Forestier-12b] FORESTIER M. (Corresp.), VELCIN J. STAVRIANOU A. ZIGHED D. A. «Extracting Celebrities from Online Discussions». In proc. of *International Conference on Advances in Social Network Analysis and Mining (ASONAM 2012)* p.322-326. Rang C.
- [Imoussaten-11] Imoussaten A. Montmain J. RICO A. RICO F. «A dynamical model for simulating a debate outcome». In proc. of *International Conference on Agents and Artificial Intelligence (ICAART 2011)* p.1. Ref HAL: [hal-00808629_v1](https://hal.archives-ouvertes.fr/hal-00808629_v1). Rang C.

- [Ioan-12] Ioan C.-A. (Corresp.), VELCIN J., Trausan-matu S. «Tagging Choreographic Data for Data Mining and Classification». In proc. of *IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2012)*. Rang A.
- [Kafafy-11] KAFAFY A., BOUNEKKAR A., BONNEVAY S. «A Hybrid Evolutionary Metaheuristics (HEMH) applied on 0/1 Multiobjective Knapsack Problems». In proc. of *Genetic and Evolutionary Computation Conference (GECCO 2011)* p.497-504. Ref HAL: [hal-00624096_v1](#). Rang A.
- [Kafafy-12] KAFAFY A. (Corresp.), BOUNEKKAR A., BONNEVAY S. «HEMH2: An Improved Hybrid Evolutionary Metaheuristics for 0/1 Multiobjective Knapsack Problems.». In proc. of *International Conference on Simulated Evolution And Learning (SEAL 2012)* p.104-116. DOI: [10.1007/978-3-642-34859-4_11](#). Rang B.
- [Kafafy-12a] KAFAFY A. (Corresp.), BOUNEKKAR A., BONNEVAY S. «Hybrid Metaheuristics based on MOEA/D for 0/1 Multiobjective Knapsack Problems : A comparative Study». In proc. of *IEEE Congress on Evolutionary Computation (IEEE CEC 2012)* p.3616-3623. DOI: [10.1109/CEC.2012.6253015](#). Rang A.
- [Kafafy-13] KAFAFY A. (Corresp.), BONNEVAY S., BOUNEKKAR A. «A Hybrid Evolutionary Approach with Search Strategy Adaptation for Multiobjective Optimization». In proc. of *Genetic and Evolutionary Computation Conference (GECCO 2013)* p.631-638. DOI: [10.1145/2463372.2463458](#). Rang A.
- [LeBras-09a] Le Bras Y. Lenca P. Moga S. LALLICH S. «All-monotony: a generalization of the all-confidence antimonotony». In proc. of *International Conference on Machine Learning and Applications (ICMLA 2009)* p.759-764. Ref HAL: [hal-00445030_v1](#). Rang B.
- [LeBras-09b] Le_bras Y. Lenca P. LALLICH S. «On optimal rule mining: a framework and a necessary and sufficient condition of existence». In proc. of *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2009)* p.705-712. DOI: [10.1007/978-3-642-01307-2_71](#). Rang A.
- [LeBras-10b] Le Bras Y. Meyer P. LALLICH S., Lenca P. «A robustness measure of association rules». In proc. of *European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2010)* p.227-242. DOI: [10.1007/978-3-642-15883-4](#). Rang A.
- [LeBras-11a] Le Bras Y. Lenca P. LALLICH S. «Mining classification rules without support: an anti-monotone property of jaccard measure». In proc. of *International Conference Discovery Science (DS 2011)* p.179-193. DOI: [10.1007/978-3-642-24477-3](#). Rang B.
- [Mayag-12] Mayag B. ROLLAND A., AH-PINE J. «Elicitation of a 2-additive bi-capacity through cardinal information on trinary actions». In proc. of *Information Processing and Management of Uncertainty (IPMU 2012)* p.238-247. DOI: [10.1007/978-3-642-31724-8_25](#). Rang B.
- [Mhamdi-09] Mhamdi F. RAKOTOMALALA R., Elloumi M. «A hierarchical n-Grams Extraction Approach for Classification Problem». In proc. of *International Conference on Signal-Image Technology and Internet-Based Systems (SITIS 2009)* p.211-222. DOI: [10.1007/978-3-642-01350-8_20](#).
- [Miguel-12] Miguel C. Dubois D. Prade H. RICO A., Tamas W. «General interpolation by polynomial functions of distributive lattices». In proc. of *Information Processing and Management of Uncertainty (IPMU 2012)* p.347-355. DOI: [10.1007/978-3-642-31718-7_36](#). Rang B.
- [Muhlenbach-09] Muhlenbach F. LALLICH S. «A New Clustering Algorithm Based on Regions of Influence with Self-Detection of the Best Number of Clusters». In proc. of *IEEE International Conference on Data Mining (ICDM 2009)* p.884-888. Ref HAL: [hal-00446155_v1](#). Rang A.
- [Muhlenbach-10] Muhlenbach F. LALLICH S. «Discovering Research Communities by Clustering Bibliographical Data». In proc. of *IEEE International Conference on Web Intelligence (WI-IAT 2010)* p.500-507. Ref HAL: [hal-00516610_v1](#). Rang B.
- [Musat-11] Musat C. VELCIN J. (Corresp.), Trausan-Matu S. RIZOIU M.-A. (Corresp.). «Improving Topic Evaluation Using Conceptual Knowledge.». In proc. of *International Joint Conference on Artificial Intelligence (IJCAI 2011)* p.1866-1871. DOI: [10.5591/978-1-57735-516-8/ijcai11-312](#). Rang A.
- [Musat-11a] Musat C. VELCIN J., RIZOIU M.-A., Trausan-matu S. «Concept-based Topic Model Improvement». In proc. of *International Symposium on Methodologies for Intelligent Systems (ISMIS 2011)* p.100. Ref HAL: [hal-00616247_v2](#). Rang B.
- [Orobinska-13] OROBINSKA O. (Corresp.), CHAUCHAT J.-H. (Corresp.), Charonova N. «Enrichissement d'une ontologie de domaine par extension des relations taxonomiques». In proc. of *International Conference on Terminology and Artificial Intelligence (TIA 2013)* p.129-136.
- [Pisetta-09b] PISETTA V., ZIGHED D. A. «Similarity and Kernel Matrix Evaluation based on Spatial Autocorrelation Analysis». In proc. of *IEEE International Conference on Intelligent Systems (IS 2009)* p.422-430. DOI: [10.1007/978-3-642-04125-9_45](#). Rang C.
- [Pisetta-10] PISETTA V. (Corresp.), Jouve P.-E. ZIGHED D. A. «Learning with Ensembles of Randomized Trees : New Insights». In proc. of *European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2010)* p.67-82. DOI: [10.1007/978-3-642-15939-8_5](#). Rang A.
- [Prade-10] Prade H. RICO A. «Describing objects by means of Sugeno integrals». In proc. of *International Conference on Soft Computing and Pattern Recognition (SoCPaR 2010)* p.6-11. DOI: [10.1109/SOCPAR.2010.5685870](#).
- [Prade-11a] Prade H. RICO A. «Possibilistic Evidence». In proc. of *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011)* p.713-722. Rang B.

- [Qureshi-09] QURESHI T. (Corresp.), ZIGHED D. A. «Using Resampling Techniques for Better Quality Discretization». In proc. of *International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2009)* p.68-81. Ref HAL: [hal-00923582_v1](#). Rang B.
- [Qureshi-09a] QURESHI T., ZIGHED D. A. «A Soft Discretization Technique for Fuzzy Decision Trees Using Resampling». In proc. of *Intelligent Data Engineering and Automated Learning (IDEAL 2009)* p.586-593. DOI: [10.1007/978-3-642-04394-9_71](#). Rang B.
- [Qureshi-09b] QURESHI T., ZIGHED D. A. «A Decision Boundary based Discretization Technique using Resampling». In proc. of *Conference of the World Academy of Science, Engineering and Technology (WASET 2009)* p.820-825.
- [Rico-12] RICO A. (Corresp.), Sugeno M. «Eliciting CPTS-integrals on bipolar scale». In proc. of *Information Processing and Management of Uncertainty (IPMU 2012)* p.268-277. DOI: [10.1007/978-3-642-31724-8_28](#). Rang B.
- [Rico-12a] RICO F. (Corresp.), ZIGHED D. A., EZZEDDINE D. «Neighborhood Random Classification». In proc. of *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2012)* p.98-108. DOI: [10.1007/978-3-642-30217-6_9](#). Rang A.
- [Rizoiu-12] RIZOIU M.-A. (Corresp.), VELCIN J., LALLICH S. «Structuring typical evolutions using Temporal-Driven Constrained Clustering». In proc. of *IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2012)* p.610-617. Ref HAL: [hal-00773398_v1](#). Rang A.
- [Stavrianou-09b] STAVRIANOU A., VELCIN J., CHAUCHAT J.-H. «Definition and Measures of an Opinion Model for Mining Forums». In proc. of *International Conference on Advances in Social Network Analysis and Mining (ASONAM 2009)* p.188-193. DOI: [10.1109/ASONAM.2009.34](#). Rang C.
- [Strauss-09a] Strauss O. Lahrech A. RICO A., Mariano-Goulart D. Telle B. «NIBART : A New Interval Based Algebraic Reconstruction Technique for error quantification of emission tomography images». In proc. of *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2009)* 148-155. Rang A.
- [Strauss-11a] Strauss O. RICO A. «Imprecise linear filtering: a second step». In proc. of *conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2011)* p.457-463. Ref HAL: [hal-00623829_v1](#).
- [Strauss-12] Strauss O. RICO A. «Towards a robust Imprecise Linear Deconvolution». In proc. of *Soft Methods in Probability and Statistics (SMPS 2012)* p.55-62.
- [Thibault-09] Thibault G. Aussem A. BONNEVAY S. «Incremental Bayesian Network Learning for Scalable Feature Selection». In proc. of *International Symposium on Intelligent Data Analysis (IDA 2009)* p.202-212. Rang B.
- [Umek-09] Umek L. Zupan B. Toplak M. Morin A. CHAUCHAT J.-H., Makovec G. Smrke D. «Subgroup Discovery in Data Sets with Multi-Dimensional Responses: A Method and a Case Study in Traumatology». In proc. of *Conference on Artificial Intelligence in Medicine (AIME 2009)* p.265-274. DOI: [10.1007/978-3-642-02976-9_39](#). Rang B.
- [Velcin-14] VELCIN J. (Corresp.), KIM Y.-M., Brun C. Dormagen J.-Y. Sanjuan E. Khouas L. Peradotto A. BONNEVAY S., Roux C. Boyadjian J. Molina A. Neihouser M. «Investigating the Image of Entities in Social Media: Dataset Design and First Results». In proc. of *Language Resources and Evaluation Conference (LREC 2014)*. Rang A.
- [Zighed-11] ZIGHED D. A., ABDESSELAM R. «Measuring the resemblance between proximity measures in a topological structure». In proc. of *International Conference on Applied Stochastic Models and Data Analysis (ASMDA 2011)*.
- [Zighed-12] ZIGHED D. A., ABDESSELAM R., Hadgu A. «Topological comparisons of proximity measures». In proc. of *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2012)* p.379-391. DOI: [10.1007/978-3-642-30217-6_32](#). Rang A.

Autres publications internationales

- [Abdeslam-10] ABDESSELAM R. Discriminant Analysis on Mixed Predictors. In *Studies in Classification, Data Analysis and Knowledge Organisation* (2010).
- [Anokhin-12] Anokhin N. Lanagan J. VELCIN J. «Social Citation: Finding Roles in Social Networks. An Analysis of TV-Series Web Forums». In proc. of *International Workshop on Mining Communities and People Recommenders (COMMPER 2012)*.
- [Bahri-09a] BAHRI E., LALLICH S., NICOLOYANNIS N., Maddouri M. «A hybrid approach of boosting against noisy data, Mining Complex Data». *Mining Complex Data* (2009) p.41-54. Ref HAL: [hal-00678867_v1](#).
- [Chemin-11] Chemin M. RICO A., Prade H. «Decomposition of possibilistic belief functions into simple support functions». In proc. of *Workshop on Fuzzy Methods for Knowledge-Based Systems (EUROFUSE 2011)*.
- [Do-10] Do T. N. Lenca P. LALLICH S., Pham N.-K. «Classifying very-high-dimensional data with random forests of oblique decision trees». *Advances in knowledge discovery and management* (2010) p.39-55. DOI: [10.1007/978-3-642-00580-0_3](#).
- [Dubois-11] Dubois D. Prade H. RICO A. «A possibilistic logic view of Sugeno integrals». In proc. of *Workshop on Fuzzy Methods for Knowledge-Based Systems (EUROFUSE 2011)*.

- [Dubois-13a] Dubois D. Prade H. RICO A. «Qualitative capacities as a basis for imprecise possibility theory». In proc. of *34th Linz Seminar on Fuzzy Set Theory (LINZ 2013)*. Rang C.
- [Elsayed-09] ELSAIED A. HACID H. ZIGHED D. A. «Exploring Validity Indices for Clustering Textual Data». *Mining Complex Data* (2009) p.281-300. DOI: [10.1007/978-3-540-88067-7_16](https://doi.org/10.1007/978-3-540-88067-7_16).
- [Guille-12a] GUILLE A. (Corresp.), HACID H. «A Predictive Model for the Temporal Dynamics of Information Diffusion in Online Social Networks». In proc. of *International Workshop on Mining Social Network Dynamics (MSND 2012)* p.1145. DOI: [10.1145/2187980.2188254](https://doi.org/10.1145/2187980.2188254).
- [Guille-13] GUILLE A. (Corresp.). «Information Diffusion in Online Social Networks». In proc. of *SIGMOD/PODS 2013 PhD Symposium (SIGMOD/PODS 2013)* p.31-36. DOI: [10.1145/2483574.2483575](https://doi.org/10.1145/2483574.2483575).
- [Hacid-09] HACID H. ZIGHED D. A. «A Machine learning-Based Model for Content-Based Image Retrieval». *Artificial Intelligence for Maximizing Content Based Image Retrieval* (2009) p.171-191. DOI: [10.4018/978-1-60566-174-2.ch008](https://doi.org/10.4018/978-1-60566-174-2.ch008).
- [Imoussaten-13] Imoussaten A. Montmain J. RICO A. RICO F. «A time-varying model to simulate a collective decisional problem.». *Communications in Computer and Information Science* (2013) p.56-71. Ref HAL: [hal-00808600_v1](https://hal.archives-ouvertes.fr/hal-00808600_v1). Rang C.
- [LeBras-12] Le Bras Y. LENCA P. LALLICH S. «Formal framework for the study of algorithmic properties of objective interestingness measures». *Data Mining: Foundations and Intelligent Paradigms* (2012) p.77-98. Ref HAL: [hal-00724795_v1](https://hal.archives-ouvertes.fr/hal-00724795_v1).
- [Pisetta-09a] PISETTA V. ZIGHED D. A. «Kernel Construction via Graph Adaptation». In proc. of *Workshop on Topological Learning*.
- [Rakotomalala-09] RAKOTOMALALA R. Mhamdi F. «Using the Text Categorization Framework for the Protein Classification». *Handbook of Research on Text and Web Mining Technologies* (2009) p.297-311. DOI: [10.4018/978-1-59904-990-8.ch008](https://doi.org/10.4018/978-1-59904-990-8.ch008), ISBN13: [9781599049908](https://doi.org/10.4018/978-1-59904-990-8).
- [Rizoiu-11] RIZOIU M.-A. (Corresp.), VELCIN J. «Topic Extraction for Ontology Learning». *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances* (2011) p.38-61. DOI: [10.4018/978-1-60960-625-1.ch003](https://doi.org/10.4018/978-1-60960-625-1.ch003).
- [Rizoiu-13] RIZOIU M.-A. (Corresp.). «Semi-Supervised Structuring of Complex Data». In proc. of *Doctoral Consortium of IJCAI '13(DC-IJCAI 2013)* p.3239-3240. Ref HAL: [hal-00866984_v1](https://hal.archives-ouvertes.fr/hal-00866984_v1).
- [Rolland-12a] ROLLAND A. RICO F. «Using set functions for multiple classifiers combination». In proc. of *Workshop From Multiple Criteria Decision Aid to Preference Learning (DA2PL 2012)* p.57-62.
- [Stavrianou-09a] STAVRIANOU A. CHAUCHAT J.-H. VELCIN J. «A Content-Oriented Framework for Online Discussion Analysis». In proc. of *IEEE International Symposium on Mining and Web Workshop of AINA (IEEE MAW 2009)* p.721-726.
- [Stavrianou-10a] STAVRIANOU A. VELCIN J. CHAUCHAT J.-H. «PROG: A Complementary Model to the Social Networks for Mining Forums». *From Sociology to Computing in Social Networks* (2010) p.59-79.
- [VanErven-14] Van Erven T., CUGLIARI J. «Making Regional Forecast add up». In *Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension* (2014). Ref HAL: [hal-00943529_v1](https://hal.archives-ouvertes.fr/hal-00943529_v1).
- [Zighed-10] ZIGHED D. A. Ritschard G. MARCELLIN S. «Asymmetric and sample size sensitive entropy measures for supervised learning». *Studies in Computational Intelligence* (2010) p.27-42. DOI: [10.1007/978-3-642-05183-8_2](https://doi.org/10.1007/978-3-642-05183-8_2).
- [Zolotaryova-09] Zolotaryova I. CHAUCHAT J.-H. «Joint Master Program 'Business Intelligence & Statistics' of Kharkiv National University of Economics (Ukraine) and Université Lumière Lyon2 (France)». In proc. of *Symposium on Business Informatics in Central and Eastern Europe. Internationalen Tagung Wirtschaftsinformatik (WI)* p.57-64.

Ouvrages ou direction d'ouvrages de portée nationale

- [Bounekkar-09] BOUNEKKAR A. Duru G. «Les nouvelles organisations des systèmes de santé : Nouvelles technologies de l'information, évaluation et financement». (2009).

Reuves nationales

- [Amor-09] Amor S. B. BONNEVAY S. Bui M. LAMURE M. «A Stochastic and Pretopological Modeling Aerial Pollution of an Urban Area». In *Santé Décision Management*, vol. 7/3 (2009) p.410-426.
- [Bonnevay-09] BONNEVAY S. «Pretopological operators for gray-level image analysis». In *Studia Informatica Universalis*, vol. 7.1 (2009) p.27-44. Ref HAL: [hal-00515396_v1](https://hal.archives-ouvertes.fr/hal-00515396_v1).
- [LeBras-12b] Le Bras Y. Meyer P. Lenca P. LALLICH S. «Mesure formelle de la robustesse des règles d'association». In *Revue des Nouvelles Technologies de l'Information* (2012). Ref HAL: [hal-00609515_v1](https://hal.archives-ouvertes.fr/hal-00609515_v1).
- [Musat-10] Musat C. (Corresp.), RIZOIU M.-A. Trausan-Matu S. «An Intra and Inter-Topic Evaluation and Cleansing Method». In *Romanian Journal of Human-Computer Interaction*, vol. 3/2 (2010) 81 - 96. Ref HAL: [hal-00552105_v1](https://hal.archives-ouvertes.fr/hal-00552105_v1).
- [Prudhomme-09] PRUDHOMME E. LALLICH S. «Représentation des données par un comité de cartes auto-organisatrices : une application aux données bruitées». In *Revue des Nouvelles Technologies de l'Information* (2009) p.177-202.

- [Stavrianou-09] STAVRIANOU A. VELCIN J. CHAUCHAT J.-H. «A combination of opinion mining and social network techniques for discussion analysis». In *Revue des Nouvelles Technologies de l'Information* (2009).

Conférences nationales avec comité de programme et actes

- [Bahri-09c] BAHRI E. LALLICH S. «FCP-Growth, une adaptation de FP-Growth pour générer des règles d'association de classe». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2009)* p.461-462. Rang C.
- [Bahri-09d] BAHRI E. LALLICH S. «Pour une classification associative plus efficace». In proc. of *Conférence francophone sur l'apprentissage artificiel (CAP 2009)*.
- [Bahri-10] BAHRI E. LALLICH S. «Proposition d'une méthode de classification associative adaptative». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2010)* p.501-512. Rang C.
- [Do-09] Do T.-N. LALLICH S. Pham N.-K. Lenca P. «Un nouvel algorithme de forêts aléatoires d'arbres obliques particulièrement adapté à la classification de données en grandes dimensions». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2009)*. Rang C.
- [Do-10a] Do T. N. Lenca P. LALLICH S. «Enhancing network intrusion classification through the Kolmogorov-Smirnov splitting criterion». In proc. of *International Conference on Theories and Applications of Computer Science (ICTACS 2010)* p.50-61. Ref HAL: [hal-00523096_v1](https://hal.archives-ouvertes.fr/hal-00523096_v1).
- [Dubois-12b] Dubois D. Prade H. RICO A. «Intégrales et désintégrales qualitatives». In proc. of *Logique Floue et ses Applications (LFA 2012)* p.24-33.
- [Imoussaten-10] Imoussaten A. Montmain J. RICO A. (Corresp.), RICO F. «Un modèle dynamique pour la simulation de l'issue d'un débat,». In proc. of *Logique Floue et ses Applications (LFA 2010)*.
- [Orobinska-14] OROBINSKA O. (Corresp.), CHAUCHAT J.-H. Charonova N. «Application des ressources linguistiques à grande échelle pour le peuplement d'une ontologie de domaine». In proc. of *Terminologie & Ontologie : Théories et Applications (TOTh 2014)* p.001.
- [Pisetta-10a] PISETTA V. Jouve P.-E. ZIGHED D. A. «Construction de noyaux pour l'apprentissage supervisé à partir d'arbres aléatoires». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2010)* p.525-536. Ref HAL: [hal-00519610_v1](https://hal.archives-ouvertes.fr/hal-00519610_v1). Rang C.
- [Rico-11a] RICO F. (Corresp.), ZIGHED D. A. «Classificateurs aléatoires topologiques à base de graphes de voisinages.». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2011)* p.83-88. Ref HAL: [hal-00625686_v1](https://hal.archives-ouvertes.fr/hal-00625686_v1). Rang C.
- [Rizoiu-10] RIZOIU M.-A. (Corresp.), VELCIN J. CHAUCHAT J.-H. «Regrouper les données textuelles et nommer les groupes à l'aide des classes recouvrantes». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2010)* p.561-572. Ref HAL: [hal-00516045_v1](https://hal.archives-ouvertes.fr/hal-00516045_v1). Rang C.
- [Rolland-12d] ROLLAND A. (Corresp.), Kasparian J. «Etude critique du palmarès de la qualité de vie dans les pays de l'OCDE». In proc. of *Journées de Statistique (JDS 2012)*. Ref HAL: [hal-00700502_v1](https://hal.archives-ouvertes.fr/hal-00700502_v1).
- [Rolland-12e] ROLLAND A. (Corresp.), Chanvillard H. «Un outil d'accompagnement des missions pour une licence professionnelle "chargé d'études statistiques" en alternance». In proc. of *Journées de Statistique (JDS 2012)*.
- [Rolland-13b] ROLLAND A. (Corresp.), Kasparian J. «Variations autour du "palmarès des villes étudiantes" du magazine l'Etudiant». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2013)* p.43-58. Ref HAL: [hal-00815006_v1](https://hal.archives-ouvertes.fr/hal-00815006_v1). Rang C.
- [Rolland-14] ROLLAND A. «Un indicateur pour mesurer la dépendance aux paramètres des palmarès». In proc. of *Journées de la statistique*.
- [Zighed-11a] ZIGHED D. A. ABDESSELAM R. BOUNEKKAR A. «Equivalence topologique entre mesures de proximité». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2011)*. Rang C.

Autres publications nationales

- [Bahri-09] BAHRI E. LALLICH S. «Introduction de l'élagage pour l'extraction des règles d'association de classe sans génération de candidats». In proc. of *Atelier Qualité des données et des connaissances EGC (QDC-EGC 2009)*.
- [Brunet-09] Brunet G. Desjardins B. LALLICH S. «Espace, temps et lignées familiales. La dynamique géographique des patronymes au Québec du XVII^e siècle au début du XIX^e siècle,». In proc. of *Colloque Démographie historique, Entretiens du Centre Jacques Cartier*.
- [Dubois-12a] Dubois D. Prade H. RICO A. «Intégrales et désintégrales qualitatives et leur contrepartie logique». In proc. of *Journées de l'Intelligence Artificielle Fondamentale (JAIF 2012)*.
- [Dubois-13] Dubois D. Prade H. RICO A. «Les capacités sont des possibilités imprécises». In proc. of *Logique Floue et ses Applications (LFA 2013)*.
- [Forestier-09] FORESTIER M. VELCIN J. Ganascia J.-G. «Un cadre formel pour la veille numérique sur la presse en ligne». In proc. of *Atelier Veille Numérique EGC (VN-EGC 2009)*.

- [Forestier-10] FORESTIER M., VELCIN J., ZIGHED D. A. «Fouille de discussions pour l'identification de rôles sociaux». In proc. of *Atelier Recherche et Recommandation d'information dans les Réseaux Sociaux INFORSID (REISO-INFORSID 2010)*.
- [Gicquel-12] GICQUEL Q., Metzger M.-H., LALLICH S. «Elaboration de règles de décision pour la détection d'infections du site opératoire à partir de données médicales textuelles». In proc. of *EPICLIN / Journées des statisticiens des CLCC (EPICLIN 2012)*.
- [LeBras-09] Le Bras Y., Lenca P., LALLICH S., Moga S. «Généralisation d'une propriété d'antimonotonie de la confiance pour l'extraction de motifs intéressants non fréquents». In proc. of *Atelier Qualité des données et des connaissances EGC (QDC-EGC 2009)*.
- [LeBras-10] Le Bras Y., Meyer P., Lenca P., LALLICH S. «Mesure de la robustesse de règles d'association». In proc. of *Atelier Qualité des données et des connaissances EGC (QDC-EGC 2010)* p.27-38. Ref HAL: [hal-00444450_v1](https://hal.archives-ouvertes.fr/hal-00444450_v1).
- [LeBras-11] Le Bras Y., Lenca P., LALLICH S. «Un cadre formel pour l'étude des mesures d'intérêt des règles d'association». In proc. of *Journée du GDR 13 sur la fouille de données*. Ref HAL: [hal-00632775_v1](https://hal.archives-ouvertes.fr/hal-00632775_v1).
- [Lenca-11] LENCA P., LALLICH S. «Le choix d'une bonne mesure de qualité, condition du succès d'un processus de fouille de données». In proc. of *atelier data mining, applications, cas d'études et success stories EGC (ACE-EGC 2011)* p.5-8. Ref HAL: [hal-00725673_v1](https://hal.archives-ouvertes.fr/hal-00725673_v1).
- [Pisetta-09] PISETTA V., ZIGHED D. A. «L'effet du voisinage sur la séparabilité des classes : Application à la sélection de noyaux pour les SVM». In proc. of *Atelier Topological Learning (TP-EGC)*.
- [Prade-11] Prade H., RICO A. «Vers une version possibiliste de la théorie des fonctions de croyance». In proc. of *Journées de l'Intelligence Artificielle Fondamentale (JAIF 2011)*.
- [Rico-11] RICO A., RICO F., ROLLAND A. (Corresp.). «Utilisation des 4-capacités pour la modélisation dynamique de la prise de décision collective». In proc. of *Congrès de la Société française de Recherche Opérationnelle et Aide à la Décision (ROADEF 2011)*.
- [Ritschard-09] Ritschard G., MARCELLIN S., ZIGHED D. A. «Arbre de décision pour données déséquilibrées : sur la complémentarité de l'intensité d'implication et de l'entropie décentrée». *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités* (2009) p.207-221.
- [Rolland-12] ROLLAND A. «Aide à la décision multicritère et apprentissage automatique pour la classification». In proc. of *Atelier aide à la Décision à tous les Etages (AIDE-EGC 2012)* p.15-24.
- [Rolland-12c] ROLLAND A. (Corresp.), Ould sidi M.-M., Quilot-turion B. «Méthodes multicritères pour le tri de fruits virtuels». In proc. of *Congrès de la Société française de Recherche Opérationnelle et Aide à la Décision (ROADEF 2012)*.
- [Rolland-13] ROLLAND A. «ROBUSTESSE DU PALMARES DES VILLES ETUDIANTES».
- [Stavrianou-09c] STAVRIANOU A., VELCIN J., CHAUCHAT J.-H. «Un cadre pour la représentation et l'analyse de débats sur le Web». In proc. of *Atelier Veille Numérique EGC (VN-EGC 2009)*.
- [Stavrianou-10b] STAVRIANOU A., VELCIN J., CHAUCHAT J.-H. «Recommendation of key messages extracted from forums». In proc. of *Atelier Recherche et Recommandation d'information dans les Réseaux Sociaux INFORSID (REISO-INFORSID 2010)*.
- [Strauss-11] Strauss O., RICO A. «Déconvolution intervalliste basée sur une représentation imprécise de la réponse impulsionnelle d'un capteur». In proc. of *Journée du GDR-ISIS sur l'Inversion et Assimilation d'Images*.
- [Thibault-10] Thibault G., Aussem A., BONNEVAY S. «Extraction de frontière de Markov dans les données de grande dimension». In proc. of *Journées Francophones sur les Réseaux Bayésiens (JFRB 2010)*. Ref HAL: [hal-00467544_v1](https://hal.archives-ouvertes.fr/hal-00467544_v1).

Publications communes SID-DMD

Ouvrages ou direction d'ouvrages de portée internationale

- [Brissaud-11] Brissaud M., LAMURE M., Milan J.-J., Auray J.-P., NICOLOYANNIS N., Duru G., Terrenoire M., Tounissoux D., ZIGHED D. A., BONNEVAY S., Le T. V., Bui M., Ben Amor S., Levorato V., KABACHI N. «Basics of pretopology». (2011). Ref HAL: [hal-00605389_v1](https://hal.archives-ouvertes.fr/hal-00605389_v1).

Revue internationale

- [Gavin-11] GAVIN G., VELCIN J., Aubertin P. «Privacy Preserving Aggregation of Secret Classifiers». In *Transactions on Data Privacy*, vol. 4/3 (2011) p.167-187.
- [Grabova-10] GRABOVA O., DARMONT J. (Corresp.), CHAUCHAT J.-H., Zolotaryova I. «Business Intelligence for Small and Middle-Sized Enterprises». In *SIGMOD Record*, vol. 39/2 (2010) p.39-50. Ref HAL: [hal-00555014_v1](https://hal.archives-ouvertes.fr/hal-00555014_v1). Rang A.
- [Guille-13b] GUILLE A. (Corresp.), HACID H., FAVRE C., ZIGHED D. A. «Information Diffusion in Online Social Networks: A Survey». In *SIGMOD Record*, vol. 42/2 (2013) p.17-28. DOI: [10.1145/2503792.2503797](https://doi.org/10.1145/2503792.2503797). Rang A.

Autres publications internationales

- [BenHassine-10a] BEN HASSINE S. (Corresp.), DARMONT J. CHAUCHAT J.-H. «Aggregation of data quality metrics using the Choquet integral». In proc. of *International Workshop on Quality in Databases VLDB (QDB-VLDB 2010)*. Ref HAL: [hal-00519690_v1](https://hal.archives-ouvertes.fr/hal-00519690).
- [Dermouche-13] DERMOUCHE M. (Corresp.), Kouas L. VELCIN J. LOUDCHER S. «How to Learn with Naive Bayes and Prior knowledge: an Application to Sentiment Analysis». In proc. of *International Workshop on Semantic Evaluation (SemEval-NAACL 2013)* p.364-368. Rang C.
- [Gavin-10] GAVIN G. (Corresp.), VELCIN J. «Quadratic error minimization in a distributed environment with privacy preserving». In proc. of *Workshop on Privacy and Security issues in Data Mining and Machine Learning ECML/PKDD (PSDML-ECML 2010)*. Ref HAL: [hal-00519963_v1](https://hal.archives-ouvertes.fr/hal-00519963).
- [Guille-13a] GUILLE A. (Corresp.), FAVRE C. HACID H. ZIGHED D. A. «SONDY : An Open Source Platform for Social Dynamics Mining and Analysis». In proc. of *ACM SIGMOD International Conference on Management of Data (SIGMOD 2013 demo session)* p.1005-1008. DOI: [10.1145/2463676.2463694](https://doi.org/10.1145/2463676.2463694).

Conférences nationales avec comité de programme et actes

- [Basileu-10] BASILEU C. BOUNEKKAR A. KABACHI N. LAMURE M. «Vers un modèle de diffusion spatiale d'une pandémie». In proc. of *Colloque international Veille Stratégique Scientifique et Technologique (VSST 2010)*.
- [Basileu-10a] BASILEU C. Ben Amor S. Bui M. BOUNEKKAR A. KABACHI N. LAMURE M. «Stochastic networks». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2010)*. Rang C.
- [Dermouche-13a] DERMOUCHE M. VELCIN J. LOUDCHER S. Khouas L. «Une nouvelle mesure pour l'évaluation des méthodes d'extraction de thématiques : la Vraisemblance Généralisée». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2013)* p.317-328. Ref HAL: [halshs-00794553_v1](https://halshs.archives-ouvertes.fr/halshs-00794553). Rang C.
- [Guille-12b] GUILLE A. (Corresp.), HACID H. FAVRE C. «Une approche multidimensionnelle basée sur les comportements individuels pour la prédiction de la diffusion de l'information sur Twitter». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2012)* p.405. Ref HAL: [hal-00665432_v1](https://hal.archives-ouvertes.fr/hal-00665432). Rang C.
- [Guille-13c] GUILLE A. (Corresp.), FAVRE C. ZIGHED D. A. «SONDY : une plateforme open-source d'analyse et de fouille pour les réseaux sociaux en ligne». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2013)* p.45-48. Ref HAL: [hal-00770557_v1](https://hal.archives-ouvertes.fr/hal-00770557). Rang C.
- [Guille-14] GUILLE A. (Corresp.), FAVRE C. «Une méthode pour la détection de thématiques populaires sur Twitter». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2014)* p.83-88. Ref HAL: [hal-00942678_v1](https://hal.archives-ouvertes.fr/hal-00942678). Rang C.
- [Guille-14a] GUILLE A. (Corresp.), FAVRE C. «Un système de détection de thématiques populaires sur Twitter». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2014)* p.605-608. Ref HAL: [hal-00942676_v1](https://hal.archives-ouvertes.fr/hal-00942676). Rang C.
- [Loudcher-13a] LOUDCHER S. VELCIN J. Forissier V. Broilliard C. Simonnot P. «Analyse des réclamations d'allocataires de la CAF :un cas d'étude en fouille de données». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2013)* p.449-460. Ref HAL: [halshs-00794544_v1](https://halshs.archives-ouvertes.fr/halshs-00794544). Rang C.

Autres publications nationales

- [Guille-12] GUILLE A. HACID H. FAVRE C. «Predicting the Temporal Dynamics of Information Diffusion in Social Networks». Ref Arxiv: [1302.5235](https://arxiv.org/abs/1302.5235).

Publications externes des personnels recrutés au cours des cinq dernières années**Revue internationale**

- [AhPine-09] AH-PINE J. Bressan M. Clinchant S. Csurka G. Hoppenot Y. Renders J.-M. «Crossing textual and visual content in different application scenarios». In *Multimedia Tools and Applications*, vol. 42/1 (2009) p.31-56. DOI: [10.1007/s11042-008-0246-8](https://doi.org/10.1007/s11042-008-0246-8). Rang B.
- [AhPine-10c] AH-PINE J. Marcotorchino J.-F. «Unifying Some Association Criteria Between Partitions by Using Relational Matrices». In *Communications in Statistics - Theory and Methods*, vol. 39/3 (2010) 531 - 542. DOI: [10.1080/03610920903140262](https://doi.org/10.1080/03610920903140262). Rang B.
- [AhPine-11a] AH-PINE J. «On data fusion in information retrieval using different aggregation operators». In *Web Intelligence and Agent Systems*, vol. 9/1 (2011) p.43-55. DOI: [10.3233/WIA-2011-0204](https://doi.org/10.3233/WIA-2011-0204). Rang C.
- [Jouve-09] Culus J.-F., JOUBE B. Convex circuit-free coloration of an oriented graph. In *European Journal of Combinatorics*, vol. 30/1 (2009) p.43-52.

Conférences internationales avec comité de programme et actes

- [AhPine-09a] AH-PINE J. «Cluster Analysis Based on the Central Tendency Deviation Principle». In proc. of *International Conference on Advanced Data Mining and Applications (ADMA 2009)* p.5-18. DOI: [10.1007/978-3-642-03348-3_5](https://doi.org/10.1007/978-3-642-03348-3_5). Rang B.
- [AhPine-09b] AH-PINE J. Jacquet G. «Cliques-Based Clustering for Improving Named Entity Recognition Systems». In proc. of *Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)* p.51-59. Rang A.
- [AhPine-10b] AH-PINE J. «On aggregating binary relations using 0-1 integer linear programming». In proc. of *International Symposium on Artificial Intelligence and Mathematics (ISAIM 2010)* p.1-10. Ref HAL: [hal-00523138_v1](https://hal.archives-ouvertes.fr/hal-00523138_v1). Rang B.
- [AhPine-10d] AH-PINE J. «Normalized Kernels as Similarity Indices». In proc. of *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2010)* p.362-373. DOI: [10.1007/978-3-642-13672-6_36](https://doi.org/10.1007/978-3-642-13672-6_36). Rang A.
- [Prade-09] Prade H. (Corresp.), RICO A. Serrurier M. Raufaste E. «Eliciting Sugeno Integrals: Methodology and a Case Study». In proc. of *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)* p.712-723. DOI: [10.1007/978-3-642-02906-6_61](https://doi.org/10.1007/978-3-642-02906-6_61). Rang B.
- [Prade-09a] Prade H. RICO A. Serrurier M. «Elicitation of Sugeno integrals: A Version Space Learning Perspective». In proc. of *International Symposium on Methodologies for Intelligent Systems (ISMIS 2009)* p.392-401. DOI: [10.1007/978-3-642-04125-9_42](https://doi.org/10.1007/978-3-642-04125-9_42). Rang B.

Autres publications internationales

- [AhPine-09c] AH-PINE J. Clinchant S. Csurka G. Liu Y. «XRCE's participation in ImageCLEF 2009». In proc. of *Workshop of the Cross-Language Evaluation Forum (CLEF 2009)*. Rang B.
- [AhPine-09d] AH-PINE J. Renders J.-M. Viaud M.-L. «A Continuum between Serendipitous Browsing and Query-based Search for Multimedia Information Access». In proc. of *Workshop on Adaptive Multimedia Retrieval (AMR 2009)*.
- [AhPine-10] AH-PINE J. Marcotorchino J.-F. «Overview of the Relational Analysis approach in Data-Mining and Multi-criteria Decision Making». *Web Intelligence and Intelligent Agents (2010)*.
- [AhPine-10a] AH-PINE J. Clinchant S. Csurka G. Perronnin F. Renders J.-M. «Leveraging Image, Text and Cross-media Similarities for Diversity-focused Multimedia Retrieval». *ImageCLEF (2010)* p.315-342. DOI: [10.1007/978-3-642-15181-1_17](https://doi.org/10.1007/978-3-642-15181-1_17).
- [AhPine-11] AH-PINE J. Clinchant S. Csurka G. «Comparison of Several Combinations of Multimodal and Diversity Seeking Methods for Multimedia Retrieval». *Multilingual Information Access Evaluation II. Multimedia Experiments (2011)* p.124-132. DOI: [10.1007/978-3-642-15751-6_13](https://doi.org/10.1007/978-3-642-15751-6_13).
- [AhPine-11b] AH-PINE J. Renders J.-M. Viaud M.-L. «A Continuum between Browsing and Query-based Search for User-Centered Multimedia Information Access». In proc. of *Workshop on Adaptive Multimedia Retrieval (AMR 2009)* p.111-123. DOI: [10.1007/978-3-642-18449-9_10](https://doi.org/10.1007/978-3-642-18449-9_10).
- [Clinchant-10] Clinchant S. Csurka G. AH-PINE J. Jacquet G. Perronnin F. Sánchez J. Minoukadeh K. «XRCE's Participation in Wikipedia Retrieval, Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2010». In proc. of *Workshop of the Cross-Language Evaluation Forum (CLEF 2010)*. Rang B.
- [Strauss-09] Strauss O. RICO A. «Towards an interval based deconvolution in signal processing». In proc. of *International Workshop of the ERCIM Working Group on Computing & Statistics (ERCIM 2009)*.

Revue nationale

- [Auray-09] Auray J.-P. BONNEVAY S. Bui M. Duru G. LAMURE M. «Prétopologie et applications : un état de l'art». In *Studia Informatica Universalis*, vol. 7.1 (2009) p.27-44.

Conférences nationales avec comité de programme et actes

- [AhPine-10e] AH-PINE J. «Une famille d'indices de similarité généralisant la mesure de cosinus». In proc. of *Congrès de la Société Francophone de Classification (SFC 2010)* p.83-86.
- [Strauss-09b] Strauss O. RICO A. «Vers une déconvolution intervalliste : un usage détourné des mesures non-additives». In proc. of *Logique Floue et ses Applications (LFA 2009)* 67-82.

RAYONNEMENT ET ATTRACTIVITÉ ACADÉMIQUES**Équipe SID**

| | Internationaux | Nationaux | Exemples les plus significatifs |
|--------------------|----------------|-----------|---|
| Comités éditoriaux | 7 | 9 | Neurocomputing, International Journal of Data Mining, Modelling and Management (IJDDMM), IGI Global Editorial Advisory Review Board, Open Journal of Databases, Ingénierie des Systèmes |

| | Internationaux | Nationaux | Exemples les plus significatifs |
|---|----------------|-----------|---|
| | | | d'Information (ISI), Journées Entrepôts de Données et Analyse en ligne (EDA, fondées par ERIC) |
| Comités de lecture (revues et ouvrages) | 38 | 12 | IEEE Transactions on Knowledge and Data Engineering (TKDE), ACM SIGMOD Record, Data & Knowledge Engineering (DKE), Journal of Decision System (JDS), Journal of Intelligent Information Systems (JIIS), Distributed and Parallel Databases (DAPD), ACM Transactions on Intelligent Systems and Technology (TIST), Revue Techniques et Sciences Informatiques (TSI), Ingénierie des Systèmes d'Information (ISI) |
| Comités de programme | 82 | 62 | DOLAP, DEXA, ADBIS, MEDI, KDIR, ICADIWT, ICIST, ICTACS, AAMAS, MMAS, MEDES, DS, EGC, BDA |
| Comités d'organisation | 17 | 18 | VLDB 2009, DS-ALT 2012, WWW 2012, MEDES 2009, Wi-AT 2011, UC4DS@ADBIS 2012, Cloud-I@VLDB 2012-2013, SASO 2012, JFSMA 2009, INFORSID 2013-2014, démos et ateliers EGC |
| Jurys de thèse et de HDR | 1 | 25 | 10 thèses en tant que rapporteur 3 HDR dont 1 en tant que rapporteur |
| Expertises | 7 | 26 | COS externes (13), ANRT (13), AUF (3) |
| Invitations | 14 | 18 | DAWAK, Dagstuhl seminar, MOSIM, Ecole d'été Web Intelligence, USTO Oran (Algérie), Université de Blida (Algérie), Université de Cantho (Vietnam) |

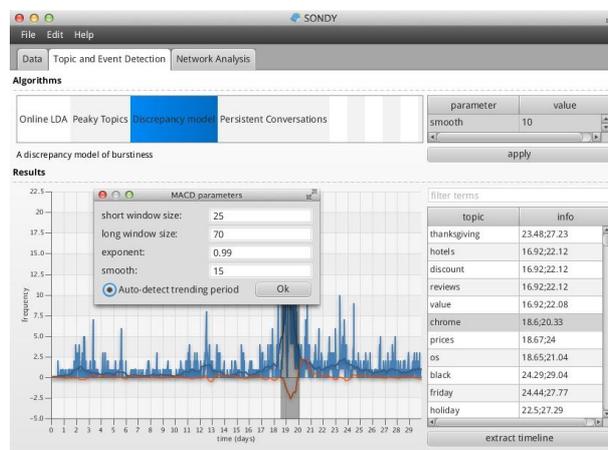
Équipe DMD

| | Internationaux | Nationaux | Exemples les plus significatifs |
|---|----------------|-----------|--|
| Comités éditoriaux | 4 | 4 | International Journal of Data Analysis Techniques and Strategies (IJDATS), Quality Issues, Measures of Interest and Evaluation of data mining models (QIMIE), International Journal of Social Network Mining, International Association for Statistical Computing, Revue des Nouvelles Technologies de l'Information (RNTI, présidence du comité), Conférence Extraction et Gestion des Connaissances (EGC, dont ERIC est un membre fondateur) |
| Comités de lecture (revues et ouvrages) | 43 | 6 | IEEE Transactions on Knowledge and Data Engineering (TKDE), ACM Transaction on Intelligent Systems (TIST), Journal of Social Network Analysis and Mining (SNAM), International Journal of Social Network Mining (IJSNM), Journal of Intelligent Information Systems (JIIS), Computational Intelligence (CI), European Journal of Operation Research (EJOR), Engineering Applications of Artificial Intelligence (EAAI), Pattern Recognition Letters, Advances in Data Analysis and Classification (ADAC), Fuzzy Sets and Systems, Computational Statistics, Machine Learning, Computational Intelligence, Journal of Applied Mathematics |
| Comités de programme | 50 | 38 | ECML-PKDD, PAKDD, ISMIS, ICTAI, COMPSTAT, DMIN, ASMDA, SMTDA, KDBI/EPIA, ICANNASAC, ASONAM, EGC, CAP |

SHSdocNET⁹ est un portail social sémantique pour l'exposition des compétences et la veille scientifique en SHS, réalisé en collaboration avec l'ISH.



SONDY (Open Source Platform for Social Dynamics Mining and Analysis)¹⁰ est une plateforme Java d'analyse et de fouille de données issues des média sociaux. Ses fonctions principales sont la détection d'événements et de sujets, ainsi que l'analyse d'influence.



⁹ <http://docnet.ish-lyon.cnrs.fr>

¹⁰ <http://mediamining.univ-lyon2.fr/people/guille/software.php>

Annexe 7 : Liste des contrats

CONTRATS INSTITUTIONNELS SUR FINANCEMENT PUBLIC

Équipe SID

| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication |
|------------|---|---|-----------|--------------------|------------------|--------------|--------------------------|
| DPC-CLO | Détection de Phénomènes Complexes dans les Corpus Linguistiques Oraux | ICAR | 2009-2011 | 20 000 € | 10 000 € | Lyon 2 (BQR) | Responsable scientifique |
| TASSILI | Mobilité internationale de chercheurs | Université Saad Dahleb, Blida (Algérie) | 2011-2014 | 39 600 € | 39 600 € | EGIDE | Responsable scientifique |
| TRAJEC-Psy | TRAJECtoires patients en PSYchiatric | Hospices civils de Lyon | 2012-2015 | 86 700 € | - | HCL | Partenaire |

Équipe DMD

| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication |
|------------|--|--|-----------|--------------------|------------------|----------------------------|--------------------------|
| NOFDSHS | Nouveaux Outils de Fouille de Données pour les SHS | ISH | 2008-2010 | 25 000 € | 25 000 € | Lyon 2 (BQR) | Responsable scientifique |
| ECHOUTCOME | Étude des systèmes de santé européens | 7 partenaires (France, Italie, Belgique...) | 2010-2013 | | 282 000 € | Union Européenne | Responsable scientifique |
| SHSdocNET | Portail Web social sémantique pour l'exposition des compétences et la veille scientifique en SHS | ISH | 2011-2012 | 20 000 € | 20 000 € | Lyon 2 (BQR) | Responsable scientifique |
| FLURESP | Cost-effectiveness assessment of European influenza human pandemic alert and response strategies | 9 partenaires (dont Italie, Hongrie, Espagne...) | 2011-2014 | 1,2 M€ | 338 202 € | Union Européenne | Responsable scientifique |
| GGN | G-Graphs and networks | ESSEC, Univ. Martinique, Univ. Bucarest | 2012 | 2 500 € | 2 500 € | GDR RO | Partenaire |
| RéseauxWWW | Modèles de grands réseaux : jeux de poursuite et décomposition modulaire. Applications au WWW | ARC6 | 2012 | 3 000 € | 3 000 € | Région Rhône-Alpes | Responsable scientifique |
| IRF | Infonomics Resource Facility | ISH, EVS, LHC | 2013-2015 | 300 000 € | - | Université de Lyon (PALSE) | Responsable scientifique |
| UTE | Modélisation des contributions non linéaires du système électrique | UdelaR (Uruguay) | 2014-2015 | 55 000 € | - | FSE-ANII (Uruguay) | Partenaire |

Collaborations SID-DMD

| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication (équipe leader) |
|----------|---|---|-----------|--------------------|------------------|-----------------------------|--------------------------------|
| RFCDP | Le rôle des forums citoyens dans le débat public. Construction et test d'outils semi-automatiques pour l'étude de la dynamique des discours | ELICO | 2010-2011 | 20 000 € | 20 000 € | Lyon 2 (BQR) | Co-responsable (DMD) |
| ImagiWeb | Images sur le Web : analyse de la dynamique des images sur le Web 2.0 | XRCE, AMI Software, EDF, LIA, CEPEL | 2012-2015 | 872 000 € | 160 000 € | ANR | Responsable scientifique (DMD) |
| REQUEST | Big Data, Big Analytics, Visual Analytics et Cloud Computing | 16 partenaires (dont Thales, Orange Labs, SNCF, LIP6, LIMSI...) | 2014-2017 | 9 M€ | 190 000 € | Investissement d'Avenir/FSN | Partenaire (DMD) |

CONTRATS INDUSTRIEL, CONTRATS SUR FINANCEMENT PRIVÉ**Équipe SID**

| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication |
|------------|---|--------------------|-----------|--------------------|------------------|--------------------|--------------------------|
| ProxAn | Proximity Analysis | Incubateur CREALYS | 2008-2009 | 30 000 € | 30 000 € | Région Rhône-Alpes | Responsable scientifique |
| CIFRE-AID | Analyse et amélioration de la qualité des données dans un environnement multi-sources | AID | 2009-2012 | 12 000 € | 12 000 € | AID | Responsable scientifique |
| Buzzinbees | Détection automatique d'incidents | Buzzinbees | 2011-2012 | 4 000 € | 4 000 € | Buzzinbees | Conseil |
| CPCAM-sécu | Détection d'intrusions | CPCAM | 2013-2016 | 15 000 € | 15 000 € | CPCAM | Responsable scientifique |

Équipe DMD

| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication |
|-------------------|--|--------------------|-----------|--------------------|------------------|--------------------|--------------------------|
| BETWEEN | Modèle de représentation et d'analyse des débats en ligne sur Internet | Incubateur CREALYS | 2007-2009 | 20 500 € | 20 500 € | Région Rhône-Alpes | Responsable scientifique |
| Routing | Optimisation de routing dans des graphes dynamiques | Visioglobe | 2010 | 5 000 € | 5 000 € | Visioglobe | Responsable scientifique |
| CIFRE-Technicolor | Latent Group Dynamics : Evolution of Emergent Roles in Social Networks | Technicolor | 2013-2015 | 30 000 € | 30 000 € | Technicolor | Responsable scientifique |
| GeoMidPoint | Recherche automatique de lieux de formation sur une carte Google Map | Feu vert | 2013-2014 | 4 000 € | 4 000 € | Feu vert | Responsable scientifique |

Collaborations SID-DMD

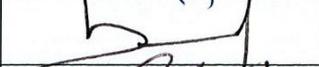
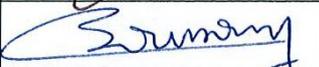
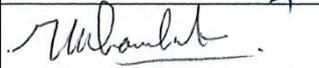
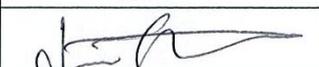
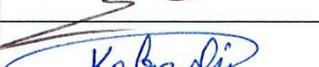
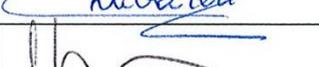
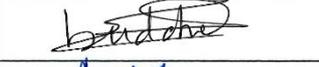
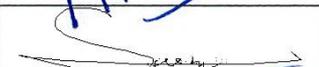
| Acronyme | Titre | Partenaires | Années | Financement global | Financement ERIC | Financier | Implication (équipe leader) |
|-----------|---|--------------|-----------|--------------------|------------------|--------------|--------------------------------|
| CAF | Etude exploratoire des réclamations des allocataires | CNEDI-CAF | 2012 | 8 000 € | 8 000 € | CNEDI-CAF | Responsable scientifique (SID) |
| CIFRE-AMI | Modélisation des controverses sur le Web et les médias publics par le contenu de la structure du réseau | AMI Software | 2012-2014 | 27 000 € | 27 000 € | AMI Software | Responsable scientifique (DMD) |

Annexe 8 : Document unique d'évaluation des risques - DUER (lorsqu'il existe)

Le DUER est établi au niveau de l'établissement à l'Université Lyon 2, qui abrite l'essentiel des locaux du laboratoire.

Annexe 9 : Liste des personnels

Liste des personnels (chercheurs, enseignants-chercheurs et assimilés) de l'unité présents au 30 juin 2014 et qui le seront toujours au 1^{er} janvier 2016.

| NOM | Prénom | Statut | Etablissement | Équipe | Signature |
|--------------|-------------|------------------------------|---------------|--------|---|
| AH-PINE | Julien | MCF 26 | Lyon 2 | DMD |  |
| BENTAYEB | Fadila | MCF HDR 27 | Lyon 2 | SID |  |
| BONNEVAY | Stéphane | MCF HDR 27 | Lyon 1 | DMD |  |
| BOUNEKKAR | Ahmed | MCF 27 | Lyon 1 | DMD |  |
| BOUSSAID | Omar | PR 27 | Lyon 2 | SID |  |
| CHAUCHAT | Jean-Hugues | PR 27 émérite | Lyon 2 | DMD |  |
| CUGLIARI | Jairo | MCF 26 | Lyon 2 | DMD |  |
| DARMONT | Jérôme | PR 27 | Lyon 2 | SID |  |
| FAVRE | Cécile | MCF 27 | Lyon 2 | SID |  |
| GAVIN | Gérald | MCF 27 | Lyon 1 | SID |  |
| HARBI | Nouria | MCF 27 | Lyon 2 | SID |  |
| KABACHI | Nadia | MCF 27 | Lyon 1 | SID |  |
| LALLICH | Stéphane | PR 27 demande éméritat | Lyon 2 | DMD |  |
| LOUDCHER | Sabine | MCF HDR 27 | Lyon 2 | SID |  |
| RAKOTOMALALA | Ricco | MCF 27 | Lyon 2 | DMD |  |
| RICO | Fabien | MCF 27 | Lyon 1 | DMD |  |
| RICO | Agnès | MCF 27 | Lyon 1 | DMD |  |
| ROLLAND | Antoine | MCF 26 | Lyon 2 | DMD |  |
| SIANI | Carole | MCF HDR 27 | Lyon 1 | DMD |  |
| VELCIN | Julien | MCF 27 | Lyon 2 | DMD |  |
| ZIGHED | Abdelkader | PR 27 | Lyon 2 | DMD |  |

Annexe 10 : Liste des thèses soutenues et en cours

HDR soutenues

| NOM | Prénom | Titre | Coordinateur | Équipe | Année de soutenance |
|----------|--------|--|--------------|--------|---------------------|
| BENTAYEB | Fadila | Entrepôts et analyse en ligne de données complexes centrés utilisateur : un nouveau défi | D.A. Zighed | SID | 2011 |
| LOUDCHER | Sabine | Vers l'OLAP sémantique pour l'analyse en ligne des données complexes | O. Boussaïd | SID | 2011 |

Thèses soutenues

| NOM | Prénom | Titre | Encadrant(s) | Équipe | Année de soutenance | Durée | Devenir du docteur |
|------------|---------------|---|------------------------------|--------|---------------------|-------|--|
| PRUDHOMME | Elie | Représentation et fouille de données volumineuses | S. Lallich | DMD | 2009 | 4 | Ingénieur de recherche et développement (CDI) |
| THOMAS | Julien | Apprentissage automatique de données dynamiques et complexes | N. Nicoloyannis, D.A. Zighed | DMD | 2009 | 3,5 | Ingénieur d'étude |
| BAHRI | Emna | Amélioration du processus d'apprentissage automatique pour des données complexes | S. Lallich | DMD | 2010 | 4 | Ingénieur d'étude en SSII (CDI) |
| STAVRIANOU | Anna | Modeling and Mining of web discussions | J.H. Chauchat, J. Velcin | DMD | 2010 | 4,5 | Ingenieur de recherche, Xerox Research Centre Europe (CDI) |
| MAIZ | Nora | Intégration de données par médiation basée sur les ontologies pour l'analyse en ligne (OLAP) à la demande | O. Boussaïd, F. Bentayeb | SID | 2010 | 4,5 | Emploi dans l'industrie |
| QURESHI | Taimur | Data Mining in domain of decision trees and complex data | D.A. Zighed | DMD | 2010 | 4 | Enseignant, Université d'Islamabad, Pakistan |
| WEI | Zihua | The research on chinese Text Multi-Label classification | J.H. Chauchat | DMD | 2010 | 3 | Enseignante, Université Tondji, Shanghai, Chine |
| MAVRIKAS | Efhtymios | Entre les mots - Méthodes d'analyse informatique du discours idéologique | N. Nicoloyannis, D.A. Zighed | DMD | 2010 | 7 | Directeur de la société Ad Axem, Syros, Grèce |
| SALEM | Rashed Khalil | Active XML data warehouses for intelligent, on-line decision support | J. Darmont, O. Boussaïd | SID | 2012 | 4 | Lecturer, Université de Minoufiya, Egypte |
| PISETTA | Vincent | New Insights into Decision Trees Ensembles | D.A. Zighed | DMD | 2012 | 4 | Ingénieur CNAM Lyon |
| FORESTIER | Mathilde | Extraction automatique de réseaux sociaux à partir de textes | D.A. Zighed, J. Velcin | DMD | 2012 | 4 | Postdoc, LSIS, Aix Marseille Université |
| HACHICHA | Marouane | Analyse OLAP de données complexes dans les entrepôts de données XML | J. Darmont | SID | 2012 | 5 | Postdoc, Université d'Aveiro, Portugal |
| NGUYEN | Huu-Hoa | Fouille de données pour la détection d'intrusions dans les systèmes informatiques | J. Darmont, N. Harbi | SID | 2013 | 2,75 | Lecturer, Université de Cantho, Vietnam |
| RIZOIU | Marian-Andrei | Structurations semi-supervisée des données complexes - application aux SI en SHS et à la fouille sur le WEB | S. Lallich, J. Velcin | DMD | 2013 | 3,75 | Postdoc, NICTA, Australie |
| KAFIFY | Ahmed | Intelligent metaheuristics for multi-objective decision support - Application to health care system | S. Bonnevey, A. Bounekkar | DMD | 2013 | 5 | Lecturer, Université du Caire, Egypte |
| TRIKI | Salah | Sécurisation des entrepôts de données : de la conception à l'exploitation | O. Boussaïd, N. Harbi | SID | 2013 | 3,5 | Assistant, Université de Kairouan, Tunisie |
| AKNOUCHE | Rachid | Entrepôts de textes ; de l'entreposage à la modélisation multidimensionnelle de données textuelles | F. Bentayeb, O. Boussaïd | SID | 2014 | 3,5 | Fonctionnaire, Ministère de la défense algérienne |

Thèses en cours

| NOM | Prénom | Encadrant principal | | Co-encadrant n° 1 | | Sujet de thèse | Équipe | Année | Financement |
|-----------------------------|--------------|---------------------|----|-------------------|----|---|------------|----------|--|
| | | Nom | % | Nom | % | | | | |
| AAZI | Fatima Zohra | ABDESSELAM | 25 | ZIGHED | 25 | Sélection des variables par SVM pour la discrimination multiclasse | DMD | janv.-12 | Cotutelle université d'Hassan 1er Settat - Financement recherche CNRST Maroc |
| ARRES | Billel | BOUSSAID | 50 | KABACHI | 50 | Utilisation du paradigme MapReduce pour la création de cubes OLAP dans un environnement de Cloud Computing | SID | nov.-12 | Financement gouvernement algérien |
| ATTASENA | Varunya | DARMONT | 50 | HARBI | 50 | Sécurité et performance des entrepôts de données dans les nuages | SID | oct.-11 | Financement gouvernement thaïlandais |
| BEN HASSINE ép. GUETARI | Soumaya | DARMONT | 25 | CHAUCHAT | 25 | Qualité des données dans les bases et entrepôts de données : vers une concrétisation du concept de système introspectif. Application aux opérations de comptage et de ciblage dans les bases de données commerciales | SID DMD | nov.-08 | CIFRE AID puis CDI |
| DEHDOUH | Khaled | BOUSSAID | 50 | BENTAYEB | 50 | Techniques d'optimisation pour les modèles de base de données NoSQL dans un environnement de cloud computing | SID | nov.-12 | Financement gouvernement algérien |
| DERMOUCHE | Mohamed | LOUDCHER | 25 | VELCIN | 25 | Modélisation des controverses sur le Web et les médias publics par le contenu et la structure du réseau | SID DMD | janv.-12 | CIFRE AMI Software |
| DESROZIERS | Katia | SIANI | 50 | TOUMI | 50 | Évaluation pharmaco-économique dans l'hyperactivité vésicale - Estimation de l'utilité en fonction de la sévérité des symptômes - Modélisation des coûts des soins de santé via les modèles à surreprésentation de zéro | DMD | oct.-11 | CIFRE Creativ Research |
| EZZEDDINE | Diala | ZIGHED | 50 | RICO-F | 50 | Apprentissage topologique | DMD | oct.-10 | Financement gouvernement libanais |
| GONCHAROVA ép. OROBINSKA | Olena | CHAUCHAT | 50 | | | Les méthodes et les modèles d'élaboration automatisés des ontologies | DMD | nov.-10 | Cotutelle Université Polytechnique de de Kharkiv (Ukraine) |

| NOM | Prénom | Encadrant principal | | Co-encadrant n° 1 | | Sujet de thèse | Équipe | Année | Financement |
|--------------------------|---------|---------------------|----|-------------------|----|---|------------|----------|--|
| GRABOVA ép. LAVIGNAC | Oksana | CHAUCHAT | 25 | DARMONT | 25 | Entrepôts des données sur le Web | SID DMD | févr.-09 | Cotutelle Université Nationale d'Economie de Kharkiv (Ukraine) |
| GUILLE | Adrien | ZIGHED | 50 | FAVRE | 50 | Propagation et qualité de l'information dans les réseaux sociaux | DMD SID | oct.-11 | CDU |
| GUIZANI | Nachoua | DARMONT | 10 | FAYN | 90 | Conception d'une architecture de services d'intelligence ambiante pour l'optimisation de la qualité de service de transmission de messages en e-santé | SID | févr.-13 | CIFRE AXON |
| JAKAWAT | Wararat | LOUDCHER | 50 | FAVRE | 50 | New models for detecting events in social or information networks: an application to research domains study | SID | oct.-12 | Financement gouvernement thaïlandais |
| JMEL épouse FERCHICHI | Jihene | ZIGHED | 25 | AH-PINE | 25 | Agrégation d'informations multisources et hétérogènes pour l'amélioration de systèmes de recommandation | DMD | oct.-12 | Cotutelle ISG Tunis |
| KHEMIRI | Rym | DARMONT | 25 | BENTAYEB | 25 | Personnalisation des analyses OLAP dans les entrepôts de données | SID | oct.-09 | Cotutelle Université de Tunis |
| LUMBRERAS | Alberto | JOUVE | 25 | VELCIN | 25 | Latent Group Dynamics: Evolution of Emergent Roles in Social Networks | DMD | févr.-13 | CIFRE TECHNICOLOR |
| ÖZTÜRK | Aybüke | DARMONT | 34 | LALLICH | 33 | Élaboration, mise en œuvre et exploitation d'un modèle de description d'objet archéologique complexe | SID DMD | oct.-13 | ADR ARC5 Région Rhône-Alpes |
| PIERROT | David | DARMONT | 25 | HARBI | 25 | Analyse et information en temps réel pour la détection d'intrusions dans les systèmes informatiques | SID | sept.-13 | PAST + contrat CTI-RAA |
| RATSIMBAZAFY | Rado | BOUSSAID | 50 | BENTAYEB | 50 | Le traitement des requêtes massives dans le cadre des très grands entrepôts de données dans des environnements de cloud computing | SID | mars-13 | Financement association DEHFI Madagascar |
| SAAB | Fares | BONNEVAY | 50 | EL KHOURY | 50 | Monte Carlo Tree Search » pour l'optimisation combinatoire et multi objectif - Application à la construction de réseaux bayésiens | DMD | avr.-13 | Financement fondation Compu-vision SARL |
| SELMANE | Sid Ali | BOUSSAID | 50 | BENTAYEB | 50 | Méthodes et outils pour la détection et la gestion des communautés dans les réseaux sociaux | SID | oct.2011 | Financement gouvernement algérien |
| SOBATI MOGHADAM | Somayeh | DARMONT | 50 | GAVIN | 50 | Confidentialité de données décisionnelles partagées dans le nuage | SID | sept.-13 | Financement entreprise Parsa Kimia (Iran) |

| NOM | Prénom | Encadrant principal | | Co-encadrant n° 1 | | Sujet de thèse | Équipe | Année | Financement |
|-----------------|---------------|---------------------|----|-------------------|----|---|------------|----------|-----------------------------|
| SORIANO MORALES | Edmundo Pavel | LOUDCHER | 50 | AH-PINE | 50 | Weakly supervised and unsupervised information extraction from text by leveraging open sources of information | DMD SID | sept.-13 | CDU |
| WANG | Xinyu | DARMONT | 50 | AH-PINE | 50 | Large scale data mining in distributed systems | DMD SID | sept.-13 | Financement projet REQUEST |
| YOUNSI | Fatima Zohra | BOUSSAID | 50 | | | Conception d'une ontologie pour un outil d'aide multicritère d'aide à la décision | SID | mars-10 | Cotutelle Université d'Oran |

Annexe 11 : Liste des postdoctorants extérieurs accueillis à ERIC

| NOM | Prénom | Statut | Provenance | Période |
|----------|-----------|------------------------------|--------------|-----------|
| HAJLAOUI | Kafil | ATER Lyon 2 | France | 2010-2011 |
| KIT | Chantola | Chercheur contractuel Lyon 2 | Cambodge | 2011-2012 |
| ASFARI | Ounas | ATER Lyon 2 | France | 2011-2013 |
| LUST | Thibaut | Chercheur contractuel Lyon 1 | Belgique | 2012-2013 |
| CHAGHERI | Samaneh | ATER Lyon 2 | France | 2013-2014 |
| KIM | Young-Min | Chercheur contractuel Lyon 2 | Corée du Sud | 2013-2014 |

Annexe 12 : Implication des membres d'ERIC dans l'enseignement et les instances universitaires

| NOM | Prénom | Établissement/composante | Fonction | Période |
|-----------|-------------|--|---|--------------|
| BENTAYEB | Fadila | Département Informatique et Statistique, FSEG Lyon 2 | Responsable, Licence L3 IDS | Depuis 2004 |
| BENTAYEB | Fadila | ICOM | Responsable des TICE | 2012-2014 |
| BENTAYEB | Fadila | Laboratoire ERIC | Responsable, séminaires du laboratoire | Depuis 2010 |
| BENTAYEB | Fadila | Laboratoire ERIC | Responsable, Equipe SID | Depuis 2012 |
| BENTAYEB | Fadila | ICOM | Membre élue, conseil de l'institut | Depuis 2013 |
| BONNEVAY | Stéphane | Laboratoire ERIC | Responsable, axe DECCO | 2010-2012 |
| BONNEVAY | Stéphane | Laboratoire ERIC | Directeur adjoint | Depuis 2010 |
| BONNEVAY | Stéphane | Polytech Lyon, Université Lyon 1 | Membre élu, conseil de gouvernance | Depuis 2010 |
| BOUNEKKAR | Ahmed | CNU section 27 | Membre suppléant | 2008-2012 |
| BOUNEKKAR | Ahmed | Consortium International e-MIAGE | Président | Depuis 2011 |
| BOUNEKKAR | Ahmed | Université Lyon 1 | Responsable, Master e-MIAGE | Depuis 2009 |
| BOUNEKKAR | Ahmed | Université Lyon 1 | Membre, comité consultatif | Depuis 2008 |
| BOUNEKKAR | Ahmed | Université Lyon 1 | Membre élu, conseil d'administration | 2010-2012 |
| BOUSSAID | Omar | Département Informatique et Statistique, ICOM Lyon 2 | Directeur | Depuis 2010 |
| BOUSSAID | Omar | Département Informatique et Statistique, ICOM Lyon 2 | Responsable, Master 2 IDS-IIDEE | Depuis 1996 |
| CHAUCHAT | Jean-Hugues | Département Informatique et Statistique, FSEG Lyon 2 | Responsable, Master 2 IDS-SISE | Jusqu'à 2011 |
| CHAUCHAT | Jean-Hugues | Département Informatique et Statistique, FSEG Lyon 2 | Responsable, Master franco-ukrainien IDS-M-Kharkov | Jusqu'à 2011 |
| DARMONT | Jérôme | Département Informatique et Statistique, FSEG Lyon 2 | Directeur | 2003-2010 |
| DARMONT | Jérôme | Département Informatique et Statistique, ICOM Lyon 2 | Coordinateur local (site de Lyon), Master Erasmus Mundus DMKM | Depuis 2010 |
| DARMONT | Jérôme | FSEG Lyon 2 | Responsable, Licence bidisciplinaire IDEA | Depuis 2006 |
| DARMONT | Jérôme | Département Informatique et Statistique, ICOM Lyon 2 | Responsable, Master franco-ukrainien IDS-M-Kharkov | Depuis 2005 |
| DARMONT | Jérôme | Laboratoire ERIC | Responsable, axe ENA-DC | 2010-2012 |
| DARMONT | Jérôme | Laboratoire ERIC | Directeur | Depuis 2012 |
| DARMONT | Jérôme | Ecole Doctorale InfoMaths | Représentant du laboratoire ERIC au conseil et au comité des thèses | Depuis 2012 |

| NOM | Prénom | Établissement/composante | Fonction | Période |
|---------------|----------|---|---|-------------|
| DARMONT | Jérôme | Université Lyon 2 | Correspondant C2i | 2008-2009 |
| DARMONT | Jérôme | Université Lyon 2 | Membre élu, conseil scientifique | Depuis 2012 |
| DARMONT | Jérôme | Université Lyon 2 | Représentant de l'ED Infomaths | Depuis 2013 |
| DARMONT | Jérôme | Institut de la Communication, Université Lyon 2 | Administrateur provisoire | 2013-2014 |
| DARMONT | Jérôme | Institut de la Communication, Université Lyon 2 | Membre élu, conseil de l'institut | Depuis 2014 |
| DARMONT | Jérôme | Université Lyon 2 | Coordonnateur du GEI 26-27-61-71 | Depuis 2014 |
| FAVRE | Cécile | Lyon 2, Faculté d'Anthropologie, de Sociologie et de Science Politique | Membre, comité pédagogique du Master Européen EGALES | Depuis 2010 |
| FAVRE | Cécile | Lyon 2, Faculté d'Anthropologie, de Sociologie et de Science Politique | Responsable Formation Continue, Master EGALITES | Depuis 2012 |
| HARBI | Nouria | Département Informatique et Statistique, ICOM Lyon 2 | Responsable, Master 2 IDS-OPSIE | Depuis 2006 |
| HARBI | Nouria | Département Informatique et Statistique, ICOM Lyon 2 | Directrice adjointe | Depuis 2012 |
| KABACHI | Nadia | Université Lyon 1 | Responsable doubles diplômes de licence-master en informatique, Université de Ho Chi Minh (Saigon, Vietnam) | Depuis 2010 |
| KABACHI | Nadia | Université Lyon1, Département Informatique | Responsable des Relations Internationales | Depuis 2008 |
| KABACHI | Nadia | Conseil National des Universités - CNU27 | Membre suppléant | 2012-2015 |
| KABACHI | Nadia | Consortium de L'AUF, pôle scientifique D (Electronique, Sciences & Technologie de l'Information et de la Communication) | Représentante de l'Université Lyon 1 | 2010-2014 |
| LALLICH | Stéphane | Département Informatique et Statistique, ICOM Lyon 2 | Responsable, Master 1 Informatique | 2004-2014 |
| LALLICH | Stéphane | Ecole Doctorale InfoMaths | Représentant du laboratoire ERIC au Conseil de l'ED | 2007-2012 |
| LALLICH | Stéphane | Ecole Doctorale InfoMaths | Représentant du laboratoire ERIC au Comité des thèses de l'ED | 2007-2012 |
| LALLICH | Stéphane | Laboratoire ERIC | Responsable, axe FODA | 2010-2012 |
| LALLICH | Stéphane | Université Lyon 2 | Responsable, GEI 26-27-61-71 | Depuis 2008 |
| LALLICH | Stéphane | Université Lyon 2 | Responsable, Doctorat d'informatique | Depuis 2007 |
| LOUDCHER | Sabine | Laboratoire ERIC | Directrice adjointe | 2003-2012 |
| RAKOTOMALAL A | Ricco | Département Informatique et Statistique, ICOM Lyon 2 | Responsable, Master 2 IDS-SISE | Depuis 2013 |
| RICO | Fabien | Département Informatique, FST Lyon 1 | Co-responsable, Licence Informatique | 2009-2010 |
| ROLLAND | Antoine | IUT Lumière, Université Lyon 2 | Responsable, licence pro CESTAT | 2010-2013 |
| ROLLAND | Antoine | IUT Lumière, Université Lyon 2 | Chef du département STID | Depuis 2013 |

| NOM | Prénom | Établissement/composante | Fonction | Période |
|--------|--------|--|--|--------------|
| VELCIN | Julien | Département Informatique et Statistique, FSEG Lyon 2 | Directeur adjoint | 2010-2012 |
| VELCIN | Julien | Département Informatique et Statistique, ICOM Lyon 2 | Responsable Master ECD | 2007-2014 |
| VELCIN | Julien | Faculté Science Economique et de Gestion, Lyon 2 | Elu, conseil de faculté | 2008-2012 |
| VELCIN | Julien | Laboratoire ERIC | Responsable, Équipe DMD | Depuis 2012 |
| VELCIN | Julien | Université Lyon 2 | Membre élu, conseil scientifique | Depuis 2013 |
| ZIGHED | Djamel | Département Informatique et Statistique, ICOM Lyon 2 | Coordinateur, Master Erasmus Mundus DMKM | Depuis 2009 |
| ZIGHED | Djamel | Département Informatique et Statistique, ICOM Lyon 2 | Co-responsable Master ECD | Depuis 1999 |
| ZIGHED | Djamel | Institut des Sciences de l'Homme de Lyon | Directeur | Depuis 2011 |
| ZIGHED | Djamel | Laboratoire ERIC | Directeur | Jusqu'à 2012 |

Annexe 13 : Liste des formations suivies par le personnel administratif

| NOM | Prénom | Formation | Année | Durée (j) |
|----------------------|---------|--|-------|-----------|
| GABRIELE PIETROFORTE | Valérie | Stage d'observation à l'Université de Kingston, UK (programme ERASMUS) | 2009 | 6 |
| GABRIELE PIETROFORTE | Valérie | Rédiger des écrits professionnels | 2009 | 2 |
| GABRIELE PIETROFORTE | Valérie | Les tableaux de bord RH | 2009 | 2 |
| GABRIELE PIETROFORTE | Valérie | Stage extensif Anglais niveau B1+ | 2010 | 6,25 |
| GABRIELE PIETROFORTE | Valérie | Word : réalisation de documents structurés | 2010 | 1 |
| GABRIELE PIETROFORTE | Valérie | Valorisez vos travaux : outils et témoignages | 2010 | 1 |
| GABRIELE PIETROFORTE | Valérie | Anglais : mise en situation et écrits professionnels | 2011 | 1,5 |
| GABRIELE PIETROFORTE | Valérie | Formation SAP SIFAC | 2011 | 4,5 |
| GABRIELE PIETROFORTE | Valérie | Préparation méthodologie concours : cas pratique | 2012 | 3 |
| OSMAN | Habiba | Mise en œuvre du portail web dans le cadre du marché voyage | 2013 | 0,5 |
| OSMAN | Habiba | SIFAC recettes Lyon 2 | 2013 | 0,5 |
| OSMAN | Habiba | SIFAC commande d'achat / budget | 2013 | 1 |
| OSMAN | Habiba | SIFAC missions | 2013 | 0,5 |
| OSMAN | Habiba | EXCEL 2010 : listes et tableaux croisés dynamiques | 2014 | 1 |
| OSMAN | Habiba | Transmission de données : du visible à l'invisible | 2014 | 0,25 |
| OSMAN | Habiba | La recherche en France | 2014 | 0,5 |
| OSMAN | Habiba | Finances publiques - Gestion financière et comptable | 2014 | 1,5 |

Annexe 14 : Liste des séminaires ERIC depuis 2010

| Année | Intervenant | Organisme |
|-----------|-----------------|-------------------------|
| 2010-2011 | Richard Chbeir | Université de Bourgogne |
| 2010-2011 | Bruno Fantino | CPAM Rhône |
| 2010-2011 | Eric Fourboul | AMI Software |
| 2010-2011 | Eric Gaussier | Université Grenoble 1 |
| 2010-2011 | Patrick Giroux | EADS |
| 2010-2011 | Michel Grabisch | Université Paris 1 |

| Année | Intervenant | Organisme |
|-----------|----------------------|--------------------------------------|
| 2010-2011 | Eric Guichard | ENSSIB Lyon |
| 2010-2011 | Aurélien Lauf | AMI Software |
| 2010-2011 | Thomas Lebarbé | Université Grenoble 3 |
| 2010-2011 | Eric Lichtfouse | INRA |
| 2010-2011 | Charles Madeira | Masa Group |
| 2010-2011 | Stan Matwin | Université d'Ottawa |
| 2010-2011 | Rokia Missaoui | Université du Québec en Outaouais |
| 2010-2011 | Ahmed Mostefaoui | Université de Franche-Comté |
| 2010-2011 | Vincent Mousseau | École Centrale Paris |
| 2010-2011 | Claudiu Musat | Université Polytechnique de Bucarest |
| 2010-2011 | Gilbert Ritschard | Université de Genève |
| 2010-2011 | Lydia Rivoli | Université de Naples |
| 2010-2011 | Agnieszka Rusinowska | Université Paris 1 |
| 2010-2011 | Mondher Toumi | Université Lyon 1 |
| 2011-2012 | Ounas Asfari | Université Lyon 2 |
| 2011-2012 | Lydia Boudjeloud | Université de Lorraine |
| 2011-2012 | Mohand Boughanem | Université Toulouse 3 |
| 2011-2012 | Caroline Brun | Xerox |
| 2011-2012 | Pascal Cuxac | INIST-CNRS Nancy |
| 2011-2012 | Laurent d'Orazio | Université Clermont-Ferrand II |
| 2011-2012 | Marine Depecker | CEA |
| 2011-2012 | Thibaut Lust | Université Lyon 2 |
| 2011-2012 | Brice Mayag | Université Paris Dauphine |
| 2011-2012 | Wei Pang | Aberdeen University |
| 2011-2012 | Ioannis Partalas | Université Grenoble 1 |
| 2011-2012 | Farouk Toumani | Université Clermont-Ferrand II |
| 2012-2013 | Antoine Rolland | Université Lyon 2 |
| 2012-2013 | Mathilde Forestier | Université Lyon 2 |
| 2012-2013 | Nguyen Dinh Thuc | Ho Chi Minh University |
| 2012-2013 | Julien Subercaze | Université de Saint-Étienne |
| 2012-2013 | Ramzi Abbess | Techlimed |
| 2012-2013 | Brahim Chaib-Draa | Université Laval, Québec |
| 2012-2013 | Alexandre Janon | Université Lyon 1 |
| 2012-2013 | Lazhar Labiod | AgroParisTech |
| 2012-2013 | Farida Enikeeva | INRIA Rhône-Alpes |
| 2012-2013 | Annette Casagrande | CNRS Grenoble |
| 2012-2013 | Farid Boussama | Université Montpellier 1 |
| 2012-2013 | Sarah Ouadah | Université Paris 6 |
| 2013-2014 | Audrey Baneyx | Sciences Po Paris |
| 2013-2014 | Badih Ghattas | Aix Marseille Université |
| 2013-2014 | Gabriela Ciuperca | Université Lyon 1 |
| 2013-2014 | Nicolas Labroche | Université Paris 6 |
| 2013-2014 | Ludovic Denoyer | Université Paris 6 |
| 2013-2014 | Julien Jacques | Université Lille 1 |
| 2013-2014 | Mustapha Lebbah | Université Paris 13 |
| 2013-2014 | Jean-Charles Lamirel | Université de Strasbourg |

Annexe 15a : Rapport d'évaluation du laboratoire (AERES, 2010)

Rapport d'évaluation

| | |
|----------------------------|---|
| Nom de l'unité : | Équipe de Recherche en Ingénierie des Connaissances |
| Acronyme de l'unité : | ERIC |
| Label demandé : | EA |
| N° actuel : | 3083 |
| Nom du directeur : | M. Djamel ZIGHED |
| Nom du porteur de projet : | |

Membres du comité d'experts

| | |
|---|--|
| Président : | M. Frédéric BENHAMOU, Université de Nantes, LINA |
| Experts : | M. Colin DE LA HIGUERA, Université de Nantes, LINA |
| | M. Philippe PREUX, Université de Lille 3, LIFL |
| | M. Michel SCHNEIDER, Université de Clermont-Ferrand-ISIMA, LIMOS |
| Expert(s) proposés par des comités d'évaluation des personnels (CNU, CoNRS, CSS INSERM, représentant INRA, INRIA, IRD...) | Mme Thérèse LIBOUREL (CNU) |
| Délégué scientifique représentant de l'AERES : | M. Michel RIVEILL |
| Représentant(s) des établissements et organismes tutelles de l'unité : | Mme N. FOURNIER (VPCS Lyon 2) |
| | Mme C. GOUTAUDIER (VPCS Lyon 1) |

1 • Introduction

La visite s'est déroulée le 8 février 2010 dans les locaux du laboratoire ERIC. La visite a débuté vers 9h et s'est terminée aux alentours de 18h30, comme prévu dans le planning préparé avec la direction du laboratoire. Après une présentation générale du laboratoire et de son bilan scientifique, les trois équipes puis le laboratoire ont exposé leurs bilans et leurs projets scientifiques. Le comité a également eu l'opportunité d'interviewer les représentantes des universités de Lyon 1 et Lyon 2 et de rencontrer des représentants des doctorants et des enseignants-chercheurs, ainsi que l'équipe de direction et les personnels BIATOS du laboratoire séparément.

Le laboratoire ERIC a été créé en 1995 comme jeune équipe de l'Université de Lyon 2, qui regroupe le droit, les lettres et les sciences humaines. Il a été labellisé équipe d'accueil par le ministère de l'éducation nationale, de la recherche et de la technologie en 1999. Le laboratoire est situé sur le campus de la Porte des Alpes de l'Université Lumière Lyon 2. Suite à l'arrivée d'une nouvelle équipe en provenance de Lyon 1, une seconde localisation a été ouverte sur le campus Laënnec.

Le domaine d'activité du laboratoire est essentiellement celui de l'extraction de connaissances à partir des données (ECD). Les trois équipes du laboratoire couvrent l'ensemble des problématiques liées à ce thème de recherche, à savoir un aspect bases de données autour des entrepôts de données (équipe ENA DC), un aspect classification et apprentissage pour la fouille de données (équipe FODA) et enfin un aspect décision, qui intègre l'utilisateur pour prendre en compte ses stratégies (équipe DECCO). Cette dernière équipe, nouvelle, est liée à l'arrivée d'une équipe de Lyon 1 qui intègre ERIC à l'occasion de ce nouveau quadriennal. La recherche est bien répartie sur le continuum théorie, logiciels, applications, en particulier vers les SHS et la santé.

L'équipe de direction est principalement composée du directeur et de la directrice adjointe. Un second directeur adjoint représente le site de Laënnec au sein de l'équipe de direction. Il est envisagé d'étendre l'équipe de direction aux responsables d'équipe si cela s'avère nécessaire.

Effectifs de l'unité :

Le bilan ne considère pas l'équipe de Lyon 1 qui vient de rejoindre le laboratoire.

| Effectifs de l'unité | Dans le bilan | Dans le projet |
|---|---------------|-----------------|
| N1 : Nombre d'enseignants-chercheurs | 11 | 21 |
| N2 : Nombre de chercheurs des EPST ou EPIC | 0 | 0 |
| N3 : Nombre d'autres enseignants-chercheurs et chercheurs | 0 | 5 ¹¹ |
| N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires | 0,5 | 1 |
| N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires | 0,2 | 0,7 |
| N6 : Nombre de doctorants | 15 | 25 |
| N6 : Nombre de personnes habilitées à diriger des recherches ou assimilées | 5 | 9 |

¹¹ Les 5 autres enseignants chercheurs sont essentiellement des médecins qui collaborent de manière très étroite avec le laboratoire.

2 • Appréciation sur l'unité

L'unité de recherche ERIC est une unité de petite taille (11 permanents dans le bilan, 21 dans le projet présenté) qui a su concentrer son développement sur une thématique bien identifiée en y intégrant l'ensemble des aspects scientifiques, technologiques et applicatifs. L'unité s'est également attachée à être un centre du développement d'activités d'enseignement sur ses compétences de recherche, en collaboration avec d'autres sites français et européens, ce qui a amené à la labellisation d'un Master Erasmus Mundus en 2009. Le développement et la maintenance de logiciels, ainsi que le développement d'applications dans deux domaines cohérents avec l'implantation et les collaborations industrielles des chercheurs sont également des points forts. Un autre point remarquable est la cohésion humaine du laboratoire, qui a semblé uni derrière un projet et un directeur, tout en ayant conscience de certaines faiblesses et en étant déterminé à y remédier.

Les principaux points faibles concernent la politique de publication internationale, actuellement principalement quantitative, le nombre insuffisant de contrats et projets sélectifs de type ANR et européens du PCRD et le positionnement stratégique de l'unité dans son environnement immédiat, qui est peu lisible. Ce dernier point concerne par exemple le positionnement du laboratoire ERIC vis à vis du LIRIS, du LIESP et de l'institut des systèmes complexes. Enfin, l'effort d'organisation prévu pour le futur quadriennal, qui voit à la fois le doublement des effectifs permanents et l'éclatement sur deux sites, pourrait s'avérer insuffisant. En particulier, son aspect organisationnel met en oeuvre une structuration qui n'a pas su choisir entre le fonctionnement par projet et celui par équipe thématique, en proposant une structure matricielle complexe à expliciter et difficile à lire.

Malgré cela, le comité juge que le laboratoire, en particulier à l'occasion de la préparation de cette première évaluation par un comité de visite, a pris conscience de ses faiblesses et s'engage résolument dans une nouvelle voie, en termes d'objectifs qualitatifs et d'organisation. L'arrivée de la nouvelle équipe est scientifiquement pertinente et il semble y avoir une volonté d'intégration réelle de part et d'autre.

Pour accompagner cette dynamique, il est important que les établissements règlent rapidement les problèmes de locaux (surfaces sur Lyon 2, mais aussi problèmes liés à la bi localisation) et augmentent les moyens humains du laboratoire en termes de personnel technique dédié au développement et au suivi de logiciels. Le comité tient à faire remarquer aux établissements que, par comparaison, ce type de personnel est aussi important pour les laboratoires d'informatique que le sont les personnels affectés aux activités d'expérimentation en physique ou en biologie.

Le comité conseille donc à l'unité de poursuivre son travail dans la direction engagée, afin de se concentrer sur ses points forts, de préciser son positionnement et éventuellement de renforcer des collaborations avec les autres unités lyonnaises connexes (par exemple Projet Web Intelligence dans le cluster régional), afin de se hisser à un niveau européen en recherche qui rejoigne celui obtenu en enseignement avec l'attribution du label Erasmus Mundus.

Données de production

| | |
|---|--------|
| A1 : Nombre de producteurs parmi les chercheurs et enseignants chercheurs référencés en N1 et N2 dans la colonne projet | 15 |
| A2 : Nombre de producteurs parmi les autres personnels référencés en N3, N4 et N5 dans la colonne projet | 4 |
| A3 : Taux de producteurs de l'unité $[A1/(N1+N2)]$ | 71,4 % |
| Nombre d'HDR soutenues (ERIC-Lyon 2) 2006-2009 | 3 |
| Nombre de thèses soutenues (ERIC-Lyon 2) 2006-2009 | 12 |

3 • Appréciations détaillées

Appréciation sur la qualité scientifique et la production

Le projet scientifique dans son ensemble est cohérent et ambitieux. Il attaque des problèmes difficiles avec le souci d'obtenir des résultats aussi bien sur le plan théorique que sur celui des algorithmes et de l'implémentation, puis des applications. On retrouve ces caractéristiques et ces qualités sur les trois équipes du laboratoire.

Le comité a cependant noté une différence notable entre la qualité des travaux menés dans les équipes et celle des publications qui en résultent et donc de l'impact général des travaux. Les résultats obtenus, alors qu'ils en ont souvent le potentiel, ne sont pas assez publiés dans les revues et les conférences de référence des différents domaines abordés. Le directeur du laboratoire, qui a effectué un remarquable travail de classification des publications à l'aide des listes CORE et ERA (celui-ci a été remis au comité lors de la visite), obtient d'ailleurs des résultats qui confirment l'appréciation après étude du dossier. De plus, environ 30% des membres du laboratoire sont non producteurs.

On ne saurait donc trop conseiller aux membres du laboratoire et aux équipes de changer de politique de publication en identifiant les workshops, les conférences et les revues internationales cibles qui sont les références indiscutables dans leurs domaines et en visant

(a) lorsque c'est nécessaire, surtout pour les jeunes doctorants et docteurs, les meilleurs workshops pour tester les résultats auprès des spécialistes internationaux incontournables puis

(b) les conférences internationales cibles (voir dans les remarques par équipe pour des exemples) pour publier les résultats rapidement et enfin

(c) les consolider par une publication dans les revues internationales cibles sélectionnées. Cela n'empêche pas de participer aux conférences et aux revues nationales, en particulier pour animer la communauté.

Appréciation sur le rayonnement, l'attractivité, et l'intégration de l'unité de recherche dans son environnement

Le laboratoire ERIC a un rayonnement national certain. Il est l'un des acteurs majeurs de l'animation de la communauté et des formations dispensées sur ses domaines de compétence. Ses membres sont très impliqués dans cette animation et dans l'organisation scientifique et pratique des conférences nationales du domaine (EGC, EDA, nombreux ateliers, etc.). La création du Master ECD, puis du Master Erasmus Mundus sont également des éléments de rayonnement. Certains membres du laboratoire ont un rayonnement international notable, par la qualité de leurs publications et en termes de participation à des comités de lecture et de programme de conférences internationales de bon niveau. Mais ils restent cependant minoritaires.

Sur le plan de l'attractivité, le nombre de recrutements sur la période a été limité, mais le rattachement d'une nouvelle équipe, dans son intégralité, montre un potentiel réel dans ce sens, si l'on prend en compte que le rattachement à une université plutôt dédiée aux Sciences Humaines et Sociales peut de ce point de vue parfois représenter un frein.

Un point faible est le manque de participation aux programmes de recherche sélectifs nationaux (ANR) et internationaux (PCRD), si l'on excepte la nouvelle équipe DECCO qui rejoint le laboratoire. Là encore, la qualité des travaux menés, les relations industrielles et la maîtrise de domaines d'application interdisciplinaire devrait permettre au laboratoire de mieux réussir sur ce point, qui devient un indicateur central pour l'évaluation des structures et des équipes. En dehors de ces appels d'offre, le laboratoire fait preuve à la fois de capacités à attirer des financements externes avec des grands groupes (Crédit Lyonnais, Sanofi-Pasteur) et des PME locales. Le montant des contrats privés est cependant en forte baisse au cours des dernières années mais est compensé par un investissement plus important du conseil Régional ces deux dernières années.

Sur le plan de la valorisation, il faut noter l'activité d'incubateur joué par le laboratoire, en relation avec CREALYS. Il ne s'agit pas de spin-offs du laboratoire mais de start-up qui s'appuient sur l'expertise des chercheurs pour développer techniquement leurs produits et leur offre commerciale. Il s'agit là d'une

participation du laboratoire à l'éclosion d'un écosystème en lien avec les recherches du laboratoire qui est original et intéressant. Ajoutons sur ce plan, la production de plusieurs logiciels en open source, dont TANAGRA, qui fait l'objet d'une très large utilisation académique.

Les relations internationales sont assez nombreuses et motivées par des collaborations sur des Masters et parfois sur des co-encadrements de thèses. Quelques co-publications sont mentionnées.

Appréciation sur la stratégie, la gouvernance et la vie de l'unité

Les objectifs de l'unité sont clairs et cohérents : s'attaquer à l'ensemble des problématiques de l'extraction de connaissances à partir des données, depuis la représentation et le traitement de données complexes jusqu'à l'aide à la décision. Cet aspect à la fois bien focalisé mais aussi suffisamment large pour couvrir la totalité du cycle doit permettre à la fois de réaliser des percées sur des points précis (en maîtrisant l'ensemble de l'environnement et la sérialisation des phases de traitement) mais aussi de pouvoir développer une offre globale, que ce soit pour la formation ou le développement de logiciels et d'applications. Un autre objectif connexe est de très fortement intégrer la recherche et le développement des spécialités de Masters, ce qui a permis entre autres l'obtention du Master Erasmus Mundus.

La stratégie utilisée pour y parvenir s'est concrétisée par le recrutement d'enseignants-chercheurs sur des thématiques peu présentes dans l'unité et par l'intégration d'une nouvelle équipe dans son intégralité qui apporte au laboratoire des compétences en amont de la chaîne (traitement des données complexes) et en aval (aide à la décision). Cette nouvelle équipe apporte également un domaine d'application très large (santé) et une compétence avérée en termes de capacité à obtenir et à gérer des contrats, en particulier européens.

La gouvernance actuelle est satisfaisante, et combine une vision scientifique claire, une bonne lisibilité en termes de stratégie, et un fonctionnement qui semble convenir à l'ensemble des membres du laboratoire.

Le comité rappelle ici que le laboratoire avec l'aide des universités Lyon 1 et 2 doit mieux affirmer le positionnement stratégique de l'unité dans son environnement immédiat, qui est peu lisible en particulier dans le recouvrement scientifique entre les travaux menés au LIRIS, au LIESP ou à l'institut des systèmes complexes.

Appréciation sur le projet

Le projet scientifique présenté s'appuie principalement sur 3 points :

- la continuité des travaux engagés dans les 3 équipes et sur les 2 domaines d'application SHS et santé,
- l'intégration de la nouvelle équipe DECCO,
- une organisation matricielle en équipes et en domaines d'applications transversaux.

Le comité est conscient que l'exercice qui consiste à rédiger un projet quadriennal de recherche pour l'unité est nouveau pour le laboratoire. Si le projet en lui-même, d'après ce que le comité en a perçu lors de la visite, est cohérent scientifiquement et bénéficie de l'adhésion des enseignants-chercheurs impliqués, le développement présenté dans le dossier souffre de nombreux défauts.

En premier lieu, le texte est extrêmement court (essentiellement 2 pages pour le projet de l'Unité, auxquelles il faut ajouter les annexes de présentation des équipes) et se contente de décrire de manière très générale les challenges du domaine de recherche sans préciser quels sont les objectifs scientifiques à quatre ans et surtout sans décrire les pistes envisagées pour atteindre ces objectifs.

En second lieu, l'organisation scientifique proposée semble à la fois refléter l'état d'une réflexion qui n'est pas encore parvenue à son terme et présenter un réel danger de confusion managériale pour les personnels. En effet, avec un rattachement scientifique principal et secondaire et un rattachement à un domaine d'application transversal, chaque enseignant-chercheur se retrouve structurellement et simultanément engagé dans trois entités a priori distinctes et ayant chacune son organisation propre. Il semble préférable de se limiter à une seule dimension, soit celle des équipes, dans un schéma traditionnel, soit, dans une volonté d'intégration de l'interdisciplinarité au coeur du projet scientifique, par domaines d'applications transversaux.

Troisièmement, si les compétences de l'équipe entrante sont présentées dans leur cohérence avec le cadre scientifique de l'unité, l'intégration elle-même est peu précisée. Cela concerne à la fois le projet scientifique, la capacité à mobiliser l'unité dans ses réponses aux appels d'offre sélectifs nationaux et internationaux et même la gestion quotidienne de la bi-localisation.

Il manque aussi, que ce soit dans le bilan ou le projet un positionnement de l'unité dans le paysage national, européen et international en précisant les complémentarités, les éléments de différenciation et une évaluation grossière des forces comparées de l'unité par rapport à ses concurrents.

Conclusion

En conclusion, ERIC est un laboratoire qui présente une grande cohérence, à la fois sur le plan de ses choix de développement scientifiques et pédagogiques (lien recherche-Masters) et sur celui de la cohésion des personnels. Le directeur a su s'approprier le projet scientifique global de son unité, qu'il a présenté de façon particulièrement convaincante et convaincre les chercheurs de conserver cap et cohésion. En conséquence, il règne dans ce laboratoire une excellente ambiance de travail et une réelle solidarité.

Les principales difficultés concernent l'écart qui subsiste entre la qualité globale du travail effectué qui est satisfaisante, et le manque de résultats de la valorisation des résultats, que ce soit en termes de publications dans des conférences et revues de référence ou de réussite à l'ANR ou au PCRD. Le comité s'est également posé des questions sur le faible positionnement régional vis-à-vis des autres acteurs impliqués, et les questions ouvertes sur la capacité des deux entités qui fusionnent à créer un nouveau projet d'intégration et une ambition commune qui aillent au-delà de la superposition des compétences. Une attention particulière devra aussi être portée à la question du renouvellement générationnel des cadres dans certaines équipes.

La visite a permis d'évaluer les marges et les capacités de progression sur ces différents points et l'avis du comité est que l'unité s'est dotée d'atouts qui devraient lui permettre de progresser de manière significative au cours du quadriennal à venir.

4 • Analyse équipe par équipe et/ou par projet

| | |
|----------------------|---|
| Équipe 1 : | Axe ENA-DC, Entrepôts et Analyse de Données Complexes |
| Nom du responsable : | M. Jérôme Darmont |

Effectifs

Pour le bilan du nouvel axe ENA-DC on se base que sur la configuration de Lyon 2.

| Effectifs de l'équipe | Dans le bilan | Dans le projet |
|---|---------------|----------------|
| N1 : Nombre d'enseignants-chercheurs | 5 | 6 |
| N2 : Nombre de chercheurs des EPST ou EPIC | | |
| N3 : Nombre d'autres enseignants-chercheurs et chercheurs | | |
| N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires | | |
| N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires | | |
| N6 : Nombre de doctorants | 7 | |
| N7 : Nombre de personnes habilitées à diriger des recherches ou assimilées | 2 | 2 |

Appréciations détaillées

Appréciation sur la qualité scientifique et la production

L'équipe ENA-DC (Entrepôts et Analyse de Données Complexes) constitue l'équipe amont de la boucle entreposage-fouille-connaissances-décision que l'unité cherche à mettre en place. Les recherches menées couvrent un large spectre couvrant la modélisation de données complexes (images, documents, etc.), leur intégration dans des structures multidimensionnelles pérennes (entrepôts), leur interaction avec les techniques de fouille de données. L'interaction avec l'équipe FODA est naturelle mais devrait être renforcée pour atteindre les objectifs annoncés d'ETL intelligent. Les problèmes abordés sont bien d'actualité et l'équipe bénéficie du dynamisme du domaine.

Les publications sont nombreuses mais de qualités inégales. Une part importante est constituée de chapitres d'ouvrages. 31 publications sont significatives et parmi celles-ci 15 sont sélectives. Globalement la productivité peut être qualifiée de convenable sans plus. 4 thèses ont été soutenues. C'est un chiffre modeste qui s'explique en partie par le fait que sur la moitié de la période l'encadrement HDR était assuré par une seule personne (pas de HDR avant fin 2006). Deux HDR ont été soutenues il y a deux ans et on note maintenant une augmentation du nombre d'inscriptions en thèse (9 thèses en cours dont 2 qui devraient être soutenues prochainement).

On note aussi une activité en termes de production et de diffusion de logiciels notamment DWEB (Data Warehouse Engineering Benchmark). Ce type d'activité est louable. Il permet une forme de validation des travaux réalisés et de reconnaissance de compétences.

Appréciation sur le rayonnement, l'attractivité, et l'intégration de l'équipe ou du projet dans son environnement

L'équipe jouit d'une bonne notoriété nationale liée à l'animation des journées Entrepôts de Données et Analyse en ligne (EDA) depuis 2005 et Atelier Systèmes Décisionnels (ASD) depuis 2006.

On note la participation à deux ACI et à 3 contrats industriels.

La notoriété internationale s'établit par le biais de plusieurs collaborations avec des universités étrangères (Université Laval et Université du Québec au Canada, École Nationale d'informatique de Tunis, Université d'Oklahoma, Université de Fianarantsoa à Madagascar) qui se traduisent par des publications en commun, des échanges d'enseignants-chercheurs ou des co-encadrements de thèses.

Appréciation sur la stratégie, la gouvernance et la vie de l'équipe ou du projet

On note un bon dynamisme au sein de cette équipe avec des réunions hebdomadaires et un intranet collaboratif.

Les membres de l'équipe ont une très forte implication dans les structures administratives de l'université Lyon 2 et dans les formations (master multi localisé en extraction des connaissances, master en ingénierie pour la décision et l'évaluation économique, master sécurité des systèmes d'information dans les organisations). Ils participent aussi à une formation d'excellence Erasmus Mundus qui regroupe six universités européennes.

Appréciation sur le projet

Pour l'essentiel le projet reprend les mêmes thèmes en affirmant plus systématiquement le croisement avec les applications dans le secteur SHS et le secteur de la Santé. Comme cela est souligné plus loin, le spectre thématique affiché est trop large.

Conclusion :

● Avis :

Il s'agit indéniablement d'une équipe qui fait preuve d'un bon dynamisme mais dont l'activité est pénalisée par une très forte implication dans les structures d'administration et d'enseignement de l'université Lyon2. Les choix thématiques sont bons mais il faudrait réduire le spectre et mener des investigations plus profondes. La synergie avec l'équipe FODA recèle des possibilités intéressantes et doit être renforcée.

- Points forts opportunités :

La proximité avec le secteur SHS et le secteur Santé apporte des problèmes intéressants en liaison directe avec le domaine de recherche.

- Points à améliorer et risques :

Le principal risque est celui de la dispersion. Les thèmes abordés sont trop nombreux pour une équipe somme toute assez réduite. Il vaudrait mieux réduire le spectre et se focaliser sur quelques points durs. Il faut améliorer la qualité de la production scientifique en privilégiant les publications dans des revues internationales sélectives. Il faut maintenir la progression du nombre de thèses soutenues.

Il est aussi apparu que les ressources financières étaient limitées pour couvrir les besoins des enseignants-chercheurs et des doctorants (malgré un financement récurrent relativement important). Il faudrait donc envisager une plus forte implication dans des contrats pour générer des ressources propres plus importantes.

Il faut améliorer la visibilité internationale.

- Recommandations :

L'objectif d'analyse en ligne intelligent est ambitieux, les recherches devraient se resserrer sur ce type d'objectif (opérateurs spécifiques en lien avec la fouille de données) en renforçant les liens avec l'équipe FODA.

L'élaboration de logiciels est une tâche très souhaitable mais qui peut se révéler très lourde. Ne serait-il pas possible de la mener en coopération avec des équipes nationales ou internationales ?

Une politique plus affirmée de coopération avec les pays du Maghreb et de l'Europe de l'Est devrait permettre d'accroître la capacité de l'équipe à attirer des doctorants financés par leur pays.

4 • Analyse équipe par équipe et/ou par projet

| | |
|----------------------|--|
| Équipe 2 : | Axe FODA (Fouille de Données et Apprentissage) |
| Nom du responsable : | M. Stéphane Lallich |

Effectifs

| Effectifs de l'équipe | Dans le bilan | Dans le projet |
|---|---------------|----------------|
| N1 : Nombre d'enseignants-chercheurs | 5 | 8 |
| N2 : Nombre de chercheurs des EPST ou EPIC | | |
| N3 : Nombre d'autres enseignants-chercheurs et chercheurs | | |
| N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires | | |
| N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires | | |
| N6 : Nombre de doctorants | 8 | |
| N7 : Nombre de personnes habilitées à diriger des recherches ou assimilées | 3 | 4 |

FODA est une équipe composée de 8 enseignants-chercheurs, tous en rattachement principal. Mais c'est aussi l'équipe « rattachement secondaire » de 8 autres enseignants-chercheurs.

Les recherches des membres de l'équipe FODA, sur la période 2005-2008 ont concerné :

- Des questions relatives à l'utilisation d'autres arguments que le support et la confiance dans les algorithmes d'extraction de règles d'association.
- La recherche de règles à la fois compréhensibles mais avec des performances (au moins) similaires à celles d'autres systèmes. La comparaison avec les forêts randomisées, en particulier, a été menée.
- Des questions de discrétisation ont été étudiées, en particulier les difficultés liées à l'obtention d'une bonne discrétisation à partir de données échantillonnées.
- Les problèmes de déséquilibres de classes ont également fait l'objet d'études. Ici encore, la difficulté étudiée concerne le biais introduit par l'échantillonnage.
- L'apprentissage topologique, où c'est l'interdépendance entre les données qui explique celles-ci, a été un thème moteur. On note que ce thème de recherche rejoint d'autres étudiés dans les communautés internationales en apprentissage automatique, parfois sous d'autres noms.

- La fouille de réseaux sociaux et de corpus est un thème qui intéresse l'équipe d'autant plus que cela permet la collaboration avec les autres équipes du laboratoire.

On notera que sur l'ensemble de ces questions, la volonté affichée est d'avoir une approche à la fois théorique (une analyse mathématique précise est effectuée) et appliquée à des cas correspondant aux collaborations d'ERIC, donc dans des contextes médicaux ou des sciences humaines.

Le point fort de l'équipe est que sur les thématiques recensées, leur potentiel basé sur la présence de collègues à la fois de la section 26 et de la section 27 est un atout majeur pour les questions traitées. Cela leur permet d'avoir un recul particulier sur ces questions.

Un autre point fort de cette équipe réside dans le travail impressionnant d'animation de la communauté francophone effectué par plusieurs chercheurs d'ERIC depuis plusieurs années. Ils ont contribué (largement) à hisser la conférence EGC à un niveau très intéressant, et leur travail permet un essaimage d'idées vers d'autres pays francophones.

Un troisième point fort concerne le travail de diffusion de logiciels de fouille de données et d'apprentissage. C'est un travail remarquable qui a conduit au développement d'un site web permettant le téléchargement de nombreux logiciels et qui correspond à plusieurs centaines de connexions par jour.

D'autres points très favorables concernent l'investissement très important dans la réalisation du continuum recherche / enseignement / parcours du master informatique, bourses Cifre, enseignements en visioconférence. Ce travail vient de se voir récompensé par l'obtention d'un label Erasmus Mundus pour une filière de Master, ce qui va sans doute permettre à l'équipe de collaborer au plus haut niveau international.

Le grand regret que l'on peut avoir tient à la stratégie de publications suivie : étant donnée la qualité des chercheurs et de certaines des idées défendues, on se serait attendu à trouver leurs résultats publiés dans les conférences et revues de référence sur ces questions : the Machine Learning Journal, JMLR, ICML, KDD, PKDD, ECML, SDM, ICDM, et, pour les questions les plus théoriques COLT, ALT, NIPS ou AISTATS. Or ce n'est pas le cas et c'est tout à fait dommage.

Un autre défaut est que les publications les plus visibles semblent concentrées entre quelques chercheurs.

Enfin, s'il est clair que l'on dénombre un nombre conséquent de doctorants dans l'équipe, leur devenir doit être précisé.

La recommandation serait de viser dès le départ le plus haut niveau de publication, y compris (voire surtout) pour les plus jeunes membres de l'équipe. Cela doit être le cas au moins quand les recherches portent sur des problèmes centraux de l'apprentissage et de la fouille de données (comme le sont la plupart de ceux répertoriés plus haut) et pour lesquels il existe donc de réelles possibilités de publication à fort impact.

La présence d'un réel savoir faire en matière de développement de logiciel de référence doit inciter les chercheurs de l'équipe à chercher la confrontation expérimentale systématique de leurs résultats avec ceux de la communauté internationale sur les bancs d'essai reconnus.

On ne peut qu'encourager l'équipe à renforcer son attractivité, diversifier les sources de recrutements dans l'avenir, tenter d'attirer des jeunes maîtres de conférence ayant un fort potentiel et déjà une certaine reconnaissance internationale.

4 • Analyse équipe par équipe et/ou par projet

| | |
|----------------------|------------------------------------|
| Équipe 3 : | Axe DECCO (Décision et Complexité) |
| Nom du responsable : | M. Stéphane Bonnevey |

Effectifs

Pour le bilan du nouvel axe DECCO on se base sur l'équipe de Lyon 1 qui a rejoint ERIC. Ce bilan ne figure pas dans le dossier déposé à l'AERES.

| Effectifs de l'équipe | Dans le bilan | Dans le projet |
|---|---------------|----------------|
| N1 : Nombre d'enseignants-chercheurs | 8 | 6 |
| N2 : Nombre de chercheurs des EPST ou EPIC | | |
| N3 : Nombre d'autres enseignants-chercheurs et chercheurs | 5 | 5 |
| N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires | | |
| N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires | | |
| N6 : Nombre de doctorants | 21 | |
| N7 : Nombre de personnes habilitées à diriger des recherches ou assimilées | 3 | 3 |

Concernant les effectifs doctorants, les sources divergent : ERIC indique 14, alors que le document « Fiches d'activités des doctorants » en recense 8. De même, les 21 doctorants du bilan proviennent d'ERIC : aucune trace dans les rapports d'activité.

L'équipe DECCO est en cours d'intégration dans le laboratoire ERIC ; elle s'est séparée du LIRIS pendant le contrat quadriennal en cours pour rejoindre ERIC et est actuellement située dans des locaux indépendants du LIRIS et d'ERIC.

Appréciations détaillées

Appréciation sur la qualité scientifique et la production

Les recherches menées dans DECCO concernent les problématiques suivantes :

- analyse et modélisation de la prise de décision collective (en environnement complexe),
- modélisation des réseaux complexes et caractérisation de leur évolution.

Un réseau complexe est ici entendu comme un réseau à structure multi-échelles, contenant des données hétérogènes. Les problématiques sont actuellement très fortement ancrées dans le domaine de la santé. Cette équipe est associée à l'école doctorale interdisciplinaire Science-Santé.

En termes de méthodologie, l'équipe met en avant à la fois l'utilisation d'outils mathématiques (logique et statistiques) et de simulations à base d'agents.

Au sein d'ERIC, l'objectif de cette équipe (complémentaire des équipes ENA-DC et FODA de l'unité ERIC) est de fermer la boucle allant des problématiques liées aux données (ENA-DC), à leur analyse (FODA), puis à la prise de décision. Il convient de noter qu'il s'agit là d'un objectif explicite ; il reste à l'atteindre, et à se donner les moyens de l'atteindre.

Globalement, l'activité de publications est relativement faible en moyenne ; en particulier, de nombreuses publications sont réalisées dans le domaine de la santé, peu en informatique et parmi celles-ci, on ne peut qu'encourager les chercheurs de l'équipe à publier dans des congrès et journaux de bonne qualité.

Le nombre de thèses soutenues n'apparaît pas dans le dossier. 8 thèses sont en cours. Il est important de noter que les doctorants sont inscrits à l'École doctorale interdisciplinaire Sciences Santé.

Appréciation sur le rayonnement, l'attractivité, et l'intégration de l'équipe ou du projet dans son environnement

L'équipe jouit d'une bonne notoriété nationale. Il est membre de l'Association d'Économétrie Appliquée et de la Société Française d'Économie de la Santé. Il assure la coresponsabilité du groupe de travail « Décision distribuée » de la société ROADEF (Recherche Opérationnelle et Aide à la Décision Française). Il a par ailleurs porté un projet public et participé à 5 autres. Il édite la revue « Santé Décision Management ».

La notoriété internationale s'établit par la coordination d'un projet européen (European Consortium for Patient Outcome) et le biais de plusieurs collaborations avec des universités et des organismes étrangers (Université de Heidelberg en Allemagne, Université de Cheikh Anta Diop au Sénégal, Institut de recherche en santé publique au Burkina-Faso, Banque Mondiale à Washington, Centre d'Études et de Documentation Économique au Caire, Bureau International du Travail à Genève) qui se concrétisent essentiellement par des projets menés en commun. L'équipe a aussi organisé la 9ème conférence internationale « On Systems Science in Health Care ».

Appréciation sur la stratégie, la gouvernance et la vie de l'équipe ou du projet

Les membres de l'équipe ont porté le master « Sciences de la santé » depuis sa création jusqu'à sa fusion avec le master AMIV. Ils s'impliquent dans différentes unités d'enseignement de l'université Lyon 1 et de l'université Lyon 2.

Appréciation sur le projet

Pour l'essentiel le projet reprend les mêmes thèmes qui ont été étudiés dans la période 2005-2008. Les objectifs visés devraient être mieux précisés. Il est dommage que dans ce projet le croisement fondements - applications que l'unité met en avant ne soit pas mieux établi. Certaines investigations théoriques sont menées sans souci des applications. Inversement certaines applications ne correspondent pas à des thématiques étudiées et s'inscrivent donc plutôt dans une perspective d'ingénierie.

Conclusion :

● Avis :

Cette équipe fait preuve d'un certain dynamisme mais le projet de recherche n'est pas assez précis et sélectif. Il faudrait l'ancrer définitivement au sein de l'unité et surmonter le risque de l'éloignement géographique. La production scientifique reste modeste et de qualité très moyenne. Les collaborations internationales sont diversifiées mais il conviendrait sans doute de privilégier quelques-unes en les renforçant et en visant notamment l'élaboration de productions communes.

● Points forts opportunités :

L'équipe jouit d'une bonne visibilité au niveau national et international. L'impulsion donnée par le projet européen devrait permettre à l'équipe d'assurer des objectifs plus clairs et de renforcer la politique de publications.

● Points à améliorer et risques :

Il faut améliorer la qualité de la production scientifique en privilégiant les publications dans des revues internationales sélectives. La mise en cohérence avec les deux équipes ENA-DC et FODA est bien affirmée, mais il faudra la mettre en oeuvre.

L'équipe doit réfléchir à une stratégie permettant d'amener les maîtres de conférences à soutenir une habilitation à diriger les recherches, ce qui s'appuie notamment, et encore, sur une stratégie de publications dans les meilleurs congrès et journaux.

L'équipe comprend un grand nombre de doctorants et elle doit veiller à leur avenir.

● Recommandations :

La première priorité est de mieux préciser les objectifs de recherche en visant la cohérence et la synergie avec les deux autres équipes.

Il faut aussi que l'équipe se donne les moyens d'améliorer sa production scientifique en visant d'abord la qualité. La production est importante dans le domaine de la santé, très insuffisante en informatique (viser systématiquement des revues telles JAIR, AI, conférences IJCAI, AAMAS pour ce qui est de la décision collective, des conférences comme WWW, SDM, KDD pour les aspects réseaux sociaux). Un meilleur pilotage de l'équipe devrait permettre d'améliorer ce point rapidement.

Il faut enfin prévoir un meilleur suivi des EC pour augmenter le nombre de HDR.

| Note de l'unité | Qualité scientifique et production | Rayonnement et attractivité, intégration dans l'environnement | Stratégie, gouvernance et vie du laboratoire | Appréciation du projet |
|-----------------|------------------------------------|---|--|------------------------|
| B | B | B | A | B |

Nom de l'équipe : ENA-DC

| Note de l'équipe | Qualité scientifique et production | Rayonnement et attractivité, intégration dans l'environnement | Stratégie, gouvernance et vie du laboratoire | Appréciation du projet |
|------------------|------------------------------------|---|--|------------------------|
| B | B | B | non noté | B |

Nom de l'équipe : FODA

| Note de l'équipe | Qualité scientifique et production | Rayonnement et attractivité, intégration dans l'environnement | Stratégie, gouvernance et vie du laboratoire | Appréciation du projet |
|------------------|------------------------------------|---|--|------------------------|
| B | B | A | non noté | B |

Nom de l'équipe : DECCO

| Note de l'équipe | Qualité scientifique et production | Rayonnement et attractivité, intégration dans l'environnement | Stratégie, gouvernance et vie du laboratoire | Appréciation du projet |
|------------------|------------------------------------|---|--|------------------------|
| B | B | A | non noté | B |

Annexe 15b : Rapport d'évaluation à mi-parcours (2012)

| Rapport d'évaluation

| | |
|--|--|
| Nom de l'unité : | Entrepôts, Représentation & Ingénierie des Connaissances |
| Acronyme de l'unité : | ERIC |
| Label demandé : | EA |
| N° actuel : | 3083 |
| Nom du directeur (2012-2013) : | Jérôme Darmont |
| Nom du porteur de projet (2014-2018) : | |

| Membres du comité d'experts

| | |
|-------------|--|
| Président : | Monsieur Jin-Kao Hao, Université d'Angers |
| Experts : | Monsieur Bernard Espinasse, Aix-Marseille Université (AMU) |
| | Monsieur Guy Melançon, Université de Bordeaux I |

| | |
|--|---|
| Représentant(s) des établissements et organismes tutelles de l'unité : | |
| | Madame Isabelle Lefort, Vice-Présidente du Conseil Scientifique, et de la recherche, de l'Université Lyon 2 |

1 • Introduction

La visite s'est déroulée le 14 janvier 2013 à l'ERIC dans les locaux de l'Université de Lyon 2. Il s'agit d'une évaluation à mi-parcours sollicitée par les établissements de tutelle (Universités de Lyon 1 et 2), la dernière évaluation AERES ayant eu lieu en février 2010.

La journée s'est déroulée selon le programme suivant qui a été défini conjointement par le comité de visite et la direction de l'ERIC:

9h30 : Présentation du bilan et du projet scientifique du laboratoire, discussion

12h : Repas

13h30 : Entretiens (direction des universités de tutelle, doctorants, personnel administratif et technique, direction du laboratoire

15h : Délibération du comité

17h : Compte-rendu du comité

Historique et localisation géographique de l'unité :

Fondé à Lyon 2 en 1995, le laboratoire ERIC a opéré à partir de 2009 un regroupement, officialisé en 2010, avec l'équipe MA²D (méthodes et algorithmes pour l'aide à la décision) de Lyon 1, sur la base de complémentarités thématiques en matière de recherche dans le domaine du décisionnel, d'une sensibilité commune au domaine des sciences humaines et sociales et de nombreuses collaborations sur les plans scientifiques et pédagogiques. Finalement, ERIC a intégré en 2012 l'Institut des Sciences de l'Homme (ISH), Unité de Service et de Recherche du CNRS et de l'Université de Lyon. La période 2009-2012 a donc été un moment charnière d'intégration de nouveaux collègues au sein d'ERIC.

Le laboratoire ERIC occupe une position originale dans le paysage informatique lyonnais, par son positionnement scientifique ciblé sur l'informatique décisionnelle, alors que les trois autres laboratoires d'informatique de la place de Lyon, le LIRIS, le LIP et le DISP, sont positionnés sur des créneaux plus généraliste pour le premier (image, données, connaissances, services) et également très spécialisés (parallélisme et systèmes de production, respectivement) pour les deux autres. Toutefois, des proximités thématiques existent avec des équipes du LIRIS et du DISP, ainsi qu'avec l'équipe MODEME du Centre Magellan (laboratoire de gestion) de Lyon 3, qui se traduisent par des participations croisées à des jurys de thèse ou l'organisation commune de manifestations scientifiques.

Équipe de Direction :

Compte-tenu de la double tutelle d'ERIC, le laboratoire est dirigé par un directeur d'un des deux établissements de tutelle (J. Darmont, Lyon 2) et un directeur adjoint de l'autre (S. Bonnevey, Lyon 1). Un conseil de direction du laboratoire, constitué du directeur, du directeur adjoint, des responsables d'équipes de recherche (F. Bentayeb et J. Velcin) et de la responsable administrative et financière (V. Gabriele Pietroforte), se réunit tous les mois entre les conseils de laboratoire pléniers.

Effectifs de l'unité :

| Effectifs de l'unité | Nombre au 30/06/2012 | Nombre au 01/01/2014 | 2014-2018 Nombre de produisants du projet |
|--|----------------------|----------------------|--|
| N1 : Enseignants-chercheurs titulaires et assimilés | 22 | 22 | 20 |
| N2 : Chercheurs des EPST ou EPIC titulaires et assimilés | 0 | 0 | 0 |
| N3 : Autres personnels titulaires (n'ayant pas d'obligation de recherche) | 0 | 0 | 0 |
| N4 : Autres enseignants-chercheurs (PREM, ECC, etc.) | 0 | 0 | 0 |
| N5 : Autres chercheurs des EPST ou EPIC (DREM, Post-doctorants, visiteurs etc.) | 0 | 0 | 0 |
| N6 : Autres personnels contractuels (n'ayant pas d'obligation de recherche) | 0 | 0 | 0 |
| TOTAL N1 à N6 | 22 | 22 | 20 |

| | |
|---------------------|------|
| Taux de producteurs | 91 % |
|---------------------|------|

| Effectifs de l'unité | Nombre au 30/06/2012 | Nombre au 01/01/2014 |
|---|---------------------------------|---------------------------------|
| Doctorants | 23 | |
| Thèses soutenues | 10 | |
| Post-doctorants ayant passé au moins 12 mois dans l'unité * | 3 | |
| Nombre d'HDR soutenues | 2 | |
| Personnes habilitées à diriger des recherches ou assimilées | 11 | 9 |

2 • Appréciation sur l'unité

Points forts et possibilités liées au contexte :

1. Un positionnement original et singulier dans l'environnement lyonnais avec de nombreuses collaborations académiques (au niveau national et international) et industrielles.
2. Une implication très forte dans la formation de niveau master, avec deux parcours dont un européen Erasmus Mundus.
3. Une amélioration très nette des supports de publication et une augmentation significative de financement sur projets nationaux et européens depuis l'évaluation AERES de 2010.
4. Une bonne visibilité du laboratoire au niveau national et international.

Points à améliorer et risques liés au contexte :

- Manque de locaux adaptés et de support technique.
- Pour certains, l'effort de publication dans les meilleurs supports doit être poursuivi.
- Surveiller que les services d'enseignement ne deviennent pas trop lourds pour certains enseignants-chercheurs.
- Consolider la vision scientifique du laboratoire pour assurer son développement pour le long terme.

Recommandations :

1. Renforcer et consolider le positionnement original du laboratoire et l'ancrage avec les SHS.
2. Poursuivre et amplifier les travaux intra- et inter-équipes.
3. Continuer l'effort de publications dans les meilleurs supports.
4. Préparer le départ à la retraite de membre(s) senior(s) et définir le plus tôt possible une politique claire concernant le maintien et/ou le renouvellement de thématiques.
5. La tutelle devrait être attentive à :
 - Apporter les moyens nécessaires (postes EC, BIATOSS) à la politique de recherche de l'unité pour que les enseignant-chercheurs ne fassent pas office d'ingénieur.
 - Améliorer les conditions de travail des membres du laboratoire notamment en terme de locaux.

3 • Appréciations détaillées

Appréciation sur la production et la qualité scientifiques

Selon le dernier rapport d'évaluation AERES (2010), une des principales difficultés concerne "l'écart qui subsiste entre la qualité globale du travail effectué qui est satisfaisante, et le manque de résultats de la valorisation des résultats, que ce soit en terme de publications dans des conférences et revues de références ou de réussite à l'ANR ou au PCRD."

Grâce à une nouvelle politique de publications, cette difficulté a été largement palliée par une augmentation significative des articles dans des revues et de conférences de qualité du domaine. Plus de 90 % des membres du laboratoire peuvent être considérés producteurs selon les critères AERES. On ne peut qu'encourager le laboratoire à continuer ses efforts dans cette direction afin de valoriser au maximum ses résultats.

Cette période est aussi marquée par l'obtention de plusieurs projets importants incluant notamment deux projets européens et un projet ANR.

Appréciation sur le rayonnement et l'attractivité académiques

L'ERIC joue un rôle important dans son environnement. L'ERIC est d'abord un laboratoire très visible sur le plan national voire international dans ses domaines de compétences (les entrepôts et la fouille de données). Ensuite, certains membres du laboratoire jouent un rôle clé dans le développement de ces domaines en France et à l'international. Le laboratoire maintient de nombreux échanges internationaux. Il joue un rôle très actif dans les principales structures nationales du domaine à travers ses responsabilités dans l'association EGC et la revue RNTI, ses actions d'organisation de conférences/workshops/journées et ses collaborations industrielles.

Enfin, malgré le nombre très réduit d'allocations ministérielles accordées par l'Ecole Doctorale, il est remarquable que le laboratoire arrive à recruter plusieurs doctorants par an sur des financements très variés (grâce à ses collaborations internationales pour des boursiers étrangers, des projets industriels/conventions CIFRE...).

Appréciation sur l'interaction avec l'environnement social, économique et culturel

L'interaction du laboratoire avec le monde économique est très bonne avec de nombreuses collaborations industrielles (contrats, conventions CIFRE...). Il faut noter également l'initiative importante de certains membres du laboratoire dans l'édition, le maintien et la diffusion de plusieurs logiciels libres dont TANAGRA, un logiciel de fouille de données reconnu, est un excellent exemple.

Appréciation sur l'organisation et la vie de l'unité :

Le laboratoire a entendu les recommandations du précédent comité appelant à abandonner l'organisation matricielle et à revenir à une structuration plus classique par équipes.

Le laboratoire apparaît être une unité cohérente et soudée. La proximité des thématiques du laboratoire et les travaux communs entre ses membres facilitent les communications internes et favorisent la production scientifique. L'organisation du laboratoire en deux équipes semble appropriée. Il faut veiller à préserver cette unité et continuer des travaux entre les deux équipes.

Appréciation sur l'implication dans la formation par la recherche

Le laboratoire est très fortement impliqué dans la spécialité "Fouille de données et gestion des connaissances" du Master d'Informatique de Lyon 2: parcours "Extraction des connaissances à partir de données (ECD)" en partenariat avec Polytech' Nantes (une trentaine d'étudiants en M2) et parcours "Erasmus Mundus DMKM" en partenariat avec 5 universités européennes (aussi une trentaine d'étudiants en M2). Un projet de Doctorat Erasmus Mundus a été également déposé en 2012. Ces formations constituent un excellent facteur pour la visibilité et l'attractivité du laboratoire et un riche vivier pour le recrutement de doctorants.

Appréciation sur la stratégie et le projet à cinq ans

Peu de laboratoires évoluant dans un environnement académique en LLASHS ont une production scientifique et une visibilité pouvant se comparer à ERIC. Il faut reconnaître à cette équipe son talent, et son

dynamisme. ERIC est, en particulier, moteur dans l'animation de la communauté nationale sur ses thèmes de prédilection. La taille du laboratoire lui permet sans doute de fédérer la totalité des forces en présence ; c'est une chance que ERIC a su saisir au fil des ans.

Pour assurer son développement, le laboratoire devrait affirmer davantage sa vision scientifique de son avenir. Dans un environnement lyonnais où plusieurs laboratoires d'informatique de taille imposante sont présents, il est indispensable de promouvoir et défendre sa "couleur" scientifique originale tout en renforçant ses liens déjà existants avec les autres laboratoires locaux en informatique et LLASHS sur la base de la complémentarité thématique.

Enfin, cette vision doit également alimenter la politique de recrutement et de renouvellement des EC, avec l'appui et l'engagement des établissements de tutelle.

Remarques additionnelles sur la facture du rapport d'activité

Ces remarques sont formulées dans le but d'aider à la préparation du prochain rendez-vous d'évaluation.

Concernant la rédaction de l'annexe détaillant la production scientifique (Annexe 5), il eut été judicieux de fournir une liste par équipe - tout en soulignant les productions émanant d'un travail inter-équipes (quitte à dupliquer les entrées). La numérotation [aa-xyz] n'est pas utile et brouille la lecture des références ; on aurait pu préférer des citations de type apacite (LaTeX) rappelant les noms des auteurs et l'année de publication.

Il faut veiller à fournir des références complètes : les maisons d'éditions et les collections des ouvrages édités sont essentielles (fournir une URL cliquable dans le document d'accompagnement peut s'avérer utile) ; les références des publications en revue doivent être complétées en fournissant les numéros de volume et de pages.

Il eut été utile de décrire en une phrase le résultat qui est au cœur de chacune des 10 publications retenues parmi la liste des publications marquantes de l'équipe. Incidemment, cette liste présente certains écarts avec les références proposés au bilan (section 2.3 « Contributions majeures ») alors qu'on s'attend à un recoupement plus fort (peut-être s'agit-il seulement de pouvoir proposer une liste de plus de 10 publications majeures ?).

Par ailleurs, il eut été utile de connaître les personnes impliquées dans chacun des axes (sections 1.3 et 2.3 « Contributions majeures ») admettant naturellement des chevauchements ; ces thèmes ne semblent pas cloisonnés - puisqu'on peut s'attendre à ce que la fouille de graphes intersecte les travaux sur la fouille d'opinions, par exemple - mais servent à structurer et à cibler les efforts de l'équipe.

4 • Analyse équipe par équipe

| | |
|----------------------|--|
| Équipe 1 : | Equipe SID (Systèmes d'Information Décisionnels) |
| Nom du responsable : | Fadila Bentayeb (MCF HDR) |

Effectifs

| Effectifs de l'équipe | Nombre au 30/06/2012 | Nombre au 01/01/2014 | 2014-2018 Nombre de produisants du projet |
|--|----------------------|----------------------|--|
| N1 : Enseignants-chercheurs titulaires et assimilés | 8 | 8 | 8 |
| N2 : Chercheurs des EPST ou EPIC titulaires et assimilés | 0 | 0 | 0 |
| N3 : Autres personnels titulaires (n'ayant pas d'obligation de recherche) | 0 | 0 | 0 |
| N4 : Autres enseignants-chercheurs (PREM, ECC, etc.) | 0 | 0 | 0 |
| N5 : Autres chercheurs des EPST ou EPIC (DREM, Post-doctorants, visiteurs etc.) | 0 | 0 | 0 |
| N6 : Autres personnels contractuels (n'ayant pas d'obligation de recherche) | 0 | 0 | 0 |
| TOTAL N1 à N6 | 8 | 8 | 8 |

| Effectifs de l'équipe | Nombre au 30/06/2012 | Nombre au 01/01/2014 |
|---|----------------------|----------------------|
| Doctorants | 11 | |
| Thèses soutenues | 2 | |
| Post-doctorants ayant passé au moins 12 mois dans l'unité | 2 | |
| Nombre d'HDR soutenues | 2 | |
| Personnes habilitées à diriger des recherches ou assimilées | 4 | 4 |

Appréciations détaillées

La thématique centrale de l'équipe SID (Systèmes d'Information Décisionnels) porte sur les entrepôts de données (ED) et l'analyse en ligne (OLAP), d'une façon générale, l'informatique décisionnelle. Avec l'avènement des données complexes (textes, images, son, vidéo, ...), l'analyse en ligne doit s'adapter à la nature spécifique de ces données tout en gardant l'esprit de l'OLAP. Ainsi les opérateurs OLAP classiques sont définis pour des données classiques et s'avèrent souvent inadaptés pour des données complexes. Ces dernières portent généralement plus de sémantique que les données classiques, sémantiques qu'il s'agit de prendre en compte dans la modélisation multidimensionnelle et dans leur analyse en ligne. Les activités de recherche de l'équipe poursuivent deux objectifs majeurs :

- La proposition de nouveaux modèles d'ED et d'analyse en ligne pour répondre aux nouveaux défis liés à l'entreposage de données complexes, à la prise en compte de l'utilisateur et à la sécurité des données et des résultats ;
- La conception et le développement d'architectures de SID dans différents domaines d'application comme la linguistique, l'histoire, la médecine, etc.

De par sa proximité avec l'équipe DMD, une des spécificités des travaux développés par l'équipe SID réside dans la combinaison de l'OLAP avec plusieurs méthodes de traitement de données comme la fouille de données, la statistique et la recherche d'information. Cette synergie doit permettre de franchir le fossé sémantique pouvant exister entre des données complexes riches en information et des analyses OLAP simples.

Pour conclure, l'activité de l'équipe SID essaie d'apporter des solutions relatives à : (1) l'intégration et la représentation des données complexes, (2) l'analyse en ligne (OLAP) de données complexes, (3) les entrepôts centrés utilisateur, et enfin (4) la sécurité et la qualité des données.

Appréciation sur la production et la qualité scientifiques

L'effort de l'équipe depuis 2010 a surtout porté sur les publications internationales, dont le volume a augmenté de 73 % (le nombre de revues internationales a, notamment, plus que doublé), tandis que le volume de publications nationales a baissé de 26 % du fait de cet effort. L'équipe SID a suivi les recommandations du comité d'évaluation AERES de 2010, en privilégiant dans le choix des revues et des conférences internationales la qualité et non la quantité, même si comme elle le reconnaît elle-même, des efforts restent à faire pour publier dans les revues et conférences de rang A.

Tout d'abord tous les permanents de l'équipe sont publiants. Ensuite les publications tant au niveau des revues que des conférences, impliquent souvent d'autres auteurs appartenant à d'autres équipes en pointe de la communauté française des ED (Toulouse-IRIT, LI-Tours, LIRMM-Montpellier), ainsi de nombreux auteurs d'universités étrangères. Enfin il y a plusieurs publications de qualité impliquant des membres des deux équipes SID et DMD, montrant ainsi un certain décloisonnement entre les deux équipes.

Notons que parmi les publications sur cette période 2009-2012, il y a eu la soutenance de 2 HDR de chercheurs de l'équipe, celle de Fadila Bentayeb (2011) et celle de Sabine Loudcher (2011). Ces HDR sont de grande qualité, et souvent référencées dans la communauté francophone des ED.

Enfin, l'équipe SID produit aussi des logiciels libres téléchargeable autour des entrepôts de données et l'analyse en ligne.

Appréciation sur le rayonnement et l'attractivité académiques

Comme déjà évoqué, l'équipe SID est bien intégrée dans la communauté des ED, de par de nombreuses collaborations cristallisées dans des publications impliquant des chercheurs de diverses équipes, ceci tant à un niveau national qu'international. Son rayonnement est ainsi effectif aux deux niveaux. Notons que l'équipe SID est fondatrice de la conférence francophone reconnue EDA (Entrepôt de Données et Analyse en ligne), dans laquelle elle est toujours fortement impliquée. Elle est par ailleurs impliquée dans l'organisation de sessions de conférences ou workshop internationaux dans le domaine des entrepôts de données. Les membres des équipes SID sont aussi présents dans les comités de programmes de nombreuses conférences internationales ou nationales reconnues dans le domaine des entrepôts, notamment DOLAP.

En ce qui concerne les doctorants, on ne peut que constater une forte attractivité de l'équipe, au vu des nombreuses thèses financées par des bourses de pays étrangers, tant du Magreb, que d'Europe (Ukraine), et d'Asie (Thaïlande).

Appréciation sur l'interaction avec l'environnement social, économique et culturel

Principalement de par son attache assumée aux Sciences Humaines et Sociales, l'équipe est très bien intégrée dans l'environnement social, économique et culturel. Signalons aussi son interaction avec le monde industriel au travers de son implication dans notamment des conventions CIFRE.

Appréciation sur l'organisation et la vie de l'équipe

L'équipe apparaît très dynamique, soudée et il semble y avoir une très bonne ambiance de travail en son sein, comme le confirme d'ailleurs les doctorants de l'équipe rencontrés lors de cette visite. Notons que l'équipe se réunit chaque vendredi après-midi, avec des réunions d'une durée de deux à trois heures. Ce qui est remarquable.

Appréciation sur l'implication dans la formation par la recherche

C'est sans nul doute un des points forts du laboratoire et aussi de l'équipe. L'équipe prend en effet sa part dans l'opérationnalisation des formations de niveau master : la spécialité "Fouille de données et gestion des connaissances" du Master d'Informatique de Lyon 2: parcours "Extraction des connaissances à partir de données (ECD)" avec Polytech' Nantes, et le parcours "Erasmus Mundus DMKM" en partenariat avec 5 universités européennes (aussi une trentaine d'étudiants en M2). Notons qu'un projet de Doctorat Erasmus Mundus a été également déposé en 2012.

Appréciation sur la stratégie et le projet à cinq ans

Le projet à cinq ans (2013-2017) de l'équipe SID porte bien évidemment sur l'informatique décisionnelle. Deux axes sont privilégiés :

Le premier axe privilégié concerne les entrepôts et l'analyse en ligne et s'intéresse plus particulièrement aux données textuelles et aux réseaux sociaux. Plus globalement, il s'agit pour l'équipe de concevoir des entrepôts d'informations et/ou de connaissance avec une analyse en ligne adaptée aux informations et/ou aux connaissances, avec des opérateurs OLAP spécifiques. L'équipe s'intéresse ainsi à l'intégration d'une ontologie dans le processus même d'entreposage. Signalons d'autres ouvertures comme l'élaboration d'entrepôts Web et la combinaison de technologies Web sémantique et OLAP, ou encore les entrepôts de données issues de simulation. Cet axe concerne aussi l'analyse en ligne collaborative avec plus particulièrement l'enrichissement des modèles d'entrepôts de données et des cubes OLAP par des connaissances (profils, annotations...).

Le second axe privilégié s'inscrit dans le « Cloud Computing », c'est l'informatique décisionnelle dans les nuages. Cette dernière doit permettre notamment de diffuser plus largement l'usage de la "Business Intelligence - BI", permettant à plus d'utilisateurs, de collaborateurs, de partager dans la limite de leur rôle et fonctions, des informations, et des résultats d'analyses. L'équipe s'intéresse ainsi à la BI à la demande ou BI collaborative devant fournir des outils d'analyse en ligne pour tous, de permettre aux citoyens de s'approprier les outils décisionnels. Associés à ces objectifs l'équipe souhaite contribuer sur : la définition de structure et de langages pour la modélisation multidimensionnelle dans les nuages, la personnalisation des analyses selon des profils d'utilisateurs, la qualité et la sécurité des données, l'utilisation d'agents intelligents pour l'intégration, l'analyse en ligne et la sécurité des données dans un environnement collaboratif dans le nuage.

Ces deux axes apparaissent d'actualité et très ambitieux. Cependant si l'équipe semble posséder toutes les compétences pour développer, dans les trois années prochaines, des contributions significatives sur chacun de ces axes, il n'est pas certain qu'elle dispose de suffisamment de forces vives. Aussi faudrait-il peut être réduire un peu l'ambition du projet.

Conclusion :

- Points forts et possibilités liées au contexte :

L'équipe possède une parfaite maîtrise de l'objet d'étude, à savoir les entrepôts de données, tant en ce qui concerne leur modélisation que leur exploitation, notamment leur analyse en ligne (OLAP). Le projet que

l'équipe s'est définie est pertinent et ambitieux. Ses collaborations tant nationales qu'internationales sont un atout dans la réalisation de ce projet. Son enracinement dans le domaine des SHS permettra aussi d'appréhender certains aspects de ce projet, notamment l'étude des réseaux sociaux, mais aussi les aspects collaboratifs.

Enfin, rappelons l'excellente ambiance de travail régnant dans cette équipe, tant au niveau des permanents que des doctorants, équipe qui arrive à se réunir une fois par semaine.

Un éventuel rapprochement avec d'autres équipes de petites tailles dans la région lyonnaise et privilégiant déjà des thématiques de recherche voisines pourrait être envisagée, on pense notamment à l'équipe MODEME de l'IAE de Lyon.

- Points à améliorer et risques liés au contexte :

Le projet est à l'égal de l'excellence des travaux de l'équipe, même s'il peut paraître en première lecture trop ambitieux au vu des forces disponibles

Les risques principaux apparaissent être tout d'abord liés à la disponibilité des chercheurs, essentiellement des enseignants-chercheurs, fortement impliqués dans des programmes de formation originaux et attractifs. Le risque est ainsi que les chercheurs n'aient pas assez de temps à consacrer à leurs recherches, étant fortement impliqués dans des charges d'enseignements et d'administration.

Un autre risque est l'implication dans des projets contractualisés, eux aussi gourmands en temps, et parfois éloignés des axes de recherche privilégiés par l'équipe, mais cependant nécessaires au fonctionnement de l'équipe sur le plan financier.

- Recommandations :

Pour son projet ambitieux, fortement associé au SHS, l'équipe SID a besoin de moyens tant matériels (locaux et matériel notamment), qu'humain. Sur le côté matériel, l'équipe nous apparaît à l'étroit dans ses locaux actuels, locaux qui nécessiteraient d'ailleurs des travaux (notamment pour l'accessibilité à des toilettes).

Sur le volet humain, la présence d'un ingénieur d'étude ou de recherche, pouvant être mutualisé avec l'équipe DMD, nous apparaît indispensable pour supporter les développements informatiques associés au développement du projet.

En ce qui concerne les postes d'enseignant-chercheur, les institutions de tutelles pourraient prioriser un poste de Professeur et/ou de Maître de Conférences en 27^{ième} section (informatique) qui seraient nécessaires au développement de l'équipe, ces enseignants-chercheurs recrutés pouvant apporter à l'équipe plus de compétences notamment dans des domaines comme les ontologies, Web sémantique, et les réseaux sociaux, compétences impliquées dans le projet de l'équipe. Il serait également souhaitable que l'équipe puisse pouvoir attirer des chercheurs à temps plein.

L'intégration dans le milieu SHS doit encore être renforcée, elle apporte à l'équipe d'inépuisables domaines d'application, ceci en interdisciplinarité.

4 • Analyse équipe par équipe

| | |
|----------------------|-------------------------------|
| Équipe 2 : | DMD (Data Mining et Décision) |
| Nom du responsable : | Julien Velcin (MCF) |

Effectifs

| Effectifs de l'équipe | Nombre au 30/06/2012 | Nombre au 01/01/2014 | 2014-2018 Nombre de produisants du projet |
|--|----------------------|----------------------|--|
| N1 : Enseignants-chercheurs titulaires et assimilés | 14 | 14 | 12 |
| N2 : Chercheurs des EPST ou EPIC titulaires et assimilés | | | |
| N3 : Autres personnels titulaires (n'ayant pas d'obligation de recherche) | | | |
| N4 : Autres enseignants-chercheurs (PREM, ECC, etc.) | | | |
| N5 : Autres chercheurs des EPST ou EPIC (DREM, Post-doctorants, visiteurs etc.) | | | |
| N6 : Autres personnels contractuels (n'ayant pas d'obligation de recherche) | | | |
| TOTAL N1 à N6 | 14 | 14 | 12 |

| Effectifs de l'équipe | Nombre au 30/06/2012 | Nombre au 01/01/2014 |
|---|----------------------|----------------------|
| Doctorants | 11 | |
| Thèses soutenues | 8 | |
| Post-doctorants ayant passé au moins 12 mois dans l'unité | 1 | |
| Nombre d'HDR soutenues | 0 | |
| Personnes habilitées à diriger des recherches ou assimilées | 7 | 5- |

Appréciations détaillées

L'appréciation de l'équipe DMD se base sur le contenu du rapport, sur la présentation du bilan et du projet (en matinée), complétée par les entretiens de l'après-midi. Les commentaires formulés ici apportent notamment des éléments de réponses au dernier rapport d'évaluation (2010).

Appréciation sur la production et la qualité scientifiques

L'équipe DMD développe ses travaux dans le domaine de la fouille de données et de l'apprentissage, et depuis plusieurs années. Les résultats obtenus se situent autant sur le volet théorique (par exemple : apprentissage et forêts aléatoires menant à une publication PKDD 2010 ; propriétés d'antimonotonie pour l'extraction de règles d'association, publié chez Wiley en 2012) que sur le volet plus « appliqué » (évaluation des modèles de « topics », publié à IJCAI en 2011) engageant la transversalité des approches de fouille et d'IA. La thématique d'« Aide à la décision », prise dans un sens large (Business Intelligence), est une cible définissant la direction dans laquelle les efforts sont fournis - et cela concerne tout autant DMD que l'équipe SID de ERIC qui en fait une priorité forte de son projet (« Devenir leader dans le domaine des SID en France »).

La présentation des travaux se déploie sur trois axes : « Modélisation, caractérisation, fouille dans les graphes » ; « Modèles pour l'aide à la décision » ; « Analyse de données complexes et fouille d'opinions ». Les thèmes de recherche de DMD sont donc bien en phase à la fois avec les défis qui se posent à cette communauté de recherche, et avec les besoins engendrés par la production croissante de données numériques (le « Big data »). L'équipe DMD est naturellement bien placée pour apporter des réponses aux questions et aux défis qui se posent aujourd'hui.

Par ailleurs, l'équipe s'est résolument attaquée aux exigences de publication à l'international qui avaient été formulées lors de la dernière campagne d'évaluation. On trouve de nombreuses publications de qualité parues dans des revues dont la visibilité est acquise et qui assurent aux travaux de l'équipe une portée scientifique accrue (on note, parmi les publications en revue internationale : Journal of Applied Statistics - Taylor & Francis ; Expert Systems with Applications - Elsevier ; Web Intelligence and Agent Systems - IOS Press ; Computational Intelligence - Wiley...). L'activité d'édition d'ouvrages est à l'égal du travail d'animation qu'assurent les membres de DMD depuis des années. Ce changement de stratégie de publication s'est fait tout en maintenant une présence dans les forums les plus sélectifs de la discipline, tout comme dans les forums nationaux.

Appréciation sur le rayonnement et l'attractivité académiques

Nous commenterons ici les aspects liés à la contractualisation de la recherche, tant ils peuvent être vu comme un prisme au travers duquel on peut lire le rayonnement scientifique de l'équipe. C'est par sa visibilité et son rayonnement qu'une équipe suscite des opportunités de contrats. DMD a su intéresser des partenaires nationaux et s'engager dans le projet ANR d'envergure ImagiWeb. Elle est aussi partie prenante dans un appel récent autour de la thématique « Big Data » (projet Investissement d'avenir). Son implication dans deux projets européens, lié aux domaines de la santé plus proche de l'équipe de Lyon 1 venue récemment grandir ses rangs, apporte une solide réponse aux critiques du dernier rapport d'évaluation. Il y a certainement une « culture de projets » à développer dans l'équipe, et dans tout ERIC, pour pérenniser cette façon de financer ses activités.

Appréciation sur l'interaction avec l'environnement social, économique et culturel

De nombreuses thématiques de recherche de DMD, qu'elles soient décrites dans le bilan ou annoncé au projet, placent l'équipe dans le contexte de l'étude des systèmes complexes. Sur ce point, le rapport dit peu de choses sur le positionnement de DMD par rapport à son environnement immédiat (on pense à l'institut des systèmes complexes lyonnais IXXI). Cela dit, le projet inscrit résolument DMD dans une démarche la rapprochant, au niveau national, de l'institut des sciences humaines et sociales InSHS du CNRS. Il faut y voir la volonté de DMD de s'accorder avec son environnement proche - c'est en partie ce contexte universitaire SHS qui lui confère sa spécificité.

Appréciation sur l'organisation et la vie de l'équipe

L'équipe adopte un rythme de réunion favorisant sa cohésion. Des séminaires thématiques lui permettent d'accueillir sur son site des collaborateurs du PRES lyonnais. L'unité thématique de DMD est

palpable - on peut lire le résultat de la réorganisation d'ERIC en deux équipes, sur un schéma plus traditionnel comme on le lui avait proposé lors de la dernière évaluation.

Appréciation sur l'implication dans la formation par la recherche

L'implication de l'équipe dans le montage et l'animation du Master d'Informatique - parcours "Extraction des connaissances à partir de données (ECD)" (Lyon 2) témoigne de son dynamisme dans la formation par la recherche - cette remarque vaut pour tout ERIC. Par ailleurs, DMD (et ERIC tout entier) a développé un réseau de collaborateurs qui lui permet de recruter ses doctorants dans le vivier international - c'est là un atout pour une équipe évoluant en contexte universitaire SHS.

Appréciation sur la stratégie et le projet à cinq ans

Le projet scientifique de DMD met en exergue un nombre de questions et de verrous à lever qui entre dans son champ thématique : données massives et scalabilité des algorithmes ; traçabilité des données, fouille d'opinions ; approche par apprentissage topologique intégrant données multi-attribuées et méta-données ; etc. Ce sont là de réelles opportunités pour DMD ; l'équipe possède les compétences pour attaquer ces questions, et se distinguer en apportant des réponses originales.

Le format de l'équipe doit lui permettre d'apporter une réponse « intégrée » à ces questions qui émanent pour beaucoup de besoins sociétaux. L'exposé nous fait moins sentir le besoin d'avancées théoriques que la nécessité de produire un effort transversal (informatique / statistiques / mathématiques appliquées) pour apporter des solutions aux problèmes évoqués. On pourrait comparer DMD à une EPI (INRIA), par sa taille et par les perspectives annoncées au projet, pour lui suggérer d'articuler le projet autour de jalons (milestones).

La vision qui se dégage du projet est en phase avec la stratégie de contractualisation qui est en marche dans DMD. L'assise scientifique de DMD, sa notoriété et les compétences in-house, sont de réelles garanties de succès. La projection à 5 ans doit servir de moteur pour les enrichir.

Conclusion :

- Points forts et possibilités liées au contexte :

L'équipe DMD (et le laboratoire ERIC tout entier) se trouve dans un environnement de proximité avec les SHS. Ses thématiques de recherche en font un acteur clé des défis posés par la numérisation grandissante des données de toute sorte, alors que tout un pan des SHS se « digitalise ». Il faut y voir une occasion unique pour DMD de s'affirmer dans son environnement proche, comme sur la place de Lyon et au niveau national.

L'équipe DMD est, dans une certaine mesure, une équipe « historique » de la fouille de données. Sa notoriété nationale, depuis longtemps acquise, doit rayonner au-delà. L'effort de publication observé aujourd'hui va dans ce sens et doit être maintenu. Aussi, il faut veiller à associer pleinement les doctorants à cette politique de publication à l'international.

- Points à améliorer et risques liés au contexte :

La place de la problématique santé dans DMD, et dans son articulation avec l'environnement lyonnais doit être réfléchi. Cette double portée Santé / SHS ne doit pas forcer une si petite unité à faire le grand écart.

L'implication dans des actions de recherche contractualisée doit être poursuivie. Les compétences et le dynamisme de DMD sont des atouts pour réussir dans cette voie. L'établissement doit en prendre la pleine mesure, puisque ce mode de fonctionnement exige un soutien administratif et logistique certain - notamment lorsqu'il s'agit de s'investir au niveau européen.

La place du logiciel TANAGRA, son développement, son évolution, doit entrer dans la stratégie du laboratoire dans la mesure où il lui apporte visibilité et notoriété. Les ressources nécessaires au maintien et au développement d'un tel outil sont à l'égal de sa qualité et de son impact scientifique (sur les communautés académiques, mais aussi sur l'industrie).

Nous avons souligné à plus d'une fois combien DMD rassemble un patrimoine scientifique unique dans son domaine d'activités. Il faut veiller à le préserver - pour assurer le succès futur de l'équipe, la bonne

implémentation de la stratégie et l'atteinte des objectifs, notamment au travers des recrutements (renouvellement ou nouveau recrutement).

- **Recommandations :**

La taille de chacune des équipes de ERIC, voire ERIC tout entier, peut être comparée à une équipe d'un plus grand laboratoire ou à une EPI INRIA. De plus, sur le plan thématique, ERIC présente à la fois un spectre de compétences et un niveau d'intégration que l'on peut envier. De ce point de vue, ERIC pourrait développer une vision plus proactive de ses forces tant auprès des collègues SHS de Lyon (et ailleurs) que vers le monde industriel en terme de transfert de compétences et de technologie. Il peut être un acteur incontournable pour affirmer sur Lyon la place de son établissement dans l'univers des SHS.

Ayant en vue la prochaine campagne d'évaluation, la formulation du projet à 5 ans saura s'articuler et mieux intégrer les aspects stratégiques que nous soulignons ici. Les objectifs de recherche sont clairement identifiés et sont en phase avec les courants de recherche nationaux et internationaux. Il faut mûrir la dimension de politique scientifique du projet, en termes de positionnement et de développement futur.