

Unité de recherche
Document d'autoévaluation

CAMPAGNE D'ÉVALUATION 2019-2020
VAGUE A

Le dossier d'autoévaluation comprend, le présent document d'autoévaluation (incluant les annexes en fin de document) ainsi que les deux fichiers Excel « Données du contrat en cours » et « Données du prochain contrat ». Tous ces documents sont téléchargeables sur le site du Hcéres.

NB : On renseignera ce document d'autoévaluation en s'appuyant sur l'« [Aide à la rédaction du document d'autoévaluation d'une unité de recherche](#) » téléchargeable sur le site du Hcéres.

INFORMATIONS GÉNÉRALES

Nom de l'unité pour le contrat en cours : Entrepôts, Représentation et Ingénierie des Connaissances

Nom de l'unité pour le prochain contrat (en cas de changement) : ERIC

Acronyme pour le contrat en cours : ERIC

Acronyme pour le prochain contrat (en cas de changement) :

Domaine scientifique (si évaluation interdisciplinaire, indiquer 2 domaines) : ST

Sous-domaines scientifiques (dans la nomenclature du Hcéres) par ordre décroissant d'importance :

ST6_1 Informatique, ST1_2 Mathématiques appliquées

Directrice / directeur pour le contrat en cours : Jérôme Darmont

Directrice / directeur (ou porteur de projet) pour le prochain contrat : Julien Jacques

Type de demande :

Renouvellement à l'identique

Fusion, scission, restructuration

Création ex *nihilo*¹

Établissements et organismes de rattachement :

Liste des établissements et organismes tutelles de l'unité de recherche **pour le contrat en cours et pour le prochain contrat.**

Contrat en cours :

- Université Lumière Lyon 2

- Université Claude Bernard Lyon 1

| Prochain contrat :

| - Université Lumière Lyon 2

| - Université Lyon 1 / Établissement cible 2020

Choix de l'évaluation interdisciplinaire de l'unité de recherche (ou d'une ou plusieurs équipes internes) :

Oui

Non

Activités de recherche clinique :

Oui

Non

¹ Les unités en création ex *nihilo* seront, dans la plupart des cas, évaluées sur dossier sur la base d'un projet.

BILAN GLOBAL DE L'UNITÉ

1- Présentation de l'unité

Introduction

Fondée en 1995 à l'[Université Lumière Lyon 2](#) comme jeune équipe, puis labélisée équipe d'accueil en 1999, ERIC a été rejointe par plusieurs collègues de l'[Université Claude Bernard Lyon 1](#) en 2010. Ses tutelles sont les universités Lyon 2 et Lyon 1 et elle est localisée sur les campus Porte des Alpes (Bron) et LyonTech-La Doua (Villeurbanne).

Le laboratoire ERIC est structuré en deux équipes de recherche rassemblant chacune des enseignant-es-chercheur-es de Lyon 1 et Lyon 2 :

- Data Mining et Décision (DMD), composée de 12 membres permanents (3 PR, 9 MCF dont 2 HDR) et 1 professeur émérite ;
- Systèmes d'Information Décisionnels (SID), composée de 8 membres permanents (3 PR, 5 MCF dont 1 HDR).

Effectifs et moyens

Début 2019, le laboratoire est composé de 48 membres, dont 20 enseignants-es-chercheur-es permanent-es (dont 9 HDR), 1 personnel BIATSS, 19 doctorant-es, 1 ATER docteur et 7 membres associé-es (dont 1 professeur émérite). L'évolution des effectifs d'ERIC (Figure 1) montre une stabilité. Globalement, les mutations et départs en retraite sont réaffectés au laboratoire. Les recrutements ont été au nombre de 4 sur la période 2014-2018 : 2 PR 27 (recrutements internes), 1 PR 26 (recrutement externe), 1 MCF 27 (recrutement interne). Toutefois, la direction de l'Université Lyon 2, prenant en compte le faible taux d'encadrement des sections CNU 26-27 et mettant en place une stratégie pour l'informatique et la statistique dans l'établissement, a ouvert au concours deux postes PR (26-27 et 27) en 2019, qui sont des créations nettes pour ERIC. L'âge moyen des permanent-es d'ERIC est d'un peu moins de 48 ans (proche de la moyenne nationale pour la section CNU 27) et la tranche d'âge la plus représentée est celle des 41-50 ans, indiquant un vieillissement par rapport à la période précédente (notamment en raison des recrutements PR), où la tranche d'âge la plus représentée était celle des 31-40 ans.

ERIC est principalement alimenté en doctorant-es par le biais de son réseau international (boursier-es d'excellence et cotutelles) et les partenariats CIFRE (Figure 2). Toutefois, on note une nette baisse des thèses financées par l'étranger (qui représentaient 60 % sur la période 2009-2014) au profit des conventions CIFRE, les besoins de l'industrie ayant fortement coïncidé avec notre politique de collaboration industrielle sur la période 2014-2019. Les CDU sont en légère baisse (16 à 11 %), mais sont compensés par des allocations de recherche de la région Auvergne-Rhône-Alpes. Chaque doctorant-e bénéficie d'un bureau, d'un espace de travail comprenant un poste informatique et d'un accès aux ressources du laboratoire (calcul et stockage de données, documentation, impression, espace web...).

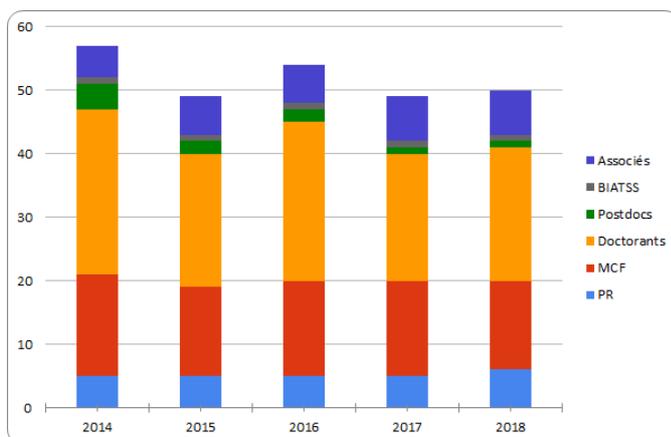


Figure 1 : Évolution des effectifs

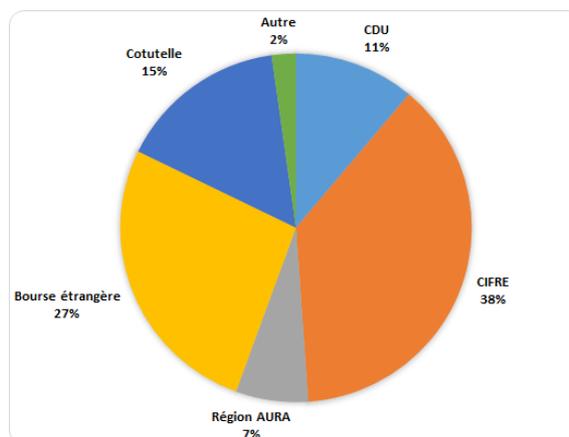


Figure 2 : Mode de financement des thèses

Les moyens financiers perçus par ERIC sont indiqués dans la Figure 3, qui montre une baisse de recettes importante en 2016, due à la fin (prévue) du projet ANR ImagiWeb et à celle (prématurée) du projet PIA REQUEST, qui a été plusieurs fois décalé avant que l'entreprise Thalès n'en abandonne finalement la coordination. Par la suite, nous n'avons pas obtenu de nouveaux financements de ce type malgré des soumissions régulières de projets ANR avec ERIC comme coordinateur scientifique ou comme partenaire (deux à trois par an). Cette baisse a toutefois majoritairement impacté la masse personnel (contrats doctoraux et

postdoctoraux) et n'a que marginalement grevé le fonctionnement du laboratoire, par ailleurs renforcé par une augmentation des contrats industriels, notamment en accompagnement des thèses CIFRE. Par ailleurs, la dotation de l'Université Lyon 2 a été exceptionnellement augmentée de 27 k€ en 2017, ce qui a permis au laboratoire d'acquies une petite infrastructure de calcul et de stockage. Lyon 2 a également revu son système de calcul des dotations aux laboratoires (incluant une part variable) en 2018, ce qui a occasionné une hausse de la DGF pour ERIC, qui s'est encore accentuée en 2019 pour atteindre 35 k€.

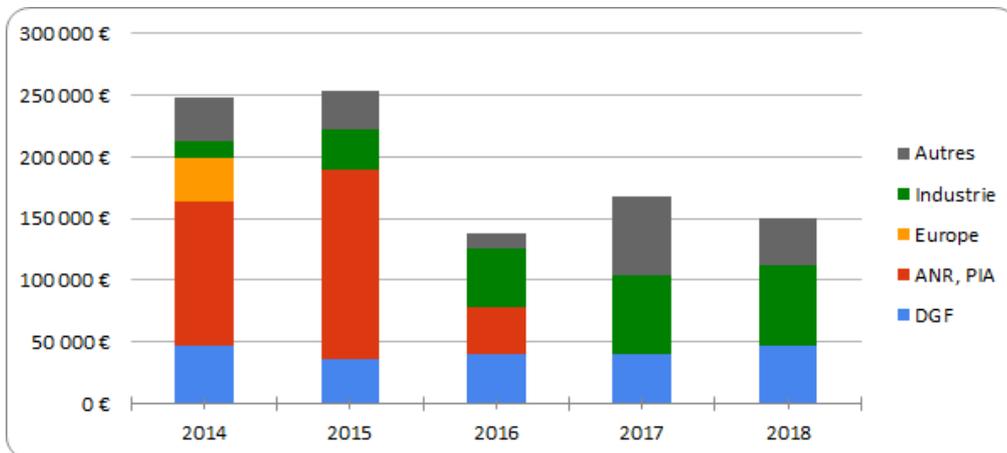


Figure 3 : Évolution des ressources financières

Politique scientifique

Les recherches d'ERIC se situent dans les domaines de la science des données et de l'informatique décisionnelle. Le laboratoire s'est donné pour objectif de valoriser les mégadonnées (*big data*), notamment dans les sciences humaines et sociales (SHS), et se situe dans les champs suivants (Figure 4) :

- entrepôts de données : intégration intelligente des mégadonnées, modélisation multidimensionnelle d'objets complexes, analyse en ligne avancée, sécurité du processus d'entrepasage (équipe SID) ;
- fouille de données et décision : apprentissage automatique, apprentissage statistique, modèle pour données complexes, fouille de données, agrégation multicritère (équipe DMD).

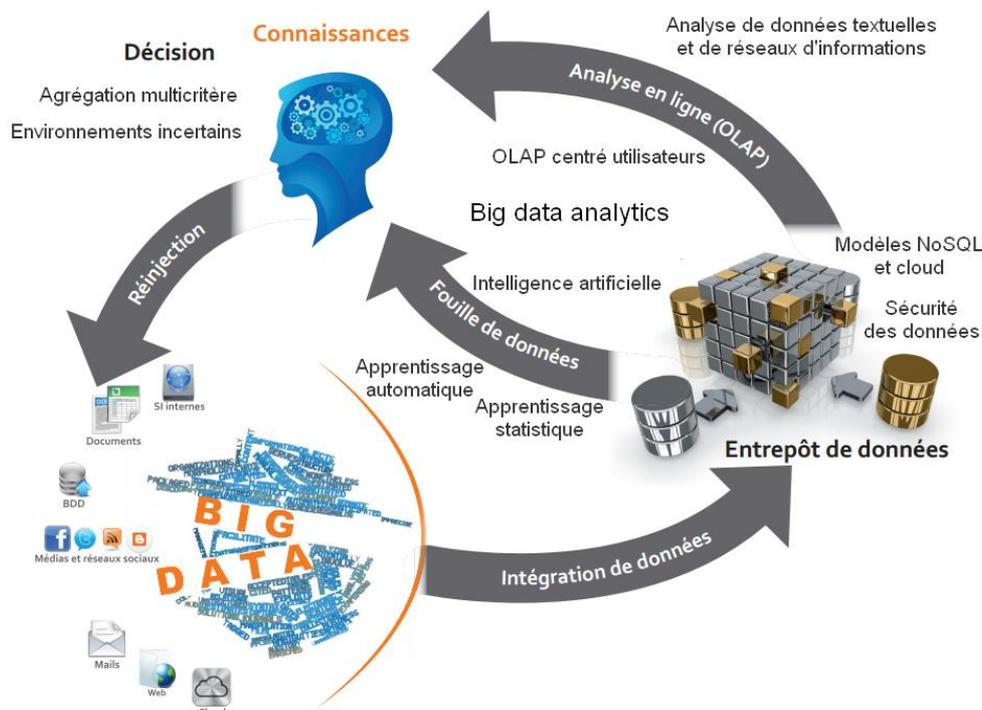


Figure 4 : Thématiques de recherche d'ERIC

De plus, ERIC étant principalement ancré à Lyon 2, université de lettres, langues et SHS (LLSHS), il a une longue tradition de collaboration avec les laboratoires de ces disciplines. Depuis la fin de la période quinquennale 2009-2014, ERIC a développé stratégiquement une activité interdisciplinaire principalement liée aux humanités

numériques (HN), qui constitue un axe de recherche transversal aux deux équipes DMD et SID. Il s'agit en fait de la mise en œuvre d'une politique scientifique visant à asseoir les activités du laboratoire sur trois piliers : informatique, mathématiques appliquées et HN. Ainsi, à la suite du bilan des équipes de recherche, nous dresserons également un bilan de cette activité interdisciplinaire, afin qu'il soit explicite (pour répondre à une recommandation de la précédente évaluation).

Toujours par suite de l'évaluation de la période 2009-2014, nous avons pris en compte les recommandations des évaluateurs en nous donnant 4 grands objectifs.

1. « Sur le long terme, il doit s'abstenir de geler sa réflexion stratégique, et effectuer un choix clair entre un ancrage STIC/SHS qui devrait pouvoir se lire dans son organigramme et dans ses effectifs et son ancrage Math/Info actuel. Dans les deux cas, il doit alors mieux préciser sa relation aux autres acteurs du site. » Notre objectif a été durant la période 2014-2019 de viser une reconnaissance CNRS à l'horizon du contrat quinquennal 2021-2025, via un projet interdisciplinaire lié aux HN. À cette fin, nous nous sommes rapprochés du laboratoire [ELICO](#), avec lequel nous avons organisé des séminaires croisés pendant deux ans. Toutefois, ELICO, qui est le seul laboratoire de sciences de l'information et de la communication sur le site Lyon-Saint-Étienne, a souhaité renforcer son ancrage et sa visibilité disciplinaires. Par la suite, malgré quelques contacts, via des projets de recherche communs, avec le laboratoire [CRTI](#) (linguistique), il ne nous a finalement pas paru opportun de poursuivre dans cette voie. Pour autant, nous souhaitons réaffirmer l'ancrage d'ERIC sur ses trois piliers informatique, mathématiques appliquées et HN, avec l'avantage de rester ouverts à tout l'éventail des collaborations possibles avec les LLSHS. Poursuivre la politique de collaboration scientifique du laboratoire avec les LLSHS permet à ERIC d'occuper un positionnement original au sein de l'informatique lyonnaise. La relation avec les autres acteurs (math/info, avec un objectif de renforcer les liens existants, et LLSHS) du site est développée dans la section suivante.
2. « Il peut être souhaitable pour le laboratoire de spécifier davantage sa politique de recrutement à terme, dans la perspective notamment d'un renforcement éventuel de son ancrage SHS. » En réponse à cette recommandation, nous avons formalisé notre politique de recrutement en faisant remonter les besoins en postes des équipes de recherche, puis en les interclassant et en les articulant avec les besoins d'enseignements de nos composantes. Ainsi, nous avons un pipeline de profils sur 5 ans que nous adaptons à mi-quinquennal. De plus, si, compte-tenu de l'ADN du laboratoire, il n'a pas été envisagé de recruter des collègues de LLSHS à ERIC, nous avons inclus depuis 2014 dans tous les profils de postes proposés au concours dans les sections CNU 26 et 27 un volet HN demandant une expérience dans le domaine, ainsi qu'une volonté de s'investir dans les projets interdisciplinaires du laboratoire.
3. « Le laboratoire entretient un volume significatif de partenariats avec le monde socio-économique (...). Il est par contre assez peu partie prenante des grands projets de site et des interactions que celui-ci entretient avec les collectivités territoriales. » En étant membre de la [MSH-Lyon Saint-Etienne](#) (anciennement ISH, depuis 2012), puis en intégrant durant la période 2014-2019 le [LabEx Intelligence des Mondes Urbains](#) (IMU) et, sans en être membre pour l'instant, en étant partenaire de la [Fédération Informatique de Lyon](#) (FIL) et du [LabEx Mathématiques et Informatique Fondamentale de Lyon](#) (MILYON), ERIC a intégré plus étroitement un réseau de laboratoires d'informatique, de mathématiques et de LLSHS qui maille le site de Lyon-Saint-Étienne qui permet de répondre à des appels à projets de ces structures et de la Région AURA. L'appartenance au LabEx IMU ouvre aussi des opportunités de collaboration avec la Métropole de Lyon, ainsi qu'avec des industriels (chaire Renault en cours de montage). Le détail de ces affiliations est donné dans la section suivante.
4. « Le laboratoire ERIC doit veiller à maintenir un équilibre adéquat entre activités pédo-administratives et activités de recherche. Il devrait aussi pouvoir être doté d'un ingénieur de développement à même de conforter ses activités relatives au logiciel en ligne. » Malgré les efforts fournis par l'Université Lyon 2 pour combler son sous-encadrement en informatique et en mathématiques appliquées, ce dernier demeure important (également à Lyon 1) et les enseignant-es-chercheur-es toujours sollicité-es par de nombreuses activités pédagogiques et administratives. Toutefois, nous évitons autant que possible d'effectuer trop d'heures complémentaires. Par ailleurs, nous demandons systématiquement de nouveaux postes d'enseignant-es-chercheur-es à nos deux tutelles. Nous avons également obtenu de Lyon 2 la création d'un poste d'ingénieur de recherche en informatique à mi-temps (partagé avec le laboratoire DISP situé sur le même campus), ouvert au concours 2019.

Enfin, le profil d'activité d'ERIC est résumé dans le Tableau 1. La production scientifique (essentiellement sous la forme de revues et de conférences internationales dans les domaines indiqués en Section 1, Politique scientifique) inclut de nombreuses copublications avec des collègues français et étrangers. Nous contribuons également à la recherche académique par l'organisation de journées thématiques et d'ateliers internationaux. Les interactions avec l'environnement prennent la forme de missions de conseil auprès d'entreprises, que nous essayons systématiquement de faire fructifier en contrats de recherche ou en thèses CIFRE. Le laboratoire invite également régulièrement des professeur-es (un-e ou deux par an) et des étudiant-es étranger-es pour des séjours courts (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 sur [notre site web](#) et l'animation de groupes de recherche (GDR MACS et ROADEF, Pretopologics, action ADOC du GDR MaDICS, GT DAHLIA de l'association EGC) et de conférences aux plans national (EGC, EDA) et international (ateliers dans des conférences internationales). Enfin, la formation par la recherche inclut l'encadrement de projets et de stages orientés recherche dès le master 1, un encadrement doctoral actif (en moyenne 23 doctorant-es par an) et l'accueil régulier de doctorant-es et postdoctorant-es étranger-es. Le directeur du laboratoire est également membre du conseil de [l'école doctorale InfoMaths](#), représentant de l'ED à Lyon 2. Deux membres d'ERIC font également partie du comité des thèses de l'ED.

Production de connaissance	Valorisation et transfert	Appui à la communauté	Formation par la recherche
40 %	15 %	15 %	30 %

Tableau 1 : Profil d'activités

2- Présentation de l'écosystème recherche de l'unité

Le positionnement scientifique d'ERIC sur l'informatique décisionnelle au sens large nous distingue des autres laboratoires d'informatique du site Lyon-Saint-Étienne, qui sont soit plus généralistes ([LaHC](#) : physique et informatique, [LIRIS](#) : image, données, connaissances, services), soit également spécialisés mais dans d'autres domaines ([CITI](#) : télécom, [CREATIS](#) : santé, [DISP](#) : systèmes de production, [LIP](#) : parallélisme). Toutefois, des proximités thématiques existent avec des équipes des laboratoires LIRIS, DISP, LaHC et le [Centre Magellan](#) (laboratoire de gestion doté d'une équipe Systèmes d'information). Elles se traduisent notamment par des participations croisées à des journées scientifiques, séminaires et jurys de thèse, par exemple.

Du point de vue des mathématiques appliquées, le positionnement d'ERIC sur l'apprentissage statistique nous distingue des autres laboratoires en mathématique du site ([ICJ](#) et [UMPA](#), pour lesquels cet axe de la statistique n'était jusqu'alors pas présent. Les récents et futurs recrutements de ces deux laboratoires sur cette thématique démontrent la pertinence de notre positionnement stratégique précoce. ERIC fait également partie de l'initiative SciDoLySE, qui a vu le jour en 2019 dans le but de fédérer les chercheur-es des laboratoires d'informatique et de mathématiques appliquées du site la thématique du *machine learning*.

De plus, nous partageons la même école doctorale (Infomaths ED 512) avec les laboratoires sus-cités (sauf CREATIS et le LaHC) et, bien que nous ne soyons pas membres de la FIL (le CNRS souhaitant n'y conserver pour l'instant que des UMR, au contraire de nos tutelles et de la COMUE Université de Lyon comprise), nous y sommes associés et pouvons bénéficier de ses appels à projets. Nous participons également, par l'intermédiaire de la FIL, à des salons professionnels qui nous permettent de montrer notre savoir-faire et de développer notre réseau industriel. De la même manière, sans être membres du LabEx MILYON, nous pouvons également participer à ses appels à projets et bénéficier de son réseau. Finalement, plusieurs collègues interviennent dans la formation doctorale centralisée au niveau de la COMUE Université de Lyon, pour des doctorant-es de différentes disciplines.

Par ailleurs, notre intégration en 2012 à la MSH-LSE (USR CNRS) a créé une plus grande proximité encore avec des unités de recherche en LLSHS et nous a permis de trouver de nouveaux terrains de collaboration (dont un avec le laboratoire de gestion COACTIS, qui a abouti à un projet Région qui a financé un doctorant pour ERIC). L'appartenance à la MSH nous a également ouvert des appels à projets communs entre la MSH et [l'Institut Rhônalpin des Systèmes Complexes](#) (IXXI, dont un membre d'ERIC fait partie du Conseil scientifique). ERIC a également rejoint le LabEx IMU en cours de quinquennal, ce qui nous a permis d'intégrer deux projets d'ampleur suffisante pour financer des doctorant-es et postdoctorant-es. La visibilité d'ERIC dans ces structures a permis à la communauté de bien nous identifier comme un acteur des humanités numériques.

Enfin, ERIC est également membre du groupe de recherche CNRS [Masse de Données, Informations et Connaissances en Sciences](#) (MaDICS, dont une action, [ADOC](#), est co-portée par un membre du laboratoire), partenaire de [l'Institut du Genre](#) (groupement d'intérêt scientifique) et co-fondateur du groupe de travail Humanités numériques au sein de l'association EGC ([DAHLIA](#)), ce qui renforce le réseau interdisciplinaire et le rayonnement d'ERIC.

3- Produits et activités de la recherche de l'unité

Bilan scientifique

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). L'activité du laboratoire est reconnue

au plan international, comme en attestent la production (revues et conférences internationales classées²) et le rayonnement scientifique de ses membres (participation aux comités de lecture des grandes revues et conférences internationales du domaine, Annexe 4), ainsi que les collaborations menées (la production scientifique d'ERIC inclut notamment de nombreuses copublications avec des collègues français et étrangers). L'évolution de la production scientifique du laboratoire est détaillée dans la Figure 5. Elle montre une augmentation globale du volume de 11 % par rapport à la période de référence précédente, avec une augmentation notable du nombre de revues internationales (+67 %) et de conférences nationales (+48 %), aux dépens des « autres publications », qui diminuent de 49 % en volume, par suite de la politique de publication du laboratoire ciblant principalement les revues et les conférences. La qualité des supports de publication (Figure 6) augmente de 12 % pour les revues internationales de rang A et B, mais baisse sensiblement (-21 %) pour les conférences internationales de rang A et B, ce qui constituera un point de vigilance pour la prochaine période quinquennale. On note également une augmentation importante (+63 %) des revues et conférences internationales non référencées, à surveiller également. Toutefois, avec l'essor des publications liées à l'axe transversal HN, le laboratoire a publié dans des supports reconnus en SHS, mais qui n'apparaissent pas dans les classements en informatique et statistique, ce qui explique une partie de ce phénomène.

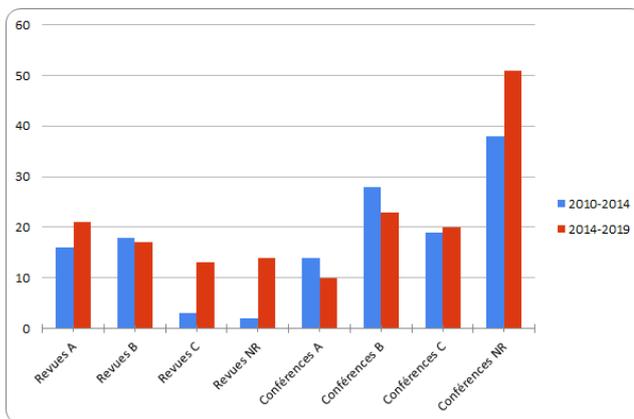
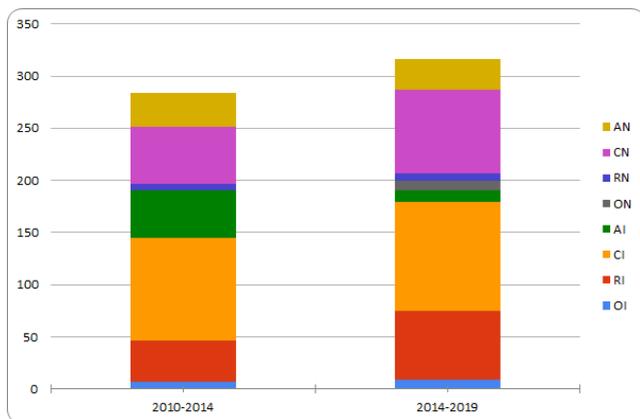


Figure 5 : Nombre de publications par type

Figure 6 : Nombre de publications classées

O – Ouvrages et direction d'ouvrages, R – Revues, C – Conférences avec comité de lecture et actes, A – Autres publications, I – Portée internationale, N – Portée nationale, NR – Support de publication non référencé

Nous contribuons également à l'animation scientifique par l'organisation/la co-organisation de journées thématiques (une à trois par an depuis 2012, rassemblant chacune 20 à 40 participant·es), ainsi que de manifestations nationales (INFORSID 2014, EDA 2017, SFC 2017, StatLearn 2017, divers ateliers) et internationales (Artificial Evolution 2015, DEXA 2017, divers workshops).

International

Outre les relations personnelles des enseignant·es-chercheur·es du laboratoire avec des collègues étrangers du monde entier, qui donnent lieu à mobilité, projets (y compris thèses en cotutelle) et copublications, le laboratoire a mis en place une stratégie d'institutionnalisation de ses relations internationales. En premier lieu, nous avons nommé une responsable des relations internationales. Par suite, nous avons signé des MOU avec le Korean Institute of Science and Technology Information ([KISTI](#), l'équivalent coréen d'INRIA) en 2014 et le Center for Research of Knowledge Application and Web Service de l'[Université Nationale de Tainan](#) (Taiwan) en 2016. Nous sommes également en lien depuis 2015 avec le European Research Center for Information Systems ([ERCIS](#)), un réseau scientifique en systèmes d'information (au sens large) que nous souhaitons intégrer avec l'ambition de devenir partenaires de projets européens.

Nous avons également une politique soutenue d'accueil de chercheur·es étranger·es de tous niveaux (master, doctorat, postdoc, collègues), en moyenne cinq personnes par an, entre autres grâce au projet de coopération franco-algérien TASSILI (mobilité). Nous avons également bénéficié deux fois du programme du [Collegium de Lyon](#), un Institut d'études avancées qui finance des séjours longs de collègues étrangers de premier plan sur la base de projets interdisciplinaires avec des laboratoires lyonnais. Cette politique s'articulant parfaitement avec notre axe transversal HN, nous avons pu faire venir Ian Davidson, de l'Université de Californie – Davis, en 2017-2018 et accueillerons Taylor Arnold, de l'Université de Richmond (USA) en 2019-2020. Ces invités

² Nous utilisons le classement [SJR](#) pour les revues et le classement [CORE 2018](#) pour les conférences, avec les correspondances suivantes : Rang A = SJR Q1 / CORE A* et A ; Rang B = SJR Q2 / CORE B ; Rang C = SJR Q3 et Q4 / CORE C.

disposent durant tout leur séjour d'un espace de travail, d'un ordinateur, d'un badge d'accès et de toutes les ressources (documentaires, de calcul...) du laboratoire.

Médiation et vulgarisation scientifique

Nous avons mis en œuvre des actions de médiation et de vulgarisation scientifique au cours de ce quinquennal. La plus emblématique est la réalisation en 2014 d'un [clip présentant les activités du laboratoire](#), avec le concours de la MSH-LSE et de tous les membres d'ERIC. D'autres actions sont plus personnelles, mais sont relayées en conseil de laboratoire pour viser à leur généralisation :

- cofondation et animation des [Cafés de la Statistique à Lyon](#) par un membre d'ERIC (quatre éditions par an depuis 2012) ;
- participation de membres du laboratoire à la [Nuit européenne des chercheurs](#) et au Festival [Pop Sciences](#) ;
- conférences de vulgarisation à la [Maison des Mathématiques et de l'Informatique de Lyon](#), à la bibliothèque de Lyon, dans des lycées ;
- animation de l'atelier « Pour l'égalité f/h dans les arts et la culture : des chiffres, des paroles... et des actes ? Freins et leviers d'une objectivation numérique de la réalité. », Université d'Automne de l'association HF #1 pour l'égalité femmes-hommes dans les arts et la culture (2016) ;
- présentation vulgarisée des entrepôts de données dans l'ouvrage [Les Big Data à découvert](#) (CNRS éditions), dont un membre d'ERIC est coauteur (2017) ;
- interview du journal le Progrès sur les sondages (2017) ;
- participation de deux membres du laboratoire au [programme Saventuriers](#) via le LabEx IMU en 2017-2018 et 2018-2019 ;
- organisation par deux membres d'ERIC, dans le cadre du « off » de la Web Conference 2018 à Lyon, des rencontres « L'informatique, pourquoi pas moi ? Les jeunes prennent la parole ! » et « L'informatique, pourquoi pas elles ? » ;
- présentation d'une mini-conférence « Lire les chiffres, les chiffres trompeurs » aux Universités d'automne du Mouvement HF #3 pour l'égalité femmes-hommes dans les arts et la culture (2018) ;
- vidéo sur l'éthique en statistique réalisée dans le cadre d'un MOOC de l'ENS de Lyon.

Faits marquants

2010-2014 : Coordination du projet EGIDE7 Tassili. Ce projet franco-algérien, dont la thématique était l'analyse en ligne (OLAP) de données textuelles et des réseaux sociaux, a permis de financer plusieurs séjours scientifiques à ERIC de jeunes chercheur-es de l'Université Saad Dahleb de Blida. Cette coopération a abouti à la codirection de plusieurs thèses et à de nombreuses copublications, et ainsi fortement contribué à la structuration de la stratégie internationale du laboratoire.

2012-2015 : Coordination scientifique du projet ANR [ImagiWeb](#). Ce projet visait à analyser la dynamique des images d'entités (hommes politiques, célébrités, entreprises...) sur le Web. Il a fortement contribué à structurer l'axe Humanités numériques d'ERIC par son caractère interdisciplinaire (collaboration, entre autres, avec le laboratoire de Science politique [CEPEL](#), Montpellier) et par la participation de membres des deux équipes d'ERIC. La production scientifique issue d'ImagiWeb, y compris la mise à disposition de données et d'une plateforme logicielle, a également contribué très positivement au rayonnement du laboratoire.

2014 : Présidence de l'organisation du 32^e congrès INFORSID. INFORSID est la manifestation scientifique francophone de référence dans le domaine des systèmes d'information. Ce congrès n'avait pas été organisé à Lyon depuis 2000. La présidence par ERIC a permis d'affirmer la position du laboratoire sur la place lyonnaise, tout en réaffirmant les liens avec les autres laboratoires d'informatique du site, qui se sont joints au comité d'organisation.

2015 : Vingtième anniversaire du laboratoire. Ce jalon marque la maturité du laboratoire et la pertinence de la stratégie scientifique qui a été menée depuis sa fondation : positionnement sur la thématique de la fouille de données commune à l'informatique et à la statistique, puis élargissement au début des années 2000 au domaine des bases de données décisionnelles, puis en 2010 à celui de la décision, qui sont tous deux très complémentaires du noyau initial.

2016 : Association du laboratoire au LabEx IMU. Lors des appels à projets LabEx, ERIC s'était investi sur une proposition liée aux humanités numériques et n'avait pas eu la possibilité d'intégrer le projet IMU, pourtant également cohérent avec nos thématiques de recherche. Le projet de LabEx HN n'ayant pas été retenu, nous n'avons pas pu bénéficier de ce type de structure jusqu'en 2016. Grâce à l'implication dans le projet IMU IDENUM (porté par le laboratoire ELICO), nous avons pu démontrer notre expertise et intégrer le LabEx. Cela

nous a ouvert de nouvelles perspectives de collaboration, qui ont notamment mené à notre participation dans le projet HyperThéséau (porté par le laboratoire Archéorient et démarré en 2018) et nous ont permis de financer plusieurs stages de master.

2017 : Installation d'une infrastructure de calcul et de stockage pour ERIC à la DSI de Lyon 2. Par suite d'un appel à projet interne à Lyon 2, nous avons, constatant une grande difficulté à accéder aux ressources de calcul existantes du site (notamment le CC IN2P3), demandé en 2015 avec quatre autres laboratoires le financement d'une infrastructure de calcul et de stockage pour un montant de 64 k€. Malgré un rapport interne très favorable, ce projet n'avait pas été retenu en raison d'un rapport externe nous enjoignant à utiliser les ressources existantes à Lyon... Les difficultés d'accès à l'infrastructure du CC IN2P3 persistant, Lyon 2 nous a octroyé en 2017 une enveloppe de 27 k€ qui nous a permis d'acquérir deux serveurs ESX et une baie de stockage. Nous avons complété ce dispositif sur les fonds du laboratoire par un troisième serveur ESX en 2018 et une extension mémoire (x2) des trois serveurs fin 2018-début 2019 (Annexe 3). Cela nous permet d'être autonomes et souples (via des machines virtuelles) pour répondre aux besoins des enseignant·es-chercheur·es et des doctorant·es.

4- Organisation et vie de l'unité

Pilotage, animation, organisation de l'unité

Dans la configuration 2014-2018 du laboratoire, ERIC est dirigé par un directeur d'un des deux établissements de tutelle et un directeur adjoint de l'autre (Annexe 2). Le conseil de laboratoire, qui se réunit tous les mois, est constitué de tous les membres permanents, d'un·e représentant·e élu·e du personnel administratif, d'un·e représentant·e élu·e des doctorant·es et postdoctorant·es et d'un·e représentant·e élu·e 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 de la direction, des responsables des équipes de recherche et de la responsable administrative et financière. Le conseil de direction se réunissait jusqu'en 2017 tous les mois, entre les conseils de laboratoire. Cependant, par suite de l'évaluation à mi-parcours que nous nous imposons statutairement, nos évaluateurs (externes) ont jugé que cette organisation était un peu complexe. En conséquence, le conseil de direction, qui *de facto* ne statuait que sur l'ordre du jour des conseils de laboratoire, est devenu virtuel (échanges de courriels).

Toutefois, le conseil de direction, dans sa formation restreinte aux enseignant·es-chercheur·es, fait également office de commission des thèses. Il valide les premières inscriptions en doctorat et les jurys de soutenance avant transmission à l'école doctorale. Il organisait aussi avant la mise en place des comités de suivi de thèse par l'ED InfoMaths une soutenance à mi-parcours (rapport et présentation orale évalués par deux collègues extérieur·es au laboratoire) pour chaque doctorant·e, en liaison avec le(s) encadrant·e(s). En cas de candidatures multiples à un contrat doctoral, le classement des projets de thèse soumis à l'école doctorale est discuté et voté en conseil de laboratoire, dans sa formation restreinte aux enseignant·es-chercheur·es. Le conseil de laboratoire établit également la politique de recrutement, sur la base des besoins en postes des équipes de recherche. Nous maintenons ainsi un *pipeline* de demandes de postes profilés, et ce pour nos deux tutelles, qui nous permet d'avoir un discours clair avec nos partenaires (tutelles lors des dialogues de gestion, composantes pour la synchronisation avec les besoins de formation lors de l'élaboration des profils de poste).

Cette organisation est complétée par plusieurs responsables : relations internationales, organisation des séminaires (une vingtaine par an en moyenne), valorisation (en lien avec la [SAT Pulsalys](#)) et correspondantes LabEx IMU et GIS Institut du Genre. Un groupe de travail permanent constitué de 4 membres du laboratoire est également chargé de l'administration et de la sécurité des serveurs informatiques d'ERIC (3 serveurs de calcul, 1 serveur web, 1 baie de stockage, 2 NAS), en lien avec la DSI de Lyon 2 où le matériel est localisé. Finalement, des missions sont confiées ponctuellement à des membres du laboratoire, par exemple pour la mise en place d'un plan de gestion de données en 2017-2018.

En termes d'animation scientifique interne, outre les séminaires et l'activité propre des équipes, le laboratoire organise également chaque année une journée consacrée aux stagiaires, qui y présentent leurs travaux, ainsi qu'une journée « hors les murs » qui allie présentations scientifiques (la plupart du temps des doctorant·es de première année) et moment convivial. Les doctorant·es du laboratoire organisent également de leur propre initiative une journée qui leur est spécifique et réservée.

Pour finir, les ressources issues des DGF et des prélèvements sur ressources propres (10 % hors ressources fléchées, comme les allocations doctorales) 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, qui sont alloués au prorata des effectifs des équipes. À titre d'exemple, la répartition entre le budget commun et celui des équipes DMD et SID en 2019 est donné dans la Figure 7. Les dépenses des équipes de recherche sont essentiellement consacrées au financement de stages de master et de missions, majoritairement pour les doctorant·es qui présentent leurs travaux dans des conférences.

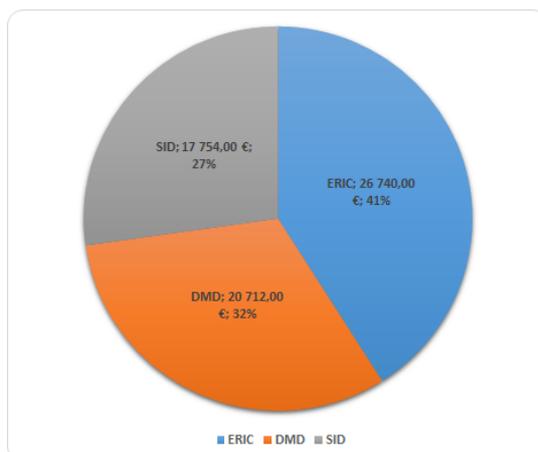


Figure 7 : Répartition 2019 du budget entre les équipes de recherche

Parité ; Intégrité scientifique ; Hygiène et sécurité ; Développement durable et prise en compte des impacts environnementaux ; Propriété intellectuelle et intelligence économique

La démographie sexuée du laboratoire en 2018-2019 est représentée dans le Tableau 2. Il montre qu'ERIC compte plus de femmes que dans la [population globale de la section CNU 27](#), notamment au niveau MCF ; la proportion étant à peu près identique à la population globale au niveau PR. Bien qu'il n'y ait pas de mesure particulière pour favoriser la parité et l'égalité professionnelle au sein du laboratoire, tou-tes ses membres sont sensibilisé-es à cette problématique du fait des recherches interdisciplinaires sur le genre et l'informatique menées à ERIC.

	MCF	PR	BIATSS	Total
Femmes	5 (36 %)	1 (17 %)	1 (100 %)	7 (33 %)
Hommes	9 (64 %)	5 (83 %)	0	14 (67 %)

Tableau 2 : Démographie d'ERIC

Le règlement intérieur du laboratoire inclut un article sur l'éthique scientifique, et notamment sur les standards de rédaction d'un article scientifique, l'ordre des auteurs, le plagiat et les publications multiples, redondantes ou simultanées. Ce règlement est accessible sur l'intranet du laboratoire et il est porté à la connaissance de tous les nouveaux membres d'ERIC.

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 dans l'installation électrique, l'isolation et l'étanchéité (air et eau). De l'amiante est également présent dans la colle des dalles du sol et dans certains plafonds mal identifiés. Toutefois, il est prévu que le laboratoire soit relocalisé dans un bâtiment neuf en 2022. Suite à des tentatives de vol, le laboratoire dispose d'un contrôle d'accès par badge 24h/24. Enfin, certain-es membres du laboratoire (direction, administration) ont suivi des formations de secourisme et de gestion des alertes incendies, et un membre du laboratoire fait partie du groupe de travail Sécurité de Lyon 2.

La politique de prévention des risques psychosociaux est essentiellement impulsée par nos tutelles (il existe notamment une cellule d'écoute et d'accompagnement à Lyon 2). Elle est relayée au sein du laboratoire sous la forme d'affichage, de mise à disposition de brochures sur un présentoir et dans notre intranet (par exemple, un guide pratique pour s'informer et se défendre contre le harcèlement sexuel dans l'ESR).

Ayant découvert tardivement dans le quinquennal que la problématique du développement durable était un critère d'évaluation, nous n'avons pas pu mettre en œuvre de politique particulière en la matière. Le laboratoire a toutefois participé au [Challenge Mobilité 2019](#) organisé par la Région Auvergne-Rhône-Alpes et s'est classé 182^e sur 370 dans sa catégorie, avec 80 % de participation.

ERIC est confronté à des problématiques de propriété intellectuelle principalement dans deux cas : dans le cadre de projets impliquant des entreprises (accompagnements de thèses CIFRE inclus), auquel cas les services de valorisation de nos tutelles travaillent et valident les conventions ; et lorsque le laboratoire produit des logiciels (essentiellement à des fins de recherche). Toutefois, ces derniers sont quasi-systématiquement proposés sous des licences libres (type Creative Commons ou ad-hoc). Dans le cas où un projet issu du laboratoire aurait un potentiel de valorisation, nous pourrions faire appel à la SAT Pulsalys à la fois pour l'accompagnement économique (plan d'affaire, etc.) et juridique. Les publications du laboratoire sont systématiquement

référencées sur [HAL](#), qui alimente notre intranet, y compris en texte intégral pour la plupart (après délai d'embargo si nécessaire).

PROJET ET STRATÉGIE À CINQ ANS DE L'UNITÉ

1- Analyse SWOT

POINTS FORTS	POINTS À AMÉLIORER
<p>Thématiques de recherche (correspondant aux équipes de recherche SID et DMD, ainsi qu'à l'axe transversal HN) sur lesquelles ERIC bénéficie d'une reconnaissance nationale et internationale.</p> <p>Nombreuses collaborations locales et nationales avec des laboratoires de LLSHS (projets de recherche) et des entreprises (thèses CIFRE).</p>	<p>Qualité des supports de publication en conférences internationales.</p> <p>Réussite aux appels à projets nationaux (ANR) et internationaux (Europe). Insertion dans des réseaux internationaux « efficaces ».</p> <p>Implication collective dans la médiation scientifique.</p>
POSSIBILITÉS OFFERTES PAR LE CONTEXTE	RISQUES LIÉS AU CONTEXTE
<p>Appartenance à des structures fédératives type MSH-LSE, LabEx IMU, Fédération Informatique de Lyon (avec objectif d'intégration à terme) : échanges scientifiques, collaborations et liens avec le tissu socio-économique facilités ; appels à projets.</p> <p>Toujours une forte demande de gestion et de valorisation des mégadonnées, tant par l'industrie que par la recherche, y compris en LLSHS.</p>	<p>Incertitudes liées à la recomposition du paysage universitaire du site Lyon-Saint-Étienne, notamment des écoles doctorales, et de ses conséquences sur les financements locaux de la recherche.</p> <p>Maintien du sous-encadrement des 26^e et 27^e sections dans nos établissements de tutelle induisant une lourde charge d'animation et d'administration pédagogique.</p>

2- Structuration, effectifs et orientations scientifiques

Pilotage, animation, organisation de l'unité

La configuration du laboratoire en fin de contrat 2014-2018, décrite plus haut, sera reconduite. Le laboratoire ERIC sera dirigé par un directeur de Lyon 2 (Julien Jacques) et d'une directrice adjointe de Lyon 1 (Nadia Kabachi).

Le conseil de laboratoire (constitué de tous les membres permanents, d'un-e représentant-e élu-e du personnel administratif, d'un-e représentant-e élu-e des doctorant-es et postdoctorant-es et d'un-e représentant-e élu-e des membres associés) continuera de se réunir tous les mois. Outre être un lieu d'information pour les membres du laboratoire, il statue notamment sur la politique de recrutement du laboratoire et, en formation restreinte, sur le classement des candidatures multiples à un contrat doctoral. Le conseil de direction, en configuration commission des thèses (direction et responsables des équipes de recherche) continuera de valider les premières inscriptions en doctorat, ainsi que les jurys de soutenance, avant transmission à l'école doctorale.

Les responsables relations internationales, valorisation et séminaire continueront leur rôle. La gestion de l'administration et de la sécurité des serveurs du laboratoire, actuellement réalisée par 4 membres du laboratoire, sera en partie déléguée à l'ingénieur de recherche qui sera recruté en 2019. Le laboratoire continuera d'organiser des événements scientifiques et festifs tels que la journée « hors les murs » et la journée des stagiaires. Les ressources du laboratoire continueront d'être partagées entre les dépenses mutualisées et les budgets propres aux équipes de recherche, laissant ainsi une liberté importante au fonctionnement de chaque équipe.

Outre ce mode de fonctionnement qui a fait ses preuves, deux nouveautés seront apportées lors du prochain contrat. Tout d'abord, un calendrier de renouvellement des responsables d'équipes sera instauré, avec une périodicité de deux ans. De plus, un échange entre le directeur et chacun des membres permanents du laboratoire sera proposé, à raison d'une à deux réunions par contrat. Cet échange informel aura pour objectif de faire le point sur les objectifs, les attentes, les besoins ainsi que l'évolution de la carrière du membre du laboratoire.

Effectifs et politique de recrutement

Les effectifs d'ERIC vont augmenter à court terme, avec deux recrutements de professeur.es des universités en 2019 à Lyon 2, ainsi qu'un poste d'ingénieur de recherche à mi-temps (partagé avec la composante Lyon 2 du

laboratoire DISP) mis au concours, également en 2019. Ce dernier recrutement était une demande de longue date, appuyée par nos évaluateur.trices successif.ves, qui permettra notamment de pérenniser les développements logiciels effectués au laboratoire. Par ailleurs, deux collègues professeurs sont susceptibles de partir à la retraite en fin du quinquennal en cours ou au début du suivant, pour lesquels nous demanderont un remplacement, même si pas nécessairement sur un poste PR suite aux deux recrutements de 2019.

La politique de recrutement d'ERIC est *bottom-up*. Elle s'appuie en effet sur les besoins hiérarchisés des deux équipes de recherche et le laboratoire, via son conseil, arbitre et interclasse les demandes de postes (Tableau 3). Il est à noter que tous les profils de poste proposés par ERIC depuis 2012 contiennent une composante humanités numériques, même si elle n'est pas systématiquement explicite dans les intitulés du Tableau 3. Cette politique de recrutement s'adresse principalement à notre tutelle Lyon 2, les membres du laboratoire y étant majoritaires et nombreux à l'Institut de la communication (ICOM), en particulier. Les membres Lyon 1 d'ERIC sont eux répartis dans trois composantes, dont seule la Faculté des Sciences et Technologies (via son Département d'informatique) reconnaît le laboratoire comme faisant partie de son périmètre. De manière à mettre en synergie les profils enseignement et recherche, ERIC est invité à l'ICOM Lyon 2 et au Département d'informatique Lyon 1. Le laboratoire participe également aux dialogues de gestion initiés par les services de la recherche de ses deux tutelles, qui comportent un volet RH.

Lyon 2	Lyon 1
1. MCF 26-27 profil DMD « machine learning »	1. PR 27 profil DMD « machine learning »
2. MCF 27 profil SID « sécurité des systèmes d'information »	2. MCF 27 profil SID « Web sémantique »
3. PR 26-27 profil DMD « machine learning pour le traitement automatique de la langue »	
4. MCF 27 profil SID « mégadonnées distribuées »	
5. MCF 26-27 profil DMD « machine learning pour les humanités numériques »	
6. PR 27 profil SID « gestion et analyse de données SHS »	
7. MCF 26-27 profil DMD « machine learning »	
8. MCF 27 profil SID « Recherche d'information dans les systèmes décisionnels »	
9. MCF 27 profil SID « Traitement automatique de la langue dans les systèmes décisionnels »	

Tableau 3 : Demandes de postes à venir (par ordre de priorité)

Objectifs et politique scientifiques

Par suite de l'échec d'un rapprochement avec un laboratoire SHS en vue d'une reconnaissance CNRS transdisciplinaire, nous souhaitons néanmoins réaffirmer notre ancrage scientifique pluri et transdisciplinaire sur trois piliers : l'informatique, les mathématiques appliquées et les humanités numériques. Plus précisément, le laboratoire ERIC dispose à travers de ses équipes de recherche de deux champs de recherche forts et complémentaires que sont les entrepôts de données et les modèles d'apprentissage pour la fouille de données, en s'appuyant pour cela sur des compétences en informatique et en mathématiques appliquées, et qui se rencontrent au service des humanités numériques. Cette spécificité du laboratoire est un avantage sur lequel nous souhaitons continuer de capitaliser.

Outre les objectifs scientifiques définis par les équipes de recherche, le laboratoire a pour visées globales de :

- maintenir sa production scientifique à un haut niveau de qualité, notamment en sélectionnant plus systématiquement des supports de publication bien évalués (de rang A et B), de manière qu'ils constituent plus de 50 % des publications internationales d'ERIC ;
- mettre en valeur et promouvoir les prototypes logiciels et *benchmarks* développés au laboratoire sur la durée. Le recrutement d'un ingénieur de recherche en 2019 devrait y concourir ;
- généraliser la participation du laboratoire (et non plus seulement de certains membres) aux actions de médiation scientifique coordonnées par l'Université de Lyon, à la fois par une dynamique collective impulsée par un.e responsable et le rappel que c'est un critère d'évaluation personnel ;
- augmenter le nombre de titulaires de l'HDR, notamment en poursuivant notre politique active d'impliquer systématiquement les maître.sses de conférences qui le souhaitent dans des coencadrements de thèse, afin de répondre aux besoins d'encadrement doctoral induits, notamment, par les sollicitations industrielles sur le dispositif CIFRE.

BILAN DE L'ÉQUIPE DATA MINING ET DECISION (DMD)

1- Présentation de l'équipe DMD

Effectifs et moyens

L'équipe DMD est composée au 30 juin 2019 de 12 membres permanents en activité, et sera renforcée par le recrutement d'un Professeur en septembre 2019, ce qui portera le nombre de membres permanents à 13 pour le prochain contrat, ce qui est constant par rapport au précédent contrat. Plus précisément, l'équipe DMD comporte 4 PR (2 en section CNU 26 et 2 en CNU 27) et 9 MCF (3 en CNU 26 et 6 en CNU 27). Dans la période de référence, l'équipe a accueilli 22 doctorants dont 12 thèses soutenances, et 37 stagiaires.

Pour assurer la continuité de l'arrivée de fonds extérieurs pour alimenter la recherche en dehors de la dotation ministérielle, l'équipe DMD a mené plusieurs actions. Développer de nouvelles collaborations avec les entreprises notamment pour des thèses CIFRE (12 sur la période) et continuer à chercher des financements sur projet : ANR, Labex, PIA ainsi que d'autres financements de plus petite envergure (PGMO, IXXI, Lyon 2...).

Politique 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 développer des outils permettant de manipuler ces données et d'en extraire de l'information utile, la stratégie de l'équipe DMD est de s'inspirer de l'apprentissage automatique, de l'apprentissage statistique et de l'intelligence artificielle, qui sont au confluent de l'informatique et des mathématiques appliquées. En effet, l'originalité et la force de l'équipe DMD, tant d'un point de vue du site Lyon Saint-Etienne que d'un point de vue national, est d'être composée d'enseignants-chercheurs issues de ces deux disciplines.

Les points faibles de l'équipe qui avaient été relevés lors de la précédente évaluation se situaient à trois niveaux : le positionnement de l'équipe vis-à-vis du laboratoire LIRIS, le trop grand nombre de lignes de recherche ainsi qu'une trop forte implication dans les structures d'enseignement. Concernant le premier point, l'originalité de l'équipe est son ancrage informatique et mathématiques appliquées, qui est plus que jamais d'actualité à l'heure où les chercheurs dans ces deux disciplines se fédèrent sur le site Lyon Saint-Etienne autour de la thématique du *machine learning* et de l'intelligence artificielle. Cette double compétence permet à l'équipe de proposer des solutions de fouilles de données complexes basées sur des modélisations innovantes, souvent à base de modèles probabilistes, des algorithmes d'inférence et d'optimisation efficaces ainsi que des implémentations de ces solutions dans des bibliothèques et logiciels libres. Concernant le second point, l'équipe a travaillé à recentrer ses lignes de recherches en un nombre d'axes restreints, deux, plus un axe transversal au laboratoire. Les deux axes propres à l'équipe sont intitulés « apprentissage pour les données complexes » et « prévision et décision ». L'axe transversal est lié aux « humanités numériques ». Enfin, concernant la forte implication de l'équipe dans les structures d'enseignement, elle est incontournable du fait des nombreuses formations de qualité portées par les membres du laboratoire. Néanmoins, nous avons au sein de l'équipe fait en sorte de recentrer un maximum des responsabilités afférentes sur les PR de l'équipe, afin de libérer du temps de recherche pour les jeunes MCF.

2- Produits et activités de la recherche de l'équipe DMD

Résumé synthétique des axes de recherche

Les thématiques de recherche de l'équipe DMD peuvent être classées en trois axes majeurs. Le premier axe concerne le développement de modèles et d'algorithmes d'*apprentissage pour les données complexes (big data)*. Le second axe, *prévision et décision*, est dédié à la mise au point de techniques de prévision et d'agrégation multicritère pour l'aide à la décision. Le troisième, relatif aux humanités numériques, est un axe transversal au laboratoire et les activités le concernant seront décrites dans une partie dédiée.

Nous résumons ici les thématiques abordées pour chacun des deux axes de recherche propres à l'équipe, et, par soucis de concision, nous ne présentons que quelques exemples de contribution à chaque fois.

1. Apprentissage pour données complexes

Les données complexes sont généralement caractérisées par leur grande hétérogénéité, leur volume ainsi que par le fait qu'elles évoluent dans le temps. Les travaux de l'équipe DMD ont porté sur ces trois aspects.

Apprentissage pour données hétérogènes. Nous avons développé un certain nombre de modèles et d'algorithmes d'apprentissage non supervisé pour des données de natures variées, lorsqu'un seul type de données est présent à la fois (données textuelles [Dermouche-15, SorianoMorales-16, SorianoMorales-16a,

SorianoMorales-17, Gourru-18, Velcin-18], données catégorielles ordinales [Biernacki-16, Jacques-18], données fonctionnelles [Bouveyron-15, Bouveyron-17, Slimen-18], graphes [Lumbreras-16], images [Rizoiu-14a]) mais également lorsque plusieurs types sont présents simultanément (graphes et textes [Davidson-18, Brochier-19], images et textes [AhPine-15b]). Nous pouvons par exemple citer les travaux faisant intervenir données textuelles et données relationnelles issues de graphes. [Davidson-18] proposent une solution au problème de description des clusters, construits à partir d'un graphe, à partir des données textuelles produites par les nœuds du graphe. En considérant ce problème comme un problème d'optimisation combinatoire, ils étudient la complexité de ce dernier et définissent des conditions suffisantes pour résoudre ce problème. [Brochier-19] définissent un espace de représentation pour les nœuds d'un réseau prenant en compte à la fois la structure de réseau et le contenu textuel associé aux nœuds. Leur approche est basée sur une formulation de type factorisation de matrice.

Apprentissage pour données dynamiques. De nombreux travaux ont été réalisés pour proposer des algorithmes de clustering où les clusters évoluent dans le temps [Kim-2015, Hasnat-2017] et où le temps est pris en compte dans le clustering [Rizoiu-2014, Rizoiu-2016, Dermouche-2014]. Par exemple, dans [Kim-2015] un modèle de mélange bayésien est proposé avec un lien entre les paramètres pour des composantes du mélange situés à deux pas de temps temporel successif. Ainsi, le modèle de mélange, et donc les clusters, à un temps donné dépend du passé de par l'impact du passé sur ses paramètres.

Apprentissage pour données volumineuses. Concernant la volumétrie des données, deux aspects ont été abordés. Le premier est le cas où le nombre de variables devient très grand, et pour lequel ont été développés des algorithmes de *clustering* de variables simultanément au clustering d'individus (*co-clustering* [Jacques-16, Slimen-18, Bouveyron17, Jacques-18, Wang-17b]) ou non [Yengo-16]. La majorité des travaux en *co-clustering* de l'équipe sont basés sur des approches probabilistes à base de modèles des blocs latents. Cette approche probabiliste a notamment l'avantage de donner une définition à la notion de (co-)cluster, et fournit également un cadre mathématique permettant de développer des critères de choix de modèles, primordiaux dans une approche non supervisée. Le second aspect est le développement d'algorithmes qui passent à l'échelle [Ah-Pine-16, Truica-16, Ah-Pine-18, Auder-18], comme par exemple [Ah-Pine-18] qui introduit un algorithme de clustering hiérarchique agglomératif basé sur des produits scalaires plutôt que des distances Euclidiennes. En procédant à une agglomération ascendante contrainte sur une matrice de produits scalaires sparsifiée, l'algorithme proposé passe mieux à l'échelle que les procédures de clustering hiérarchique classiques.

II. Préviation et décision

Cet axe de recherche traite du développement d'outils d'aide à la décision, via des approches de prédiction de séries temporelles et des approches d'agrégation multicritère.

Préviation de séries temporelles. De nouveaux modèles de prévisions ont été développés, utilisant des approches de types données fonctionnelles [Nedellec-14, Antoniadis-16, Auder-18, Nagbe-18, Martinez-Alvarez19] et des réseaux bayésiens dynamiques [Roos-16a, Roos-17]. Par exemple, [Nedellec-14] proposent une approche multi-échelle combinant différents types de modèles (lissage, modèles additifs généralisés, forêts aléatoires) à chaque niveau d'échelle de la série temporelles. Le choix de techniques de prévision adaptées à la régularité du niveau d'échelle permet d'obtenir un modèle de prévision très efficace. Un développement innovant et particulièrement attendu fût la mise au point d'intervalle de prédiction pour série temporelles [Antoniadis-16].

Agrégation multicritère. D'un côté, un certain nombre de travaux s'intéresse à l'agrégation de critères quantitatifs à l'aide de l'intégrale de Choquet [Rolland-13, AhPine-16, Brison-17, Lust-17, Cugliari-18]. Par exemple, [Ah-Pine-16] définit un nouvel opérateur d'agrégation basé sur des normes triangulaires pour modéliser les dépendances entre critères. D'un autre côté, ce sont des travaux basés sur l'intégrale de Sugeno [Dubois-14, Dubois-16, Rico-18, Dubois-18a] ou sur d'autres opérateurs d'agrégation [Dubois-15, Dubois-16-a, Dubois-16b, Dubois-17a, Couceiro-17, Dubois-18] qui portent sur l'agrégation de critères qualitatifs. Nous pouvons citer par exemple [Dubois-17a] qui porte sur les intégrales de Sugeno qui sont des opérations d'agrégation impliquant un schéma de pondération de critères basé sur l'utilisation de fonctions appelées capacités ou mesures floues. Dans cet article, des versions généralisées ont été définies sur des chaînes totalement ordonnées, en étendant l'opération qui combine la valeur de la capacité sur chaque sous-ensemble de critères et la valeur de la fonction d'utilité sur des éléments du sous-ensemble.

Bilan scientifique

I. Production de connaissance

Les 20 % des publications que l'équipe met en avant sont celles dont la référence est en gras (Annexe 4) et couvrent toutes nos thématiques de recherche. Le premier critère de sélection est celui des conférences et revues les mieux classées, puis en second terme de tri le respect de la diversité des courants thématiques de l'équipe. Les 20 % des articles dans des revues mis en avant sont au nombre de 8 (parmi 44), toutes dans des revues de rang A, dans les champs disciplinaires de l'informatique, de la statistique et du machine learning. Les

20 % des articles publiés dans les conférences sont au nombre de 16 parmi 111, 10 dans des conférences de rang A et 6 dans des conférences de rang B.

L'évolution de la production scientifique de l'équipe DMD est détaillée dans les Figures 8 et 9. Ces dernières montrent une augmentation du nombre de publications dans les revues et les conférences internationales, avec une nette progression des revues de rang A par rapport à la précédente période. On notera notamment les 17 revues de rangs A parmi les 44 publications dans des revues. La politique de publication de l'équipe visant à privilégier les revues et conférences de haut niveau a commencé à porter ses fruits.

On notera également une augmentation des revues non classées, correspondant à des publications dans des domaines autres que l'informatique et la statistique, suite à des collaborations académiques ou industrielles (publications en SHS mais également en biologie, médecine, dans le domaine de l'énergie, des transports ...). De même, on peut noter une part forte des conférences non classées, qui correspondent à des publications dans des domaines spécialisés (comme pour les revues non classées) mais également à des publications dans des conférences nationales (EGC, Journées de Statistique, Logique Floue et ses Applications) avec la volonté de participer à faire vivre la communauté.

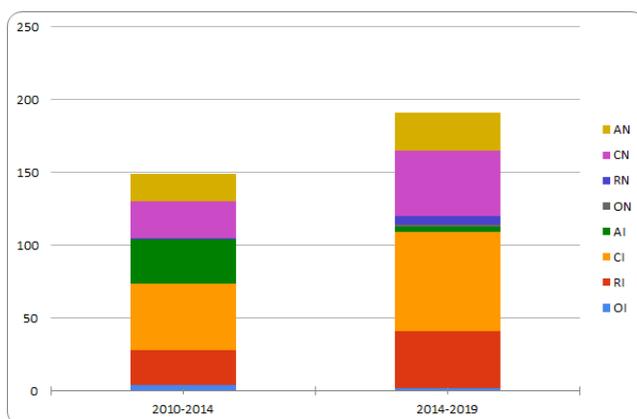


Figure 8 : Nombre de publications par type

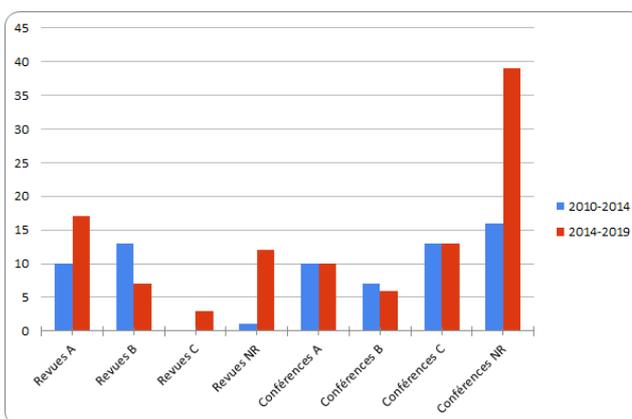


Figure 9 : Nombre de publications classées

O – Ouvrages et direction d'ouvrages, R – Revues, C – Conférences avec comité de lecture et actes, A – Autres publications, I – Portée internationale, N – Portée nationale, NR – Support de publication non référencé

L'équipe contribue également au développement d'outils informatiques spécialisés à destination des informaticiens et statisticiens mais également d'outils « grand publics » destinés aux praticiens non spécialistes. Cela comprend des packages R pour les le clustering et la prévision de données fonctionnelles, pour les données ordinales, des bibliothèques Python pour l'analyse textuelles et la détection d'événement sur Twitter, mais également des logiciels libre destinés à un publics large pour la collecte et l'analyse de corpus textuels.

II. Interaction avec l'environnement académique

Contrats de recherche. Durant la période de référence, l'équipe a participé à 24 projets de recherche. L'équipe a porté un projet ANR (Imagiweb) sur l'analyse des représentations (au sens des images ou des opinions) qui circulent sur le Web et un projet PIA (Openmiage) à visée essentiellement pédagogique. L'équipe a porté un projet de l'agence nationale uruguayenne pour la recherche (ANII), intitulé « Modèles de prévision pour la demande électrique. Elle a également piloté le projet européen FLURESP dont l'objectif était l'évaluation coût-efficacité des stratégies européennes d'alerte et de réaction en cas de pandémie de grippe humaine en Europe. Elle a été partenaire d'un projet PIA (Request) qui a malheureusement été interrompu avant son terme. L'équipe a porté 6 projets PGMO-IRSDI en collaboration avec des partenaires industriels (EDF et Thalès). Enfin, l'équipe a porté ou participé à 9 autres projets (financement Labex, IXXI, ISTE...). On peut trouver la liste complète des projets de l'équipe dans la section Projets de l'Annexe 4.

Collaborations. Les membres de l'équipe DMD ont de nombreuses collaborations locales, avec les différents laboratoires en informatique et mathématiques appliquées de la métropole lyonnaise. Elle est notamment dans le bureau de la récente initiative SciDoLySE visant à regrouper les chercheurs en machine learning du site Lyon Saint-Etienne. L'équipe a également des collaborations avec plusieurs laboratoires en sciences humaines et sociales. Elle a également de multiples collaborations nationales tant avec des laboratoires en informatique qu'en mathématiques appliquées (LJD – Nice, LIRMM – Montpellier, IMAG – Montpellier, LIUM – Le Mans, LMO – Orsay, INRIA Modal – Lille...) et internationales (Canada, Espagne, Maroc, Roumanie, Uruguay, etc.) et accueille régulièrement des chercheurs étrangers pour des séjours au sein du laboratoire (Brésil, Uruguay, Espagne, USA). Notamment, elle vient d'accueillir Ian Davidson (Université de Californie Davis, USA) pour un séjour d'une année (2017-2018) et accueillera Taylor Arnold (Université de Richmond, USA) pour une année également (2018-2019), tous deux dans le cadre du Collegium de Lyon, un institut d'études avancées qui finance des séjours longs de collègues étrangers de premier plan.

Activités d'évaluation et d'expertise. Les membres de l'équipe DMD sont régulièrement sollicités pour évaluer des dossiers de projets de recherche pour l'ANR et pour l'ANRT. Un MCF de l'équipe est membre du CNU 27 et participe à ce titre à l'évaluation de dossiers de demandes de qualification, PEDR, de CRT, Promotions et de suivi de carrières. Les membres de l'équipe participent à l'évaluation de nombreux articles dans des conférences en informatique (IJCAI, ECML-PKDD, AAAI...) et des revues de statistique et machine learning (JASA, JMLR, AoAS, STCO, CSDA, JSS, Neurocomputing, ...) de premier plan. Elle participe également à la relecture d'article pour de nombreuses conférences nationales (EGC, Congrès de la Société Francophone de Classification, Journées de Statistique de la SFdS, ...)

Rayonnement et animation scientifique. L'équipe comprend des éditeurs associés des revues Computation Intelligence, Fuzzy Sets and Systems, Statistique et Enseignement, ainsi qu'un membre du comité scientifique de la revue Management & Data Science. L'équipe comprend également un membre du comité de pilotage de la conférence Extraction et de Gestion des Connaissances (EGC) et un du workshop StatLearn. Un des membres de l'équipe est membre du bureau (trésorier) de la SFdS (Société Française de Statistique). Deux autres membres sont responsable de groupe de travail de l'association EGC. Plusieurs membres de l'équipe sont régulièrement invités dans des conférences ou des laboratoires à l'étrangers. Enfin, un membre de l'équipe est Professeur associé de l'Université Dalhousie (Canada) et un autre membre est chercheur associé au Laboratoire Sistema Nacional de Investigadores (Uruguay).

III. Interaction avec l'environnement non académique

L'équipe DMD est très active dans les partenariats avec les entreprises privées, comme en attestent les 14 contrats de recherche en partenariat avec des grands groupes (EDF R&D, ENEDIS, Michelin) ou de plus petites entreprises et les 12 contrats de thèses CIFRE en informatique ou en mathématiques appliquées, là encore en collaboration avec de grands groupes (Orange, EDF R&D, RATP) ou de plus petites entreprises.

L'équipe DMD a également une activité importante de vulgarisation, en participant à de nombreux événements sur la place lyonnaise. Elle est notamment organisatrice des Cafés Statistique de Lyon qui organisent un débat public sur un sujet de société lié à la Statistique, à raison de 4 débats par an.

IV. Implication dans la formation par la recherche

L'équipe DMD a été coordinateur du Master Erasmus Mundus DMKM (Data Mining and Knowledge Management) qui s'est arrêté en 2016. Elle est impliquée dans la direction et le fonctionnement du parcours Data Mining du Master 2 Informatique de l'Université de Lyon.

Les membres de l'équipe DMD participent très régulièrement à l'évaluation de travaux de thèses en France ou à l'étranger (environ 15 thèses par PR sur la période, dont les $\frac{3}{4}$ comme rapporteur).

L'équipe a fait soutenir 12 thèses et compte actuellement 10 doctorants. La durée moyenne des thèses avoisine les 4 ans. Celle-ci a été notamment impactée par quelques thèses en cotutelle qui ont duré très longtemps. Les doctorants ont un bon rythme de publications avec une moyenne de 2,7 publications par doctorant.

Un des membres de l'équipe est très actif dans la diffusion d'outils pédagogiques (plus de 200 tutoriels ou cours en ligne et 4 ebooks), ce qui constitue une vitrine importante pour le laboratoire.

Faits marquants

2014 : renforcement de l'axe statistique de l'équipe avec le recrutement de Julien Jacques sur un poste PR. La spécificité et la force de l'équipe DMD est son double ancrage informatique et mathématiques appliquées (statistique). Le recrutement d'un Professeur en section CNU 26 a permis d'inculquer une dynamique dans le développement de la recherche en apprentissage statistique au sein de l'équipe, et ainsi à affirmer cette double compétence informatique et statistique.

2015 : clôture du projet ANR Imagiweb par l'organisation d'un workshop sur l'étude du web politique. Le succès de ce projet ANR et cet événement ont permis de mettre en avant les activités de l'équipe au cœur des humanités numériques. Cette année a aussi été marquée par l'organisation de la conférence internationale *Artificial Evolution* à Lyon.

2016 : organisation d'un workshop sur l'intelligence artificielle et les big data à Lyon, sous l'égide de l'AFIA et du GdR MaDICS, a été un grand succès. Cette organisation sur un thème phare de l'actualité témoigne du dynamisme de l'équipe et contribue à sa reconnaissance.

2017 : Organisation de StatLearn à Lyon, conférence annuelle sur l'apprentissage statistique, qui a accueilli 12 orateurs internationaux et plus de 250 participants. Dans la lignée du développement de la recherche en apprentissage statistique au sein du laboratoire ERIC, cet événement a contribué à affirmer la position stratégique du laboratoire sur cette thématique dans la place lyonnaise.

2018 : Séjour de Ian Davidson (UC Davis) dans le cadre du Collegium de Lyon. Chercheur de haut niveau, reconnu internationalement pour ses travaux en *machine learning*, Ian Davidson a inculqué une dynamique sur

des sujets nouveaux au sein de l'équipe. Des premiers résultats prometteurs ont été publiés dans une conférence de tout premier plan [Davidson-18].

3- Organisation et vie de l'équipe DMD

Pilotage, animation, organisation de l'équipe

L'équipe DMD est pilotée par un responsable d'équipe, Julien Velcin avant septembre 2015 puis Julien Jacques. Le responsable d'équipe anime la vie scientifique de l'équipe et répartit le budget en fonction des besoins de chacun. Ces besoins sont essentiellement des besoins de déplacements pour des conférences, ainsi que le financement d'un stage chaque année, attribué à l'un des membres de l'équipe suite à un appel à sujet de stage interne.

L'équipe se réunit les premiers lundis de chaque mois, pour une réunion organisée en trois parties : un tour de table où chacun parle de ses activités du mois écoulé, une partie information et discussion, où des points importants qui concernent l'équipe sont débattus (par exemple la stratégie scientifique), et enfin un court exposé scientifique où un des membres de l'équipe présente une avancée récente.

Depuis 2017, un groupe de travail se réunit à raison d'une à deux fois par mois, sur la thématique du text mining, et de ses sujets connexes. A chaque séance, un des membres du groupe présente en détail un article sur une thématique récente et innovante d'un chercheur extérieur au laboratoire. Ce groupe de travail est important pour la cohésion de l'équipe et contribue à faire émerger des thématiques de recherche d'intérêt commun.

Chaque année depuis 2017, au mois de juillet, une journée est consacrée aux présentations des travaux des étudiants en stage. C'est un moment convivial où l'on peut discuter et débattre autour des activités de recherche des stagiaires.

Enfin, un événement d'une journée est organisée chaque année, sous la forme d'un workshop ou d'une école de printemps, sur des thématiques phares de l'équipe : clustering (2014), analyse statistique des données fonctionnelles (2015), clustering de données textuelles dynamiques (2016), analyse relationnelle mathématique (2017), école de printemps text mining (2018), Representation Learning for Mining Complex Data (2019). Cette journée regroupe entre 30 et 40 personnes, principalement du site lyonnais, et permet de mettre en avant les thématiques d'intérêts et les points forts de l'équipe DMD.

Par ailleurs, l'équipe partage l'utilisation de serveurs de calcul haute performance acquis sur différents appels à projets.

PROJET ET STRATÉGIE À CINQ ANS DE L'ÉQUIPE DMD

1- Analyse SWOT

POINTS FORTS	POINTS À AMÉLIORER
<p>Thématiques de recherche au confluent de l'informatique et des mathématiques appliquées : machine learning (apprentissage automatique, apprentissage statistique), optimisation, agrégation multicritère.</p> <p>Un modèle économique qui fonctionne sans reposer sur les financements de grands projets de types ANR ou européens.</p> <p>Nombreuses collaborations avec des entreprises (grands groupes, PME, startups) et des laboratoires en SHS.</p>	<p>Augmenter le nombre de nos publications dans les grandes conférences de machine learning (ICML, NIPS...).</p> <p>Développer des thématiques de recherche plus fondamentales en collaboration avec des laboratoires en SHS.</p> <p>Réussite aux appels à projets nationaux (ANR) et internationaux (Europe).</p> <p>Améliorer le fonctionnement du laboratoire du fait de sa localisation multi-sites.</p>
POSSIBILITÉS OFFERTES PAR LE CONTEXTE	RISQUES LIÉS AU CONTEXTE
<p>À l'heure du développement massif et de la démocratisation de l'intelligence artificielle et du machine learning, nos thématiques de recherche sont très sollicitées.</p> <p>Une demande extrêmement importante en outils, méthodes et savoir-faire en matière d'analyse de données, que ce soit de la part de partenaires académiques ou d'entreprises privées, notamment dans le bassin régional.</p> <p>Présence de chercheurs des deux communautés informatique et mathématiques appliquées dans une même équipe de recherche.</p>	<p>Thématique de recherche en machine learning très compétitive, qui demande un investissement matériel et technique très important et qui évolue très vite.</p> <p>Les réponses aux besoins de collaboration avec des laboratoires en SHS passent très souvent par une première phase d'ingénierie très chronophage et non directement valorisable comme une activité de recherche.</p> <p>Maintien d'une cohérence thématique de l'équipe tout en gardant les spécificités de chacun.</p>

2- Structuration, effectifs et orientations scientifiques

Structuration et politique de recrutement

Le mode de fonctionnement actuel avec une réunion mensuelle animée par un responsable d'équipe est pertinent et nous souhaitons le conserver. Néanmoins, quelques aménagements seront mis en place. Tout d'abord, un calendrier de renouvellement du responsable d'équipe sera instauré, avec une périodicité de deux années. De plus, une réflexion sera menée afin de faciliter la participation aux réunions d'équipe des membres en poste à l'Université Lyon 1 (par la mise en place d'une visioconférence par exemple).

L'effectif de l'équipe DMD va augmenter à court terme avec le recrutement en 2019 d'un professeur des universités sur un profil « statistique et machine learning appliquées aux SHS ». L'orientation machine learning de ce poste est due à une volonté de l'équipe d'être plus présent sur cette thématique, qui est au confluent de l'informatique et des mathématiques appliquées. Les demandes de postes d'enseignant-chercheurs qui seront formulées au cours du prochain contrat seront sur cette lignée, avec un profil machine learning et une ouverture systématique des postes aux sections CNU 26 et 27.

Par ailleurs, nous avons comme objectif d'améliorer notre transfert de connaissance, notamment en professionnalisant et maintenant nos développements logiciels actuels. Ceci permettra de gagner un temps précieux lors des collaborations notamment avec des laboratoires de SHS. Ces collaborations demandent très souvent une phase préliminaire d'ingénierie qui pourra être réalisée directement au sein de ces laboratoires de SHS grâce aux outils que nous avons développés. À l'heure actuelle, du fait du manque de personnel pour maintenir ces logiciels, il est nécessaire à chaque nouvelle collaboration d'investir un temps précieux pour une remise à niveau informatique de ces logiciels. Le recrutement en 2019 d'un poste d'ingénieur à mi-temps pour le laboratoire permettra de pallier ce manque.

Orientations et objectifs scientifiques

Notre objectif lors du prochain contrat sera de continuer à proposer des modèles d'apprentissage (automatique, statistique) pour les données complexes (données textuelles structurées, données hétérogènes, données dynamiques...), la prévision et l'aide à la décision. La ligne directrice de nos activités de recherche est la suivante : des modèles interprétables, estimés par des algorithmes efficaces, un transfert efficace. En effet, nous allons focaliser nos recherches sur des modèles interprétables et explicables plutôt que sur des modèles à fort pouvoir prédictif. C'est ce type de modèle qui est attendu en tout premier lieu par les SHS. Par ailleurs, nous chercherons également à développer des stratégies pour permettre l'interaction entre l'humain et le modèle. Ces modèles interprétables nécessitent des techniques d'inférence efficace et efficiente. Nous avons comme objectif d'accentuer nos recherches sur cette partie, et notamment dans le domaine de l'optimisation, en lien avec le développement d'une direction de recherche en machine learning fondamentale, notamment appuyée par les futurs recrutements. Enfin, nous ambitionnons de mettre en place une politique de transfert de technologie, en nous basant notamment sur l'ingénieur qui sera recruté en 2019. Un modèle interprétable estimé par un algorithme efficace n'a en effet d'intérêt que s'il est accompagné d'un transfert efficace.

Un autre objectif de l'équipe sur le prochain contrat sera d'améliorer la synergie entre les thématiques fortes de l'équipe que sont l'élaboration de modèles d'apprentissage et l'aide à la décision basée sur des approches multicritères. Cela s'appuiera notamment sur des travaux de recherche à l'interface entre ces thématiques, pilotés par des membres de l'équipe ayant contribué à la fois à l'une et l'autre.

Enfin, comme nous l'avons explicité auparavant, il est important que nous arrivions à extraire des nombreuses collaborations en SHS des problématiques de recherche fondamentales en informatique et mathématiques appliquées. Les moyens mis en œuvre pour remplir cet objectif seront entre autres le maintien des outils logiciels développés par l'équipe, ce qui permettra d'avoir une première réponse aux problématiques SHS avec des modèles et outils existants, et ainsi de pouvoir plus rapidement mettre à jour une nouvelle problématique de recherche fondamentale pour l'équipe.

Partenariats et choix stratégiques

D'un point de vue partenariat, nous souhaitons dans un premier temps accentuer nos partenariats locaux en informatique et mathématiques appliquées : en participant activement au réseau Scidolyse qui fédère les chercheurs en machine learning sur le site Lyon Saint-Etienne, en répondant régulièrement aux appels à projets de la Fédération Informatique de Lyon (FIL) et du Labex MiLyon, en développant nos collaborations avec les laboratoires LIRIS et ICJ. Nous souhaitons également poursuivre notre dynamique de développement de notre réseau de partenariats internationaux, en continuant notamment à nous appuyer sur les invitations de chercheur étrangers : via le Collegium de Lyon, via les Professeurs invités au sein des universités Lyon 1 et Lyon 2, via des dispositifs plus spécifiques comme l'IFUMI avec l'Uruguay. Nous souhaitons également développer notre mobilité sortante, en incitant et soutenant les chercheurs permanents et non permanents à effectuer des mobilités à l'étranger.

D'un point de vue production scientifique, nous allons viser plus systématiquement les conférences de rang A (notamment en machine learning, mais pas seulement), même si la compétition est très rude. Pour autant, nous ne délaisserons pas les conférences qui sont importantes pour le développement de nos réseaux de collaborations, même si ces dernières sont souvent des conférences nationales ou non classées.

Afin d'améliorer la visibilité de notre équipe de recherche, nous allons essayer d'être plus présents dans l'organisation de grandes conférences nationales et internationales, et ce en partenariat avec les autres acteurs du site Lyon Saint-Etienne.

Enfin, un des points importants que nous souhaitons améliorer au sein de l'équipe est notre taux de succès sur les appels à projets de grande envergure (projets régionaux, projets nationaux de type ANR, projets européens). Pour cela, nous travaillons à mettre en place une politique de réponses aux appels à projets qui soit fédérative et qui permette de partager entre les différents membres de l'équipe les efforts que demande la soumission de tels projets. Ainsi, un effort de soumission mieux réparti sur l'ensemble des membres de l'équipe (et même du laboratoire) accroîtra les chances de succès en lissant les efforts de chacun.

BILAN DE L'ÉQUIPE SYSTÈMES D'INFORMATION DÉCISIONNELS (SID)

1- Présentation de l'équipe SID

Effectifs et moyens

L'équipe SID a été créée en 2001 au laboratoire ERIC dans la mouvance de la *Business Intelligence (BI)*, des entrepôts de données – ED - (*Data Warehouses*) et de l'analyse en ligne OLAP (*On-Line Analytical Processing*) sous l'impulsion d'enseignants-chercheurs. Elle est composée de 8 permanents dont 3 PR (recrutement d'un PR en interne en 2015) et 5 MCF dont 1 HDR. De ce fait, l'effectif de l'équipe n'a pas bougé. Dans la période de référence, SID compte 23 doctorants dont 13 thèses soutenues, 1 ATER et 50 stagiaires. L'équipe attire de nombreux étudiants étrangers qui souhaitent préparer leur thèse. Sur la période de référence, SID a accueilli des doctorants de plusieurs nationalités : Algérie, Iran, Maroc, Thaïlande, Tunisie, Turquie, Vietnam. Pour assurer la continuité de l'arrivée de fonds extérieurs pour alimenter la recherche en dehors de la dotation ministérielle, l'équipe SID a mené plusieurs actions. Développer de nouvelles collaborations avec les entreprises notamment pour des thèses CIFRE (SWORD, Bertin IT) et continuer à chercher des financements sur projet (financement Lyon2, financement Région ou encore les ANR même si l'équipe n'a pas pu obtenir ce type de financement sur la période de référence malgré plusieurs tentatives).

Politique scientifique

Objectifs scientifiques. L'objectif principal de la recherche menée au sein de l'équipe SID s'inscrit totalement dans l'objectif scientifique global du laboratoire ERIC qui se traduit par la valorisation des mégadonnées (*big data*), notamment dans les sciences humaines et sociales. Il s'agit pour l'équipe SID de concevoir une nouvelle génération d'architectures décisionnelles centralisées et/ou distribuées incluant un processus d'intégration intelligent des mégadonnées et des ED plus adaptés aux big data tout en assurant une bonne performance, une analyse en ligne avancée pour tenir compte de la variété des données, et enfin une sécurité du processus d'entreposage nécessaire pour la protection des données et des accès.

Stratégie de l'équipe. Jusqu'en 2016, l'équipe SID était structurée en quatre axes de recherche : (1) *big data analytics*, (2) analyse en ligne de données textuelles et de réseaux d'information, (3) OLAP centré utilisateur et (4) sécurité des données. À l'issue de l'évaluation mi-parcours organisée par l'Université Lyon 2 en 2016, l'équipe SID a suivi les recommandations des experts pour réduire ses axes de recherche en deux axes majeurs : (1) *big data management* et (2) *BI & analytics*. Dans ce contexte, les données à considérer sont produites par des sources diverses (sources de données classiques, données issues de réseaux sociaux, données SHS, données capteurs, etc.) et les applications s'appuient sur des modèles et des supports de stockage variés (SGBD relationnels, SGBD NoSQL, Data Lake, etc.). Une particularité des travaux de l'équipe est de proposer des solutions innovantes pour la gestion et l'analyse en ligne des big data en combinant des concepts issus du domaine des ED et de l'OLAP avec des méthodes et des outils spécifiques aux bases de données tant au niveau théorique (modèles conceptuels, logiques et physiques de données) que système (techniques de stockage, indexation et optimisation de requêtes). L'équipe SID mobilise également d'autres thématiques de recherche telles que la fouille de données, la recherche d'information, le *cloud computing* et la sécurité afin de répondre au mieux aux attentes des nouvelles applications big data. Nous utilisons par exemple des techniques issues de la recherche d'information pour modéliser de façon multidimensionnelle et analyser en ligne des données textuelles ; ou encore, nous nous appuyons sur des méthodes de classification automatiques pour trouver la meilleure stratégie de regroupement de données dans des entrepôts NoSQL afin d'améliorer leur performance.

Ambition affichée de l'équipe. L'affichage des deux axes de recherche autour du big data management et de la *BI & analytics* est très cohérent et en parfaite adéquation avec l'ambition et les priorités fixées par la stratégie de l'équipe : être et rester leader en France sur la thématique des ED. En effet, SID est devenue aujourd'hui l'une des équipes de recherche leaders en France dans la communauté scientifique qui s'est constituée autour de la thématique des ED depuis déjà plusieurs années. Elle a depuis étendu ses travaux à une nouvelle génération d'entrepôts notamment avec l'avènement des *big data* : *Big data warehouse*, NoSQL, OLAP, Text-OLAP, lacs de données (Data Lake), Cloud BI, etc. Il constitue également un positionnement original clair et assumé d'abord en interne puis dans le paysage lyonnais, et enfin au niveau national et international. Parallèlement, SID est impliquée de manière très forte dans des projets transversaux sur les humanités numériques qui s'inscrivent dans le projet global du laboratoire ERIC. Dans ce contexte, plusieurs coopérations se sont développées entre les deux équipes du laboratoire ERIC et avec d'autres laboratoires autour de projets relevant de nouveaux défis liés aux applications BI notamment dans le domaine des SHS.

2- Produits et activités de la recherche de l'équipe SID

Bilan scientifique

I. Production de connaissance

Axes de recherche. L'objectif de la BI est de produire des indicateurs de performance permettant de comprendre le passé, d'analyser le présent afin d'extrapoler une vision à long terme pour définir par exemple les avantages compétitifs futurs d'une entreprise ou d'une collectivité. Avec l'avènement des big data, l'équipe SID a renforcé ses thématiques de recherche autour de la donnée suivant les deux axes affichés de de l'équipe sur des sujets aussi bien académiques qu'industriels.

Axe 1 - Big Data Management. L'équipe SID a fait de cet axe un axe prioritaire durant ces cinq dernières années, puisque plusieurs thèses ont déjà été soutenues et plusieurs autres sont en cours. Les travaux ont porté essentiellement sur deux thèmes phares : (A) Nouvelles architectures décisionnelles et (B) Sécurité des données.

A. Nouvelles architectures décisionnelles. Une partie de l'équipe SID est très active dans cet axe de recherche en s'intéressant plus particulièrement à l'élaboration de nouvelles architectures décisionnelles distribuées notamment pour la conception logique et physique d'entrepôts de données distribués ainsi que l'optimisation de leur performance.

Entrepôts de données distribués. Les premiers travaux de l'équipe dans ce domaine ont porté sur la modélisation logique des entrepôts pour les big data. Ces travaux ont permis de définir de nouveaux modèles logiques d'entrepôts en tenant compte à la fois de la spécificité des modèles en étoile et de l'OLAP dans un environnement distribué. Un premier modèle logique NoSQL en colonnes pour les big ED a été ainsi proposé avec un nouvel opérateur d'agrégation OLAP [Dehdouh-14e, Dehdouh-14d]. Ces travaux ont été poursuivis par la suite pour définir des modèles logiques et physiques plus performants pour tenir compte de l'aspect volumétrie des big data et assurer le passage à l'échelle. Plusieurs stratégies de regroupement puis de placement des données dans les différents nœuds d'un cluster ont été alors proposées en s'appuyant à la fois sur la modélisation multidimensionnelle des entrepôts et les méthodes de classification automatiques [Boussahoua-17, Boussahoua-18, Ramdane-18]. D'autres travaux se sont intéressés à l'aspect vitesse en cherchant à modéliser l'évolution temporelle des requêtes soumises à un ED distribué. La solution apportée est le partage de données pour optimiser le traitement des requêtes massives [Ratsimbazafy-17, Ratsimbazafy-18]. La variété des données a été également étudiée notamment pour les documents textuels. Dans ce contexte, le modèle d'entrepôt en constellation a été étendu aux données textuelles pour y intégrer les dimensions lexicales (termes) produites à partir de la phase d'indexation de documents, les dimensions sémantiques (thèmes) extraites à partir d'une encyclopédie, ainsi qu'une sélection des faits inspirée des modèles statistiques de langue [Aknouche-14]. D'autres travaux sur l'analyse en ligne de textes ont été développés au sein de l'équipe et seront présentés dans l'Axe 2. Enfin, un intérêt particulier a été porté récemment sur les lacs de données (*data lakes*) qui répondent aux problématiques de stockage des big data, mais requièrent une gouvernance efficace pour ne pas devenir des marécages (*data swamps*) inopérants. Pour y concourir, des modèles multidimensionnels évolutifs de métadonnées ont été proposés permettant de lier et d'interroger efficacement des ensembles de données hétérogènes [Nogueira-18a, Sawadogo-19].

Optimisation des performances. L'optimisation des performances des ED relationnels ou non relationnels a toujours été un thème fort de l'équipe SID. Dans le cadre relationnel, des stratégies efficaces de placement, de distribution et de colocalisation des données ont été proposées en s'appuyant sur les systèmes multi-agents [Arres-15, Arres-15a, Arres-15b]. Dans le cadre non-relationnel, des modèles physiques des big ED NoSQL en colonnes ont été proposés pour permettre le passage à l'échelle en s'appuyant sur des stratégies de placement des données efficaces puis une étude de performance a été réalisée selon plusieurs critères : volume de données, nombre de familles de colonnes, etc. [Ramdane-18, Boussahoua-17]. Pour mener à bien ces travaux, l'équipe SID a développé des bancs d'essais décisionnels pour les big data avec les SGBD NoSQL en colonnes [Dehdouh-14c, Boussahoua -18].

B. Sécurité des données. Dans un environnement ouvert (données dans le nuage) où l'exploitation des données est devenue un enjeu économique et avec l'avènement des big data, la sécurité des données est devenue une priorité. Dans ce contexte, l'équipe SID a engagé des travaux de recherche à la fois sur la sécurité des ED dans le cloud (accès, confidentialité...) et la détection des intrusions.

Sécurité des entrepôts dans le cloud. La sécurité des données est le problème le plus prégnant dans le stockage en nuage. Crypter les données est une solution, mais toutes les méthodes d'encryptage ne permettent pas d'effectuer des calculs sur les données cryptées ; et celles qui le permettent sont actuellement coûteuses. Plusieurs types de schémas d'encryption de moindre complexité ont donc été proposés répondant à différentes problématiques de sécurité : (a) des approches basées sur le partage de clés secrètes (*secret sharing*) permettant de mettre en œuvre non seulement la confidentialité des données, mais aussi leur disponibilité et leur intégrité [Attasena-14, Attasena-15, Attasena-17] ; (b) des méthodes visant à optimiser les performances de

certaines types de requêtes sur des données cryptées (requêtes d'agrégation et nécessitant un tri, respectivement) sans trop dégrader le niveau de sécurité [SobatiMoghadam-16a, Sobati Moghadam-17, SobatiMoghadam-17a] et un processus de chiffrement avant stockage basé sur les systèmes multi-agents qui consiste à altérer les données sensibles [Rhazlane-16b]. Parallèlement, il a été proposé un cryptosystème homomorphique basé sur le problème de la factorisation efficace en pratique [Gavin-16b].

Détection des intrusions. La plupart des méthodes proposées dans la littérature utilisent la fouille de données pour la détection d'intrusions. Celles-ci travaillent sur des trames réseau et sont testées sur de vieux jeux de données (années 90) qui ne reflètent plus les attaques actuelles. Dans ce domaine, l'équipe a donc proposé une solution [Pierrot-16] : (1) basée sur des logs de systèmes et de pare-feu, qui sont beaucoup plus légers et faciles à traiter que des trames réseau ; (2) conçu un système de détection hybride (visualisation des flux, apprentissage supervisé et non-supervisé) capable de fonctionner en temps réel ; (3) validé cette proposition avec des données réelles et actuelles, ainsi que par des retours d'experts en sécurité informatique.

Axe 2 - BI & Analytics. Dans ce deuxième axe, l'équipe SID a privilégié l'élaboration de nouvelles techniques d'analyse en ligne des big data dans un contexte centralisé et/ou distribué. L'objectif étant d'intégrer les caractéristiques des big data comme le volume, la variété et la vitesse dans le processus analytique des données autour de deux thèmes : Nouvelles approches OLAP pour les big data et Analyse de données textuelles et de réseaux d'information.

Nouvelles approches OLAP. Les systèmes OLAP fournissent de puissants outils d'exploration, de navigation interactive et de visualisation rapide des informations au sein de la structure multidimensionnelle. Avec l'avènement des big data, l'OLAP doit s'adapter aux caractéristiques des mégadonnées tout en gardant l'esprit de l'OLAP. De nouveaux modèles multidimensionnels pour les big data ont été alors proposés permettant de construire plus facilement des cubes OLAP. Deux nouveaux opérateurs d'agrégation OLAP ont alors été définis : C-Cube (*Columnar Cube*) qui exploite le stockage en colonnes du SGBD relationnel MonetDB [Dehdouh-14d] et CN-Cube (*Columnar NoSQL Cube*) qui utilise la notion de famille de colonnes du SGBD NoSQL Hbase [Dehdouh-14e]. Ces travaux constituent une avancée significative dans l'analyse en ligne des big data.

Analyse de données textuelles et de réseaux d'information. L'équipe SID s'est toujours intéressée à la prise en compte du critère « variété » des données dans les systèmes d'aide à la décision [Boukraa-14] notamment pour les données textuelles et les réseaux sociaux. Dans ce contexte, plusieurs travaux ont été menés et des résultats prometteurs ont été obtenus notamment pour la structuration des documents textuels. Ainsi, une architecture de base de données générique qui permet une structure de document flexible, intègre de nombreuses techniques de prétraitement et indexe automatiquement les textes [Truica-16]. Pour valider et comparer ces systèmes, des bancs d'essais ont été conçus à cet effet [Truica-17, Truica-18]. D'autres travaux de l'équipe se sont intéressés à la définition de cubes de textes en combinant des techniques issues de la recherche d'information, de la fouille de données et des graphes avec l'OLAP. Des mesures textuelles sous forme de vecteurs de termes et des opérateurs d'agrégation de documents textuels basés sur la notion de propagation de pertinence ont été proposés [Oukid-15, Oukid-16] ainsi que deux nouvelles fonctions d'agrégation basées sur les graphes [Bouakkaz-16]. Par ailleurs, des contributions importantes ont été réalisées dans le domaine de l'analyse en ligne des réseaux sociaux. Dans cette mouvance, l'équipe a présenté une approche qui permet d'enrichir les graphes avec des cubes de données qui viennent décrire les nœuds et/ou les arêtes du réseau selon les besoins d'analyse [Jakawat-16, Jakawat-16a, Hannachi-15]. D'autres travaux ont proposé de modéliser conjointement et dynamiquement les thématiques et les opinions de documents issus du Web (blogs, tweets, rapports, ...) en utilisant une approche basée sur les modèles de thématiques probabilistes (*topic models*) [Dermouche-15].

Production scientifique. Les 20 % des publications que l'équipe met en avant sont celles dont la référence est en gras (Annexe 4) et couvrent toutes nos thématiques de recherche. Le premier critère de sélection est celui des conférences et revues les mieux classées, puis en second terme de tri le respect de la diversité des courants thématiques de l'équipe. Les 20 % des revues mis en avant sont au nombre de 5 (parmi 26) dont 4 de rang A et 1 de rang B et les 20 % des articles publiés dans les conférences sont au nombre de 20 parmi 99. Ce sont des conférences internationales de bon niveau dans le domaine de recherche de l'équipe dont 2 sont de rang A (ICDM, FUZZ-IEEE) et 18 de rang B (ADBS, DAWAK, DEXA, IDEAS, etc.).

L'évolution de la production scientifique de l'équipe SID est détaillée dans les Figures 10 et 11. Elle montre une légère diminution par rapport à la période de référence précédente. Dans le même temps, le nombre de publications dans des revues internationales a augmenté même si un effort reste à faire pour les publications de rang A. Un point de vigilance est toutefois à noter concernant le nombre de revues non classées qui est également en augmentation même si elles sont bien reconnues dans le domaine et l'équipe souhaite les soutenir (*International Journal of Business Intelligence and Data Mining, International Journal of Data Warehousing and Mining*).

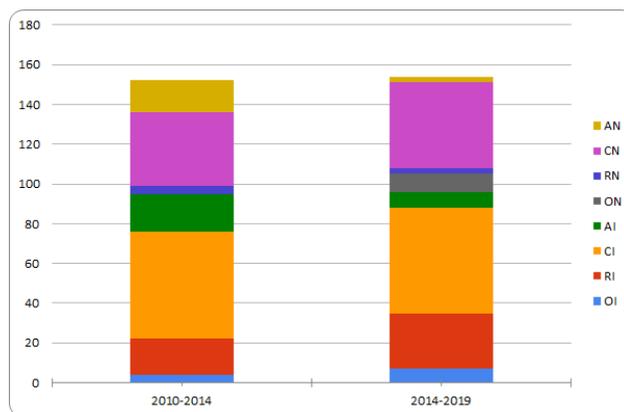


Figure 10 : Nombre de publications par type

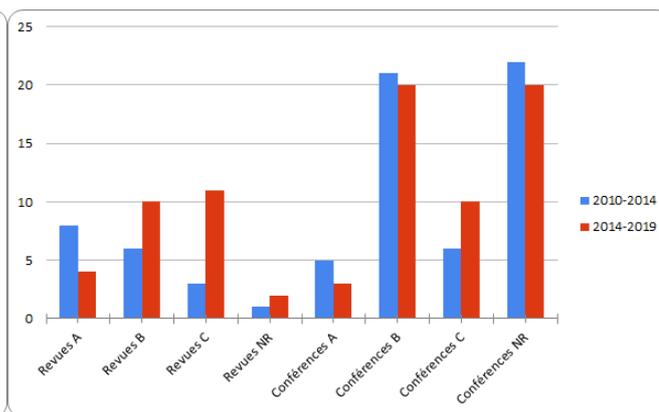


Figure 11 : Nombre de publications classées

O – Ouvrages et direction d'ouvrages, R – Revues, C – Conférences avec comité de lecture et actes, A – Autres publications, I – Portée internationale, N – Portée nationale, NR – Support de publication non référencé

Durant la période de référence, l'équipe a également développé des outils logiciels pour valider, tester et évaluer ses travaux de recherche. On peut citer les bancs d'essai décisionnels ou encore la plateforme pour l'intégration, la modélisation et l'analyse de documents textuels.

II. Interaction avec l'environnement académique.

La reconnaissance de l'équipe dans le domaine de la Business Intelligence et son important réseau de coopération national et international lui permettent d'être impliquée dans un grand nombre d'activités visibles liées à la recherche.

Contrats de recherche. Durant la période de référence, l'équipe a participé à 17 projets de recherche. Elle a porté un projet Tassili Hubert-Curien avec l'Algérie (Text-OLAP, 2011-2014), un projet PEPS (GéoNum, 2017-2019) et co-porté un projet financé par la région (BigData SHS.0, 2017-2019). Elle a également participé à un projet d'investissement d'avenir (REQUEST, 2014-2017), à plusieurs projets financés par le LabEx IMU (Hyper Thesaurus et lacs de données, 2018-2020 ; IDENUM, 2016-2019 ; CAPIMU, 2017-2018) et un projet financé par l'AUF (Projet RCCSO, 2013-2015). Enfin, l'équipe a obtenu plusieurs projets locaux financés par Lyon 2 ainsi que des projets en collaboration avec des partenaires industriels dans le cadre de thèses CIFRE. On peut trouver la liste complète des projets de l'équipe dans la section Projets de l'Annexe 4.

Collaborations. Les membres de l'équipe SID ont de nombreuses collaborations locales, avec les différents laboratoires en informatique et mathématiques appliquées de la métropole lyonnaise mais également avec plusieurs laboratoires en sciences humaines et sociales. Elle a également de multiples collaborations nationales (IRIT - Toulouse, LIRMM - Montpellier, LIMOS - Clermont-Ferrand, LIAS - Poitiers, LI - Tours, IRSTEA - Clermont-Ferrand) et internationales (Algérie, Angleterre, Canada, Espagne, Maroc, Roumanie, Ukraine, etc.) et accueille régulièrement des chercheurs étrangers pour des séjours au sein du laboratoire (Algérie, Canada, Pologne).

Activités d'évaluation et d'expertise. Les membres de l'équipe SID sont régulièrement sollicités pour évaluer des dossiers de projets de recherche (par exemple, pour l'ANR en France, COFECUB au Brésil, STIC-AMSUD (France Amérique du Sud), la FCI (Fondation Canadienne de l'innovation), le CRSNG (Conseil de Recherche en Sciences Naturelles et en Génie du Canada) et des projets de thèse CIFRE pour l'ANRT ainsi qu'à des comités de prix de thèses en France. Trois enseignants-chercheurs de l'équipe SID sont membres du CNU 27 et participent à ce titre à l'évaluation de dossiers de demandes de qualification, PEDR, de CRT, Promotions et de suivi de carrières. Un membre de SID a participé à la campagne d'évaluation individuelle des chercheurs et ingénieurs de recherche d'Irstea en 2018-2019, et un membre de l'équipe est au conseil d'orientation stratégique du laboratoire CESI LINEACT Région Sud-Est et au comité technique du projet [PIA ISTEEX](#). Un membre de SID est chercheur associé au laboratoire Surrey University, Royaume Uni. Notons également que trois des membres de SID sont membres du Labex IMU. L'un de ces membres est aussi co-responsable de l'axe scientifique « Sociétés et humanités numériques de la MSH Lyon-Saint-Etienne ». Enfin, de façon régulière, plusieurs membres de l'équipe participent à des comités de sélection (COS) et à l'évaluation d'articles dans plusieurs revues et conférences nationales et internationales du domaine.

Rayonnement et animation scientifique. 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. Elle coanime également l'action ADOC (Entreposage et analyse de documents) du groupe de recherche CNRS MaDICS, dont l'objectif est de rapprocher les communautés de gestion et d'entreposage des données de celles des SHS

impliquées dans les humanités numériques. L'équipe SID a mené des actions d'animation scientifique en organisant ou en coorganisant des conférences ou ateliers nationaux comme EDA2018 au Maroc, Data-Intelligence@INFORSID2018, EDA2017 et INFORSID 2014 à Lyon, Forum Jeunes Chercheurs à INFORSID2016 et des conférences ou ateliers internationaux comme DEXA 2017 à Lyon, WS@AICCSA2018, BigDataMAPS@ADBIS2018, BBIGAP@ADBIS2019, etc. L'équipe est investie également dans des actions liées aux humanités numériques en animant des ateliers au niveau national et international (Belgique, Bucarest). Mentionnons également la session démonstrations de logiciels d'EGC animée depuis 2010 par un membre de l'équipe SID. Plusieurs membres de l'équipe sont impliqués dans les conférences du domaine de très bon niveau en étant présents par exemple en tant que membres du comité de programme (ADBIS, DAWAK, DOLAP, DEXA, AICCSA, IDEAS, etc.). Les membres de l'équipe sont également très sollicités pour intégrer des comités de lecture de journaux internationaux (IJDSST, Future Internet, IJDWM, Data Semantics, etc.) ou intégrer des comités éditoriaux de certains journaux comme « International Journal of Biomedical and engineering and technology », « International journal of data mining, modeling and management » et « Open journal of databases ». Sur la période, les membres de l'équipe étaient actifs dans l'édition des actes de conférences ou de numéros spéciaux en tant que rédacteurs invités dans des journaux de leur domaine de recherche tels que (IJDWM, IGI Global). Plusieurs membres de SID sont invités pour des séminaires nationaux et internationaux : journée « Big data and cloud analytics », Algérie (2014) ; colloque international en sciences humaines et sociales », Roumanie (2017), etc.

III. Interaction avec l'environnement socio-économique

L'équipe SID affiche une volonté forte de partenariat et de transfert vers les acteurs socio-économiques. Elle a ainsi construit un tissu de relations, en particulier dans les domaines de la business intelligence, du big data, des réseaux sociaux et des humanités numériques principaux domaines d'exploitation des travaux de l'équipe. Ces collaborations prennent la forme de contrats de recherche ou de thèses CIFRE. Sur la période de référence, l'équipe collabore avec des entreprises [SWORD](#) Lyon ; [BialX](#) Lyon ; [Wizacha](#) Lyon, [Bertin IT](#) Montpellier et [Go Albert](#) Montpellier. Par ailleurs, l'ANRT sollicite régulièrement l'équipe SID pour des travaux d'expertise pour les demandes de thèses CIFRE.

IV. Implication dans la formation par la recherche

Responsabilité de masters. L'équipe SID est très impliquée dans la direction et le fonctionnement de masters : Responsabilité de la 1ère année de Master, de 2 parcours Master 2 (Business Intelligence et Big Data ; Organisation et Protection des Systèmes d'Information) et du Master Humanités Numériques avec la participation à la conception et à l'accréditation de ce dernier.

Participation à des réseaux de formation nationaux ou internationaux. L'équipe SID est également impliquée dans la responsabilité d'un master franco-ukrainien Informatique et Data Science pour le Management à Lyon 2 et des relations internationales à Lyon 1. Par ailleurs, l'équipe collabore de façon importante à l'échelle internationale : cotutelle et coencadrement de thèses avec des universités étrangères, accueil de stagiaires étrangers de niveau master, Interventions dans différents séminaires à l'étranger (Algérie, Espagne, Maroc, Roumanie, Tunisie, Vietnam, etc.).

Responsabilités administratives et pédagogiques importantes. Sur la période de référence, l'équipe était engagée dans des responsabilités administratives et pédagogiques de grande envergure telles que la responsabilité de la mention Master Informatique, la coordination locale du master Erasmus Mundus Data Management and Knowledge Discovery, la direction provisoire de l'ICOM - Institut de la communication (6 mois) et les TICE (1 an). Par ailleurs, l'équipe est impliquée au CNU 27 (3 membres), aux conseils centraux de Lyon 2 (2 membres) et de Lyon 1 (1 membre), au conseil de l'ICOM (4 membres), etc.

Jurys de thèse et de HDR. Les membres de l'équipe SID participent très régulièrement à l'évaluation de travaux de thèses en France ou à l'étranger. Sur la période de référence, l'équipe a participé à 49 jurys de thèses dont 12 à l'étranger et 3 HDR avec environ 50 % comme rapporteur globalement.

Faits marquants

- Coordination du projet EGIDE7 Tassili (2010-2014).** Ce projet franco-algérien, dont la thématique était l'analyse en ligne (OLAP) de données textuelles et des réseaux sociaux, a permis de financer plusieurs séjours scientifiques à ERIC de jeunes chercheurs de l'Université Saad Dahleb de Blida. Cette coopération a abouti à la codirection de plusieurs thèses (4 thèses soutenues) et à de nombreuses copublications et a fortement contribué à la reconnaissance de l'équipe à l'international.
- De **nombreux projets financés** sur la période notamment pour les projets en collaboration avec des laboratoires en SHS (financement Labex IMU, Région, Lyon 2, etc.). L'équipe est ainsi reconnue sur le site Lyon-Saint-Etienne comme partenaire privilégié pour les projets liés aux humanités numériques.
- Soutien de l'équipe aux conférences nationales.** Présidence de l'organisation du 32^e congrès INFORSID (2014) qui est la référence francophone dans le domaine des systèmes d'information ; Présidence du

comité de programme d'EGC (2015), ce qui a contribué à renforcer la reconnaissance et la visibilité des thèmes de l'équipe dans la communauté EGC ; Présidence de l'organisation de la 13^e édition de la conférence francophone EDA (2017) qui la référence francophone dans le domaine des systèmes d'information décisionnels.

3- Organisation et vie de l'équipe SID

Pilotage, animation, organisation de l'équipe

Organisation de l'équipe. L'équipe SID est composée de 6 enseignants-chercheurs de Lyon 2 (3 PR, 1 MCF-HDR, et 2 MCF) et 2 enseignants-chercheurs de Lyon 1 (2 MCF). Durant ce quinquennat, l'équipe SID n'a pas évolué en termes d'effectif (1 MCF promue PR en recrutement interne). En suivant les recommandations des experts, SID s'est réorganisée en deux axes de recherche pour permettre à chacun de ses membres de mieux se positionner et accroître sa visibilité à l'extérieur. Du fait de sa petite taille, l'organisation interne de l'équipe est relativement simple. Elle repose essentiellement sur la responsable (F. Bentayeb) qui assure l'essentiel des tâches administratives et de gestion, d'animation de la recherche et arbitre les différents choix, généralement après consultation de l'ensemble de l'équipe. La responsable assure également un lien de qualité avec la direction, ainsi qu'un ancrage et une représentativité scientifique tant interne qu'externe au laboratoire ERIC.

Vie de l'équipe. Pendant les 3 premières années du quinquennat, l'équipe se réunissait 2 fois par mois (1^{er} et 2^e vendredi de chaque mois) pour permettre au plus grand nombre d'assister aux réunions d'équipe. Depuis 2 ans, un seul créneau mensuel a été retenu (le 3^e lundi de chaque mois) pour la réunion mensuelle plus d'autres créneaux fixés selon les actualités de l'équipe. Ces réunions comportent une partie scientifique ouverte à toute l'équipe et une partie réservée aux permanents ciblée sur la gestion de l'équipe. Sur ces créneaux, l'équipe organise les séminaires où les orateurs sont les membres de l'équipe ; les orateurs externes sont souvent programmés dans le séminaire du laboratoire ERIC. Les réunions internes de l'équipe permettent aux doctorants de présenter l'état d'avancement de leurs travaux, les pré-soutenances de thèses et aux permanents de développer, par exemple, un point précis de la recherche. Par ailleurs l'équipe encourage ses MCF à préparer une HDR ; d'ailleurs l'un des membres de l'équipe est en CRCT pour préparer une HDR. Durant ce quinquennat, l'équipe était très active en termes d'animation scientifique notamment dans l'organisation de conférences, d'ateliers et de journées de travail tant au niveau national qu'international.

Profil d'activités. L'équipe SID a un profil caractérisé par (1) une recherche académique très forte de l'ordre de 40 % qui se matérialise par une production scientifique assez importante, (2) une interaction importante avec l'environnement socio-économique de l'ordre de 15 % qui se traduit par les nombreux projets locaux, projet région, et les contrats CIFRE, (3) un appui à la recherche de l'ordre de 15 % qui se matérialise par les activités diverses telles que l'édition de numéros spéciaux, de conférences et de workshops, et l'implication dans des comités de pilotage, de lecture et de programme, et (4) une participation importante à la formation par la recherche de l'ordre de 30 % visible au travers d'encadrements de doctorants et d'une implication forte dans des responsabilités lourdes et responsabilités de masters.

Politique de publication. La politique de publication de l'équipe vise et privilégie les revues et les conférences importantes du domaine, les ateliers internationaux spécialisés, les conférences nationales afin de maintenir et consolider les liens au niveau national. Il est important de noter que les supports de publication les plus pertinents dans le domaine de recherche de l'équipe sont classés B (JJDWM, DAWAK, DOLAP). Dans sa stratégie, l'équipe finance en priorité les missions pour les publications d'au moins de rang B à l'international mais encourage néanmoins les déplacements de doctorants dans les conférences du domaine pour leur permettre une meilleure intégration dans le monde de la recherche.

Parité

L'équipe SID compte 5 femmes (62 %) et 3 hommes (38 %) dans son effectif sachant que le CNU 27 affiche 23,8 % pour les femmes et 76,2 % pour les hommes. SID est sensible à la question de la parité et l'égalité professionnelle entre les femmes et les hommes ; l'un de ses membres est coresponsable de la mention de master Etudes sur le Genre à Lyon 2.

PROJET ET STRATÉGIE À CINQ ANS DE L'ÉQUIPE SID

1- Analyse SWOT

POINTS FORTS	POINTS À AMÉLIORER
<p>Évolution pertinente des thématiques de recherche en <i>Business Intelligence</i> (BI) et mégadonnées (<i>big data</i>) compte tenu de l'expertise de l'équipe, et en prise avec les défis scientifiques et sociétaux du moment.</p> <p>Positionnement scientifique original et clair dans le paysage lyonnais, régional et national.</p> <p>Reconnaissance des travaux de l'équipe au niveau international.</p> <p>Implication de l'équipe dans les réseaux nationaux et internationaux du domaine.</p> <p>Bonne attractivité de doctorants étrangers.</p>	<p>Poursuivre l'effort de publication d'articles dans les conférences et les revues les plus sélectives du domaine.</p> <p>Persévérer dans l'effort de montage de projets nationaux type ANR.</p> <p>S'appuyer sur la visibilité de l'équipe pour intégrer des consortiums internationaux afin de participer à des projets internationaux.</p>
POSSIBILITÉS OFFERTES PAR LE CONTEXTE	RISQUES LIÉS AU CONTEXTE
<p>L'équipe a de nombreuses relations avec son environnement social et économique, que ce soit au niveau local, régional ou national notamment par le biais de thèses CIFRE.</p> <p>Implication dans le LabEx pluridisciplinaire IMU (Intelligence des Mondes Urbains) avec des applications en sciences humaines et sociales.</p>	<p>Les tâches liées aux responsabilités administratives (7 enseignants-chercheurs sur 8 sont responsables de formations ou impliqués dans des instances locales et nationales) génèrent une surcharge de travail pour les membres de l'équipe.</p>

2- Structuration, effectifs et orientations scientifiques

Structuration de l'équipe et effectifs. Actuellement, l'équipe SID est composée de 8 permanents rassemblant 2 enseignants-chercheurs de l'Université Lyon 1 et 6 de l'Université Lyon 2. Pendant le prochain quinquennal, 2 départs à la retraite sont prévus (1 PR et 1 MCF). SID est dirigé par une responsable, membre du conseil de direction du laboratoire. Un renouvellement de direction de l'équipe est prévu pour le prochain quinquennal. L'équipe dispose d'un budget propre pris sur la dotation générale de fonctionnement du laboratoire et établi au prorata des effectifs. SID renforce ses ressources par des contrats et projets. La vie collective de l'équipe est riche, notamment autour des séminaires de doctorants et de journées d'étude. L'animation de l'équipe soutient une vie collective attachée à l'implication de ses membres dans toute leur diversité. Au cours de l'exercice, différentes périodicités ont été pratiquées (hebdomadaire, bimensuelle ou mensuelle sur des créneaux fixes) et les prises de décision sont collégiales et nourries d'échanges.

Orientations scientifiques. La BI est un domaine de recherche qui a connu de profondes mutations ces dernières années, dues notamment à la facilité d'accès aux outils décisionnels pour tous les usagers, ainsi qu'à l'avènement des *big data*, combiné aux innovations technologiques. L'équipe SID, grâce à son expertise et sa notoriété depuis près de 20 ans dans le domaine de la BI, a comme objectif de développer des travaux innovants pour permettre la gestion et l'analyse en ligne des *big data* en prise avec les nouveaux défis scientifiques et sociétaux actuels.

Objectifs scientifiques. Pour le prochain quinquennal, les objectifs de l'équipe sont de conforter le positionnement leader de l'équipe SID dans le domaine de la BI au plan national, et d'asseoir sa visibilité à l'international. L'équipe a pour objectif d'affirmer un positionnement scientifique clair et assumé sur les axes de recherche *big data management* et *BI & analytics*, en faisant évoluer ses thématiques de recherche au sein de ces deux axes. Par ailleurs, l'équipe compte poursuivre sa politique de collaboration scientifique avec les laboratoires SHS, notamment via des projets et des applications en humanités numériques.

Choix stratégiques (politique de recrutement). Pour le prochain quinquennal 2021-2026, l'équipe souhaite acquérir de nouvelles expertises dans des thématiques liées à son projet scientifique. Cela passe par des demandes de postes d'enseignants-chercheurs (soit en remplacement de départs à la retraite, soit en création) sur des profils liés à ces thématiques : 5 postes MCF et un poste PR avec un profil majeur BI & big data. Les postes MCF impliquent tous une compétence complémentaire afin de couvrir les thématiques actuelles ou nouvelles affichées par l'équipe : sécurité, parallélisme et données distribuées, TAL, Web sémantique, RI.

Moyens pour atteindre les objectifs. SID affiche une forte attractivité internationale, comme en témoignent ses multiples collaborations et ses nombreux doctorants étrangers. L'équipe compte poursuivre le développement de son réseau international pour accroître sa visibilité et recruter de nouveaux doctorants. En outre, les partenariats industriels (thèses CIFRE) seront aussi recherchés de manière proactive. Les efforts de l'équipe pour cibler des supports de publications de qualité les mieux référencés du domaine seront également reconduits pour maintenir le leadership de l'équipe dans le domaine de la BI. Par ailleurs, SID est fortement impliqué dans les projets liés aux humanités numériques, que ce soit localement, dans le cadre du LabEx IMU ou encore dans celui d'un projet COST soumis en 2018. Pour terminer, l'équipe doit tirer parti du soutien appuyé des tutelles et des relations existantes avec des partenaires nationaux ou étrangers potentiels pour répondre à des appels d'offres nationaux et internationaux tels que les projets ANR, COST ou EGIDE.

Renouvellement des partenariats. SID a un positionnement original dans l'environnement lyonnais du fait de ses travaux de recherche en BI. L'engagement de l'équipe dans son projet scientifique *big data management* et *BI & analytics* va lui permettre de renforcer les liens avec les autres laboratoires traitant de la BI en France et d'initier de nouvelles collaborations avec les entreprises. Les partenariats internationaux de l'équipe seront également confortés, que ce soit en termes d'animation scientifique, de coencadrement de thèses ou de projets de recherche.

Nouvelles thématiques scientifiques. Le projet scientifique de l'équipe s'inscrit dans la nouvelle mouvance de la BI, mais toujours dans la lignée des deux axes de recherche qu'elle a développés récemment : *big data management* et *BI & analytics*. L'objectif est de renforcer et d'investir en profondeur ces deux axes de recherche en tenant compte à la fois des big data, des technologies nouvelles et de l'évolution de la BI, qui sont en train de fondamentalement bouleverser le paysage des architectures décisionnelles.

Axe 1 - Big data management. Dans cet axe de recherche, l'équipe souhaite développer de nouvelles architectures décisionnelles massivement parallèles plus appropriées aux *big data*, en investissant des sujets novateurs tels que les lacs de données, avec la gestion et la génération de métadonnées pour gagner en agilité dans le processus d'aide à la décision ; les modèles d'entrepôts distribués en s'appuyant sur des stratégies de placement et de distribution de données pour gagner en performance, notamment avec les SGBD NoSQL ; et la sécurité des données pour la détection dynamique et en temps réel des altérations des données afin de garantir leur véracité et leur intégrité. Les données peuvent être de type numérique, textuel ou multimédia.

Axe 2 - BI & analytics. L'équipe s'intéresse 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. L'équipe s'intéresse également à la BI à la volée dans le contexte des nouvelles architectures décisionnelles agiles. SID prévoit également l'utilisation de méthodes issues de la fouille de données, de la RI, des services web, de la théorie des graphes ou encore de l'analyse d'images pour l'acquisition, l'intégration et l'analyse en ligne des *big data*. Les données étudiées peuvent être des corpus de données non structurées, des données textuelles issues des réseaux d'information, des réseaux sociaux ou encore des données multimédia. L'équipe s'intéresse à plusieurs types d'analyse en ligne, comme par exemple text-OLAP, graph-OLAP et NoSQL-OLAP, suivant le type des données ; mais aussi à la classification et à l'étiquetage de données textuelles ou multimédia, à la détection d'opinions dans les réseaux sociaux ou encore à l'analyse de ces derniers pour la veille et la prédiction des cybermenaces.

BILAN DES ACTIVITÉS DE L'AXE TRANSVERSAL HUMANITÉS NUMÉRIQUES

1- Présentation

À la fois par son appartenance à des structures essentiellement SHS (MSH-LSE, où une membre d'ERIC coanime l'axe scientifique Sociétés et humanités numériques ; Institut du Genre) et interdisciplinaires (LabEx IMU, programme IDEX Sociétés numériques), par son implication forte dans le master Humanités numériques coaccrédité par Lyon 2, Lyon 3, l'ENS Lyon et l'ENSSIB (dont la coordination de la mention), et par sa participation à de nombreux projets interdisciplinaires avec divers laboratoires de LLSHS, ERIC est devenu un acteur reconnu des humanités numériques (HN) sur le site de Lyon-Saint-Étienne. Cela nous a notamment valu d'être sollicités comme partenaire dans deux projets d'EUR déposés en 2019 : PASS – *Past Societies Studies* et GENDERING – *School of Gender Studies*.

Politique scientifique

En réponse à une demande des évaluateurs du laboratoire en 2009 de clarifier le positionnement d'ERIC dans le paysage lyonnais, nous avons souhaité affirmer une identité pluridisciplinaire informatique et mathématiques appliquées présente depuis la création du laboratoire, et également interdisciplinaire, en structurant et en développant nos collaborations avec les disciplines des LLSHS en un axe stratégique du laboratoire. Lors de l'évaluation HCERES de 2014, cet axe avait été jugé trop implicite dans notre autoévaluation, aussi le développons nous de manière séparée dans le présent rapport.

Scientifiquement, notre objectif n'est pas seulement de trouver des terrains d'application à nos recherches, mais surtout d'hybrider les méthodologies informatiques et statistiques à celles des LLSHS pour aboutir à des approches originales. Nous nous inscrivons également dans des collaborations sur le temps long, nécessaire à un travail interdisciplinaire qui porte réellement des fruits.

Par ailleurs, nous sommes en train d'intégrer les infrastructures nationales et européennes liées aux HN, soit à travers le master HN (CLARIN et DARIAH), soit à travers des projets de recherche où l'on a commencé à travailler, par exemple, avec la TGIR Huma-Num (dans les projets GéoNum et HyperThéséau), soit en adhérant aux sociétés savantes Humanistica et ADHO.

2- Produits et activités de la recherche liés aux humanités numériques

Bilan scientifique

En premier lieu, des projets de recherche (associés ou non à des thèses financées – deux codirections avec les laboratoires ARAR et EVS, respectivement) fortement liés aux HN ont été menés. Ces projets ont donné lieu à des collaborations très riches avec des laboratoires du site et extérieurs à Lyon, dans des disciplines variées.

- Archéologie et Archéométrie (ARAR)
- Centre Max Weber (CMW, sociologie)
- Centre de recherche en terminologie et traduction (CRTT)
- Coactis (Sciences de gestion)
- Éducation, Cultures, Politiques (ECP, pluridisciplinaire)
- Environnements et sociétés de l'Orient ancien (Archéorient)
- Équipe de recherche de Lyon en sciences de l'information et de la communication (ELICO)
- Histoire, Archéologie, Littératures, des mondes chrétiens et musulmans médiévaux (CIHAM)
- Histoire et Sources des Mondes Antiques (HISoMa)
- Institut de Recherches Géographiques du laboratoire Environnement, Ville, Société (EVS)
- Interactions, corpus, apprentissages, représentations (ICAR)
- Laboratoire de recherche historique Rhône-Alpes (LARHRA)
- MARGE (pluridisciplinaire littérature et information-communication)
- Centre d'Études Politiques de l'Europe Latine (CEPEL), Montpellier
- Centre Régional Francophone de Recherches Avancées en Sciences Sociales (CentreCeReFre), Bucarest
- Dauphine Recherche en Management (DRM), Paris
- Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication (GERiCO), Lille
- Institut de Linguística Aplicada (ULA, linguistique), Barcelone
- Laboratoire Image Ville Environnement (LIVE, géographie), Strasbourg
- Savoirs, Textes, Langages (STL), Lille
- Territoires, Environnement, Télédétection et Information Spatiale (TETIS), Montpellier

Notre production scientifique liée aux HN (Annexe 4) est quantitativement moins abondante que notre production strictement disciplinaire. Elle souffre également d'un « effet retard » en raison des temps plus longs

induits par les recherches interdisciplinaires. Toutefois, le nombre et la fréquence de ces publications augmente d'un facteur 6 entre les périodes 2010-2014 et 2014-2018 (Figure 12), notamment avec le recrutement de collègues sur des profils incluant les HN et le développement des projets déjà mentionnés. Il faut également noter que, la nature des articles évolue : de travaux en informatique applicables à des problématiques SHS à de plus en plus de publications dans des revues interdisciplinaires ou de SHS (qui sont celles que nous avons valorisées en premier lieu dans les 20 %, puis ensuite les publications en informatique liées ou applicables aux HN parues dans des supports référencés). Enfin, des logiciels conçus à ERIC sont susceptibles d'être exploités dans un contexte HN et certains, comme NewsBrowsers, qui permet d'annoter et d'analyser des articles de presse, ont été conçus spécifiquement à cette fin (Annexe 4). NewsBrowsers a par exemple été utilisé dans le cadre de la thèse de S. Kurpiel (Centre Max Weber).

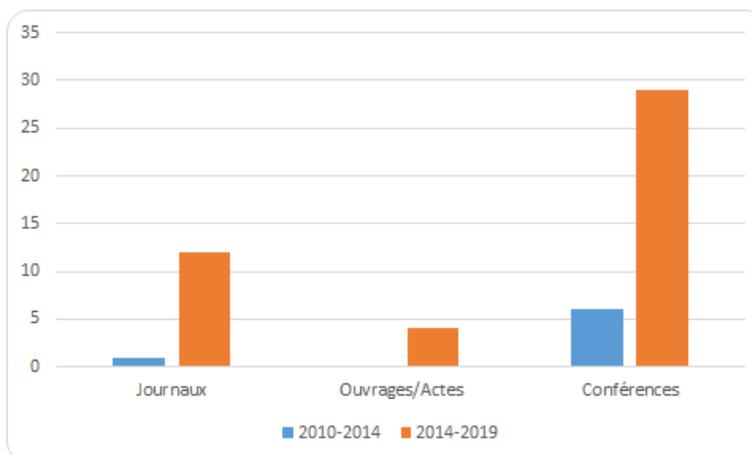


Figure 12 : Évolution des publications liées aux HN

ERIC est également actif dans l'animation scientifique autour des HN, notamment à travers la coanimation de l'action ADOC ([Entrepôts et analyse de documents](#)) du GdR MaDICS, qui vise à mettre en relation des scientifiques en SHS et en informatique, autour de la notion de gestion et d'analyse de documents (action débutée en 2016 et renouvelée pour deux ans en 2018) ; par l'implication dans le GT DAHLIA ([Digital Humanities and cultural heritage](#)) de l'association EGC (initié en 2019) ; ainsi que par (co)organisation de journées d'étude et colloques interdisciplinaires, dont la [4^e édition des journées Big Data Mining and Visualization – Focus sur les Humanités Numériques dans le Big Data](#) en 2015, qui a rassemblé une centaine de participants (Annexe 4).

Par ailleurs, le laboratoire ERIC s'inscrit aussi dans une démarche favorisant l'interdisciplinarité du point de vue du champ par essence interdisciplinaire (notamment avec la sociologie, la linguistique, l'histoire et la science politique) que sont les études de genre et de l'égalité femmes-hommes. Ceci s'ancre à la fois en termes de recherches qui ont donné lieu à des communications et des articles sur le sujet de la masculinisation de l'informatique et la question des carrières de femmes dans ce domaine, mais aussi en termes d'animation scientifique, avec l'organisation ou la participation à des événements et des projets dont l'ancrage informatique sera à développer. Le rattachement du laboratoire en 2016 au Groupe d'Intérêt Scientifique « Institut du Genre » a institutionnalisé cette démarche.

Enfin, ERIC a lancé conjointement avec le laboratoire ELICO (information-communication) une série de séminaires croisés, avec à chaque fois une présentation scientifique d'un.e enseignant.e-chercheur.e de chaque laboratoire, l'objectif étant de dégager des objets de recherche communs et de réfléchir à des thématiques scientifiques interdisciplinaires. Quatre séminaires de ce type ont eu lieu en 2016-2017.

Formation à et par la recherche

Trois thèses (dont une soutenue) et 15 stagiaires de master ont travaillé au laboratoire dans des projets directement liés aux humanités numériques, pratiquement systématiquement en collaboration avec un laboratoire de SHS. La durée longue de la seule thèse soutenue à ce jour (57 mois) est liée à la difficulté de mener un projet interdisciplinaire en informatique et archéologie/archéométrie, ainsi qu'à un long travail (plus d'un an) de nettoyage et de restructuration des données qui n'était pas initialement prévu.

Faits marquants

2016 : Association du laboratoire au LabEx IMU. Lors des appels à projets LabEx, ERIC s'était investi sur une proposition liée aux humanités numériques et n'avait pas eu la possibilité d'intégrer le projet IMU, pourtant également cohérent avec nos thématiques de recherche. Le projet de LabEx HN n'ayant pas été retenu, nous n'avons pas pu bénéficier de ce type de structure jusqu'en 2016. Grâce à l'implication dans le projet IMU IDENUM (porté par le laboratoire ELICO), nous avons pu démontrer notre expertise et intégrer le LabEx. Cela nous a ouvert de nouvelles perspectives de collaboration, qui ont notamment mené à notre participation dans

le projet HyperThéséau (porté par le laboratoire Archéorient et démarré en 2018) et nous ont permis de financer plusieurs stages de master.

2017-2018 : Séjour scientifique de Ian Davidson (University of California, Davis). Ian Davidson a travaillé avec plusieurs membres du laboratoire sur des problématiques d'interprétabilité des modèles usuellement développés en fouille de données, ainsi qu'avec des chercheurs en SHS dont J.C. Soulages du Centre Max Weber. Ces échanges ont permis le développement de techniques visant à expliquer les catégories obtenues automatiquement par les méthodes de classification non supervisée (*clustering*). Ces travaux ont donné lieu à deux publications, dont une dans la prestigieuse conférence NeurIPS, et la collaboration avec Ian Davis se poursuit, élargissant ainsi le réseau international d'ERIC.

2017 : Publication « Le Web politique au prisme de la science des données » dans la revue Réseaux. Cette publication rend compte de contributions réalisées à l'occasion des journées d'étude "Etudier le Web politique : Regards croisés", organisées conjointement par les laboratoires ERIC, CEPEL (science politique) et ELICO (information-communication). Ces journées étaient ouvertes à la fois aux chercheurs en SHS (sociologues, politistes, économistes, sémiologues, chercheurs en sciences de l'information et de la communication) s'intéressant aux marques d'expressions politiques exprimées sur le Web, mais également aux chercheurs en informatique et en mathématiques qui conçoivent de nouveaux modèles et développent des techniques de recherche d'information, d'apprentissage automatique, de visualisation et de navigation qui prennent une place de plus en plus prépondérante dans ces recherches. Cette publication dans une revue reconnue en SHS témoigne des efforts investis par les membres laboratoire dans l'interdisciplinarité et l'axe transversal HN.

2018 : Événements grand public lors du Off de la Web conference, en collaboration avec la MSH-LSE (axes Sociétés et Humanités numériques et Genre), le collectif L Digital et l'INSA de Lyon. Le 1^{er} événement (avril), « L'informatique, pourquoi pas moi ? Les jeunes prennent la parole ! », s'est adressé aux collégiens de 3^e et a mêlé une conférence « La place de l'informatique dans la société d'aujourd'hui » à des ateliers interactifs sur les thèmes « Comment être mieux informé sur les études d'informatique, à l'école et en famille ? », « Comment susciter plus d'intérêt pour l'informatique auprès des filles ? » et « Comment développer la pratique de l'informatique, à l'école et en dehors ? ». Le 2^e événement (juin), « L'informatique, pourquoi pas elles ? », une conférence suivie d'une table ronde, a permis d'analyser comment l'informatique était devenue un métier d'hommes et de débattre des mesures inclusives à mettre en œuvre pour attirer des femmes dans les métiers liés au numérique.

PROJET ET STRATÉGIE À CINQ ANS DE L'AXE TRANSVERSAL HN

1- Analyse SWOT

POINTS FORTS	POINTS À AMÉLIORER
<p>Nombreuses collaborations locales et nationales avec des laboratoires de LLSHS, notamment à travers des projets de recherche financés.</p> <p>Implication dans la coordination de l'axe Sociétés et humanités numériques de la MSH-LSE et du master HN de site, ainsi que dans les collèges de l'Université de Lyon concernés (ingénierie et SHS).</p>	<p>Insertion plus marquée dans les réseaux nationaux et internationaux explicitement liés aux HN.</p> <p>Malgré la progression lors du quinquennal écoulé, copublication avec les partenaires LLSHS.</p>
POSSIBILITÉS OFFERTES PAR LE CONTEXTE	RISQUES LIÉS AU CONTEXTE
<p>Appartenance à des structures par essence pluri/transdisciplinaires comme la MSH-LSE, le LabEx IMU et le GIS Institut du Genre, qui induisent une proximité avec les LLSHS.</p> <p>Développement de plus en plus marqué d'un pan des HN lié à la gestion et à l'analyse des données, au côté du pan édition numérique, plus ancien.</p>	<p>Paradoxalement, la grande richesse de l'écosystème lyonnais en HN peut provoquer des effets de concurrence involontaires et induire des problèmes de coordination à l'échelle du site.</p>

2- Orientations scientifiques

Les travaux scientifiques collaboratifs liés aux humanités numériques menés lors du contrat quinquennal écoulé ont principalement mobilisé et fait progresser à ERIC les connaissances liées à la gestion et l'analyse de documents textuels de tailles et de types divers (du tweet à l'ouvrage numérisé). La tendance qui se précise dès à présent et pour le prochain quinquennal implique toute la variété des mégadonnées (structurées, semi-structurées et non structurées), qu'il faut gérer et interroger de concert. C'est une thématique tout juste initiée au laboratoire avec, par exemple, nos travaux sur les lacs de données, et qui structurera notre réflexion. De plus, l'accessibilité de ces types d'organisation, de requêtage et d'analyse des données à des non spécialistes de l'informatique ou de la science des données, tels que nos partenaires chercheur.es en LLSHS, pose des problèmes de fond qu'il faudra traiter.

Dans la même veine, une autre tendance essentielle souligne l'importance de l'explicabilité. À partir du moment où l'on donne l'opportunité à des utilisateurs non informaticiens d'utiliser, voire d'interagir, avec le résultat de modèles et d'algorithmes de fouille de données, il ne s'agit plus uniquement de fournir le résultat d'un algorithme de *clustering*, par exemple, mais aussi une description concise des *clusters* produits et, si possible, de revenir aux données d'origine afin que l'utilisateur comprenne pour quelles raisons des objets (par exemple des textes) ont été placés dans une même catégorie. En classification supervisée, pour donner un autre exemple, il faut être capable de fournir les descripteurs, ou leur combinaison, qui justifient de classer un objet dans une classe et pas dans une autre. Sans cela, nous courons le risque que l'utilisateur ne fasse pas confiance au système, aussi sophistiqué qu'il puisse être.

Par ailleurs, au-delà de notre volonté d'hybrider les méthodologies informatiques et statistiques à celles des LLSHS pour aboutir à des approches originales (Section 1), nous prévoyons d'initier lors du prochain contrat quinquennal une réflexion méta, voire épistémologique, en collaboration avec nos partenaires LLSHS, afin de déterminer s'il existe des régularités dans les méthodologies adoptées dans les différents projets de recherche HN que nous menons.

D'un point de vue plus opérationnel, bien que le nombre de copublications avec nos partenaires LLSHS ait augmenté au cours du quinquennal écoulé, nous allons poursuivre nos efforts pour développer encore ce volet collaboratif. De même, nous souhaitons contribuer à la formation des chercheur.es (doctorant.es, jeunes chercheur.es ou chercheur.es confirmé.es) en LLSHS aux méthodes et outils numériques, par l'organisation récurrente d'une école d'été ou d'hiver organisée en collaboration avec la MSH-LSE.

Par ailleurs, il nous apparaît que nous sommes insuffisamment insérés dans les réseaux internationaux liés aux HN. Ces réseaux comptent essentiellement dans leurs membres des chercheur.es en LLSHS et des ingénieur.es en informatique, mais la recherche en sciences des données, peu présente et donc peu représentée, y a complètement sa place. La coorganisation avec l'UMS PERSEE de la conférence internationale de référence *Digital Humanities* à Lyon en 2022, en coopération plus large avec les unités de recherche de l'Université Lyon 2 et de l'ENS de Lyon, devrait nous permettre de progresser grandement dans ce domaine. Nous prévoyons également de développer un partenariat de recherche avec la chaire *Digital Humanities* de l'université d'Ottawa, dont les chercheurs sont reconnus pour faire partie des leaders canadiens en HN. Notre objectif est de rendre le laboratoire ERIC encore plus visible dans ces réseaux pour en devenir un acteur national et international capable d'intégrer et de mener des projets financés de grande envergure.

Enfin, il est à noter que tous les recrutements à ERIC intègrent dans les profils de poste une composante Humanités numériques. Qu'elle soit centrale comme pour les postes de professeur ouverts au concours en 2012 et 2015, implicite comme pour le poste de professeur ouvert au concours dans la filière Mathématiques et Informatique appliquées aux Sciences Humaines et Sociales (MIASHS) en 2019 ou plus de l'ordre de la coloration (expérience ou ouverture à des problématiques HN, selon le niveau de recrutement), elle est toujours présente depuis 2012.

ANNEXES

Annexe 1 : Lettre d'engagement

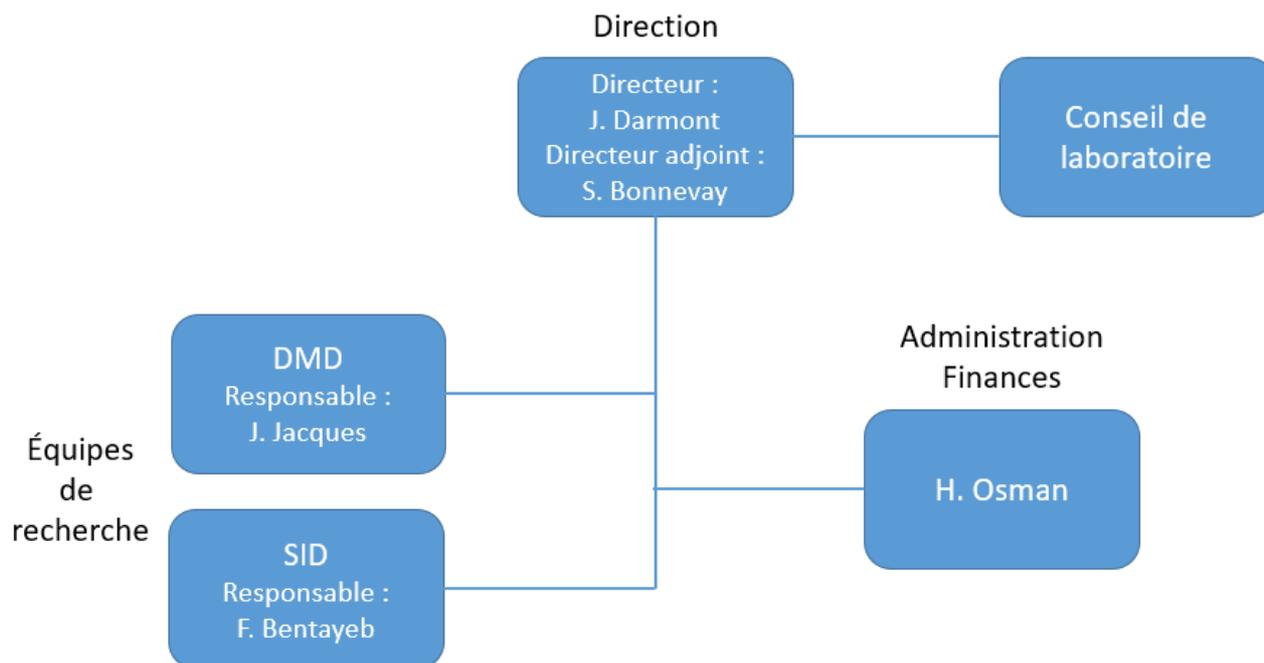
Annexe 2 : Organigramme fonctionnel

Annexe 3 : Équipements, plateformes

Annexe 4 : Produits et activités de la recherche

1. Équipe DMD
2. Équipe SID
3. Axe transversal interdisciplinaire Humanités numériques

Annexe 2 : Organigramme fonctionnel



Annexe 3 : Équipements, plateformes

Infrastructure de calcul et de stockage hébergée à la DSI Lyon 2 :

- 3 serveurs Dell PowerEdge R640 avec 384, 256 et 256 Go de RAM, respectivement ; virtualisation VMware Vsphere
- 1 baie de stockage de 18 To

Au laboratoire :

- 1 PC serveur web (en cours de migration sur une machine virtuelle)
- 1 NAS de 18 To pour stockage lent

ANNEXE 4 - Produits et activités de la recherche

En cohérence avec les données chiffrées de l'onglet 4 du fichier Excel « Données du contrat en cours », on remplira ce document destiné à l'évaluation du critère 1 du référentiel de l'évaluation « Produits et activités de la recherche », pour l'ensemble de l'unité et pour chaque équipe / thème.

CAMPAGNE D'ÉVALUATION 2019-2020 VAGUE A

Nom de l'équipe : Data Mining et Décision

Acronyme : DMD

Responsable d'équipe pour le contrat en cours : Julien Jacques

Responsable d'équipe pour le contrat à venir :

I- PRODUCTION DE CONNAISSANCES ET ACTIVITÉS CONCOURANT AU RAYONNEMENT ET À L'ATTRACTIVITÉ SCIENTIFIQUE DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME

Les 20 % les plus significatifs de la production scientifique sont indiqués par une référence en gras et séparés des 80 % restants. Les membres du laboratoire sont indiqués en majuscules soulignées et les doctorant·es de plus distingués en gras.

1- Journaux / Revues

Articles scientifiques (44 dont 38 dans des revues internationales).

1. **[AhPine-18]** AH-PINE J. «An Efficient and Effective Generic Agglomerative Hierarchical Clustering Approach». In *Journal of Machine Learning Research* (2018). Ref HAL: [hal-01981827 v1](https://hal.archives-ouvertes.fr/hal-01981827). Rang A.
2. **[Dubois-17a]** Dubois D., Prade H., RICO A., Teheux B. «Generalized qualitative Sugeno integrals». In *Information Sciences*, vol. 415 - 416 (2017) p.429-445. DOI: [10.1016/j.ins.2017.05.037](https://doi.org/10.1016/j.ins.2017.05.037). Rang A.
3. **[Dubois-17]** Dubois D., Prade H., RICO A. «Graded cubes of opposition and possibility theory with fuzzy events». In *International Journal of Approximate Reasoning*, vol. 84 (2017) p.168-185. DOI: [10.1016/j.ijar.2017.02.006](https://doi.org/10.1016/j.ijar.2017.02.006). Rang A.
4. **[AhPine-16b]** AH-PINE J. «On aggregation functions based on linguistically quantified propositions and finitely additive set functions». In *Fuzzy Sets and Systems*, vol. 287 (2016) p.1-21. DOI: [10.1016/j.fss.2015.06.002](https://doi.org/10.1016/j.fss.2015.06.002). Rang A.
5. **[Rizoïu-16]** RIZOIU M.-A., VELCIN J., BONNEVAY S., LALLICH S. «ClusPath: a temporal-driven clustering to infer typical evolution paths». In *Data Mining and Knowledge Discovery*, vol. 30/5 (2016) 1324 - 1349. DOI: [10.1007/s10618-015-0445-7](https://doi.org/10.1007/s10618-015-0445-7). Rang A.
6. **[Rico-15]** RICO F., Muhlenbach F., ZIGHED D. A., LALLICH S. «Comparison of two topological approaches for dealing with noisy labeling». In *Neurocomputing*, vol. 160 (2015) 3 - 17. DOI: [10.1016/j.neucom.2014.10.087](https://doi.org/10.1016/j.neucom.2014.10.087). Rang A.
7. **[Biernacki-16]** Biernacki C., JACQUES J. «Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm». In *Statistics and Computing*, vol. 26/5 (2016) p.929-943. Ref HAL: [hal-01052447 v2](https://hal.archives-ouvertes.fr/hal-01052447). Rang A.
8. **[Bouveyron-15]** Bouveyron C., Côme E., JACQUES J. «The discriminative functional mixture model for a comparative analysis of bike sharing systems». In *The Annals of Applied Statistics* (2015). Ref HAL: [hal-01024186 v3](https://hal.archives-ouvertes.fr/hal-01024186). Rang A.
9. [BenSlimen-18] **BENSLIMEN Y.**, Allio S., JACQUES J. «Model-Based Co-clustering for Functional Data». In *Neurocomputing* (2018). Ref HAL: [hal-01422756 v1](https://hal.archives-ouvertes.fr/hal-01422756). Rang A.
10. [Jacques-18] JACQUES J., Biernacki C. «Model-Based Co-clustering for Ordinal Data». In *Computational Statistics & Data Analysis*, vol. 123 (2018) p.101-115. DOI: [10.1016/j.csda.2018.01.014](https://doi.org/10.1016/j.csda.2018.01.014). Rang A.
11. [Bouveyron-17] Bouveyron C., Bozzi L., JACQUES J., Jollois F.-X. «The Functional Latent Block Model for the Co-Clustering of Electricity Consumption Curves». In *Journal of the Royal Statistical Society Series C Applied Statistics* (2017). Ref HAL: [hal-01533438 v1](https://hal.archives-ouvertes.fr/hal-01533438). Rang A.
12. [Antoniadis-16] Antoniadis A., Brossat X., CUGLIARI J., Poggi J.-M. A prediction interval for a function-valued forecast model : Application to load forecasting. *International Journal of Forecasting*, vol. 32 (2016). Rang A.
13. [Dubois-16a] Dubois D., Prade H., RICO A. «Residuated variants of Sugeno integrals: Towards new weighting schemes for qualitative aggregation methods». In *Information Sciences*, vol. vol. 329 (2016). DOI: [10.1016/j.ins.2015.09.034](https://doi.org/10.1016/j.ins.2015.09.034). Rang A.
14. [Nguyen-15] Nguyen T. H., Shirai K., VELCIN J. «Sentiment analysis on social media for stock movement prediction». In *Expert Systems with Applications* (2015). DOI: [10.1016/j.eswa.2015.07.052](https://doi.org/10.1016/j.eswa.2015.07.052). Rang A.
15. [Dubois-15] Dubois D., Prade H., RICO A. «Representing qualitative capacities as families of possibility measures». In *International Journal of Approximate Reasoning*, vol. 58 (2015). Ref HAL: [hal-01145911 v1](https://hal.archives-ouvertes.fr/hal-01145911). Rang A.
16. [Nedellec-14] Nedellec R., CUGLIARI J., Goude Y. GEFCom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, vol. 30 (2014). Rang A.
17. [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.

18. [Cugliari-18a] [CUGLIARI J.](#), [ROLLAND A.](#) «Simulation of multicriteria data». In *EURO Journal on Decision Processes*, vol. 6/1-2 (2018) p.21-37. Ref HAL: [hal-02053025 v1](#). Rang B.
19. [Lumbreras-17] [LUMBRERAS A.](#), [JOUVE B.](#), [VELCIN J.](#), Guégan M. «Role detection in online forums based on growth models for trees». In *Social Network Analysis and Mining*, vol. 7/1 (2017) p.49. DOI: [10.1007/s13278-017-0472-z](#). Rang B.
20. [Lumbreras-16] [LUMBRERAS A.](#), [VELCIN J.](#), Guégan M., [JOUVE B.](#) «Non-parametric clustering over user features and latent behaviors with dual-view mixture models». In *Computational Statistics* (2016). DOI: [10.1007/s00180-016-0668-0](#). Rang B.
21. [AhPine-15b] [AH-PINE J.](#), Csurka G., Clinchant S. «Unsupervised Visual and Textual Information Fusion in CBMR Using Graph-Based Methods». In *ACM Transactions on Information Systems*, vol. 33/2 (2015) p.9. DOI: [10.1145/2699668](#). Rang B.
22. [Rolland-15] [ROLLAND A.](#), [AH-PINE J.](#), Mayag B. «Elicitation of 2-additive bi-capacity parameters». In *EURO Journal on Decision Processes*, vol. 3/1 (2015) p.5-28. DOI: [10.1007/s40070-015-0043-3](#). Rang B.
23. [Guille-15] [GUILLE A.](#), [FAVRE C.](#) «Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach». In *Social Network Analysis and Mining*, vol. 5/1 (2015) 18:1-18:18. DOI: [10.1007/s13278-015-0258-0](#). Rang B.
24. [Memmah-14] Memmah M.-M., Quilot-Turion B., [ROLLAND A.](#) «Multicriteria sorting methods to select virtual peach ideotypes». In *International Journal of Multicriteria Decision Making(IJMCDM)*, vol. 4/4 (2014) 348 - 366. DOI: [10.1504/IJMCDM.2014.066874](#). Rang B.
25. [Bonnevay-16] [BONNEVAY S.](#), [GAVIN G.](#), [AUBERTIN P.](#) «Comparison of two metaheuristics to solve a 2-D cutting stock problem with set-up cost in the paper industry». In *International Journal of Metaheuristics*, vol. 5/1 (2016) p.31. Ref HAL: [hal-02051744 v1](#). Rang C.
26. [Rizoiu-14] [RIZOIU M.-A.](#), [VELCIN J.](#), [LALLICH S.](#) «How to Use Temporal-Driven Constrained Clustering to Detect Typical Evolutions». In *International Journal on Artificial Intelligence Tools*, vol. 23/04 (2014) p.1460013. Ref HAL: [hal-02052405 v1](#). Rang C.
27. [Rizoiu-14a] [RIZOIU M.-A.](#), [VELCIN J.](#), [LALLICH S.](#) «Semantic-enriched Visual Vocabulary Construction in a Weakly Supervised Context». In *Intelligent Data Analysis*, vol. 19/1 (2014). Ref HAL: [hal-00959355 v1](#). Rang C.
28. [MartinezAlvarez-19] Martínez-Álvarez F., [SCHMUTZ A.](#), Asencio-Cortés G., [JACQUES J.](#) «A Novel Hybrid Algorithm to Forecast Functional Time Series Based on Pattern Sequence Similarity with Application to Electricity Demand». In *Energies*, vol. 12/1 (2019) p.94. Ref HAL: [hal-02051183 v1](#).
29. [Auder-18] Auder B., [CUGLIARI J.](#), Goude Y., Poggi J-M. «Scalable Clustering of Individual Electrical Curves for Profiling and Bottom-Up ». In *Energies*, vol. 11/7 (2018).
30. [Nagbe-18a] [NAGBE K.](#), [CUGLIARI J.](#), [JACQUES J.](#) «Electricity Demand Forecasting Using a Functional State Space Model». In *Energies*, vol. 11/5 (2018) p.1120. Ref HAL: [hal-01701258 v2](#).
31. [Bacou-17] Bacou E., Haurogne K., Mignot G., Allard M., De Beaupaire L., Marchand J., Terenina E., Billon Y., [JACQUES J.](#), Bach M., Mormède P., Hervé J., Lieubeau B. «Acute social stress-induced immunomodulation in pigs high and low responders to ACTH». In *Physiology Behavior*, vol. 169 (2017) p.1-8. Ref HAL: [hal-02051272 v1](#).
32. [Chen-17] Chen E. C. H., Morin A., [CHAUCHAT J.-H.](#), Sankoff D. «Statistical analysis of fractionation resistance by functional category and expression.». In *BMC genomics*, vol. 18/Suppl 4 (2017) p.366. DOI: [10.1186/s12864-017-3736-0](#).
33. [Jacques-16] [JACQUES J.](#), Ruckebusch C. «Model-based co-clustering for hyperspectral images». In *Journal of Spectral Imaging*, vol. 5 (2016). DOI: [10.1255/jsi.2016.a3](#).
34. [Hasnat-17] [HASNAT M. A.](#), [HASNAT M. A.](#), [VELCIN J.](#), [BONNEVAY S.](#), [JACQUES J.](#) «Evolutionary clustering for categorical data using parametric links among multinomial mixture models». In *Econometrics and Statistics* (2017). Ref HAL: [hal-01204613 v3](#).
35. [Herbert-15] Herbert F., Tchitchek N., Bansal D., [JACQUES J.](#), Pathak S., Becavin C., Fesel C., Dalko E., Cazenave P.-A., Preda C., Ravindran B., Sharma S., Das B., Pied S. «Evidence of IL-17, IP-10, and IL-10 involvement in multiple-organ dysfunction and IL-17 pathway in acute renal failure associated to Plasmodium falciparum malaria». In *Journal of Translational Medicine*, vol. 13/1 (2015). Ref HAL: [hal-02051283 v1](#).
36. [Yengo-16] Yengo L., [JACQUES J.](#), Biernacki C., Canouil M. «Variable Clustering in High-Dimensional Linear Regression: The R Package clere». In *The R Journal*, vol. 8/1 (2016) p.92-106. Ref HAL: [hal-00940929 v1](#).
37. [Makkhongkaew-15] Makkhongkaew R., [BONNEVAY S.](#), Aussem A., Benabdeslem K. «Classification model for performance diagnosis of dry port by rail». In *Journal of Society for Transportation and Traffic Studies*, vol. 6/2 (2015) p.31-40. Ref HAL: [hal-01134826 v2](#).

38. [Do-15] DO T. N., Lenca P., LALLICH S. «Classifying Many-Class High Dimensional Fingerprint Datasets Using Random Forest of Oblique Decision Trees». In *Vietnam journal of computer science*, vol. 2/1 (2015) p.3-12. DOI: [10.1007/s40595-014-0024-7](https://doi.org/10.1007/s40595-014-0024-7).

Revue nationale

39. [Rolland-18a] ROLLAND A., Spennato M. «L'attractivité des départements STID - Etude à partir des données APB». In *Statistique et Enseignement* (2018). Ref HAL: [hal-02053033 v1](https://hal.archives-ouvertes.fr/hal-02053033).
40. [Rolland-17a] ROLLAND A. «L'alternance en DUT STID, l'exemple de l'IUT Lumière Lyon II». In *Statistique et Enseignement* (2017). Ref HAL: [hal-02053036 v1](https://hal.archives-ouvertes.fr/hal-02053036).
41. [Rolland-17] ROLLAND A., Jollois F.-X., Muri F., Ottenwaelter E., Sylvie V.-P. «Concours DATAVIZ : retour d'expérience en 1ère année de DUT STID». In *Statistique et Enseignement* (2017). Ref HAL: [hal-02053035 v1](https://hal.archives-ouvertes.fr/hal-02053035).
42. [Boyadjian-17] Boyadjian J., VELCIN J. «L'analyse quantitative des médias sociaux, une alternative aux enquêtes déclaratives ?». In *Questions de communication* (2017) 111 - 135. DOI: [10.4000/questionsdecommunication.11078](https://doi.org/10.4000/questionsdecommunication.11078).
43. [Dubois-15d] Dubois D., Prade H., RICO A. «Capacités qualitatives et information incomplète». In *Revue d'Intelligence Artificielle (Revue des Sciences et Technologies de l'Information)*, vol. vol. 29/n° 5 (2015). DOI: [10.3166/RIA.29.493-514](https://doi.org/10.3166/RIA.29.493-514).
44. [Antoniadis-14] Antoniadis A., Brossat X., CUGLIARI J., Poggi J-M. Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité. *Journal de la Société Française de Statistique* (2014).

Articles de synthèse / revues bibliographiques (1)

1. [Cugliari-18b] CUGLIARI J., Poggi J-M. Electricity Demand Forecasting. *Wiley StatsRef: Statistics Reference Online*, 2018.

2- Ouvrages

Direction et coordination d'ouvrages scientifiques / édition scientifique (2 dont 1 en anglais)

1. [Olivesi-17] Olivesi A., Boyadjian J., VELCIN J. «Le web politique au prisme de la science des données». In *Réseaux*, vol. 204/4 (2017). DOI: [10.3917/res.204.0009](https://doi.org/10.3917/res.204.0009).
2. [Lenca-15] Lenca P., LALLICH S. «Guest editor's introduction: special issue on quality issues, measures of interestingness and evaluation of data mining models - Journal of Intelligent Information Systems». In *Journal of Intelligent Information Systems*, vol. 45(3) (2015). Ref HAL: [hal-01243257 v1](https://hal.archives-ouvertes.fr/hal-01243257).

Chapitres d'ouvrage (5 en anglais)

1. [Couceiro-18] Couceiro M., Dubois D., Fargier H., Grabisch M., Prade H., RICO A. «New directions in ordinal evaluation: Sugeno integrals and beyond». *New Perspectives in Multiple Criteria Decision Making* (2018). Ref HAL: [hal-01941776 v1](https://hal.archives-ouvertes.fr/hal-01941776).
2. [Gauthier-16] Gauthier M., GUILLE A., RICO F., Deseille A. «Text Mining and Twitter to Analyze British Swearing Habits». *Handbook of Twitter for Research* (2016) p.27-46. DOI: [10.5281/zenodo.44882](https://doi.org/10.5281/zenodo.44882).
3. [VanErven-16] Van Erven T., CUGLIARI J. «Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts». Ref HAL: [hal-00920559 v1](https://hal.archives-ouvertes.fr/hal-00920559).
4. [Jacques-14] JACQUES J., Fraix-Burnet D. «Linear Regression in High Dimension and/or for Correlated Inputs». *Statistics for Astrophysics- Methods and Applications of the Regression*, vol. 66 (2014) p.149-165. Ref HAL: [hal-02051292 v1](https://hal.archives-ouvertes.fr/hal-02051292).
5. [Lanagan-14a] Lanagan J., Anokhin N., VELCIN J. «Early Stage Conversation Catalysts on Entertainment-Based Web Forums». *State of the Art Applications of Social Network Analysis* (2014) p.97-118. Ref HAL: [hal-02052410 v1](https://hal.archives-ouvertes.fr/hal-02052410).

3- Production dans des colloques / congrès, séminaires de recherche

Éditions d'actes de colloques / congrès (1)

1. [Bonnevay-16b] **BONNEVAY S.**, Legrand P., Monmarché N., Lutton E., Schoenauer M. «Artificial Evolution 2015». , vol. 9554 (2016). DOI: [10.1007/978-3-319-31471-6](https://doi.org/10.1007/978-3-319-31471-6).

Articles publiés dans des actes de colloques / congrès (113 dont 68 internationaux)

1. **[Brochier-19] BROCHIER R.**, **GUILLE A.**, **VELCIN J.** «Global Vectors for Node Representations». In proc. of *International World Wide Web Conference(WWW)*. DOI: [10.1145/3308558.3313595](https://doi.org/10.1145/3308558.3313595). Rang A.
2. **[Davidson-18] Davidson I.**, **GOURRU A.**, Ravi S. «The Cluster Description Problem - Complexity Results, Formulations and Approximations». In proc. of *Advances in Neural Information Processing Systems (NIPS 2018)*. [NIPS Proceedings](https://proceedings.nips.cc/). Rang A.
3. **[Velcin-18] VELCIN J.**, **GOURRU A.**, Giry-Fouquet E., Gravier C., Roche M., Poncelet P. «Readitopics: Make Your Topic Models Readable via Labeling and Browsing». In proc. of *International Joint Conference on Artificial Intelligence (IJCAI 2018)*. Ref HAL: [lirmm-01910611 v1](https://hal.archives-ouvertes.fr/lirmm-01910611/v1). Rang A.
4. **[Dubois-18a] Dubois D.**, Fargier H., **RICO A.** «Sugeno Integrals and the Commutation Problem». In proc. of *International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2018)* p.48-63. Ref HAL: [hal-02053021 v1](https://hal.archives-ouvertes.fr/hal-02053021/v1). Rang B.
5. **[Wang-17b] WANG X.**, **AH-PINE J.**, **DARMONT J.** «SHCoClust, a Scalable Similarity-based Hierarchical Co-clustering Method and its Application to Textual Collections». In proc. of *IEEE International Conference on Fuzzy Systems(FUZZ-IEEE 2017)* ID F-0529. Ref HAL: [hal-02053700 v1](https://hal.archives-ouvertes.fr/hal-02053700/v1). Rang A.
6. **[AhPine-16] AH-PINE J.**, **WANG X.** «Similarity Based Hierarchical Clustering with an Application to Text Collections». In proc. of *International Symposium on Intelligent Data Analysis (IDA 2016)* 320 - 331. DOI: [10.1007/978-3-319-46349-0](https://doi.org/10.1007/978-3-319-46349-0). Rang A.
7. **[Dubois-16b] Dubois D.**, **RICO A.** «Axiomatisation of discrete fuzzy integrals with respect to possibility and necessity measures». In proc. of *International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2016)*. Ref HAL: [hal-01445244 v1](https://hal.archives-ouvertes.fr/hal-01445244/v1). Rang B.
8. **[SorianoMorales-16] SORIANO MORALES E. P.**, **AH-PINE J.**, **LOUDCHER S.** «Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation». In proc. of *Conference on Intelligent Text Processing and Computational Linguistics(CICLING 2016)*. Ref HAL: [halshs-01384657 v1](https://hal.archives-ouvertes.fr/halshs-01384657/v1). Rang B.
9. **[Truica-16a] TRUICA C.-O.**, **DARMONT J.**, **VELCIN J.** «A Scalable Document-based Architecture for Text Analysis». In proc. of *International Conference on Advanced Data Mining and Applications (ADMA 2016)* p.481-494. Ref Arxiv: [1612.06195](https://arxiv.org/abs/1612.06195). Rang B.
10. **[Dermouche-16] DERMOUCHE M.**, **VELCIN J.**, Flicoteaux R., Chevret S., Tairght N. «Supervised Topic Models for Diagnosis Code Assignment to Discharge Summaries». In proc. of *Conference on Intelligent Text Processing and Computational Linguistics(CICLING)*. Rang B.
11. **[Dermouche-15] DERMOUCHE M.**, Kouas L., **VELCIN J.**, **LOUDCHER S.** «A Joint Model for Topic-Sentiment Modeling from Text». In proc. of *Annual ACM Symposium on Applied Computing (SAC 2015)*. Ref HAL: [halshs-01099984 v1](https://hal.archives-ouvertes.fr/halshs-01099984/v1). Rang B.
12. **[Hasnat-15] HASNAT M. A.**, **VELCIN J.**, **BONNEVAY S.**, **JACQUES J.** «Simultaneous Clustering and Model Selection for Multinomial Distribution: A Comparative Study». In proc. of *International Symposium on Intelligent Data Analysis (IDA)*. Ref HAL: [hal-01203561 v1](https://hal.archives-ouvertes.fr/hal-01203561/v1). Rang A.
13. **[Dubois-15c] Dubois D.**, Prade H., **RICO A.** «The Cube of Opposition: A Structure Underlying Many Knowledge Representation Formalisms». In proc. of *International Joint Conference on Artificial Intelligence (IJCAI)*. Ref HAL: [hal-01192705 v1](https://hal.archives-ouvertes.fr/hal-01192705/v1). Rang A.
14. **[Bonnevay-15] BONNEVAY S.**, **AUBERTIN P.**, **GAVIN G.** «A Genetic Algorithm to Solve a Real 2-D Cutting Stock Problem with Setup Cost in the Paper Industry». In proc. of *Genetic and Evolutionary Computation Conference(GECCO)*. Ref HAL: [hal-01134195 v1](https://hal.archives-ouvertes.fr/hal-01134195/v1). Rang A.
15. **[Kim-15] KIM Y.-M.**, **VELCIN J.**, **BONNEVAY S.**, **RIZOIU M.-A.** «Temporal Multinomial Mixture for Instance-oriented Evolutionary Clustering». In proc. of *European Conference on Information Retrieval(ECIR)*. Ref HAL: [hal-01134393 v1](https://hal.archives-ouvertes.fr/hal-01134393/v1). Rang A.
16. **[Dermouche-14] DERMOUCHE M.**, **VELCIN J.**, **LOUDCHER S.**, Khous L. «A Joint Model for Topic-Sentiment Evolution over Time». In proc. of *IEEE International Conference on Data Mining (ICDM 2014)*. Ref HAL: [halshs-01099990 v1](https://hal.archives-ouvertes.fr/halshs-01099990/v1). Rang A.
17. [Brochier-19b] **BROCHIER R.**, **GUILLE A.**, **VELCIN J.** «Link Prediction with Mutual Attention for Text-Attributed Networks». In proc. of *International World Wide Web Conference(WWW)*. DOI: [10.1145/3308560.3316587](https://doi.org/10.1145/3308560.3316587).
18. [Brochier-19a] **BROCHIER R.** «Representation Learning for Recommender Systems with Application to the Scientific Literature». In proc. of *International World Wide Web Conference(WWW)*. DOI: [10.1145/3308560.3314195](https://doi.org/10.1145/3308560.3314195).

19. [Gavin-19] GAVIN G., BONNEVAY S. «Securely aggregating testimonies with Threshold Multi-key FHE». In proc. of *International Conference on Codes, Cryptology and Information Security (C2SI)*. Ref HAL: [hal-02051956 v1](https://hal.archives-ouvertes.fr/hal-02051956).
20. [Castrillejo-18] Castrillejo A., CUGLIARI J., Massa F., Ramirez I. «Electricity Demand Forecasting: the Uruguayan Case». In Proc. *Renewable Energy: Forecasting and Risk Management (FRM 2017)*, Springer Proceedings in Mathematics & Statistics, vol 254, p.119-136. Ref HAL: [hal-01787143 v1](https://hal.archives-ouvertes.fr/hal-01787143).
21. [Dubois-18b] Dubois D., Faux F., Prade H., RICO A. «Separable Qualitative Capacities». In proc. of *International Conference on Scalable Uncertainty Management (SUM 2018)* p.124-139. Ref HAL: [hal-02053031 v1](https://hal.archives-ouvertes.fr/hal-02053031).
22. [Rico-18] RICO A. «Discrete Sugeno Integrals and Their Applications». In proc. of *International Conference on Scalable Uncertainty Management (SUM 2018)* p.18-32. Ref HAL: [hal-02053027 v1](https://hal.archives-ouvertes.fr/hal-02053027).
23. [Ghabach-18] Ghabach E., Blay-Fornarino M., EL KHOURY F., Baz B. «Guiding Clone-and-Own When Creating Unplanned Products from a Software Product Line». In proc. of *International Conference on Software Reuse (ICSR)*. Ref HAL: [hal-01903003 v1](https://hal.archives-ouvertes.fr/hal-01903003).
24. [AlNatsheh-18] AL-NATSHEH H., MARTINET L., Muhlenbach F., RICO F., ZIGHED D. A. «Metadata Enrichment of Multi-disciplinary Digital Library: A Semantic-Based Approach». In proc. of *International Conference on Theory and Practice of Digital Libraries (TPDL 2018)* p.32-43. Ref HAL: [hal-02018793 v1](https://hal.archives-ouvertes.fr/hal-02018793).
25. [Connes-18] Connes V., Dugué N., GUILLE A. «Is Community Detection Fully Unsupervised? The Case of Weighted Graphs.». In proc. of *Complex Networks and Their Applications VII*. Ref HAL: [hal-01982660 v1](https://hal.archives-ouvertes.fr/hal-01982660).
26. [Selosse-18d] SELOSSE M., JACQUES J., Biernacki C. «Analyzing large matrices of ordinal data». In proc. of *International Conference of the ERCIM working group on Computational and Methodological Statistics (CMStatistics)*. Ref HAL: [hal-01949095 v1](https://hal.archives-ouvertes.fr/hal-01949095).
27. [Brochier-18] BROCHIER R., GUILLE A., Rothan B., VELCIN J. «Impact of the Query Set on the Evaluation of Expert Finding Systems». In proc. of *BIRNDL 2018 (SIGIR 2018) (SIGIR)*. Ref Arxiv: [1806.10813](https://arxiv.org/abs/1806.10813).
28. [Ozturk-18a] OZTÜRK A., LALLICH S., DARMONT J., Waksman S. Y. «MaxMin Linear Initialization for Fuzzy C-Means». In proc. of *International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2018)*. Ref HAL: [hal-01771204 v1](https://hal.archives-ouvertes.fr/hal-01771204).
29. [Dubois-18] Dubois D., Prade H., RICO A. «Fuzzy Extensions of Conceptual Structures of Comparison». In proc. of *Information Processing and Management of Uncertainty (IPMU)*. Ref HAL: [hal-02018672 v1](https://hal.archives-ouvertes.fr/hal-02018672). Rang C.
30. [Brabant-18a] Brabant Q., Couceiro M., Dubois D., Prade H., RICO A. «Extracting Decision Rules from Qualitative Data via Sugeno Utility Functionals». In proc. of *Information Processing and Management of Uncertainty (IPMU 2018)* p.253-265. Ref HAL: [hal-01670924 v1](https://hal.archives-ouvertes.fr/hal-01670924). Rang C.
31. [Gourru-18] GOURRU A., VELCIN J., Roche M., Gravier C., Poncelet P. «United we stand: Using multiple strategies for topic labeling». In proc. of *Natural Language Processing and Information Systems (NLDB 2018)* p.352-363. DOI: [10.1007/978-3-319-91947-8_37](https://doi.org/10.1007/978-3-319-91947-8_37). Rang C.
32. [Ozturk-18] OZTÜRK A., LALLICH S., DARMONT J. «A Visual Quality Index for Fuzzy C-Means». In proc. of *International Conference on Artificial Intelligence Applications and Innovations (IAI 2018)*. Ref HAL: [hal-01727426 v1](https://hal.archives-ouvertes.fr/hal-01727426).
33. [Jacques-17] JACQUES J. «Model-based co-clustering for functional data». In proc. of *CM Statistics (CMS)*. Ref HAL: [hal-01705301 v1](https://hal.archives-ouvertes.fr/hal-01705301).
34. [Couceiro-17] Couceiro M., Dubois D., Prade H., RICO A. «Enhancing the expressive power of Sugeno integrals for qualitative data analysis». In proc. of *conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2017)* p.534-547. DOI: [10.1007/978-3-319-66830-7_48](https://doi.org/10.1007/978-3-319-66830-7_48).
35. [BenSlimen-17] BEN SLIMEN Y., Allio S., JACQUES J. «Anomaly Prevision in Radio Access Networks Using Functional Data Analysis». In proc. of *IEEE GlobeCom 2017*. Ref HAL: [hal-01613475 v1](https://hal.archives-ouvertes.fr/hal-01613475).
36. [Delacroix-17] Delacroix T., Lenca P., LALLICH S. «Computing the Mutual Constrained Independence Model». In proc. of *International Conference on Applied Stochastic Models and Data Analysis (ASMDA)*. Ref HAL: [hal-01582632 v1](https://hal.archives-ouvertes.fr/hal-01582632).
37. [SorianoMorales-17] SORIANO MORALES E. P., AH-PINE J., LOUDCHER S., LOUDCHER S. «Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation.». In proc. of *International Conference Discovery Science (DS 2017)* p.340-355. Ref HAL: [halshs-01577043 v1](https://hal.archives-ouvertes.fr/halshs-01577043).
38. [Rolland-17b] ROLLAND A., Pirlot M., valérie B. «Synthesizing a set of rules by a non-compensatory sorting model. An application to environmental evaluation.». In proc. of *Conférence of the international Federation of Operational Research Societs (IFORS)*. Ref HAL: [hal-02053432 v1](https://hal.archives-ouvertes.fr/hal-02053432).
39. [Cugliari-17] CUGLIARI J., BONNEVAY S., Achaïcha P. «Predictive Maintenance of Smartmeter's Concentrators». In proc. of *European Network for Business and Industrial Statistics*. Ref HAL: [hal-01807072 v1](https://hal.archives-ouvertes.fr/hal-01807072).

40. [Roos-17] **ROOS J.**, **BONNEVAY S.**, **GAVIN G.** «Dynamic Bayesian Networks with Gaussian Mixture Models for Short-Term Passenger Flow Forecasting». In proc. of *International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. Ref HAL: [hal-01807067](https://hal.archives-ouvertes.fr/hal-01807067). v1.
41. [Gavin-17] **GAVIN G.**, **BONNEVAY S.** «Combining imperfect knowledge of maybe corrupted sources». In proc. of *2017 ISEA Asia Security and Privacy (ISEASP)*. DOI: [10.1109/ISEASP.2017.7976997](https://doi.org/10.1109/ISEASP.2017.7976997).
42. [Dubois-16] Dubois D., Prade H., **RICO A.**, Teheux B. «Generalized Sugeno Integrals». In proc. of *Information Processing and Management of Uncertainty (IPMU 2016)*. Ref HAL: [hal-01445233](https://hal.archives-ouvertes.fr/hal-01445233). v1. Rang C.
43. [SorianoMorales-16a] **SORIANO MORALES E. P.**, **AH-PINE J.**, **LOUDCHER S.** «Hypergraph Modelization of a Syntactically Annotated English Wikipedia Dump». In proc. of *Language Resources and Evaluation Conference (LREC 2016)*. Ref HAL: [halshs-01384647](https://halshs.archives-ouvertes.fr/halshs-01384647). v1. Rang C.
44. [Roos-16a] **ROOS J.**, **BONNEVAY S.**, **GAVIN G.** «Short-Term Urban Rail Passenger Flow Forecasting: A Dynamic Bayesian Network Approach». In proc. of *International Conference on Machine Learning and Applications (ICMLA)* p.1034-1039. Ref HAL: [hal-02051746](https://hal.archives-ouvertes.fr/hal-02051746). v1. Rang C.
45. [Dermouche-16a] **DERMOUCHE M.**, Looten V., Flicoteaux R., Chevret S., **VELCIN J.**, Taright N. «ECSTRA-INSERM @ CLEF eHealth2016-task 2: ICD10 Code Extraction from Death Certificates». In proc. of *CLEF (Working Notes) 2016* p.61-68.
46. [Roos-16b] **ROOS J.**, **GAVIN G.**, **BONNEVAY S.** «A dynamic bayesian network approach to forecast short-term urban rail passenger flows with incomplete data». In proc. of *European Transport Conference (ETC)*. Ref HAL: [hal-01807053](https://hal.archives-ouvertes.fr/hal-01807053). v1.
47. [Dubois-16c] Dubois D., **RICO A.** «An axiomatisation of discrete possibilistic Choquet integrals». In proc. of *Foundations of Utility and Risk (FUR)*.
48. [Jacques-16b] **JACQUES J.**, Biernacki C. «Model-based co-clustering for ordinal data». In proc. of *Summer Working Group on Model-Based Clustering of the Department of Statistics of the University of Washington*. Ref HAL: [hal-01383912](https://hal.archives-ouvertes.fr/hal-01383912). v1.
49. [BenSlimen-16a] **BEN SLIMEN Y.**, Allio S., **JACQUES J.** «Model-based co-clustering for functional data». In proc. of *Conference for Computational Statistics (COMPSTAT 2016)*. Ref HAL: [hal-01383920](https://hal.archives-ouvertes.fr/hal-01383920). v1.
50. [Jacques-16a] **JACQUES J.**, Ruckebusch C. «Co-clustering for hyperspectral images». In proc. of *International Conference in Spectral Imaging*. Ref HAL: [hal-01383918](https://hal.archives-ouvertes.fr/hal-01383918). v1.
51. [Roos-16] **ROOS J.**, **GAVIN G.**, **BONNEVAY S.** «A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data». In *European Transport Conference (ETC)*, vol. 26 (2017) 53 - 61. DOI: [10.1016/j.trpro.2017.07.008](https://doi.org/10.1016/j.trpro.2017.07.008).
52. [Velcin-16] **VELCIN J.**, Roche M., Poncelet P. «Shallow Text Clustering Does Not Mean Weak Topics: How Topic Identification Can Leverage Bigram Features». In proc. of *DMNLP: Data Mining and Natural Language Processing (DMNLP 2016)*. Ref HAL: [lirmm-01362434](https://hal.archives-ouvertes.fr/lirmm-01362434). v1.
53. [Truica-16] **TRUICA C.-O.**, **GUILLE A.**, Gauthier M. «CATS: Collection and Analysis of Tweets Made Simple». In proc. of *ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW Companion)* p.41-44. DOI: [10.1145/2818052.2874320](https://doi.org/10.1145/2818052.2874320).
54. [Cugliari-16] **CUGLIARI J.**, Y. Goude, J. M. Poggi. Disaggregated electricity forecasting using wavelet-based clustering of individual consumers. 2016 IEEE International Energy Conference (ENERGYCON 2016).
55. [Dubois-15e] Dubois D., Prade H., **RICO A.** «The Cube of Opposition and the Complete Appraisal of Situations by Means of Sugeno Integrals». In proc. of *International Symposium on Methodologies for Intelligent Systems (ISMIS 2015)* p.197-207. Ref HAL: [hal-01291624](https://hal.archives-ouvertes.fr/hal-01291624). v1.
56. [Delacroix-15] Delacroix T., Boubekki A., Lenca P., **LALLICH S.** «Constrained Independence for Detecting Interesting Patterns». In proc. of *IEEE International Conference on Data Science and Advanced Analytics (DSAA 2015)* 1 - 10. DOI: [10.1109/DSAA.2015.7344897](https://doi.org/10.1109/DSAA.2015.7344897).
57. [Jacques-15a] **JACQUES J.** «Clustering multivariate ranking data». In proc. of *World Statistics Congress (ISI 2015)*. Ref HAL: [hal-01241256](https://hal.archives-ouvertes.fr/hal-01241256). v1.
58. [Jacques-15] **JACQUES J.** «The discriminative functional mixture model for a comparative analysis of bike sharing systems». In proc. of *International Conference of the ERCIM working group on Computational and Methodological Statistics (CMStatistics 2015)*. Ref HAL: [hal-01241254](https://hal.archives-ouvertes.fr/hal-01241254). v1.
59. [Dubois-15a] Dubois D., **RICO A.**, Teheux B., Prade H. «Characterizing variants of qualitative Sugeno integrals in a totally ordered Heyting algebra». In proc. of *conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*. Ref HAL: [hal-01179574](https://hal.archives-ouvertes.fr/hal-01179574). v1.
60. [Bonnevay-15a] **BONNEVAY S.**, **AUBERTIN P.**, Lazert T. «A Simulated Annealing algorithm for real-world 2-D Cutting Stock Problem with Setup Cost». In proc. of *Metaheuristics International Conference (MIC)*. Ref HAL: [hal-01134198](https://hal.archives-ouvertes.fr/hal-01134198). v1.

61. [Ozturk-15] **OZTÜRK A.**, Eyango L., Waksman S. Y., **LALLICH S.**, **DARMONT J.** «Warehousing Complex Archaeological Objects». In proc. of *International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2015)* p.226-239. Ref Arxiv: [1608.06462](https://arxiv.org/abs/1608.06462). Rang C.
62. [Truica-15] **TRUICA C.-O.**, **VELCIN J.**, Boicea A. «Automatic Language Identification for Romance Languages using Stop Words and Diacritics». In proc. of *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. Ref HAL: [hal-01193158 v1](https://hal.archives-ouvertes.fr/hal-01193158). Rang C.
63. [Dubois-15b] Dubois D., Durrieu C., Prade H., **RICO A.**, Ferro Y. «Extracting Decision Rules from Qualitative Data Using Sugeno Integral: A Case-Study». In proc. of *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)* p.14-24. Ref HAL: [hal-01179575 v1](https://hal.archives-ouvertes.fr/hal-01179575). Rang C.
64. [Rolland-14c] **ROLLAND A.**, **LUST I.** «2-additive Choquet Optimal Solutions in Multiobjective Optimization Problems». In proc. of *Information Processing and Management of Uncertainty (IPMU)*. Rang C.
65. [Dubois-14a] Dubois D., Prade H., **RICO A.** «On the Informational Comparison of Qualitative Fuzzy Measures». In proc. of *Information Processing and Management of Uncertainty (IPMU 2014)*. Ref HAL: [hal-01147279 v1](https://hal.archives-ouvertes.fr/hal-01147279). Rang C.
66. [Velcin-14] **VELCIN J.**, **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 C.
67. [Rolland-14a] **ROLLAND A.**, **LUST I.** «Choquet optimal solutions set in multicriteria optimization problems». In proc. of *Information Processing and Management of Uncertainty (IPMU 2014)*. Rang C.
68. [Guille-14b] **GUILLE A.**, **FAVRE C.** «Mention-anomaly-based Event Detection and Tracking in Twitter». In proc. of *International Conference on Advances in Social Network Analysis and Mining (ASONAM 2014)* p.375-382. DOI: [10.1109/ASONAM.2014.6921613](https://doi.org/10.1109/ASONAM.2014.6921613).

Colloques / congrès nationaux

69. [Strauss-19] Strauss O., **RICO A.** «Extension signée de la domination des noyaux maxitifs.». In proc. of *Logique Floue et ses Applications (LFA)*. Ref HAL: [hal-02053041 v1](https://hal.archives-ouvertes.fr/hal-02053041).
70. [Dubois-18d] Dubois D., Faux F., Prade H., **RICO A.** «Capacités qualitatives séparables». In proc. of *Logique Floue et ses Applications (LFA)*. Ref HAL: [hal-02053046 v1](https://hal.archives-ouvertes.fr/hal-02053046).
71. [Dubois-18c] Dubois D., Fargier H., **RICO A.** «Commutation des intégrales de Sugeno discrètes». In proc. of *Logique Floue et ses Applications (LFA)*. Ref HAL: [hal-02053039 v1](https://hal.archives-ouvertes.fr/hal-02053039).
72. [Brochier-18a] **BROCHIER R.**, **GUILLE A.**, **VELCIN J.**, Rothan B., Cioccio D. «Peerus Review: Un outil de recherche d'experts scientifiques». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref Arxiv: [1807.03719](https://arxiv.org/abs/1807.03719).
73. [Christophe-18] **CHRISTOPHE C.**, **VELCIN J.**, Boumghar M. «Utilisation de techniques de modélisation thématiques pour la détection de nouveauté dans des flux de données textuelles». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2018)*. Ref HAL: [hal-01803649 v1](https://hal.archives-ouvertes.fr/hal-01803649).
74. [Selosse-18a] **SELOSSE M.**, **JACQUES J.**, Biernacki C. «Co-clustering de données textuelles et continues». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01797493 v1](https://hal.archives-ouvertes.fr/hal-01797493).
75. [Nagbe-18] **NAGBE K.**, **CUGLIARI J.**, **JACQUES J.** «Prévision journalière de la consommation d'électricité à l'aide d'un modèle à espace d'états fonctionnel». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01785221 v1](https://hal.archives-ouvertes.fr/hal-01785221).
76. [Schmutz-18a] **SCHMUTZ A.**, **JACQUES J.**, Bouveyron C., Cheze L., Martin P. «funHDDC : extension du package R pour le clustering de courbes fonctionnelles multivariées». In proc. of *Rencontres R*. Ref HAL: [hal-01784304 v1](https://hal.archives-ouvertes.fr/hal-01784304).
77. [Schmutz-18] **SCHMUTZ A.**, **JACQUES J.**, Bouveyron C., Cheze L., Martin P. «Données fonctionnelles multivariées issues d'objets connectés : une méthode pour classer les individus». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01784279 v1](https://hal.archives-ouvertes.fr/hal-01784279).
78. [HuyghuesDespointes-18] **HUYGHUES-DESPOINTES C.**, **VELCIN J.**, Khouas L., **LOUDCHER S.** «Extraction de chaînes cohérentes en vue de reconstruire la trajectoire de l'information». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref HAL: [hal-01674547 v1](https://hal.archives-ouvertes.fr/hal-01674547).
79. [Martinet-18] **MARTINET L.**, **AL-NATSHEH H.**, **RICO F.**, Muhlenbach F., **ZIGHED D. A.** «Étiquetage thématique automatisé de corpus par représentation sémantique». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)* p.1-6. Ref HAL: [hal-01659639 v1](https://hal.archives-ouvertes.fr/hal-01659639).
80. [AhPine-17a] **AH-PINE J.** «Sur la normalisation de la matrice Laplacienne en partitionnement spectral». In proc. of *Congrès de la Société Francophone de Classification (SFC)*. Ref HAL: [hal-01981818 v1](https://hal.archives-ouvertes.fr/hal-01981818).

81. [Roos-17a] **ROOS J.**, **BONNEVAY S.**, **GAVIN G.** «Prévision à court terme des flux de voyageurs du réseau ferré urbain : une approche par les réseaux bayésiens dynamiques». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2017)*. Ref HAL: [hal-01807063](https://hal.archives-ouvertes.fr/hal-01807063) v1.
82. [Schmutz-17b] **SCHMUTZ A.**, **JACQUES J.**, Cheze L., Martin P. «Prédiction de la vitesse du cheval de sport à partir de données accélérométriques et gyroscopiques». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01705298](https://hal.archives-ouvertes.fr/hal-01705298) v1.
83. [Nagbe-17] **NAGBE K.**, **CUGLIARI J.**, Thebault A., **JACQUES J.** «Prévision de génération d'électricité à partir de sources renouvelables». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01705287](https://hal.archives-ouvertes.fr/hal-01705287) v1.
84. [Couceiro-17a] Couceiro M., Dubois D., Prade H., **RICO A.** «Intégrales de Sugeno généralisées en analyse de données». In proc. of *Logique Floue et ses Applications(LFA 2017)*. Ref HAL: [hal-01668232](https://hal.archives-ouvertes.fr/hal-01668232) v1.
85. [Delacroix-17a] Delacroix T., Lenca P., **LALLICH S.** «Du local au global : un nouveau défi pour l'analyse statistique implicative». In proc. of *Colloque Analyse Statistique Implicative (ASI 2017)* 103 - 116. Ref HAL: [hal-01616110](https://hal.archives-ouvertes.fr/hal-01616110) v1.
86. [Castrillejo-17] Castrillejo A., **CUGLIARI J.**, Massa F., Ramirez I. «Modèles de prévision de demande d'électricité. Application au système uruguayenne.». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-01539443](https://hal.archives-ouvertes.fr/hal-01539443) v1.
87. [Beligne-17a] **BELIGNE M.**, Campar A., **CHAUCHAT J.-H.**, Lefeuvre M., Lefort I., **LOUDCHER S.**, **VELCIN J.** «Retour d'expérience sur la détection automatique de métaphores dans des textes de Géographie». In proc. of *journées(ADOC - EDA 2017)* p.159-160. Ref HAL: [halshs-01577051](https://halshs.archives-ouvertes.fr/halshs-01577051) v1.
88. [Beligne-17] **BELIGNE M.**, Campar A., **CHAUCHAT J.-H.**, Lefeuvre M., Lefort I., **LOUDCHER S.**, **VELCIN J.** «Détection automatique de métaphores dans des textes de Géographie : une étude prospective». In proc. of *Conférence sur le Traitement Automatique des Langues Naturelles(TALN 2017)* p.86-93. Ref HAL: [halshs-01577050](https://halshs.archives-ouvertes.fr/halshs-01577050) v2.
89. [Velcin-17] **VELCIN J.**, Soulages J.-C., Kurpiel S., Dias L., Del Vecchio M., Aubrun F. «Fouille de textes pour une analyse comparée de l'information diffusée par les médias en ligne : une étude sur trois éditions du Huffington Post». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2017)*. Ref HAL: [hal-01571265](https://hal.archives-ouvertes.fr/hal-01571265) v1.
90. [AhPine-17] **AH-PINE J.**, **WANG X.** «Classification ascendante hiérarchique à noyaux et une application aux données textuelles». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref HAL: [hal-01525446](https://hal.archives-ouvertes.fr/hal-01525446) v1.
91. [Wang-17a] **WANG X.**, **AH-PINE J.**, **DARMONT J.** «A New Test of Cluster Hypothesis Using a Scalable Similarity-Based Agglomerative Hierarchical Clustering Framework». In proc. of *Conférence en Recherche d'Information et Applications (CORIA)*. Ref HAL: [hal-01504961](https://hal.archives-ouvertes.fr/hal-01504961) v1.
92. [HuyghuesDespointes-16] **HUYGHUES-DESPOINTES C.**, Bui-Xuan B.-M., Magnien C. «Forte Δ -connexité dans les flots de liens». In proc. of *Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications(ALGOTEL)*. Ref HAL: [hal-01305128](https://hal.archives-ouvertes.fr/hal-01305128) v1.
93. [Rolland-16c] **ROLLAND A.** «L'alternance en DUT STID, l'exemple de l'IUT Lumière Lyon II». In proc. of *Journées de Statistique (JDS)*. Ref HAL: [hal-02053435](https://hal.archives-ouvertes.fr/hal-02053435) v1.
94. [Rolland-16] **ROLLAND A.** «A note on top-k lists: average distance between two top-k lists». In proc. of *Journées de la statistique*. Ref HAL: [hal-02053436](https://hal.archives-ouvertes.fr/hal-02053436) v1.
95. [Rolland-16a] **ROLLAND A.** «L'ALTERNANCE EN DUT STATISTIQUE ET INFORMATIQUE DECISIONNELLE (STID) EXEMPLE DE L'IUT LUMIERE LYON II». In proc. of *Journées de Statistique*.
96. [Bouveyron-16] Bouveyron C., Côme E., **JACQUES J.** «Le modèle DFM pour une analyse comparative des systèmes de vélos en libre service». In proc. of *Journées de Statistique*. Ref HAL: [hal-01383938](https://hal.archives-ouvertes.fr/hal-01383938) v1.
97. [BenSlimen-16] **BEN SLIMEN Y.**, Allio S., **JACQUES J.** «Model-based co-clustering for functional data». In proc. of *Journées de Statistique*. Ref HAL: [hal-01383936](https://hal.archives-ouvertes.fr/hal-01383936) v1.
98. [Guille-16a] **GUILLE A.**, **SORIANO MORALES E. P.**, **TRUICA C.-O.** «Topic modeling and hypergraph mining to analyze the EGC conference history». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2016)* p.383-394. Ref HAL: [hal-01442858](https://hal.archives-ouvertes.fr/hal-01442858) v1.
99. [Guille-16] **GUILLE A.**, **SORIANO MORALES E. P.** «TOM: A library for topic modeling and browsing». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2016)*. Ref HAL: [hal-01442868](https://hal.archives-ouvertes.fr/hal-01442868) v1.
100. [AhPine-15a] **AH-PINE J.**, **WANG X.** «Classification ascendante hiérarchique à noyaux et pistes pour un meilleur passage à l'échelle». In proc. of *Journées de la statistique*. Ref HAL: [hal-01504642](https://hal.archives-ouvertes.fr/hal-01504642) v1.
101. [Dubois-15f] Dubois D., **RICO A.** «Axiomatization de l'intégrale de Choquet possibiliste». In proc. of *Logique Floue et ses Applications(LFA)*.
102. [Bouveyron-15b] Bouveyron C., **JACQUES J.** «funFEM: an R package for functional data clustering». In proc. of *Rencontres R*. Ref HAL: [hal-01383951](https://hal.archives-ouvertes.fr/hal-01383951) v1.

103. [Bouveyron-15a] Bouveyron C., JACQUES J. «Un algorithme EM pour une version parcimonieuse de l'analyse en composantes principales probabiliste». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2015)*. Ref HAL: [hal-01241262 v1](https://hal.archives-ouvertes.fr/hal-01241262).
104. [Lallich-15] LALLICH S., Lenca P. «Indices de qualité en clustering». In proc. of *Journée thématique : clustering et co-clustering*. Ref HAL: [hal-01230854 v1](https://hal.archives-ouvertes.fr/hal-01230854).
105. [Loingeville-15] Loingeville F., JACQUES J., Preda C., Guarini P., Molinier O. «Modèle Linéaire Généralisé Hiérarchique Gamma-Poisson à 3 facteurs aléatoires - Application au contrôle de qualité». In proc. of *Journées de Statistique*. Ref HAL: [hal-01152840 v1](https://hal.archives-ouvertes.fr/hal-01152840).
106. [Dermouche-15a] DERMOUCHE M., Khouas L., LOUDCHER S., VELCIN J., Fourboul E. «Analyse et visualisation d'opinions dans un cadre de veille sur leWeb». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2015)*. Ref HAL: [halshs-01100005 v1](https://halshs.archives-ouvertes.fr/halshs-01100005).
107. [Lallich-14] LALLICH S., Lenca P. «Evaluation de la qualité en clustering». In proc. of *Séminaire*. Ref HAL: [hal-01185095 v1](https://hal.archives-ouvertes.fr/hal-01185095).
108. [Dubois-14b] Dubois D., Prade H., RICO A. «La structure des capacités qualitatives». In proc. of *Logique Floue et ses Applications(LFA)*. Ref HAL: [hal-01145915 v1](https://hal.archives-ouvertes.fr/hal-01145915).
109. [Orbinska-14] OROBINSKA O., 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.
110. [Rolland-14] ROLLAND A. «Un indicateur pour mesurer la dépendance aux paramètres des palmarès». In proc. of *Journées de la statistique*.
111. [Guille-14a] GUILLE A., 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).
112. [Guille-14] GUILLE A., 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).
113. [Auder-14] Auder B., CUGLIARI J. Parallélisation de l'algorithme des k-médoides. Application au clustering de courbes. In proc. of *Journées de Statistique (2014)*.

Autres produits présentés dans des colloques / congrès et des séminaires de recherche (25)

1. [Rolland-18b] ROLLAND A., CUGLIARI J. «Interaction or Correlation? An analysis of Choquet integral approach in MCDA real situations». In proc. of *Workshop From Multiple Criteria Decision Aid to Preference Learning(DA2PL)*. Ref HAL: [hal-02053427 v1](https://hal.archives-ouvertes.fr/hal-02053427).
2. [Brabant-18] Brabant Q., Couceiro M., Dubois D., Prade H., RICO A. «Sugeno Integral for Rule-Based Ordinal Classification». In proc. of *Workshop on Learning and Reasoning: Principles and Applications to Everyday Spatial and Temporal Knowledge (IJCAI - ECAI)(IJCAI-ECAI)*. Ref HAL: [hal-01889785 v1](https://hal.archives-ouvertes.fr/hal-01889785).
3. [Sherkat-18] Sherkat E., VELCIN J., Milios E. E. «Fast and Simple Deterministic Seeding of KMeans for Text Document Clustering». In proc. of *Workshop of the Cross-Language Evaluation Forum (CLEF 2018)* p.76-88. DOI: [10.1007/978-3-319-98932-7_7](https://doi.org/10.1007/978-3-319-98932-7_7).
4. [Selosse-18c] SELOSSE M., JACQUES J., Biernacki C. «mixedClust: an R package for mixed data classification, clustering and co-clustering». In proc. of *Summer Session Working Group on Model-Based Clustering*. Ref HAL: [hal-01949171 v1](https://hal.archives-ouvertes.fr/hal-01949171).
5. [Selosse-18b] SELOSSE M., JACQUES J., Biernacki C. «Model-based co-clustering for mixed type data». . Ref HAL: [hal-01893457 v1](https://hal.archives-ouvertes.fr/hal-01893457).
6. [Bonnevay-18] BONNEVAY S., CUGLIARI J., Granger V. «Predictive maintenance from event logs using wavelet-based features: an industrial application». . Ref HAL: [hal-01856309 v1](https://hal.archives-ouvertes.fr/hal-01856309).
7. [Bourel-18] Bourel M., CUGLIARI J. «Bagging of Density Estimators». . Ref Arxiv: [1808.03447](https://arxiv.org/abs/1808.03447).
8. [Cugliari-18] CUGLIARI J., Gómez J. «Assessing Imputation of Extreme Data on Climatological Time Series». In proc. of *Conference on non-stationarity*. Ref HAL: [hal-01812715 v1](https://hal.archives-ouvertes.fr/hal-01812715).
9. [Rolland-18] ROLLAND A., CUGLIARI J. «Ranking and top-k ranking robustness index to measure dependence to parameters». . Ref HAL: [hal-01811360 v1](https://hal.archives-ouvertes.fr/hal-01811360).
10. [Selosse-18] SELOSSE M., JACQUES J., Biernacki C. «ordinalClust: an R package for analyzing ordinal data». Ref HAL: [hal-01678800 v2](https://hal.archives-ouvertes.fr/hal-01678800).
11. [Schmutz-17] SCHMUTZ A., JACQUES J., Bouveyron C., Cheze L., Martin P. «Clustering multivariate functional data in group-specific functional subspaces». . Ref HAL: [hal-01652467 v2](https://hal.archives-ouvertes.fr/hal-01652467).

12. [Selosse-17] **SELOSSE M.**, JACQUES J., Biernacki C., Cousson-Gélie F. «Analyzing quality of life survey using constrained co-clustering model for ordinal data and some dynamic implication». . Ref HAL: [hal-01643910 v2](#).
13. [AlNatsheh-17a] **AL-NATSHEH H.**, MARTINET L., Muhlenbach F., ZIGHED D. A. «UdL at SemEval-2017 Task 1: Semantic Textual Similarity Estimation of English Sentence Pairs Using Regression Model over Pairwise Features». In proc. of *International Workshop on Semantic Evaluation (SemEval-NAACL)(SemEval)* p.115-119. Ref HAL: [hal-02018770 v1](#).
14. [AlNatsheh-17] **AL-NATSHEH H.**, MARTINET L., Muhlenbach F., RICO F., ZIGHED D. A. «Semantic Search-by-Examples for Scientific Topic Corpus Expansion in Digital Libraries». In proc. of *Workshop on Semantics-Enabled Recommender Systems (ICDM - SERecSys)(ICDM - SERecSys 2017)* p.747-756. DOI: [10.1109/ICDMW.2017.103](#).
15. [Schmutz-17a] **SCHMUTZ A.**, JACQUES J., Cheze L., Martin P. «Speed prediction of the sport horse from accelerometer and gyroscope data». In proc. of *Workshop on Functional and Operatorial Statistics* 1 p. Ref HAL: [hal-01770173 v1](#).
16. [AhPine-16a] AH-PINE J., **SORIANO MORALES E. P.** «A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis». In proc. of *Workshop on Interactions between Data Mining and Natural Language Processing (DMNLP 2016)(DMNLP)*. Ref HAL: [hal-01504684 v1](#).
17. [Dubois-16d] Dubois D., Prade H., RICO A. «Organizing families of aggregation operators into a cube of opposition». *Granular, Soft and Fuzzy Approaches for intelligent Systems dedicated to Pr. Ronald R. Yager (2016)* p.27-47.
18. [Rolland-16b] **ROLLAND A.**, CUGLIARI J. «Simulating new multicriteria data from a given data set». In proc. of *Workshop From Multiple Criteria Decision Aid to Preference Learning(DA2PL)*. Ref HAL: [hal-02053434 v1](#).
19. [Dubois-15g] Dubois D., Prade H., RICO A. «Le cube des oppositions - Une structure à la base de nombreux formalismes de représentation des connaissances». In proc. of *Journées de l'Intelligence Artificielle Fondamentale (JIAF)*. Ref HAL: [hal-01179578 v1](#).
20. [Rolland-14d] **ROLLAND A.**, CUGLIARI J., Tran T. «On the use of copulas to simulate multicriteria data». In proc. of *Workshop From Multiple Criteria Decision Aid to Preference Learning(DA2PL)*. Ref HAL: [hal-02053440 v1](#).
21. [Rolland-14b] **ROLLAND A.**, CUGLIARI J., Tran T.-M.-T. «On the use of copulas to simulate multicriteria data». In proc. of *Workshop From Multiple Criteria Decision Aid to Preference Learning(DA2PL)*.
22. [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](#).
23. [Guille-14c] **GUIILLE A.** «Diffusion de l'information dans les médias sociaux : modélisation et analyse». . Ref HAL: [tel-01100255 v1](#).
24. [Antoniadis-14] Antoniadis A., Brossat X., CUGLIARI J., Poggi J.-M. «A prediction interval for a function-valued forecast model». . Ref Arxiv: [1412.4222](#).
25. [Bentayeb-14] BENTAYEB F., VELCIN J., BONNEVAY S., DARMONT J. «Data Science and Decision Support at ERIC». SIGMOD Record, 43(4), 37-42, December 2014. Ref HAL: [hal-01165623 v1](#).

4- Produits et outils informatiques

Logiciels (11)

L'équipe contribue au développement d'outils spécialisés à destination des informaticiens et statisticiens mais également d'outils « grand publics » destinés aux praticiens non spécialistes :

- Packages R pour le clustering de données fonctionnelles (funLBM, funFEM, funHDDC) ainsi que leur prévision (enercast, gtop). Package R pour l'analyse de données ordinales (ordinalClust). Bibliothèques Matlab (CLustPath) pour le clustering temporel.
- Bibliothèques Python pour l'analyse thématique de corpus textuel (TOM) et la détection d'événements sur Twitter (pyMABED)
- Logiciels libres destinés à un public large pour la collecte et l'analyse de tweets (CATS), l'analyse thématique de corpus textuel (ReadiTopics), l'analyse et l'annotation d'article de presse (NewsBrowsers).

Corpus (1)

L'équipe a mis au point la base Imagiweb qui regroupe des données Twitter d'opinion politique.

5- Activités éditoriales

Participation à des comités éditoriaux (journaux scientifiques, revues, collections, etc.) (4)

L'équipe comprend des éditeurs associés des revues Computation Intelligence, Fuzzy Sets and Systems, Statistique et Enseignement ainsi qu'un membre du comité scientifique de la revue Management & Data Science.

L'équipe comprend également un membre du comité de pilotage de la conférence Extraction et de Gestion des Connaissances (EGC) et un du workshop StatLearn.

6- Activités d'évaluation

Évaluation d'articles et d'ouvrages scientifiques (379)

Les membres de l'équipe participent à l'évaluation de nombreux articles dans des conférences en informatique (IJCAI, ECML-PKDD, AAAI...) et des revues de statistique et machine learning (JASA, JMLR, AoAS, STCO, CSDA, JSS, Neurocomputing, ...) de premier plan.

Elle participe également à la relecture d'article pour de nombreuses conférences nationales (EGC, Congrès de la Société Francophone de Classification, Journées de Statistique de la SFdS, ...)

Évaluation de projets de recherche (26)

Les membres de l'équipe évaluent régulièrement des projets pour l'ANR, l'ANRT, pour la région ou encore des projets interne aux universités Lyon 1 et Lyon 2.

7- Contrats de recherche financés par des institutions publiques ou caritatives

Autres contrats européens en tant que partenaire

L'équipe a piloté le projet européen FLURESP (2011-2014) dont l'objectif était l'évaluation coût-efficacité des stratégies européennes d'alerte et de réaction en cas de pandémie de grippe humaine en Europe. Nous avons dans ce cadre mis en place un outil d'analyse multicritère des stratégies de réponse lors de pandémies de grippe pour aider à la prise de décision.

Contrats nationaux (ANR, PHRC, FUI, INCA, etc.) en tant que porteur

L'équipe a porté le projet ANR Imagiweb (2012-2015) qui visait à mettre au point un système d'analyse des représentations (au sens des images ou des opinions) qui circulent sur le Web. Il regroupait un consortium de trois laboratoires universitaires et trois entreprises. Le projet a produit plusieurs livrables, dont un outil libre d'annotation, un jeu de données politiques sur Twitter, plusieurs algorithmes de classification supervisée et non supervisée, un prototype de démonstration. Il a également donné lieu à plusieurs journées d'étude dont les dernières, pluridisciplinaires entre informatique et science politique, ont vu leurs communications étendues pour être publiés dans la revue Réseaux.

L'équipe a porté un projet de l'agence nationale uruguayenne pour la recherche (ANII), intitulé « Modèles de prévision pour la demande électrique », et financé pour un montant de 100k\$ sur 2014-2016.

Contrats nationaux (ANR, PHRC, FUI, INCA, etc.) en tant que partenaire

L'équipe a également participé à trois projets de l'agence de recherche nationale uruguayenne (ANII) : Simulation de contributions hydriques aux barrages ; Imputation de valeurs extrêmes ; Détection de signaux anormaux dans de séries électriques.

Contrats financés dans le cadre du PIA en tant que porteur

L'équipe a porté le projet OPENMIAGE (2016-2020) qui vise à construire un dispositif numérique de formation continue « e-MIAGE » (version à distance de la formation MIAGE – Méthodes Informatiques Appliquées à la Gestion des Entreprises) dispensée dans 7 universités françaises.

Contrats financés dans le cadre du PIA en tant que partenaire

L'équipe a participé au PIA REQUEST (2014-2017) qui visait à développer des technologies innovantes en Big Data, Big Analytics, Visual Analytics et Cloud Computing. Il vise également à développer une architecture open source combinant de façon novatrice les technologies sus-mentionnées afin de répondre de façon efficace aux défis des Big Data. Ce projet qui a permis de créer de nouvelles communautés (Académiques/PME/Grandes entreprises), a également permis de favoriser une production scientifique et technologique de qualité dans le domaine des Big Data. Malheureusement du fait de l'abandon du projet par certains partenaires industriels, ce projet a été interrompu avant son terme.

Contrats financés par des associations caritatives et des fondations (ARC, FMR, FRM, etc.) en tant que porteur / partenaire

L'équipe a porté ou été partenaire de 6 projets PGMO-IRSDI en collaboration avec des partenaires industriels (EDF, Thalès) : « Model-Based Functional Co-Clustering for the Analysis and the Prediction of Electric Power Consumption » ; « Exploring and summarizing large scale functional data for prediction purpose with statistical learning tools » ; « Analyzing large datasets of multivariate functional data » ; « DyNoFlu: information dynamics and novelty in textual data streams » . Ces projets, d'un montant de 15 à 25k€ chacun, ont permis le développement de plusieurs modèles et algorithmes d'analyse et de prédiction de la consommation en électricité.

Autres contrats financés par des institutions publiques

L'équipe a également porté ou participé aux projets suivants :

- CAODRA en collaboration avec des chercheurs en information communication du laboratoire ELICO, dont le but était de visualiser le paysage des open data dans la réunion Rhône-Alpes. Financement Maison des Sciences de l'Homme (MSH) et Institut des Systèmes Complexes (IXXI).
- Géographie numérique : pour une analyse épistémologique de la Géographie par un traitement automatique des discours de géographes. Projet avec le laboratoire Environnement Ville Société (EVS, Lyon 2).
- Mesurer la politique éditoriale d'un média en ligne international : en collaboration avec les sociologues du laboratoire Max Weber (Lyon 2), ce projet vise à comparer les informations transmises par le HuffingtonPost dans différents pays et langues (anglais, français, brésilien) au moyen de techniques de text mining. Financement Institut des Systèmes Complexes (IXXI)
- Néo : cyberobservatoire de la néologie. Projet interne à l'université Lyon 2 en collaboration avec les laboratoires CRTT (Lyon 2), LIUM (Le Mans université) et IULA (UPF Barcelona).
- Idenum : analyse de la représentation de la ville à travers les médias sociaux. Financement Labex IMU de Lyon.
- ISIPA : itegrale de Sugeno interpolation et proportion analogique. Financement Labex CIMI de Toulouse,
- ESNET : futur des réseaux de services écosystémiques dans la région urbaine de Grenoble. En collaboration avec le Laboratoire d'Ecologie Alpine (LECA-CNRS). Financement ONEMA (Office National de l'Eau et des Milieux Aquatiques).
- Cartographie du Web littéraire francophone : en collaboration avec le laboratoire MARGE (Lyon 3), ce projet a été financé l'IDEX Lyon-St Etienne (80k€ sur 2019-2020) dans le cadre d'un appel à projets du Collège académique "Lettres, langues, philosophie" de l'Université de Lyon.
- 3ST : construire un outil de recommandation d'article scientifique utilisable sur le corpus collecté par l'ISTEX. Financement ISTEX (Initiative d'excellence de l'Information Scientifique et Technique).

8- Post-doctorants et chercheurs accueillis

Post-doctorant accueilli (1)

L'équipe a accueilli un post-doctorant indien pour un séjour d'une année.

Chercheurs accueillis (6)

L'équipe a accueilli 6 chercheurs, tous étrangers (Brésil, Uruguay, Espagne, USA), dont deux dans le cadre du Collegium de Lyon : Ian Davidson (Université de Californie Davis, USA) pour un séjour d'une année (2017-2018), Taylor Arnold (Université de Richmond, USA) pour une année également (2018-2019).

9- Indices de reconnaissance

Responsabilités dans des sociétés savantes

Un des membres de l'équipe est membre du bureau (trésorier) de la SfdS (Société Française de Statistique). Deux autres membres sont responsable de groupe de travail de l'association EGC.

Invitations à des colloques / congrès à l'étranger

Plusieurs membres du laboratoires sont régulièrement invités dans des conférences de statistique à l'étranger :

- Conférence Model-Based Clustering and Classification, Italy, 2018
- Workshop Model-Based Clustering, USA, 2018
- Conférence ERCIM, Royaume-Uni, 2017 et 2015
- Conférence de la société roumaine de probabilités et statistiques, Roumanie, 2017
- World Statistics Congress, Brésil, 2015
- Workshop on ordinal data, Italy, 2015

Séjours dans des laboratoires étrangers

Plusieurs membres du laboratoires ont effectué des séjours dans des universités étrangères :

- Université de Can-Tho, Vietnam, 2016
- University Politehnica of Bucharest, Roumanie, 2017
- Universidad de la República, Uruguay, 2015, 2016, 2017, 2018 (x2)
- Instituto Tecnológico de Monterrey, Mexique, 2017

Un membre de l'équipe est Professeur associé de l'Université Dalhousie (Canada) depuis 2018.

Un autre membre est chercheur associé au Laboratorio Sistema Nacional de Investigadores (Uruguay).

II - INTÉRACTION DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME AVEC L'ENVIRONNEMENT NON ACADÉMIQUE, IMPACTS SUR L'ÉCONOMIE, LA SOCIÉTÉ, LA CULTURE, LA SANTÉ

1- Interactions avec les acteurs socio-économiques

Contrats de R&D avec des industriels

L'équipe a mené 14 contrats de recherche en partenariats avec des grands groupes (EDF R&D, ENEDIS, Michelin) ou de plus petites entreprises (LIZEO, Sword, Seripress, Feu Vert, Search XPR, Galderma, MasterK).

Bourses Cifre

L'équipe a encadré 12 thèses CIFRE en informatique ou en mathématiques appliquées, là encore en collaboration de grands groupes (Orange, EDF R&D, RATP) ou de plus petites entreprises (MeetSys, DSRT, Enercoop, LIM Group, AMI Software).

2- Activités d'expertise scientifique

Activités de consultant

Un membre de l'équipe a une activité permanente de consultant avec la société Lizeo, et un autre a eu deux activités ponctuelles de consultant avec les entreprises privées Euroeka Marketing Conseil et Accurate.

3- Produits destinés au grand public

Produits de vulgarisation

Un des membres de l'équipe est le fondateur et l'organisateur des Café Statistique de Lyon, qui organisent un débat public sur un sujet de société lié à la Statistique, à raison de 4 débats par an.

Plusieurs membres de l'équipe participent tous les ans au Festival Pop Sciences, à la Nuit de la Recherche, et ont donné des conférences de vulgarisation à la Maison des Mathématiques et de l'Informatique de Lyon, à la bibliothèques de Lyon, dans des lycées. Un interview au Progrès sur les sondages a été donné en 2017, et une vidéo sur l'éthique en Statistique a été réalisée dans le cadre d'un MOOC de l'ENS de Lyon.

III - IMPLICATION DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME DANS LA FORMATION PAR LA RECHERCHE

1- Produits des activités pédagogiques et didactiques

Ouvrages

Un des membres de l'équipe est très actif dans la diffusion d'outils pédagogiques (plus de 200 tutoriels ou cours en ligne), et a notamment publié 4 ebooks. Ces outils très visibles sur internet sont une vitrine importante pour le laboratoire.

E-learning, moocs, cours multimédia

D'autres membres de l'équipe ont réalisés 4 cours multimedia en ligne, dont 2 dans le cadre du PIA OPENMIAGE.

2- Productions scientifiques (articles, ouvrages, etc.) issues des thèses

Productions scientifiques issues des thèses

Les doctorants de l'équipe ont publiés 60 articles dans des revues ou des conférences, ce qui fait une moyenne de 2,7 articles par doctorant.

Les noms des doctorants sont indiqués en gras dans les listes de publications précédentes.

3- Formation

Nombre de personnes Habilitées à Diriger des Recherches (HDR) : 5

Nombre d'HDR soutenues : 1 (Julien Velcin en 2015)

Doctorants (nombre total) : 22

Doctorants bénéficiant d'un contrat spécifique au doctorat : 22

Nombre de thèses soutenues : 12

Durée moyenne des thèses : 49 mois

Stagiaires accueillis (M1, M2) : 37

Personnes responsables d'une mention ou d'un parcours de master (nombre total) : 5

Personnes responsables d'une mention ou d'un parcours de master à labellisation internationale (Erasmus Mundus, par ex.) : 2



Département d'évaluation
de la recherche

ANNEXE 4 - Produits et activités de la recherche

En cohérence avec les données chiffrées de l'onglet 4 du fichier Excel « Données du contrat en cours », on remplira ce document destiné à l'évaluation du critère 1 du référentiel de l'évaluation « Produits et activités de la recherche », pour l'ensemble de l'unité et pour chaque équipe / thème.

CAMPAGNE D'ÉVALUATION 2019-2020
VAGUE A

Nom de l'équipe : Systèmes d'Information Décisionnels

Acronyme : SID

Responsable d'équipe pour le contrat en cours : Fadila Bentayeb

Responsable d'équipe pour le contrat à venir :

I- PRODUCTION DE CONNAISSANCES ET ACTIVITÉS CONCOURANT AU RAYONNEMENT ET À L'ATTRACTIVITÉ SCIENTIFIQUE DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME

1- Journaux / Revues

Les 20 % des articles publiés dans les revues mis en avant par l'équipe (6 articles parmi 32) sont ceux dont la référence est en gras. Ils couvrent globalement toutes les thématiques de recherche de l'équipe et comptent 4 revues de rang A dont 2 articles de synthèse et 2 articles de rang B.

Articles scientifiques (28)

1. **[Ratsimbazafy-18] RATSIMBAZAFY R., BOUSSAID O.** «Multiple Decisional Query Optimization in Big Data Warehouse». In *International Journal of Data Warehousing and Mining*, vol. 14/3 (2018) p.22-43. Ref HAL: [hal-02051408_v1](https://hal.archives-ouvertes.fr/hal-02051408_v1). Rang B.
2. **[Truica-18] Truica C.-O., DARMONT J., Boicea a., Radulescu F.** «Benchmarking Top-K Keyword and Top-K Document Processing with T2K2 and T2K2D2». In *Future Generation Computer Systems*, Vol. 85 (2018) p.60-75. Ref HAL: [hal-01717121_v1](https://hal.archives-ouvertes.fr/hal-01717121_v1). Rang A.
3. **[Oukid-16] Oukid L., Benblidia N., BOUSSAID O., BENTAYEB F.** «Tlabel : A new OLAP aggregation operator in text cubes». In *International Journal of Data Warehousing and Mining*, vol. 12/4 (2016) 54–74. Ref HAL: [hal-02054192_v1](https://hal.archives-ouvertes.fr/hal-02054192_v1). Rang B.
4. **[Boukraa-14] Boukraa D., BOUSSAID O., BENTAYEB F.** «Complex Object-Based Multidimensional Modeling and Cube Construction». In *Fundamenta Informaticae*, vol. 132/2 (2014) p.203-238. Ref HAL: [hal-01484443_v1](https://hal.archives-ouvertes.fr/hal-01484443_v1). Rang A.
5. [Zaouai-19] Zaouai M., Kazar o., Ortiz Bellot g., Haba b., KABACHI N., krishnamyrthty M., *Ambiance Intelligence Approach Using IoT and Multi-Agent System*. In *International Journal of Distributed Systems and Technologies (IJ DST)*. January 2019; 10(1):19, pp. 15-32. DOI:[10.4018/IJ DST.2019010104](https://doi.org/10.4018/IJ DST.2019010104). Rang C.
6. [Belouaar-19] Belouuar H., Kasar O., KABACHI N., *Web Service Selection Approach Based on Agent and Fuzzy Logic*. In *Journal of Digital Information Management (JDIM)*, Volume 17 Issue 1, February 2019. DOI: <https://doi.org/10.6025/jdim/2019/1/1-12>. Rang C.
7. [Zaouai-19] Zaouai M., Kazar O., Haba B., Ortiz Bellot G, KABACHI N., *New approach using an IoT robot to oversight the smart home environment*. In *Internet of Things*, Issue 1, Volume 3, 2019/02/7, DOI : [10.21494/ISTE.OP.2019.0323](https://doi.org/10.21494/ISTE.OP.2019.0323), ISSN 2514-8273. Éditeur ISTE OpenScience.
8. [Ouret-19] Ouret Z., Boukraa D., BOUSSAID O., Challal R., *AuMixDw: Towards an automated hybrid approach for building XML data warehouses*. In *Data & Knowledge Engineering* (2019), <https://doi.org/10.1016/j.datak.2019.01.004>, in press. Rang B.
9. [Bimonte-16] Bimonte S., Schneider M., BOUSSAID O. «Business intelligence indicators : Types, models and implementation». In *International Journal of Data Warehousing and Mining*, vol. 12/4 (2016) 75–98. Ref HAL: [hal-02054113_v1](https://hal.archives-ouvertes.fr/hal-02054113_v1). Rang B.
10. [Jakawat-16a] **JAKAWAT W., FAVRE C., LOUDCHER S.** «Graphs enriched by cubes for OLAP on bibliographic networks». In *International Journal of Business Intelligence and Data Mining*, vol. 11/1 (2016) p.85-107. DOI: [10.1504/IJBIDM.2016.076435](https://doi.org/10.1504/IJBIDM.2016.076435). Rang B.
11. [Oukid-15] Oukid L., Benblidia N., Asfari O., BENTAYEB F., BOUSSAID O. «Contextualized Text OLAP Based on Information Retrieval». In *International Journal of Data Warehousing and Mining*, vol. 11/2 (2015) p.1-21. Ref HAL: [hal-01484437_v1](https://hal.archives-ouvertes.fr/hal-01484437_v1). Rang B.
12. [Bouakkaz-16] Bouakkaz M., LOUDCHER S., Ouinten Y. «OLAP Textual Aggregation Approach using the Google Similarity Distance». In *International Journal of Business Intelligence and Data Mining*, vol. 11/1 (2016) p.31-48. Ref HAL: [halshs-01231490_v1](https://hal.archives-ouvertes.fr/halshs-01231490_v1). Rang B.
13. [Guille-15] GUILLE A., FAVRE C. «Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach». In *Social Network Analysis and Mining*, vol. 5/1 (2015) 18:1-18:18. DOI: [10.1007/s13278-015-0258-0](https://doi.org/10.1007/s13278-015-0258-0). Rang B.
14. [Pierrot-14] Perriot R., Pfeifer J., D'orazio L., Bachelet B., Bimonte S., DARMONT J. «Cost Models for Selecting Materialized Views in Public Clouds». In *International Journal of Data Warehousing and Mining*, vol. 10/4 (2014) p.1-25. Ref HAL: [hal-01060585_v1](https://hal.archives-ouvertes.fr/hal-01060585_v1). Rang B.

15. [Attasena-15] **ATTASENA V.**, **HARBI N.**, **DARMONT J.** «A Novel Multi-Secret Sharing Approach for Secure Data Warehousing and On-Line Analysis Processing in the Cloud». In *International Journal of Data Warehousing and Mining*, vol. 11/2 (2015) p.21-42. Ref HAL: [hal-01059693_v1](#). Rang B.
16. [Ghebghoub-16] Ghebghoub Y., Oukid S., **BOUSSAID O.** «An MDA approach to secure access to data on cloud using implicit security». In *International Journal of Computer Science and Information Security*, vol. 8/2 (2016) 107–120. Ref HAL: [hal-02054156_v1](#). Rang C.
17. [Bonnevay-16] **BONNEVAY S.**, **GAVIN G.**, AUBERTIN P. «Comparison of two metaheuristics to solve a 2-D cutting stock problem with set-up cost in the paper industry». In *International Journal of Metaheuristics*, vol. 5/1 (2016) p.31. Ref HAL: [hal-02051744_v1](#). Rang C.
18. [Attaf-18] Attaf S., Benblidia B., **BOUSSAID O.** «Warehousing and analysing textual data». In *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 8/3 (2018) p.216-244. Ref HAL: [hal-02051395_v1](#). Rang C.
19. [Dounya-18] Kassimi D., Kazar O., **BOUSSAID O.**, Merizig A. «New approach for intrusion detection in Big Data as a service in the Cloud». In *Journal of Digital Information Management*, vol. 16/6 (2018) p.216-244. Ref HAL: [hal-02051289_v1](#). Rang C.
20. [Beligne-18a] **BELIGNE M.**, **LOUDCHER S.**, Lefort I. «Mise en perspective d'une variation du varia». In *Géocarrefour - Revue de géographie de Lyon* (2018). Ref HAL: [halshs-01851673_v1](#).
21. [Beligne-17b] **BELIGNE M.** «Des espaces dans la détection de métaphores». In *Nouvelles perspectives en sciences sociales* (2017). Ref HAL: [halshs-01619360_v1](#).
22. [Bouakkaz-17] Bouakkaz M., Ouinten Y., **LOUDCHER S.**, Fournier-Viger P. «Efficiently mining frequent itemsets applied for textual aggregation». In *Applied Intelligence* (2017). DOI: [10.1007/s10489-017-1050-9](#).
23. [Bala-16] Bala M., **BOUSSAID O.**, alimazighi Z. «Extracting-transforming-loading modeling approach for big data analytics». In *International Journal of Decision Support System Technology (IJDSST)*, vol. 8/4 (2016) 50–69. Ref HAL: [hal-02054097_v1](#). Rang C.
24. [Hannachi-15] Hannachi L., Benblidia N., **BOUSSAID O.**, **BENTAYEB F.** «Community Cube: a semantic framework for analysing social network data». In *International Journal of Metadata, Semantics and Ontologies*, vol. 10/3 (2015) p.155-169. Ref HAL: [hal-01484427_v1](#). Rang C
25. [Selmane-15] **SELMANE S.-A.**, **BOUSSAID O.**, **BENTAYEB F.** «Towards Collaborative Multidimensional Query Recommendation with Triadic Association Rules». In *International Journal of Decision Support System Technology (IJDSST)*, vol. 7/3 (2015) p.17-35. Ref HAL: [hal-01484342_v1](#). Rang C.
26. [Fareh-15] Fareh M., **BOUSSAID O.**, Chalal R. Reconciliation model of heterogeneous information in decisional information systems, in *International Journal of Metadata, Semantics and Ontologies (IJMSO)*, 10 (1), 37–54. Rang C.
27. [Younsi-15] **YOUNSI F.-Z.**, Hamdadou D., **BOUSSAID O.** A surveillance and spatiotemporal visualization model for infectious diseases using social network, in *International Journal of Decision Support System Technology (IJDSST)*, 7 (4), 1–19. Rang C.
28. [Aknouche-14] **AKNOUCHE R.**, **BOUSSAID O.**, **BENTAYEB F.**, Asfari O. «Entrepôts de textes. Proposition d'un processus ETL et d'un modèle multidimensionnel TWM approprié». In *Ingénierie des Systèmes d'Information*, vol. 19/5 (2014) p.45-73. Ref HAL: [hal-01484448_v1](#).

Articles de synthèse / revues bibliographiques (3 ; les deux premiers sont dans les 20%)

1. [Attasena-17] **ATTASENA V.**, **DARMONT J.**, **HARBI N.** «Secret Sharing for Cloud Data Security». In *The International Journal on Very Large Databases*, vol. 26/5 (2017) p.657-681. Ref Arxiv: [1712.10155](#). Rang A.
2. [Loudcher-15] **LOUDCHER S.**, **JAKAWAT W.**, **FAVRE C.** «Combining OLAP and Information Networks for Bibliographic Data Analysis: A Survey». In *Scientometrics*, vol. 103/2 (2015) p.471-487. Ref HAL: [halshs-01136559_v1](#). Rang A.
3. [Bouakkaz-17a] Bouakkaz M., Ouinten Y., **LOUDCHER S.**, Strelalova Y. «Textual aggregation approaches in OLAP context: A survey». In *International Journal of Information Management*, vol. 37/6 (2017) p.684-69. DOI: [10.1016/j.ijinfomgt.2017.06.005](#).

2- Ouvrages

Direction et coordination d'ouvrages scientifiques / édition scientifique (5 dont 4 en anglais)

1. [Darmont-18] DARMONT J., LOUDCHER S. «Utilizing Big Data Paradigms for Business Intelligence». IGI Global (2018). DOI: [10.4018/978-1-5225-4963-5](https://doi.org/10.4018/978-1-5225-4963-5).
2. [Bentayeb-18] BENTAYEB F., Novelli N. «Special Issue on the 12th French Speaking Conference on data warehousing and OLAP». In *International Journal of Data Warehousing and Mining*, vol. 14, no. 3 (2018).
3. [Favre-17c] FAVRE C., Artaud C., Duffau C., Fraiser O., Kotto-Kombi R. «Forum Jeunes Chercheurs à Inforsid 2016». In *Ingénierie des Systèmes d'Information*, vol. 22/2 (2017) p.121-147. Ref HAL: [hal-01914115 v1](https://hal.archives-ouvertes.fr/hal-01914115).
4. [Darmont-15c] DARMONT J., Pedersen T. B. «Special section on Cloud Intelligence». In *Information Sciences*, vol. 54 (2015) p.309-356. Ref HAL: [hal-01426055 v1](https://hal.archives-ouvertes.fr/hal-01426055).
5. [Darmont-15a] DARMONT J., Pedersen T. B. «Cloud Intelligence». In *Information Sciences*, vol. 48 (2015) p.179-223. Ref HAL: [hal-01165612 v1](https://hal.archives-ouvertes.fr/hal-01165612).

Chapitres d'ouvrage (3 dont 2 en anglais)

1. [Darmont-17a] DARMONT J. «Data-Centric Benchmarking». In *Encyclopedia of Information Science and Technology, Fourth Edition* (2017) p.1772-1782. DOI: [10.4018/978-1-5225-2255-3.ch154](https://doi.org/10.4018/978-1-5225-2255-3.ch154).
2. [Darmont-17] DARMONT J., Marcel P. «Entrepôts de données et OLAP, analyse et décision dans l'entreprise». In *Les big data à découvert* (2017) p.132-133. Ref HAL: [hal-01493948 v1](https://hal.archives-ouvertes.fr/hal-01493948).
3. [Darmont-14] DARMONT J. «Data Processing Benchmarks». In *Encyclopedia of Information Science and Technology, Third Edition* (2014) p.146-152. Ref HAL: [hal-00978026 v1](https://hal.archives-ouvertes.fr/hal-00978026).

3- Production dans des colloques / congrès, séminaires de recherche

Éditions d'actes de colloques / congrès (10)

1. [Darmont-19] DARMONT J., Grabar N., Teste O. «About Variety in Humanities Big Data». *Recherche d'information, document et web sémantique*, vol. 19/1 (2019). Ref HAL: [hal-02049343 v1](https://hal.archives-ouvertes.fr/hal-02049343).
2. [Badir-18] Badir H., BENTAYEB F., BOUSSAID O. « Actes des 14^e journées EDA Business Intelligence & Big Data (EDA 2018) ». In *RNTI*, vol. B-14 (2018).
3. [Badir-18] Badir H., Boulmakoul A., BOUSSAID O. : Actes de la 12^{ème} édition de la conférences maghrébine sur les Avancées des Systèmes Décisionnels, Big Data & Applications, ASD 2018, Marrakech, Maroc, 2-3 mai 2018.
4. [Bub-18] BUB U., Dahanayake A., DARMONT J., Diamantini C., Fassetti F., Ferme E., KABACHI N., Matteucci I., MOLNAR B., Navathe S., Oro E., Petrocchi M., Rombo S., Ruffolo M., Spognardi A., Thalheim B., Ursino D. «Contributions from ADBIS 2018 Workshops». In *New Trends in Databases and Information Systems - ADBIS 2018 Short Papers and Workshops, Communications in Computer and Information Science*, vol. 909 (2018) p.91-102. DOI: [10.1007/978-3-030-00063-9](https://doi.org/10.1007/978-3-030-00063-9).
5. [Kirikova-17] Kirikova M., Kjetil N., Papadopoulos G., Johann G., Robert W., DARMONT J., Rizzi S. «ADBIS 2017 Short Papers and Workshops». In *New Trends in Databases and Information Systems - ADBIS 2017 Short Papers and Workshops*, vol. 767 (2017). Ref HAL: [hal-01578926 v1](https://hal.archives-ouvertes.fr/hal-01578926).
6. [Bentayeb-17] BENTAYEB F., BOUSSAID O., DARMONT J. « Actes des 13^e journées EDA Business Intelligence & Big Data (EDA 2017) ». *RNTI*, vol. B-13 (2017). Ref HAL: [hal-01519389 v1](https://hal.archives-ouvertes.fr/hal-01519389).
7. [Zghal-17] Zghal S., BOUSSAID O. : Actes de la 11^{ème} édition de la conférence maghrébine sur les Avancées des Systèmes Décisionnels, ASD 2017, 27-29 avril 2017 Tabarka, Tunisie.
8. [Darmont-15] DARMONT J., Otjacques B., Tamisier T. «Actes de la 15^e conférence internationale sur l'Extraction et la Gestion des Connaissances (EGC 2015)». *RNTI*, vol. E-28 (2015). Ref HAL: [hal-01165635 v1](https://hal.archives-ouvertes.fr/hal-01165635).
9. [Badir-15] Badir H., Boulmakoul A., BOUSSAID O. : Actes de la 9^{ème} édition de la conférences maghrébine sur les Avancées des Systèmes Décisionnels, Fondements et Applications, ASD 2015, Tanger Maroc, 10-12 septembre 2015.
10. [Feki-14] Feki J., BOUSSAID O. : Actes de la 8^{ème} édition de la conférences maghrébine sur les Avancées des Systèmes Décisionnels, ASD 2014, Hammamet, Tunisie, 29-31 mai 2014.

Articles publiés dans des actes de colloques / congrès (96)

Les 20% des articles publiés dans les conférences que l'équipe met en avant sont au nombre de 20 parmi 96. Ce sont des conférences internationales de bon niveau dans le domaine de recherche de l'équipe dont 3 sont de rang A (ICDM, FUZZ-IEEE, GECCO) et 17 de rang B (ADBIS, DAWAK, DEXA, DOLAP, IDEAS, etc.).

1. **[Ramdane-18]** RAMDANE Y., BOUSSAID O., KABACHI N., BENTAYEB F. « Partitioning and Bucketing Techniques to Speed up Query Processing in Spark-SQL ». In proc. of *International Conference on Parallel and Distributed Systems (ICPADS 2018)* p.142-151. **Rang B**
2. **[Boussahoua-18]** BOUSSAHOUA M., BENTAYEB F., BOUSSAID O., KABACHI N. «A Data Partitioning Optimization Approach for Distributed Data Warehouses on Column family NoSQL Systems». In proc. of *International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2018)* p.54-60. **Rang B.**
3. **[Wang-17c]** WANG X., AH-PINE J., DARMONT J. « SHCoClust, a Scalable Similarity-based Hierarchical Clustering Method and its Application to Textual Collections ». In proc. of *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2017)* ID F-0529. Ref HAL: [hal-01504986_v1](https://hal.archives-ouvertes.fr/hal-01504986_v1). **Rang A.**
4. **[Challal-17]** Challal Z., Boukhalfa K., BOUSSAID O. «Minimizing Negative Influence in Social Networks A Graph OLAP Based Approach». In proc. of *International Conference on Database and Expert Systems Applications (DEXA)* p.378-386. Ref HAL: [hal-02053863_v1](https://hal.archives-ouvertes.fr/hal-02053863_v1). **Rang B.**
5. **[Boussahoua-17]** BOUSSAHOUA M., BOUSSAID O., BENTAYEB F. «Logical Schema for Data Warehouse on Column-Oriented NoSQL Databases». In proc. of *International Conference on Database and Expert Systems Applications (DEXA 2017)* p.247-256. Ref HAL: [hal-01946758_v1](https://hal.archives-ouvertes.fr/hal-01946758_v1). **Rang B.**
6. **[Ratsimbazafy-17]** RATSIMBAZAFY R., BOUSSAID O., BENTAYEB F. «S2D: Shared Distributed Datasets, Storing Shared Data for Multiple and Massive Queries Optimization in a Distributed Data Warehouse». In proc. of *International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2017)* p.42-50. Ref HAL: [hal-01946753_v1](https://hal.archives-ouvertes.fr/hal-01946753_v1). **Rang B.**
7. **[Truica-17]** Truica C.-O., DARMONT J. «T?K?: The Twitter Top-K Keywords Benchmark». In proc. of *East-European Conference on Advances in Databases and Information Systems (ADBIS 2017)* p.21-28. DOI: [10.1007/978-3-319-67162-8](https://doi.org/10.1007/978-3-319-67162-8). **Rang B.**
8. **[SobatiMoghadam-17]** SOBATI MOGHADAM S., DARMONT J., GAVIN G. «Enforcing Privacy in Cloud Databases». In proc. of *International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2017)* p.53-73. Ref Arxiv: [1708.09171](https://arxiv.org/abs/1708.09171). **Rang B.**
9. **[SorianoMorales-16]** SORIANO MORALES E. P., AH-PINE J., LOUDCHER S. «Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation». In proc. of *Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2016)*. Ref HAL: [halshs-01384657_v1](https://hal.archives-ouvertes.fr/halshs-01384657_v1). **Rang B.**
10. **[Gavin-16a]** GAVIN G. «An Efficient Somewhat Homomorphic Encryption Scheme Based on Factorization.». In proc. of *International Conference on Cryptology and Network Security (CANS 2016)*. Ref HAL: [hal-02051969_v1](https://hal.archives-ouvertes.fr/hal-02051969_v1). **Rang B.**
11. **[Truica-16a]** Truica C.-O., DARMONT J., VELCIN J. «A Scalable Document-based Architecture for Text Analysis». In proc. of *International Conference on Advanced Data Mining and Applications (ADMA 2016)* p.481-494. Ref Arxiv: [1612.06195](https://arxiv.org/abs/1612.06195). **Rang B.**
12. **[Jakawat-16]** JAKAWAT W., FAYRE C., LOUDCHER S. «OLAP Cube-based Graph Approach for Bibliographic Data». In proc. of *International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2016)* p.87-99. Ref HAL: [halshs-01231492_v1](https://hal.archives-ouvertes.fr/halshs-01231492_v1). **Rang B.**
13. **[Bonnevay-15]** BONNEVAY S., AUBERTIN P., GAVIN G. «A Genetic Algorithm to Solve a Real 2-D Cutting Stock Problem with Setup Cost in the Paper Industry». In proc. of *Genetic and Evolutionary Computation Conference (GECCO)*. Ref HAL: [hal-01134195_v1](https://hal.archives-ouvertes.fr/hal-01134195_v1). **Rang A.**
14. **[Arres-15a]** ARRES B., KABACHI N., BOUSSAID O., BENTAYEB F. «Intentional Data Placement Optimization for Distributed Data Warehouses». In proc. of *IEEE International Conference on Systems, Man, and Cybernetics (SMC 2015)* p.80-86. Ref HAL: [hal-01488980_v1](https://hal.archives-ouvertes.fr/hal-01488980_v1). **Rang B.**
15. **[Arres-15b]** ARRES B., KABACHI N., BOUSSAID O., BENTAYEB F. «A Data Mining-based Blocks Placement Optimization for Distributed Data Warehouses». In proc. of *International Database Engineering and Application Symposium (IDEAS 2015)* p.146-153. Ref HAL: [hal-01488941_v1](https://hal.archives-ouvertes.fr/hal-01488941_v1). **Rang B.**
16. **[Dermouche-15]** DERMOUCHE M., Khouas L., VELCIN J., LOUDCHER S. «A Joint Model for Topic-Sentiment Modeling from Text». In proc. of *Annual ACM Symposium on Applied Computing (SAC 2015)*. Ref HAL: [halshs-01099984_v1](https://hal.archives-ouvertes.fr/halshs-01099984_v1). **Rang B.**

17. [Dermouche-14] **DERMOUCHE M.**, VELCIN J., LOUDCHER S., Khouas L. «A Joint Model for Topic-Sentiment Evolution over Time». In proc. of *IEEE International Conference on Data Mining (ICDM 2014)*. Ref HAL: [halshs-01099990_v1](#). Rang A.
18. [Attasena-14] **ATTASENA V.**, HARBI N., DARMONT J. «fVSS: A New Secure and Cost-Efficient Scheme for Cloud Data Warehouses». In proc. of *ACM International Workshop on Data Warehousing and OLAP (DOLAP 2014)* p.81-90. DOI: [10.1145/2666158.2666173](#). Rang B.
19. [Dehdouh-14d] **DEHDOUH K.**, BENTAYEB F., BOUSSAID O., KABACHI N. «Towards an OLAP Environment for Column-Oriented Data Warehouses». In proc. of *International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2014)* p.221-232. Ref HAL: [hal-01492725_v1](#). Rang B.
20. [Nogueira-18a] **NOGUEIRA I.**, ROMDHANE M., DARMONT J. «Modeling Data Lake Metadata with a Data Vault». In proc. of *International Database Engineering and Application Symposium (IDEAS 2018)* p.253-261. Ref HAL: [hal-01788036_v1](#). Rang B.
21. [Dehdouh-14e] **DEHDOUH K.**, BENTAYEB F., BOUSSAID O., KABACHI N. «Columnar NoSQL CUBE: Agregation operator for columnar NoSQL data warehouse». In proc. of *IEEE International Conference on Systems, Man, and Cybernetics (SMC 2014)* p.3828-3833. Ref HAL: [hal-01492731_v1](#). Rang B.
22. [Selmane-14f] **SELMANE S.-A.**, BENTAYEB F., BOUSSAID O. «P-TRIAR: Personalization Based on TRIadic Association Rules». In proc. of *East-European Conference on Advances in Databases and Information Systems (ADBIS 2014)* p.234-247. Ref HAL: [hal-01492847_v1](#). Rang B.
23. [Selmane-14e] **SELMANE S.-A.**, BOUSSAID O., BENTAYEB F. «P-TRIAR: Personalization Based on TRIadic Association Rules». In proc. of *East-European Conference on Advances in Databases and Information Systems (ADBIS 2014)* p.234-247. DOI: [10.1007/978-3-319-10933-6_18](#). Rang B.
24. [Sawadogo-19] **SAWADOGO P.**, Kibata T., DARMONT J. «Metadata Management for Textual Documents in Data Lakes». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2019)*. To appear. Ref HAL: [hal-02012092_v1](#). Rang C.
25. [R hazlane-16b] **RHAZLANE S.**, HARBI N., KABACHI N., Badir H. «Intelligent Multi Agent System based Solution for Data Protection in the Cloud». In proc. of *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*. Ref HAL: [hal-02055772_v1](#). Rang C.
26. [Bouakkaz-16a] **Bouakkaz M.**, LOUDCHER S., Ouinten Y. «A New Tool for Textual Aggregation in OLAP Context». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2016)*. Ref HAL: [halshs-01384654_v1](#). Rang C.
27. [SorianoMorales-16a] **SORIANO MORALES E. P.**, AH-PINE J., LOUDCHER S. «Hypergraph Modelization of a Syntactically Annotated English Wikipedia Dump». In proc. of *Language Resources and Evaluation Conference (LREC 2016)*. Ref HAL: [halshs-01384647_v1](#). Rang C.
28. [Roos-16a] **ROOS J.**, BONNEVAY S., GAVIN G. «Short-Term Urban Rail Passenger Flow Forecasting: A Dynamic Bayesian Network Approach». In proc. of *International Conference on Machine Learning and Applications (ICMLA)* p.1034-1039. Ref HAL: [hal-02051746_v1](#). Rang C.
29. [Ozturk-15] **OZTÜRK A.**, Eyango L., Waksman S. Y., LALLICH S., DARMONT J. «Warehousing Complex Archaeological Objects». In proc. of *International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2015)* p.226-239. Ref Arxiv: [1608.06469](#). Rang C.
30. [Arres-15] **ARRES B.**, KABACHI N., BOUSSAID O. «Optimizing OLAP Cubes Construction by Improving Data Placement on Multi-nodes Clusters». In proc. of *23rd EuroPDP International Conference on Parallel, Distributed, and Network-Based Processing (PDP 2015)* 520 - 524. DOI: [10.1109/PDP.2015.45](#). Rang C.
31. [Bouakkaz-15] **Bouakkaz M.**, LOUDCHER S., Ouinten Y. «GOTA: Using the Google Similarity Distance for OLAP Textual Aggregation». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2015)* p.121-127. Ref HAL: [halshs-01136581_v1](#). Rang C.
32. [Arres-14b] **ARRES B.**, KABACHI N., BENTAYEB F., BOUSSAID O. «An Original Approach for Processing Public Open Data with MapReduce: A Case Study». In proc. of *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2014)* 138- 143. DOI: [10.1109/AICCSA.2014.7073190](#). Rang C.
33. [Selmane-14b] **SELMANE S.-A.**, BENTAYEB F., Missaoui R., BOUSSAID O. «An 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. DOI: [10.1007/978-3-319-08326-1_7](#). Rang C.
34. [Gavin-19] **GAVIN G.**, BONNEVAY S. «Securely aggregating testimonies with Threshold Multi-key FHE». In proc. of *International Conference on Codes, Cryptology and Information Security (C2SI)*. Ref HAL: [hal-02051956_v1](#).
35. [Benaissa-18] **BENAISSA R.**, Benhammadi F., BOUSSAID O., Mokhtari A. «Clustering Approach for Data Lake Based on Medoid's Ranking Strategy». In proc. of *International Conference on Computing Systems and Applications (CSA 2018)* p.250-260. DOI: [10.1007/978-3-319-98352-3](#).

36. [Ozturk-18a] **OZTÜRK A.**, **LALLICH S.**, **DARMONT J.**, Waksman S. Y. «MaxMin Linear Initialization for Fuzzy C-Means». In proc. of *International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2018)* p.1-15. Ref HAL: [hal-01771204_v1](https://hal.archives-ouvertes.fr/hal-01771204_v1).
37. [Rhzalane-17b] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Rising Cloud Data Security and Privacy Challenges using Multi Agent Systems». In proc. of *Data Engineering in Bioinformatics, Image and Data Analysis (SMC)*. Ref HAL: [hal-02055913_v1](https://hal.archives-ouvertes.fr/hal-02055913_v1).
38. [Rhzalane-17a] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Alteration Agent for Cloud Data Security». In proc. of *ACM International Conference on Management of Emergent Digital EcoSystems (MEDES 2017)*. Ref HAL: [hal-02055789_v1](https://hal.archives-ouvertes.fr/hal-02055789_v1).
39. [Roos-17] **ROOS J.**, **BONNEVAY S.**, **GAVIN G.** «Dynamic Bayesian Networks with Gaussian Mixture Models for Short-Term Passenger Flow Forecasting». In proc. of *International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. Ref HAL: [hal-01807067_v1](https://hal.archives-ouvertes.fr/hal-01807067_v1).
40. [Gavin-17] **GAVIN G.**, **BONNEVAY S.** «Combining imperfect knowledge of maybe corrupted sources». In proc. of *2017 ISEA Asia Security and Privacy (ISEASP)*. DOI: [10.1109/ISEASP.2017.7976997](https://doi.org/10.1109/ISEASP.2017.7976997).
41. [Ozturk-18] **OZTÜRK A.**, **LALLICH S.**, **DARMONT J.** «A Visual Quality Index for Fuzzy C-Means». In proc. of *International Conference on Artificial Intelligence Applications and Innovations (AIAI 2018)* p.546-555. Ref HAL: [hal-01727426_v1](https://hal.archives-ouvertes.fr/hal-01727426_v1).
42. [SorianoMorales-17] **SORIANO MORALES E. P.**, **AH-PINE J.**, **LOUDCHER S.** «Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation.». In proc. of *International Conference Discovery Science (DS 2017)* p.340-355. Ref HAL: [halshs-01577043_v1](https://hal.archives-ouvertes.fr/halshs-01577043_v1).
43. [SobatiMoghadam-17a] **SOBATI MOGHADAM S.**, **DARMONT J.**, **GAVIN G.** «S4: A New Secure Scheme for Enforcing Privacy in Cloud Data Warehouses». In proc. of *International Conference on Information Systems and Technologies (ICIST 2017)* p.9-16. Ref Arxiv: [1708.06574](https://arxiv.org/abs/1708.06574).
44. [ElOuazzani-16] EL Ouazzani A., **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Dynamic management of data warehouse security levels based on user profiles». In proc. of *IEEE International Colloquium on Information Science and Technology (CIST)*. Ref HAL: [hal-02055928_v1](https://hal.archives-ouvertes.fr/hal-02055928_v1).
45. [Kichou-16] Kichou S., Mellah H., **BOUSSAID O.**, Meziane A. «Handicraft women recommendation approach based on user's social tagging operations». In proc. of *IEEE/WIC/ACM International Conference on Web Intelligence (WI)* 618–621. Ref HAL: [hal-02054180_v1](https://hal.archives-ouvertes.fr/hal-02054180_v1).
46. [Roos-16b] **ROOS J.**, **GAVIN G.**, **BONNEVAY S.** «A dynamic bayesian network approach to forecast short-term urban rail passenger flows with incomplete data». In proc. of *European Transport Conference(ETC)*. Ref HAL: [hal-01807053_v1](https://hal.archives-ouvertes.fr/hal-01807053_v1).
47. [Pierrot-16] **PIERROT D.**, **HARBI N.**, **DARMONT J.** «Hybrid Intrusion Detection in Information Systems». In proc. of *International Conference on Information Science and Security (ICISS 2016)*. Ref HAL: [hal-01380026_v1](https://hal.archives-ouvertes.fr/hal-01380026_v1).
48. [SobatiMoghadam-16a] **SOBATI MOGHADAM S.**, **GAVIN G.**, **DARMONT J.** «Secure Order-Preserving Indexing Schemes for Outsourced Data». In proc. of *IEEE International Carnahan Conference on Security Technology (ICCST 2016)* p.297-303. Ref HAL: [hal-01380020_v1](https://hal.archives-ouvertes.fr/hal-01380020_v1).
49. [Dehdouh-14c] **DEHDOUH K.**, **BOUSSAID O.**, **BENTAYEB F.** «Columnar NoSQL Star Schema Benchmark». In proc. of *International Conference on Model & Data Engineering (MEDI 2014)* p.281-288. Ref HAL: [hal-01492733_v1](https://hal.archives-ouvertes.fr/hal-01492733_v1).
50. [Bouakkaz-14] Bouakkaz M., **LOUDCHER S.**, Ouiten Y. «Automatic Textual Aggregation Approach of Scientific Articles in OLAP context». In proc. of *International Conference on Innovations in Information Technology (Innovations 2014)* p.30-35. Ref HAL: [halshs-01099997_v1](https://hal.archives-ouvertes.fr/halshs-01099997_v1).
51. [Arres-14a] **ARRES B.**, **BOUSSAID O.**, **KABACHI N.**, **BENTAYEB F.** «Parallel Processing of Public Open Data with the MapReduce Paradigm: A Case Study». In proc. of *Big Spatial Data* p.132-141. Ref HAL: [hal-01023308_v1](https://hal.archives-ouvertes.fr/hal-01023308_v1).
52. [Guille-14b] **GUILLE A.**, **FAVRE C.** «Mention-anomaly-based Event Detection and Tracking in Twitter». In proc. of *International Conference on Advances in Social Network Analysis and Mining (ASONAM 2014)* p.375-382. DOI: [10.1109/ASONAM.2014.6921613](https://doi.org/10.1109/ASONAM.2014.6921613).
53. [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).

Conférences et ateliers nationaux

54. [Ramdane-19] **RAMDANE Y.**, **BOUSSAID O.**, **KABACHI N.**, **BENTAYEB F.** : «Conception physique d'un entrepôt de données distribuées basée sur K-means équilibré». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2019)* p.177-188. Ref HAL: [hal-02019782_v1](#).
55. [Beligne-19] **BELIGNE M.**, **LOUDCHER S.**, Lefort I. «Transformation de documents numérisés par la plateforme Persée dans un objectifs d'analyses textuelles diachroniques». In proc. of *Colloque international : Histoire, Langues et Textométrie*. Ref HAL: [halshs-01851656_v1](#).
56. [Beligne-18] **BELIGNE M.**, **LOUDCHER S.**, Lefort I. «Analyse critique d'un changement de paradigme en géographie à partir de réflexion entre discontinuités scientifiques, lexicales et sémantiques». In proc. of *Colloque Penser avec les discontinués en géographie*. Ref HAL: [halshs-01851647_v1](#).
57. [Boussahoua-18a] **BOUSSAHOUA M.**, **BOUSSAID O.**, **BENTAYEB F.**, **KABACHI N.** «Règles d'Association pour le Partitionnement Horizontal des Entrepôts de Données dans un Système NoSQL en Colonnes». In proc. of *EDA 2018* p.89-104.
58. [Rhazlane-18] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Les Systèmes Multi Agents au Service de la Sécurité des Données Entreposées dans le Cloud». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA)*. Ref HAL: [hal-02055812_v1](#).
59. [Pierrot-18] **PIERROT D.**, **HARBI N.**, **DARMONT J.** «Détection des intrusions et aide à la décision». In proc. of *Conférence Maghrébine sur les Avancées des Systèmes Décisionnels (ASD 2018)*. Ref HAL: [hal-01761914_v1](#).
60. [HuyghuesDespointes-18] **HUYGHUES-DESPOINTES C.**, **VELCIN J.**, Khouas L., **LOUDCHER S.** «Extraction de chaînes cohérentes en vue de reconstruire la trajectoire de l'information». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref HAL: [hal-01674547_v1](#).
61. [Nogueira-18] Nogueira I., Romdhane M., **DARMONT J.** «Modélisation des métadonnées d'un data lake en data vault». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2018)* p.257-262. Ref HAL: [hal-01665547_v1](#).
62. [Truica-18a] Truica C.O., **DARMONT J.**, Boicea A., Radulescu F. «Le calcul des top-k mots clés et documents au banc d'essais avec T²K² et T²K²D²». In proc. of *34e Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA 2018)*. Ref HAL: [hal-02054922_v1](#).
63. [Zeghdaoui-18] **ZEGHDAOUI W.**, Joly F., **BOUSSAID O.**, **BENTAYEB F.** «Segmentation automatique des rapports médicaux en utilisant des réseaux de neurones convolutionnels». In proc. of *EDA 2018* p.17-30.
64. [Rhazlane-17] **RHAZLANE S.**, Elouazzani A., **HARBI N.**, **KABACHI N.**, Badir H. «Data Alteration: A Better Approach to Securing Cloud Data with Encryption». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA)*. Ref HAL: [hal-02055780_v1](#).
65. [Belkacem-17] Belkacem S., Boukhalfa K., **BOUSSAID O.** «Tri des actualités sociales: État de l'art et Pistes de recherche». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA)* p.85-100. Ref HAL: [hal-02054056_v1](#).
66. [Boussahoua-17a] **BOUSSAHOUA M.**, **BOUSSAID O.**, **BENTAYEB F.** «Optimisation des performances dans les entrepôts de données NoSQL orientés colonnes». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)* p.69-80. Ref HAL: [hal-02053993_v1](#).
67. [Roos-17a] **ROOS J.**, **BONNEVAY S.**, **GAVIN G.** «Prévision à court terme des flux de voyageurs du réseau ferré urbain : une approche par les réseaux bayésiens dynamiques». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2017)*. Ref HAL: [hal-01807063_v1](#).
68. [Favre-17b] **FAVRE C.**, **JAKAWAT W.**, **LOUDCHER S.** «Jeux de cubes pour les graphes... Ou comment des graphes enrichis par des cubes (GreC) peuvent contribuer à l'analyse de données textuelles ?». In proc. of *journées (ADOC - EDA)* p.159-160. Ref HAL: [halshs-01577053_v1](#).
69. [Beligne-17a] **BELIGNE M.**, Campar A., **CHAUCHAT J.-H.**, Lefeuvre M., Lefort I., **LOUDCHER S.**, **VELCIN J.** «Retour d'expérience sur la détection automatique de métaphores dans des textes de Géographie». In proc. of *journées (ADOC - EDA 2017)* p.159-160. Ref HAL: [halshs-01577051_v1](#).
70. [Beligne-17] **BELIGNE M.**, Campar A., **CHAUCHAT J.-H.**, Lefeuvre M., Lefort I., **LOUDCHER S.**, **VELCIN J.** «Détection automatique de métaphores dans des textes de Géographie : une étude prospective». In proc. of *Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2017)* p.86-93. Ref HAL: [halshs-01577050_v2](#).
71. [Favre-17a] **FAVRE C.**, **JAKAWAT W.**, **LOUDCHER S.** «Graphes enrichis par des Cubes (GreC) : une approche innovante pour l'OLAP sur des réseaux d'information». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 2017)* p.293-308. Ref HAL: [halshs-01577047_v1](#).
72. [Favre-17] **FAVRE C.** «Les données de la recherche vues au travers des lunettes du genre : quand l'informatique rencontre les sciences humaines et sociales pour rendre visible le non visible - Le cas de

- données de la recherche en informatique». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 2017)*. Ref HAL: [hal-01532778_v1](https://hal.archives-ouvertes.fr/hal-01532778).
73. [Wang-17a] **WANG X.**, **AH-PINE J.**, **DARMONT J.** «A New Test of Cluster Hypothesis Using a Scalable Similarity-Based Agglomerative Hierarchical Clustering Framework». In proc. of *Conférence en Recherche d'Information et Applications (CORIA 2017)*. Ref HAL: [hal-01504961_v1](https://hal.archives-ouvertes.fr/hal-01504961).
 74. [Ratsimbazafy-16] **RATSIMBAZAFY R.**, **BENTAYEB F.** «Nouvelle stratégie pour le traitement distribué des processus décisionnels massifs dans un Big Data Warehouse». In proc. of *EDA 2016* p.81-94.
 75. [Rhazlane-16a] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Toward Cloud Data Storage Security: State of the Art and a Multi Agent System Based Solution». In proc. of *Workshop International sur l'Innovation et Nouvelles Tendances dans les Systèmes d'Information (INTIS)*. Ref HAL: [hal-02055766_v1](https://hal.archives-ouvertes.fr/hal-02055766).
 76. [Rhazlane-16] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, Badir H. «Vers une solution de Protection des Données Entreposées dans la Cloud basée sur les Systèmes Multi Agents». In proc. of *Conférence Maghrébine sur les Avancées des Systèmes Décisionnels(ASD)*. Ref HAL: [hal-02055730_v1](https://hal.archives-ouvertes.fr/hal-02055730).
 77. [Khemiri-16] **KHEMIRI R.**, **BENTAYEB F.** «Recommandation de chemins de navigation dans un cube OLAP». In proc. of *EDA 2016* p.25-40.
 78. [Cabanac-16] Cabanac G., Hubert G., Tran H. D., **FAVRE C.**, Labbé C. «Un regard lexico-scientométrique sur le défi EGC 2016». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2016)*. Ref HAL: [hal-01567021_v1](https://hal.archives-ouvertes.fr/hal-01567021).
 79. [Ratsimbazafy-15] **RATSIMBAZAFY R.**, **BOUSSAID O.**, **BENTAYEB F.** «D-WorM : Middleware pour le traitement des requêtes massives dans les entrepôts de données massives». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2015)* p.193-196. Ref HAL: [hal-01493203_v1](https://hal.archives-ouvertes.fr/hal-01493203).
 80. [Oukid-15a] Oukid L., **BOUSSAID O.**, Benblidia N., **BENTAYEB F.** «TLabel: Nouvel opérateur d'agrégation par catégorisation dans les cubes de textes». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2015)* p.129-144. Ref HAL: [hal-01493198_v1](https://hal.archives-ouvertes.fr/hal-01493198).
 81. [Hannachi-15a] Hannachi L., **BOUSSAID O.**, Benblidia N., **BENTAYEB F.** «On-Line Analytical Processing on Graphs Generated from Social Network Data». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2015)* p.97-112. Ref HAL: [hal-01493169_v1](https://hal.archives-ouvertes.fr/hal-01493169).
 82. [Brice-15] Brice O., **FAVRE C.**, **LOUDCHER S.** «Analyse OLAP sur des tweets et des blogs : un retour d'expérience». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2015)*. Ref HAL: [halshs-01100008_v1](https://hal.archives-ouvertes.fr/halshs-01100008).
 83. [Dermouche-15a] **DERMOUCHE M.**, Khouas L., **LOUDCHER S.**, **VELCIN J.**, Fourboul E. «Analyse et visualisation d'opinions dans un cadre de veille sur le Web». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2015)*. Ref HAL: [halshs-01100005_v1](https://hal.archives-ouvertes.fr/halshs-01100005).
 84. [Dehdouh-14] **DEHDOUH K.**, **BENTAYEB F.**, **KABACHI N.**, **BOUSSAID O.** «C-CUBE: Un nouvel opérateur d'agrégation pour les entrepôts de données en colonnes». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID)* p.45-59. Ref HAL: [hal-01493249_v1](https://hal.archives-ouvertes.fr/hal-01493249).
 85. [Selmane-14] **SELMANE S.-A.**, **BENTAYEB F.**, **BOUSSAID O.** «Règles d'Association Triadiques pour la recommandation et l'enrichissement de requêtes décisionnelles». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014)* p.1-16. Ref HAL: [hal-01493208_v1](https://hal.archives-ouvertes.fr/hal-01493208).
 86. [Bala-14] Bala M., **BOUSSAID O.**, Alimazighi Z., **BENTAYEB F.** «PF-ETL : vers l'intégration de données massives dans les fonctionnalités d'ETL». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID)* p.61-76. Ref HAL: [hal-01493251_v1](https://hal.archives-ouvertes.fr/hal-01493251).
 87. [Pierrot-14a] **PIERROT D.**, **HARBI N.**, **DARMONT J.** «Détection des Intrusions, du monitoring des Systèmes d'Information au Graph Mining». In proc. of *Atelier International sur l'Innovation et Nouvelle Tendances dans les Systèmes d'Information (INTIS 2014)*. Ref HAL: [hal-01165641_v2](https://hal.archives-ouvertes.fr/hal-01165641).
 88. [Selmane-14d] **SELMANE S.-A.**, **BOUSSAID O.**, Missaoui R., **BENTAYEB F.** «Personalization for Three-mode Data In Social Networks». In proc. of *JS(JS)*.
 89. [Dehdouh-14b] **DEHDOUH K.**, **BENTAYEB F.**, **KABACHI N.**, **BOUSSAID O.** «L'opérateur CUBE pour les entrepôts de données NoSQL orientés colonnes». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014)* p.27-36. Ref HAL: [hal-01493211_v1](https://hal.archives-ouvertes.fr/hal-01493211).
 90. [Dehdouh-14a] **DEHDOUH K.**, **BENTAYEB F.**, **KABACHI N.**, **BOUSSAID O.** «C-CUBE: Un nouvel opérateur d'agrégation pour les entrepôts de données en colonnes». In proc. of *Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 2014)* p.45-59.
 91. [Selmane-14c] **SELMANE S.-A.**, **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)* p.1-16. Ref HAL: [hal-01018764_v1](https://hal.archives-ouvertes.fr/hal-01018764).

92. [Selmane-14a] **SELMANE S.-A.**, **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.374-378. Ref HAL: [hal-01018766_v1](#).
93. [Dehdouh-14f] **DEHDOUH K.**, **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)* p.525-528. Ref HAL: [hal-01493271_v1](#).
94. [Arres-14] **ARRES B.**, **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](#).
95. [Guille-14a] **GUILLE A.**, **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](#).
96. [Guille-14] **GUILLE A.**, **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](#).

Autres produits présentés dans des colloques / congrès et des séminaires de recherche (8)

1. [Rhzlane-17c] **RHAZLANE S.**, **HARBI N.**, **KABACHI N.**, **BADIR H.** «Implementation and Exploitation of Data Alteration Agent within a Multi-Agent System Architecture.». In proc. of *Workshop International sur l'Innovation et Nouvelles Tendances dans les Systèmes d'Information (INTIS 2017)*. Ref HAL: [hal-02055920_v1](#).
2. [Gavin-16] **GAVIN G.**, **SOBATI MOGHADAM S.** «A SWHE based on factorization.». In proc. of *International Conference on Cryptology (AfricaCrypt 2016)*.
3. [Darmont-15b] **DARMONT J.** «Collaborative BI for dummies.». ODBMS.org (2016). Ref HAL: [hal-01186041_v1](#).
4. [Arres-14c] **ARRES B.**, **KABACHI N.**, **BOUSSAID O.**, **BENTAYEB F.** «Intentional Data Placement Policy for Improving OLAP Cube Construction on Hadoop Clusters.». In proc. of *Conférence Base Données Avancées(BDA)* p.33-34. Ref HAL: [hal-01166222_v1](#).
5. [Bentayeb-14] **BENTAYEB F.**, **VELCIN J.**, **BONNEVAY S.**, **DARMONT J.** «Data Science and Decision Support at ERIC.». SIGMOD Record, 43(4), 37-42, December 2014. Ref HAL: [hal-01165623_v1](#).
6. [Grabova-14] **GRABOVA O.**, **SOBATI MOGHADAM S.**, Chagheri S., **DARMONT J.** «BI4people : le décisionnel pour tous.». In proc. of *10es journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014)* p.155-158. Ref HAL: [hal-00980915_v1](#).
7. [Selmane-14] **SELMANE S. A.**, **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)*.
8. [Darmont-14c] **DARMONT J.**, Pedersen T.B. «Report on the Second International Workshop on Cloud Intelligence (Cloud-I 2013)». SIGMOD Record, 43(1), 77-79, March 2014.

4- Produits et outils informatiques

Logiciels (6)

- 2 bancs d'essais décisionnels distribués, 2014 et 2018
- 1 plateforme de recherche d'information RICSCH : Recherche d'information contextuelle par segmentation thématique de documents), 2014
- 1 Librairie LibS4 (méthodes de crypto pour la sécurité des entrepôts dans le cloud)
- 1 application opérationnelle : D113, détection des intrusions regroupant deux méthodes de fouilles de données sur des journaux Pare-Feu.
- 1 jeu de données : RicHack, 1 jeu de données reposant des événements Pare-feu regroupant différents types d'attaques (4 giga de données).

5- Activités éditoriales

Participation à des comités éditoriaux (journaux scientifiques, revues, collections, etc.) (9)

Comité Editorial

- International Journal of Big Data Security Intelligence (IJBDSI), 2015-2018
- Bulletin of Mathematics, Informatics, Physics, 2010-2015
- International Journal of Data Mining, Modelling and Management (IJDMMM), depuis 2008
- International Journal of Biomedical Engineering and Technology (IJBET), depuis 2006

Comité de Rédaction

- Ingénierie des Systèmes d'Information, depuis 2012

Comité de Pilotage

- Conférence Extraction et de Gestion des Connaissances (EGC), depuis 2015
- Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA), depuis 2005
- Conférence sur les Avancées des Systèmes Décisionnels (ASD), depuis 2006
- ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), depuis 2018

6- Activités d'évaluation

Évaluation d'articles et d'ouvrages scientifiques (relecture d'articles / reviewing) (393)

- *Revues internationales (52)* : DATAK - Data & Knowledge Engineering, Future Internet, IJDSST - International Journal of Decision Support System Technology, IJDWM – International Journal of Data Warehousing and Mining, TDSC – IEEE Transactions on dependable and secure Computing, TKDE - IEEE Transactions on Knowledge and Data Engineering, Information Systems, etc.
- *Revues nationales (21)* : ISI - Revue Ingénierie des Systèmes d'Information, TSI - Techniques et Sciences Informatiques
- *Conférences internationales (200)* : ADBIS - East-European Conference on Advances in Databases and Information Systems, AICCSA - ACS/IEEE International Conference on Computer Systems and Applications, DAWAK - International Conference on Big Data Analytics and Knowledge Discovery, DEXA – International Conference on Database and expert Systems Applications, DOLAP - International Workshop On Design, Optimization, Languages and Analytical Processing of Big Data, etc.
- *Conférences nationales (120)* : ASD – Conférence sur les Avancées des Systèmes Décisionnels, BDA – Conférence sur la Gestion de Données – Principes, Technologies et Applications, COSI - Colloque sur l'Optimisation et les Systèmes d'Information, EDA – Conférence sur les Entrepôts de Données et l'Analyse en ligne, EGC – Conférence sur l'Extraction et la Gestion des connaissances, INFORSID - congrès INformatique des ORganisations et Systèmes d'Information et de Décision.

Évaluation de projets de recherche (16)

- ANR pour l'appel à projet Chaires industrielles Edition 2018, et l'appel à projets génériques en 2015
- ANRT pour plusieurs demandes de financement de thèses CIFRE
- COFECUB : Comité français d'évaluation de la coopération universitaire et scientifique avec le Brésil)
- CAPES : Coordination pour le perfectionnement du personnel de l'enseignement supérieur, 2017.

- CRSNG : Conseil de Recherche en Sciences Naturelles et en Génie, Evaluation d'une demande de subvention à la découverte, 2018
- Prix de l'Université Val-de-Marne, éditions 2017, 2018
- Comité technique projet PIA ISTEEX (invité), CNRS, 2018
- STIC-AmSud, 2018
- Fédération de Recherche Agorantic, 2017
- Présidence d'un comité scientifique de la Fondation Canadienne pour l'Innovation (FCI), 2017

Évaluation de laboratoires (1)

1. Évaluation des chercheurs et des ingénieurs Irstea, campagne d'évaluation 2018-2019

7- Contrats de recherche financés par des institutions publiques ou caritatives

Contrats avec les collectivités territoriales en tant que partenaire

1. 2017-2021 : AURA PMI (Transformation numérique, servicisation et mutations des modèles d'affaires des PME industrielles). Ce projet, financé par la Région Auvergne-Rhône-Alpes, vise à repenser les conditions dans lesquelles les PME peuvent adopter la digitalisation et servicisation, considérées ici comme des formes d'innovation disruptive, afin de lever les verrous scientifiques, techniques, managériaux et financiers qui les empêcheraient d'exploiter pleinement les effets de levier permis par ces processus. Collaboration avec le laboratoire COACTIS (gestion). Financement : 273000 €

Contrats financés dans le cadre du PIA en tant que partenaire

1. 2014-2017 : REQUEST (REcursive QUERy and Scalable Technology). Ce projet vise la création, structuration et l'animation de la communauté française « Big Data », ainsi que le traitement du problème de la dualité big data/big analytics. Il est porté par l'entreprise Thalès et financé par le programme d'investissements d'avenir (PIA). ERIC intervient sans la sous-tâche « Requêtage intelligent ». Financement : 191000 €

Autres Contrats (15)

1. 2018-2020 - Hyper Thesaurus et lacs de données : Fouiller la ville et ses archives. Projet interdisciplinaire financé par le Labex IMU visant à constituer une plateforme pour la pratique archéologique, par l'élargissement des référentiels de la communauté scientifique, la conception et le prototypage de modes opératoires d'indexation/archivage pérennes de données moissonnables et par l'inclusion d'amateurs qualifiés dans les opérations d'enrichissement de données partagées. Collaboration avec les laboratoires Archéorient, ArAr (Environnements et sociétés de l'Orient ancien), CESCO Paris, SAVL, Archeodunum, la FR Maison de l'Orient et de la Méditerranée, l'UMS PERSEE, Bibracte, l'Université autonome de Barcelone et les Musées d'archéologie de Catalogne. Financement : 19450 €.
2. 2017- 2019 : BigData SHS.0 - Les mégadonnées et les sciences sociales. Projet international financé par la région Auvergne-Rhône-Alpes. Soutien aux coopérations universitaires et scientifiques internationales (18 mois). Partenaires : Université Lyon 2 (porteur du projet) et Université de Bucarest. 17580 euros.
3. 2017-2019 : GéoNum (Détection automatique de métaphores dans le discours scientifique en Géographie). Ce projet PEPS financé par le CNRS vise à réaliser une étude numérique de l'ensemble des textes publiés dans les revues françaises de Géographie depuis 120 ans en vue d'une relecture épistémologique des savoirs géographiques. L'un des objectifs est de mieux comprendre l'utilisation de métaphores et notamment son évolution au cours du temps dans la production universitaire française en Géographie. Responsable du projet. Collaboration avec le laboratoire EVS. Financement: 12772 €.
4. 2016-2019. IDENUM : Identités numériques urbaines. Projet financé par le Labex IMU avec le laboratoire ELICO (Equipe de recherche de Lyon en sciences de l'Information et de la COmmunication, Lyon), le laboratoire LIRIS (Laboratoire d'informatique en image et systèmes d'information, Lyon) et le laboratoire EVS (laboratoire Environnement Ville Société, Lyon). Projet interdisciplinaire financé par le Labex IMU visant à décrire et caractériser des pratiques de documentation et d'archivage du territoire urbain par la photographie, en les reliant aux projets individuels et collectifs, privés et publics qui lui donnent statut et qualité. La recherche est focalisée sur des corpus complexes constitués par les documents photographiques produits et diffusés par des usagers de la ville ainsi que par des acteurs collectifs (collectivités territoriales, institutions patrimoniales, associations, réseau de transport public...) et collectés dans des espaces médiatiques

- semi-privés, les réseaux sociaux. Collaboration avec les laboratoires ELICO (information-communication), EVS (géographie) et LIRIS (informatique). Financement : 19600 €.
5. 2017-2019. Collaboration avec l'entreprise Sword dans le cadre d'une thèse CIFRE. Ce projet vise à identifier et caractériser les patients résistants aux traitements d'oncologie ». L'un des défis majeurs de la médecine de précision est d'orienter la recherche et le développement de solutions thérapeutiques spécifiques en exploitant les informations contenues dans les rapports médicaux. Ces rapports sont souvent maintenus sous forme de textes non structurés et constituent un volume important de données. Ce projet a pour objectif de d'exploiter ces rapports médicaux en utilisant des techniques d'extraction de connaissances, de machine learning et d'aide à la décision afin d'apporter une aide et une amélioration des soins de santé et faciliter la prise de décisions médicales. Financement : 30000 €.
 6. 2016-2019. Identification et caractérisation des trajectoires de l'information dans les médias sociaux. Contrat avec la société Bertin IT dans le cadre d'une thèse CIFRE.
 7. 2017-2018. CAPIMU : Capitalisation scientifique dans le labex IMU. Projet financé par le Labex IMU avec le laboratoire ELICO (Equipe de recherche de Lyon en sciences de l'Information et de la COmmunication, Lyon), le laboratoire LIRIS (Laboratoire d'informatique en image et systèmes d'information, Lyon) et le laboratoire EVS (laboratoire Environnement Ville Société, Lyon).
 8. 2017-2018 : COREL (Au cœur de la relation client). Projet interdisciplinaire financé par l'Université Lyon 2 et visant à mobiliser l'ensemble du corpus collecté par le Center for Customer Management pour répondre à différentes questions liées à la stratégie client des organisations (entreprises mais aussi services publics), afin de mieux comprendre l'élaboration et la mise en œuvre de la stratégie client au sein des organisations. Collaboration avec les laboratoires COACTIS et DRM-ERMES (gestion). Financement : 13000 €.
 9. 2016. Projet YETI (Youth Evidence To Immunize). L'objectif de ce projet est d'adapter une plateforme permettant de simuler le processus de diffusion, de propagation d'un virus respiratoire à transmission interhumaine directe, de la France à l'Angleterre. Il est réalisé en collaboration avec le Public Health England (PHE), le Royal College of General Practitioners (RCGP/RSC), l'Université du Surrey, l'EPHE-Paris 8, l'Institut Pasteur, l'Université d'Istanbul, Cyklad et Lyon 1.
 10. 2015-2016 : CERAM 3D (Numérisation, modélisation et exploitation d'images 3D d'objets archéologiques modèles). Projet mené en collaboration avec les laboratoires Archéologie et Archéométrie et Hubert Curien, financé par PALSE IPem et visant à créer de nouveaux outils de description et d'identification d'objets archéologiques (céramiques), basés sur des traitements d'images 3D, à des fins de formation, de recherche et de valorisation.
 11. 2015-2016 : TECTONIQ (TEchnologies de l'information et de la communication au Cœur du Territoire Numérique pour la valorisation du patrimoine). Projet interdisciplinaire PEPS CNRS en collaboration avec les laboratoires GERIICO (Lille), STL (Lille), TETIS (Montpellier) et LIVE (Strasbourg) visant à étudier les dispositifs numériques mis en place pour diffuser et échanger les informations relatives au patrimoine naturel et culturel du territoire, ainsi que leurs usages par les différents acteurs impliqués (citoyens, entreprises, scientifiques, collectivités, etc.). Il se décompose en deux axes. Le premier consiste à dresser un état du patrimoine valorisé dans les contenus numériques. Le second axe s'attache à analyser les usages de ces informations patrimoniales par les différents acteurs. Financement : 1500 €.
 12. 2013-2015. Géographie numérique : pour une analyse épistémologique de la Géographie par un traitement automatique des discours de géographes. Projet avec le laboratoire Environnement Ville Société (UMR EVS). Porteur du projet. Financement Université Lyon 2.
 13. 2013-2015. Projet RCCSO - Réseaux de Chercheurs et Contenus Sociaux Ouverts. Partenaires : CentreCeReFreA6 (Centre Régional Francophone de Recherches Avancées en Sciences Sociales), Université de Bucarest (Roumanie), Laboratoire ERIC et l'Institut de la Communication de Lyon2. Financement par l'AUF. Résumé : Il s'agit d'étudier les interactions entre les chercheurs en sciences humaines et sociales. L'un des objectifs est de créer un réseau social collaboratif dédié à la recherche en SHS, puis de détecter des communautés de recherche, d'identifier les thèmes récurrents et les thèmes émergents, etc. Partenaire du projet. Financement AUF.
 14. 2012-2015. Collaboration avec la société Go Albert (Montpellier), thèse CIFRE.
 15. 2010-2014. Projet Tassili, France-Algérie, (EGIDE7). Dans ce projet, on s'est intéressé à l'analyse en ligne de données textuelles et des réseaux sociaux. Il a permis de financer plusieurs séjours scientifiques en France (laboratoire ERIC) de jeunes chercheurs de l'Université partenaire Saad Dahleb de Blida (Algérie). Porteur du projet. Financement : 38 000 euros.

8- Post-doctorants et chercheurs accueillis

Chercheurs accueillis (14)

Dont Julien Aligon, MCF à l'Université de Toulouse (1 mois).

Chercheurs étrangers accueillis (13)

Dont Robert Wrembel, MCF à l'Université Polytechnique de Poznan, Pologne (1 mois) et 10 doctorant-es.

9- Indices de reconnaissance

Responsabilités dans des sociétés savantes

- Membre du bureau de l'association INFORSID, depuis 2018

Organisations de colloques / congrès à l'étranger (9)

- ADBIS-Workshops 2018, 2019 : European Conference on Advances in Databases and Information Systems Workshops (Hongrie en 2018 et Slovénie en 2019)
- AICCSA 2017, 2018 : ACS/IEEE International Conference on Computer Systems and Applications (Tunisie en 2017, Jordanie en 2018)
- EDA 2018 : Journées francophones sur les Entrepôts de Données et l'Analyse en Ligne, Business Intelligence & Big Data (Maroc).
- AICCSA-WorkShop 2016, 2017 (Maroc en 2016, Tunisie en 2017)
- ASD 2015 : Conférence maghrébine sur les Avancées des Systèmes Décisionnels (Maroc)

Invitations à des colloques / congrès à l'étranger (11)

- Conference on Computing Systems and Applications, Alger, Algérie, 2018
- Colloque international sur les sciences sociales et la société, Bucarest, Roumanie, 2017
- Ecole de printemps « Big Data Analytics and Stream Processing for IoT", Tabarka, Tunisia, 2017
- Colloque GEDI (Genre, égalité de droit et inégalités de fait), intervention intitulée « Les quotas pour l'égalité en questions : des chiffres, des débats, des actes et des vécus ! - Retour sur une enquête préliminaire dans l'enseignement supérieur et la recherche dans le domaine de l'informatique », 2017
- International Conference on Business Intelligence and Applications, ICBA, Blida, Algérie, 2016
- International Symposium on Networks, Computers and Communications (ISNCC), Tunisie, 2016
- International Conference on Networking and Advanced Systems (ICNAS), Annaba, Algérie, 2015
- International Workshop on Artificial Intelligence and Information and Communication Technologies (IWAICT), Biskra, Algérie, novembre 2015
- Centre de Recherche sur l'information scientifique et Technique (CERIST), Alger, Algérie, novembre 2014
- Conférence sur les Avancées des Systèmes Décisionnels (ASD), Hammamet, Tunisie, 2014

Séjours dans des laboratoires étrangers (18)

- Université USTO d'Oran – Algérie
- Université de Annaba – Algérie
- Université Saad Dahleb, Blida – Algérie
- Université de Québec à Montréal - Canada
- Barcelona Tech, Database Technologies and Information Management Group - Espagne
- Université Tanger - Maroc
- Université de Tétouane - Maroc
- Université de Bucarest, Roumanie
- Université de Genève - Suisse
- Université de Can Tho - Vietnam
- Université de Ho Chi Minh – Vietnam
- IFI (Institut Francophone d'Informatique) à Hanoi - Vietnam
- Université de Heilbronn (Institut für Gesundheitsökonomie und Medizinische Informatik)
- Université de Heidelberg (Institute for medical biometry and informatics)
- Université du Québec à Montréal, Institut de Recherche en Etudes Féministes – Canada
- Université de Genève – Suisse

II - INTÉRACTION DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME AVEC L'ENVIRONNEMENT NON ACADÉMIQUE, IMPACTS SUR L'ÉCONOMIE, LA SOCIÉTÉ, LA CULTURE, LA SANTÉ

1- Interactions avec les acteurs socio-économiques

Contrats de R&D avec des industriels. L'équipe SID a construit un tissu de relations, en particulier dans les domaines de la business intelligence, du big data, des réseaux sociaux et des humanités numériques principaux domaines d'exploitation des travaux de l'équipe. Ces collaborations prennent principalement la forme de thèses CIFRE.

Bourses CIFRE (4)

- Depuis 2018 : [BialX](#) Lyon.
- 2018 : [Wizacha](#) Lyon (abandon)
- Depuis 2017 : [SWORD](#) Lyon.
- Depuis 2016 : [Bertin IT](#) Montpellier.
- 2012-2015 : [Go Albert](#) Montpellier.

Expertises auprès de l'ANRT (3). L'ANRT sollicite de façon régulière les membres de l'équipe SID pour des travaux d'expertise concernant les demandes de thèses CIFRE.

2- Produits destinés au grand public

Produits de vulgarisation : articles, interviews, éditions, vidéos, produits de médiation scientifique, débats science et société, etc.

1. Présentation vulgarisée des entrepôts de données dans l'ouvrage Les Big Data à découvert (CNRS éditions), dont un membre d'ERIC est coauteur (2017)
2. Organisation par deux membres d'ERIC, dans le cadre du « off » de la Web Conference 2018 à Lyon, des rencontres grand public « L'informatique, pourquoi pas moi ? Les jeunes prennent la parole ! » et « L'informatique, pourquoi pas elles ? »

III - IMPLICATION DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME DANS LA FORMATION PAR LA RECHERCHE

1- Productions scientifiques (articles, ouvrages, etc.) issues des thèses

Productions scientifiques (articles, ouvrages, etc.) issus des thèses (88)

Les publications de l'équipe issus des travaux de thèses sont celles dont l'un des auteurs est un doctorant et dont le nom est mis en gras (partie I, sections 1, 2 et 3).

Nombre moyen d'articles par doctorant (3,7)

2- Formation

Nombre de personnes Habilitées à Diriger des Recherches (HDR) : 4

Nombre d'HDR soutenues : 0

Doctorants (nombre total) : 23

Doctorants bénéficiant d'un contrat spécifique au doctorat : 22

Nombre de thèses soutenues : 12

Durée moyenne des thèses : 52 mois

Stagiaires accueillis (M1, M2) : 50

Personnes responsables d'une mention ou d'un parcours de master (nombre total) : 7

Personnes responsables d'une mention ou d'un parcours de master à labellisation internationale (Erasmus Mundus, par ex.) : 2

ANNEXE 4 - Produits et activités de la recherche

En cohérence avec les données chiffrées de l'onglet 4 du fichier Excel « Données du contrat en cours », on remplira ce document destiné à l'évaluation du critère 1 du référentiel de l'évaluation « Produits et activités de la recherche », pour l'ensemble de l'unité et pour chaque équipe / thème.

CAMPAGNE D'ÉVALUATION 2019-2020 VAGUE A

Nom du thème : Axe transversal Humanités numériques

Acronyme : HN

Directeur de l'unité pour le contrat en cours : Jérôme Darmont

Directeur de l'unité thème pour le contrat à venir : Julien Jacques

I- PRODUCTION DE CONNAISSANCES ET ACTIVITÉS CONCOURANT AU RAYONNEMENT ET À L'ATTRACTIVITÉ SCIENTIFIQUE DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME

Les 20 % les plus significatifs de la production scientifique sont indiqués par une référence en gras et séparés des 80 % restants. Les membres du laboratoire sont indiqués en majuscules soulignées et les doctorant·es de plus distingués en gras.

1- Journaux / Revues

Articles scientifiques (10)

1. **[Beligne-19a] BELIGNE M., LOUDCHER S.,** Lefort I. «Potentialités et difficultés d'une relecture épistémologique d'une discipline de SHS par les outils des Humanités numériques et un corpus de revues françaises». In *Revue Humanités numériques* (2019). À paraître. Ref HAL: [halshs-01851673_v1](https://halshs.archives-ouvertes.fr/halshs-01851673).
2. **[Boyadjian-17] Boyadjian J., VELCIN J.,** «L'analyse quantitative des médias sociaux, une alternative aux enquêtes déclaratives ?». In *Questions de communication* (2017) p.111-135. DOI: [10.4000/questionsdecommunication.11078](https://doi.org/10.4000/questionsdecommunication.11078).
3. [Beligne-18] **BELIGNE M., LOUDCHER S.,** Lefort I. «Mise en perspective d'une variation du varia». In *Géocarrefour - Revue de géographie de Lyon* (2018). Ref HAL: [halshs-01851673_v1](https://halshs.archives-ouvertes.fr/halshs-01851673).
4. [Bouakkaz-17] Bouakkaz M., Ouinten Y., **LOUDCHER S.,** Fournier-Viger P. «Efficiently mining frequent itemsets applied for textual aggregation». In *Applied Intelligence* (2017). DOI: [10.1007/s10489-017-1050-9](https://doi.org/10.1007/s10489-017-1050-9).
5. [Bouakkaz-16] Bouakkaz M., **LOUDCHER S.,** Ouinten Y. «OLAP Textual Aggregation Approach using the Google Similarity Distance». In *International Journal of Business Intelligence and Data Mining*, vol. 11/1 (2016) p.31-48. Ref HAL: [halshs-01231490_v1](https://halshs.archives-ouvertes.fr/halshs-01231490). Rang C.
6. [Jakawat-16a] **JAKAWAT W., FAVRE C., LOUDCHER S.,** «Graphs enriched by cubes for OLAP on bibliographic networks». In *International Journal of Business Intelligence and Data Mining*, vol. 11/1 (2016) p.85-107. DOI: [10.1504/IJBIDM.2016.076435](https://doi.org/10.1504/IJBIDM.2016.076435). Rang B.
7. [Oukid-16] Oukid L., Benblidia N., **BOUSSAID O., BENTAYEB F.,** «(Label : A new OLAP aggregation operator in text cubes». In *International Journal of Data Warehousing and Mining*, vol. 12/4 (2016) 54-74. Ref HAL: [hal-02054192_v1](https://hal.archives-ouvertes.fr/hal-02054192). Rang B.
8. [Boussaid-15] Oukid L., Benblidia N., **ASFARI O., BENTAYEB F., BOUSSAID O.,** «Contextualized Text OLAP Based on Information Retrieval». In *International Journal of Data Warehousing and Mining*, vol. 11/2 (2015) p.1-21. Ref HAL: [hal-01484437_v1](https://hal.archives-ouvertes.fr/hal-01484437). Rang B.
9. [Guille-15] **GUILLE A., FAVRE C.,** «Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach». In *Social Network Analysis and Mining*, vol. 5/1 (2015) 18:1-18:18. DOI: [10.1007/s13278-015-0258-0](https://doi.org/10.1007/s13278-015-0258-0). Rang B.
10. [Velcin-14a] **VELCIN J.,** Peradotto A., Khouas L., Cossu J.V., Dormagen J.Y., Brun C. «Analyser l'image de marque d'entités sur le Web : revue du projet ImagiWeb». In *Ingénierie des Systèmes d'Information (ISI)*, numéro spécial sur le "big data", vol.19(3) (2014) p.159-162.

Articles de synthèse / revues bibliographiques (2)

1. **[Loudcher-15] LOUDCHER S., JAKAWAT W., FAVRE C.,** «Combining OLAP and Information Networks for Bibliographic Data Analysis: A Survey». In *Scientometrics* (2015) p.20. Ref HAL: [halshs-01136559_v1](https://halshs.archives-ouvertes.fr/halshs-01136559). Rang A.
2. [Bouakkaz-17a] Bouakkaz M., Ouinten Y., **LOUDCHER S.,** Strelakova Y. «Textual aggregation approaches in OLAP context: A survey». In *International Journal of Information Management*, vol. 37/6 (2017) p.684-69. DOI: [10.1016/j.ijinfomgt.2017.06.005](https://doi.org/10.1016/j.ijinfomgt.2017.06.005).

2- Ouvrages

Direction et coordination d'ouvrages scientifiques / édition scientifique (2)

1. [Akoka19] Akoka J., **DARMONT J.,** Du Mouza C. «Systèmes d'information et Humanités numériques». In *Ingénierie des systèmes d'Information* (2019). Numéro spécial à paraître.

- [Olivesi-17] Olivesi A., Boyadjian J., VELCIN J. «Le web politique au prisme de la science des données». In *Réseaux*, vol. 204/4 (2017). DOI: [10.3917/res.204.0009](https://doi.org/10.3917/res.204.0009).

3- Production dans des colloques / congrès, séminaires de recherche

Éditions d'actes de colloques / congrès (2)

- [Darmont-19] DARMONT J., Grabar N., Teste O. «About Variety in Humanities Big Data». *Recherche d'information, document et web sémantique*, vol. 19/1 (2019). Ref HAL: [hal-02049343 v1](https://hal.archives-ouvertes.fr/hal-02049343).
- [Favre-17] FAVRE C. «Les données de la recherche vues au travers des lunettes du genre : quand l'informatique rencontre les sciences humaines et sociales pour rendre visible le non visible - Le cas de données de la recherche en informatique». Actes du 1^{er} atelier Valorisation et Analyse des Données de la Recherche (VADOR 2017). In *proc. of Congrès Informatique des organisations et systèmes d'information et de décision (INFORSID 2017)* p.58-66. Ref HAL: [hal-01532778 v1](https://hal.archives-ouvertes.fr/hal-01532778).

Articles publiés dans des actes de colloques / congrès (29)

- [Favre-18] FAVRE C., Tain L. «Les quotas : levier ou frein au déroulement des carrières des femmes ? Analyse suite à une enquête préliminaire dans le cas de l'enseignement supérieur et la recherche en France dans le domaine de l'informatique». In *proc. 6th International Interdisciplinary Conference of Political Research (SCOPE 2018)* p.37-54.
- [Beligne-17] BELIGNE M., Campar A., CHAUCHAT J.-H., Lefeuvre M., Lefort I., LOUDCHER S., VELCIN J. «Détection automatique de métaphores dans des textes de Géographie : une étude prospective». In *proc. of Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2017)* p.86-93. Ref HAL: [halshs-01577050 v2](https://halshs.archives-ouvertes.fr/halshs-01577050).
- [Wang-17c] WANG X., AH-PINE J., DARMONT J. «SHCoClust, a Scalable Similarity-based Hierarchical Clustering Method and its Application to Textual Collections». In *proc. of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2017)* ID F-0529. Ref HAL: [hal-01504986 v1](https://hal.archives-ouvertes.fr/hal-01504986). Rang A.
- [Truica-16] Truica C.-O., GUILLE A., Gauthier M. «CATS: Collection and Analysis of Tweets Made Simple». In *proc. of ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW Companion)* p.41-44. DOI: [10.1145/2818052.2874320](https://doi.org/10.1145/2818052.2874320). Rang A.
- [Ozturk-15] OZTÜRK A., Eyango L., Waksman S. Y., LALLICH S., DARMONT J. «Warehousing Complex Archaeological Objects». In *proc. of International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2015)* p.226-239. Ref Arxiv: [1608.06469](https://arxiv.org/abs/1608.06469). Rang C.
- [Velcin-14] VELCIN J., 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 C.
- [Sawadogo-19] SAWADOGO P., Kibata T., DARMONT J. «Metadata Management for Textual Documents in Data Lakes». In *proc. of International Conference on Enterprise Information Systems (ICEIS 2019)*. To appear. Ref HAL: [hal-02012092 v1](https://hal.archives-ouvertes.fr/hal-02012092). Rang C.
- [HuyghuesDespointes-18] HUYGHUES-DESPOINTES C., VELCIN J., Khouas L., LOUDCHER S. «Extraction de chaînes cohérentes en vue de reconstituer la trajectoire de l'information». In *proc. of Conférence Extraction des Connaissances à partir des Données (EGC 2018)*. Ref HAL: [hal-01674547 v1](https://hal.archives-ouvertes.fr/hal-01674547).
- [Nogueira-18] Nogueira I., Romdhane M., DARMONT J. «Modélisation des métadonnées d'un data lake en data vault». In *proc. of Conférence Extraction des Connaissances à partir des Données (EGC 2018)* p.257-262. Ref HAL: [hal-01665547 v1](https://hal.archives-ouvertes.fr/hal-01665547).
- [Nogueira-18a] Nogueira I., Romdhane M., DARMONT J. «Modeling Data Lake Metadata with a Data Vault». In *proc. of International Database Engineering and Application Symposium (IDEAS 2018)* p.253-261. Ref HAL: [hal-01788036 v1](https://hal.archives-ouvertes.fr/hal-01788036). Rang B.
- [Ozturk-18] OZTÜRK A., LALLICH S., DARMONT J. «A Visual Quality Index for Fuzzy C-Means». In *proc. of International Conference on Artificial Intelligence Applications and Innovations (IAI 2018)* p.546-555. Ref HAL: [hal-01727426 v1](https://hal.archives-ouvertes.fr/hal-01727426).
- [Ozturk-18a] OZTÜRK A., LALLICH S., DARMONT J., Waksman S. Y. «MaxMin Linear Initialization for Fuzzy C-Means». In *proc. of International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2018)* p.1-15. Ref HAL: [hal-01771204 v1](https://hal.archives-ouvertes.fr/hal-01771204).

13. [Beligne-17a] **BELIGNE M.**, Campar A., CHAUCHAT J.-H., Lefeuvre M., Lefort I., LOUDCHER S., VELCIN J. «Retour d'expérience sur la détection automatique de métaphores dans des textes de Géographie». In proc. of *journées(ADOC - EDA 2017)* p.159-160. Ref HAL: [halshs-01577051 v1](https://halshs.archives-ouvertes.fr/halshs-01577051).
14. [SorianoMorales-17] **SORIANO MORALES E. P.**, AH-PINE J., LOUDCHER S., LOUDCHER S. «Fusion Techniques for Named Entity Recognition and Word Sense Induction and Disambiguation.». In proc. of *International Conference Discovery Science (DS 2017)* p.340-355. Ref HAL: [halshs-01577043 v1](https://halshs.archives-ouvertes.fr/halshs-01577043).
15. [AhPine-16a] AH-PINE J., **SORIANO-MORALES E.-P.** «A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis». In proc. of *Workshop on Interactions between Data Mining and Natural Language Processing (DMNLP 2016)*. Ref HAL: [hal-01504684 v1](https://hal.archives-ouvertes.fr/hal-01504684).
16. [Bouakkaz-16a] Bouakkaz M., LOUDCHER S., Ouinten Y. «A New Tool for Textual Aggregation in OLAP Context». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2016)*. Ref HAL: [halshs-01384654 v1](https://halshs.archives-ouvertes.fr/halshs-01384654). Rang C.
17. [Guille-16] GUILLE A., **SORIANO-MORALES E.-P.** «TOM: A library for topic modeling and browsing». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC 2016)*. Ref HAL: [hal-01442868 v1](https://hal.archives-ouvertes.fr/hal-01442868).
18. [Jakawat-16] **JAKAWAT W.**, FAVRE C., LOUDCHER S. «OLAP Cube-based Graph Approach for Bibliographic Data». In proc. of *International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2016)* p.87-99. Ref HAL: [halshs-01231492 v1](https://halshs.archives-ouvertes.fr/halshs-01231492). Rang B.
19. [SorianoMorales-16] **SORIANO-MORALES E.-P.**, AH-PINE J., LOUDCHER S. «Using a Heterogeneous Linguistic Network for Word Sense Induction and Disambiguation». In proc. of *Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2016)*. Ref HAL: [halshs-01384657 v1](https://halshs.archives-ouvertes.fr/halshs-01384657). Rang B.
20. [SorianoMorales-16a] **SORIANO MORALES E. P.**, AH-PINE J., LOUDCHER S. «Hypergraph Modelization of a Syntactically Annotated English Wikipedia Dump». In proc. of *Language Resources and Evaluation Conference (LREC 2016)*. Ref HAL: [halshs-01384647 v1](https://halshs.archives-ouvertes.fr/halshs-01384647). Rang C.
21. [Truica-16a] TRUCA C.-O., DARMONT J., VELCIN J. «A Scalable Document-based Architecture for Text Analysis». In proc. of *International Conference on Advanced Data Mining and Applications (ADMA 2016)* p.481-494. Ref Arxiv: [1612.06195](https://arxiv.org/abs/1612.06195). Rang B.
22. [Bentayeb-15d] Oukid L., BOUSSAID O., Benblidia N., BENTAYEB F. «TLabel: Nouvel opérateur d'agrégation par catégorisation dans les cubes de textes». In proc. of *Journées sur les Entrepôts de Données et l'Analyse en ligne (EDA 2015)* p.129-144. Ref HAL: [hal-01493198 v1](https://hal.archives-ouvertes.fr/hal-01493198).
23. [Bouakkaz-15] Bouakkaz M., LOUDCHER S., Ouinten Y. «GOTA: Using the Google Similarity Distance for OLAP Textual Aggregation». In proc. of *International Conference on Enterprise Information Systems (ICEIS 2015)* p.121-127. Ref HAL: [halshs-01136581 v1](https://halshs.archives-ouvertes.fr/halshs-01136581). Rang C.
24. [Dermouche-15] **DERMOUCHE M.**, Kouas L., VELCIN J., LOUDCHER S. «A Joint Model for Topic-Sentiment Modeling from Text». In proc. of *Annual ACM Symposium on Applied Computing (SAC 2015)*. Ref HAL: [halshs-01099984 v1](https://halshs.archives-ouvertes.fr/halshs-01099984). Rang B.
25. [Dermouche-15a] **DERMOUCHE M.**, Kouas L., LOUDCHER S., VELCIN J., Fourboul E. «Analyse et visualisation d'opinions dans un cadre de veille sur le Web». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref HAL: [halshs-01100005 v1](https://halshs.archives-ouvertes.fr/halshs-01100005).
26. [Olivier-15] Olivier B., FAVRE C., LOUDCHER S. «Analyse OLAP sur des tweets et des blogs : un retour d'expérience». In proc. of *Conférence Extraction des Connaissances à partir des Données (EGC)*. Ref HAL: [halshs-01100008 v1](https://halshs.archives-ouvertes.fr/halshs-01100008).
27. [Arres-14] **ARRES B.**, 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](https://hal.archives-ouvertes.fr/hal-00941510).
28. [Bouakkaz-14] Bouakkaz M., LOUDCHER S., Ouinten Y. «Automatic Textual Aggregation Approach of Scientific Articles in OLAP context». In proc. of *International Conference on Innovations in Information Technology (Innovations 2014)* p.30-35. Ref HAL: [halshs-01099997 v1](https://halshs.archives-ouvertes.fr/halshs-01099997).
29. [Dermouche-14] **DERMOUCHE M.**, VELCIN J., LOUDCHER S., Kouas L. «A Joint Model for Topic-Sentiment Evolution over Time». In proc. of *IEEE International Conference on Data Mining (ICDM 2014)*. Ref HAL: [halshs-01099990 v1](https://halshs.archives-ouvertes.fr/halshs-01099990). Rang A.

Autres produits présentés dans des colloques / congrès et des séminaires de recherche (7)

1. [**Favre-17d**] FAVRE C. «Quotas in higher education and research: leverage or brake for gender equality ? Study in the field of computer science in France». Communication orale, *3th Women's Worlds & 11th Fazenda Gênero*, Florianopolis (Brazil).

2. [Beligne-19] **BELIGNE M., LOUDCHER S.**, Lefort I. «Transformation de documents numérisés par la plateforme Persée dans un objectifs d'analyses textuelles diachroniques». *Colloque international : Histoire, Langues et Textométrie*. Ref HAL: [halshs-01851656 v1](https://halshs.archives-ouvertes.fr/halshs-01851656).
3. [Beligne-18] **BELIGNE M., LOUDCHER S.**, Lefort I. «Analyse critique d'un changement de paradigme en géographie à partir de réflexion entre discontinuités scientifiques, lexicales et sémantiques». *Colloque Penser avec les discontinués en géographie*. Ref HAL: [halshs-01851647 v1](https://halshs.archives-ouvertes.fr/halshs-01851647).
4. [Favre-17e] **FAVRE C.** «Les statistiques sexuées comme miroirs des inégalités de genre : apports et limites de leur production en démocratie». Communication orale, *5th International Interdisciplinary Conference of Political Research (SCOPE 2017)*, Bucarest (Roumanie).
5. [Favre-17f] **FAVRE C.**, Cabanac G., Hubert G., Labbé C. «Du bon usage de l'interdisciplinarité pour l'analyse de données sexuées : le cas des données bibliographiques de la conférence EGC». Communication orale, *6ème édition des journées "Big Data Mining and Visualization" - Regards croisés sur les data*, Lille.
6. [Favre-16] **FAVRE C.** «Femmes et recherche en informatique : d'une analyse sexuée d'une communauté scientifique aux questions de genre». Communication orale, *20e congrès international des sociologues en langue française (AISLF 2016)*, Montréal.
7. [Khouas15] Khouas L., Brun C., Peradotto A., Cossu J.V., Boyadjian J., **VELCIN J.** «Étude de l'image de marque d'entités dans le cadre d'une plateforme de veille sur le Web social». In proc. *22ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)*. Démo.

Séminaires croisés ELICO-ERIC (8) :

- 04/03/2016
Aurélie Olivesi (ELICO) : L'analyse du discours médiatique
Julien Jacques (ERIC) : Introduction au data mining
- 29/04/2016
Eva-Marie Goepfert (ELICO) : Le discours et ses objets
Sabine Loudcher (ERIC) : Tout ce que vous avez toujours voulu savoir sur les entrepôts de données et l'analyse en ligne sans jamais avoir osé le demander
- 09/12/2016
Chérifa Boukacem-Zhegmouri (ELICO) : Ce que Google et les plateformes numériques font à la recherche scientifique
Jérôme Darmont (ERIC) : Le décisionnel pour tout es
- 10/02/2017
Sarah Cordonnier (ELICO) : Données composites, corpus hétérogènes et créativité méthodologique
Julien Velcin (ERIC) : Bref aperçu sur la fouille de textes et quelques travaux réalisés au laboratoire ERIC

4- Produits et outils informatiques

Logiciels (7)

1. **NewsBrowsers** : Logiciel d'annotation et d'analyse de documents courts (ex. articles de presse)
2. CATS : collecte et analyse faciles de tweets
3. CommentWatcher : extraction automatique de commentaires postés sur les forums de discussion
4. MABED : détection et suivi d'événements dans les médias sociaux
5. RICSH : recherche d'information contextuelle par segmentation thématique de documents
6. SONDY : fouille de données dans les médias sociaux
7. TOM : librairie Python pour le Topic MOdeling

Corpus (1)

1. [Données pseudonymisées du projet ImagiWeb](#), disponibles sur demande mais soumises à charte d'utilisation.

5- Activités éditoriales

Participation à des comités éditoriaux (journaux scientifiques, revues, collections, etc.) (1)

1. Comité scientifique de la revue Management & Data Science

6- Activités d'évaluation

Évaluation d'articles et d'ouvrages scientifiques (relecture d'articles / reviewing) (2)

1. Numéro spécial Humanités numériques et systèmes d'information de la revue ISI, 2019
2. Comité scientifique du groupe de travail DAHLIA adossée à la conférence EGC, 2019

Responsabilités au sein d'instances d'évaluation (2)

1. Présidence du comité scientifique Humanités numériques de la Fondation Canadienne pour l'Innovation (FCI), 2017
2. Conseil scientifique du numérique de l'académie de Lyon, 2019

7- Contrats de recherche financés par des institutions publiques ou caritatives

Autres contrats européens en tant que partenaire (1)

1. PASSAGE (Professionalisation Aux SavoirS Autour du Genre et de l'Egalité ; 2014-2017). présente des aspects à la fois pédagogiques en lien avec le réseau du master EGALES Lyon 2 et scientifiques sur l'analyse des métiers du genre et de l'égalité, leur reconnaissance. Financement : ERASMUS+ Partenariats stratégiques.

Contrats nationaux (ANR, PHRC, FUI, INCA, etc.) en tant que porteur (1)

1. Imagiweb (2012-2015) vise à analyser de la dynamique des images d'entités (hommes politiques, célébrités, entreprises...) sur le Web. Partenaires : AMI Software, CEPEL (science politique), EDF, LIA (informatique), XEROX.

Contrats avec les collectivités territoriales en tant que porteur (1)

1. Élaboration, mise en œuvre et exploitation d'un modèle de description d'objets archéologiques complexes (2013-2017). Thèse. Partenaire : ArAr (archéologie). Financement : Région Rhône-Alpes.

Contrats avec les collectivités territoriales en tant que partenaire (3)

1. Pour une relecture épistémologique des savoirs géographiques : Analyses numériques et expériences d'acteurs des revues françaises de géographie de 1892 à 2014 (2016-2019). Thèse. Partenaire : EVS (géographie). Financement : Région Auvergne-Rhône-Alpes.
2. AURA PMI (2017-2021) : Transformation numérique, servicisation et mutations des modèles d'affaires des PME industrielles. Inclut une thèse à ERIC. Partenaire : COACTIS (gestion). Financement : Région Auvergne-Rhône-Alpes.
3. BigData SHS.0 – Les mégadonnées et les sciences sociales (2017-2019) : Soutien aux coopérations universitaires et scientifiques internationales Partenaires : laboratoire ECP (SHS pluridisciplinaire), Université de Bucarest. Financement : Région Auvergne-Rhône-Alpes.

Contrats financés par des associations caritatives et des fondations (ARC, FMR, FRM, etc.) en tant que partenaire (1)

1. Territoires d'excellence LEVIER (Leviers pour Expérimenter Vivre Innover l'Egalité Réelle. D'un diagnostic formel à l'égalité réelle dans l'entreprise ; 2013-2015). Ce projet porte sur une recherche-action au sein d'entreprises. Notre contribution concernait des aspects relevant de l'analyse des bilans chiffrés des entreprises. Financement : Fond Social Européen.

Autres contrats en tant que porteur (2)

1. Géonum (2017-2019) a pour objectif la détection automatique de métaphores dans le discours scientifique en Géographie et montre comment des méthodes et outils numériques permettent une relecture épistémologique d'une science sociale. Partenaires : EVS (géographie), MSH-LSE, UMS Persée. Financement : PEPS CNRS.
2. #néo (2017-2018) vise la mise au point d'un système permettant la détection et l'interprétation des néologismes sémantiques et morphologiques à partir d'un flux textuel, basé sur l'apprentissage de représentations vectorielles des mots. Partenaires : CRTT (linguistique), LIUM (informatique, Le Mans) et IULA (linguistique, Barcelone). Financement : Lyon 2.

Autres contrats en tant que partenaire (14)

1. Enquête exploratoire sur l'innovation culinaire : définitions, parcours et réseaux d'acteurs (2013-2014). Partenaire : CMW (sociologie). Financement : ISH et IXXI.
2. Géographie numérique (2013-2015) a pour objectif de construire un corpus de documents décrivant la géographie en tant que discipline, puis de l'analyser automatiquement par fouille de textes pour mettre en évidence les pratiques et l'évolution de la discipline. Partenaire : EVS (géographie). Financement : BQR Lyon 2.
3. Infonomics Resource Facility (2013-2015) vise à 1) mieux caractériser et comprendre les lois qui gouvernent l'information comme un objet à part entière ; 2) concevoir, déployer et mettre en œuvre de nouvelles méthodologies capables de mieux appréhender les phénomènes complexes liés aux objets de recherche des SHS. Partenaires : EVS (géographie), LaHC (informatique). Financement : PALSE.
4. RCCSO – Réseaux de Chercheurs et Contenus Sociaux Ouverts (2013-2015) : étude des interactions entre les chercheurs en SHS, création d'un réseau social collaboratif, détection de communautés de recherche, identification de thèmes récurrents et émergents. Partenaires : laboratoires ECP (SHS pluridisciplinaire) et CentreCeReFre (Roumanie). Financement : AUF.
5. Pour une épistémologie numérique de la géographie (2014-2016). Partenaires : EVS (géographie), Habiter (Reims). Financement : ISH et IXXI
6. TECTONIQ (2015) vise à étudier les dispositifs numériques mis en place pour diffuser et échanger les informations relatives au patrimoine naturel et culturel du territoire, ainsi que leurs usages par les différents acteurs impliqués (citoyens, entreprises, scientifiques, collectivités, etc.). Partenaires : GERiCO (information-communication), STL (linguistique), TETIS (informatique), LIVE (géographie). Financement : PEPS CNRS.
7. CERAM 3D (2015-2016) : Numérisation, modélisation et exploitation d'images 3D d'objets archéologiques modèles (céramiques), à des fins de formation, de recherche et de valorisation. Partenaires : laboratoires LaHC (informatique) et ArAr (archéologie). Financement : PALSE IPEm.
8. IDENUM (2015-2017) vise à décrire et caractériser des pratiques de documentation et d'archivage du territoire urbain par la photographie, en les reliant aux projets individuels et collectifs, privés et publics qui lui donnent statut et qualité. Partenaires : ELICO (information-communication), EVS (géographie), LIRIS (informatique). Financement : PALSE, LabEx IMU.
9. Détection automatique de métaphores dans le discours scientifique en Géographie (2016-2017). Partenaire : EVS (géographie). Financement : BQR Lyon 2.
10. COREL – Au cœur de la relation client (2017-2018) vise à mobiliser l'ensemble du corpus collecté par le Center for Customer Management pour répondre à différentes questions liées à la stratégie client des organisations (entreprises mais aussi services publics). Partenaires : COACTIS (gestion), DRM-ERMES (gestion). Financement : Lyon 2.
11. CAODRA – CARTographie de l'Open Data en région Rhône-Alpes (2017-2018). Visualisation du paysage des données ouvertes dans la réunion Rhône-Alpes. Partenaire : ELICO (information-communication). Financement : MSH-LSH et IXXI.
12. JADN – Evolution du journalisme à l'heure du numérique (2017-2018) vise à mesurer la politique éditoriale d'un média en ligne international (le HuffingtonPost) dans différents pays et langues. Partenaire : CMW (sociologie). Financement : MSH-LSE et IXXI.
13. HyperThésau – Hyper thésaurus et lacs de données : fouiller la ville et ses archives archéologiques (2018-2020). Partenaires : Archéorient (archéologie), ArAr (archéologie), CESCO Paris, SAVL, Archeodunum, FR

Maison de l'Orient et de la Méditerranée, UMS PERSEE, Bibracte, Université autonome de Barcelone et Musées d'archéologie de Catalogne. Financement : LabEx IMU (inclut un postdoc à ERIC).

14. Cartographie du Web littéraire francophone (2019-2020). Partenaire : MARGE (littérature, Lyon 3). Financement : IDEX Lyon-Saint Etienne.

8- Post-doctorants et chercheurs accueillis

Chercheurs étrangers accueillis (6)

1. 2014-2015 : Fabiana Alexandre Feirreira (MCF), Brésil, Validation of happiness index of Pemberton for Brazil (6 mois)
2. 2016 : Ciprian-Octavian Truica (doctorant), Roumanie, Big data and textual analytics (6 mois)
3. 2016-2017 : Mustapha Bouakkaz (doctorant), Algérie, Agrégation textuelle dans l'OLAP (7 mois)
4. 2016-2017 : Zakia Challal (doctorante), Algérie, Minimisation de l'influence négative dans les réseaux sociaux (11 mois)
5. 2017-2018 : Ian Davidson (PR), USA, Adding Humans to Machine Learning (10 mois)
6. 2019-2020 : Taylor Arnold (MCF), USA (6 mois)

9- Indices de reconnaissance

Organisations de colloques / congrès (14)

1. Étudier le Web politique : Regards croisés (Lyon, 2015)
2. Big Data Mining and Visualization: Humanités Numériques et Big Data (Lyon, 2015)
3. Humanités numériques et données ouvertes (Lyon, 2016)
4. Journée MaDICS-ADOC : Données patrimoniales (Toulouse, 2016)
5. Journée MaDICS-ADOC : Données environnementales (Lille, 2016)
6. Colloque Les données urbaines, quelles pratiques et quels savoirs ? (Lyon, 2016)
7. Atelier « Pour l'égalité f/h dans les arts et la culture : des chiffres, des paroles... et des actes ? Freins et leviers d'une objectivation numérique de la réalité. » Université d'Automne de l'association HF #1 pour l'égalité femmes-hommes dans les arts et la culture (2016)
8. Journées MaDICS-ADOC : Analyse de données textuelles (Lyon, 2017)
9. Journée MaDICS-ADOC : Variété des données SHS (Nantes, 2018)
10. L'informatique, pourquoi pas moi ? Les jeunes prennent la parole ! (2018)
11. L'informatique, pourquoi pas elles ? (2018)
12. Atelier PRISME-G@EGC2018 : Penser la Recherche en Informatique comme pouvant être Située, Multidisciplinaire Et Génée
13. Défi données MaDICS-ADOC sur la base d'un corpus de données historiques (2018-2019)
14. Journée d'étude Les figures de style, entre détection automatique et épistémologie (2019)

Invitations à des colloques / congrès à l'étranger (2)

1. Colloque international sur les sciences sociales et la société : Les mégadonnées et les sciences sociales. Villa Noël, Bucarest, 2017
2. Journées sur le thème Big Data et SHS, Université catholique de Louvain, Belgique, 2018

Séjours dans des laboratoires étrangers (3)

1. CEREFREA Villa Noël (Centre Régional Francophone de Recherches Avancées en Sciences Sociales), 2015
2. Université du Québec à Montréal, Institut de Recherche en Etudes Féministes, Canada, 2018

3. Université de Genève, collaboration avec Isabelle Collet, Suisse, 2018

II - IMPLICATION DE L'UNITÉ ET DE CHAQUE ÉQUIPE / THÈME DANS LA FORMATION PAR LA RECHERCHE

1- Productions scientifiques (articles, ouvrages, etc.) issues des thèses

Productions scientifiques (articles, ouvrages, etc.) issus des thèses : 24

Nombre moyen d'articles par doctorant : 2,7

Seuls les doctorant·es inscrit·es à Lyon sont décompté·es ici. Si l'on y ajoute les codirections avec l'étranger (Bouakkaz et Oukid), la production s'élève à 29 articles et la moyenne par doctorant·e reste pratiquement inchangée (2,6). Il faut également noter que la grande majorité des doctorant·es publie aussi des articles sans rapport avec des projets ou applications SHS ou humanités numériques.

2- Formation

Doctorants (3)

1. Aybüke Öztürk, « Design, implementation and analysis of a description model for complex archaeological objects », Université Lumière Lyon 2, 2013-2018 (co-encadrement avec laboratoire ArAr – Archéologie)
2. Max Béliigné, « Approche numérique des revues françaises de géographie (1892-2014) : pour une relecture épistémologique des savoirs géographiques », Université Lumière Lyon 2, en cours depuis 2016 (co-encadrement avec le laboratoire EVS – Géographie)
3. Pegdwendé Sawadogo, « Des lacs de données à l'analyse de documents textuels », Université Lumière Lyon 2, en cours depuis 2018 (projet AURA-PMI avec le laboratoire Coactis – Gestion)

Doctorants bénéficiant d'un contrat spécifique au doctorat (3 bourses régionales)

Nombre de thèses soutenues (1)

Durée moyenne des thèses : 57 mois

Stagiaires accueillis (M1, M2) : 15

En co-encadrement ou en collaboration via un projet avec la MSH-LSE (archéologie du paysage sonore), les laboratoires ArAr (archéologie), Coactis (gestion), CRTT (linguistique), ECP (pluridisciplinaire SHS), ELICO (information-communication), EVS (géographie) et GERiCO Lille (information-communication).

Personnes responsables d'une mention ou d'un parcours de master non informatique (nombre total) : 2