

Conformément à l'article 11 de ses statuts, le laboratoire ERIC produit un rapport d'activité à mi-parcours du contrat quinquennal. Élaboré en 2016-2017 et portant sur la période 2012-2016, il a été soumis à l'évaluation de quatre rapporteurs :

- Pr. Pavlos Moraitis, directeur du laboratoire LIPADE (Université Paris Descartes), équipe d'accueil de taille similaire à celle d'ERIC, en charge d'une évaluation globale du laboratoire ;
- Pr. Avner Bar-Hen, responsable de l'équipe « Méthodes statistiques de data mining et apprentissage » (MSDMA) du laboratoire CEDRIC (CNAM Paris), plus particulièrement en charge de l'évaluation de l'équipe Data Mining et Décision ;
- Pr. Esteban Zimányi, directeur du département « Computer and Decision Engineering » de l'Université Libre de Bruxelles, plus particulièrement en charge de l'évaluation de l'équipe Systèmes d'Information Décisionnels ;
- Pr. Patrice Bellot, responsable de l'équipe « Data, Information & content Management Group » du laboratoire LSIS (Aix-Marseille Université) et chargé de mission « Fouille de textes » à l'INS2i CNRS, plus particulièrement en charge de l'évaluation des activités du laboratoire liées aux humanités numériques.

Leurs quatre rapports d'évaluation sont annexés au présent document.

La direction du laboratoire ERIC remercie vivement les rapporteurs pour leur travail bénévole et leurs retours circonstanciés et constructifs.

Lyon, le 8 février 2018.

Paris 6/12/2017

Rapport d'évaluation mi-parcours du laboratoire ERIC

(demandé par son directeur Prof. J. Darmont)

Le laboratoire ERIC est une unité de recherche, fondée en 1995. Ses tutelles sont l'université Lumière Lyon 2 et l'Université Claude Bernard Lyon 1. ERIC est également membre de l'institut des sciences de l'Homme (ISH, unité CNRS) et rattaché à l'Ecole Doctorale Informaths (ED 512) et partenaire du Labex Intelligences des Mondes Urbains (IMU) et de l'Institut du Genre (groupement d'intérêt scientifique).

Le laboratoire ERIC est dirigé par un directeur et un directeur adjoint. La structure de direction comporte aussi un conseil de direction et un conseil du laboratoire. En 2016 le laboratoire était composé de 54 membres dont 20 enseignants chercheurs permanents, 1 BIATSS, 25 doctorants, 2 post-doctorants ou ATER et 6 membres associés.

Le laboratoire est composé de deux équipes à savoir *Data Mining et Décision (DMD)* et *Systèmes d'Information Décisionnels (SID)*. Les thématiques de recherche associées aux deux équipes sont l'apprentissage automatique pour la fouille de données, la modélisation, caractérisation, fouille dans les graphes, les modèles d'aide à la décision multicritères et l'analyse des données complexes et fouille d'opinions pour l'équipe Data Mining et Décision (DMD) et les « big data analytics », l'analyse en ligne de données textuelles et de réseaux d'information, l'OLAP centre utilisateur et la sécurité des données pour l'équipe Systèmes d'Information Décisionnels (SID). Chaque équipe est dirigée par un responsable et dispose d'un budget propre pris sur la dotation générale de fonctionnement du laboratoire et établi au prorata des effectifs mais chaque équipe renforce ses ressources par des contrats et projets.

Le but de ce rapport est de donner un point de vue général sur le laboratoire. Ainsi mes commentaires seront présentés selon quatre axes à savoir, *l'identité, la structuration et le pilotage, les activités du laboratoire et le rapport proprement dit.*

- a) *L'identité* : le laboratoire semble avoir une identité claire autour de la thématique large de l'informatique décisionnelle. Les thématiques plus spécifiques développées au sein de deux équipes autour des sciences de données (au sens large), l'apprentissage automatique et la prise de décision (essentiellement multicritères) présentent une cohérence globale, dans un domaine d'informatique bien identifiable au niveau national et international. Ils permettent aussi un positionnement scientifique clair d'ERIC par rapport à d'autres laboratoires du site de Lyon (et de la région plus largement). Néanmoins, on pourrait considérer que la partie

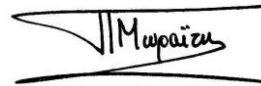
« données » semble être plus fortement représentée (et développée) que la partie « décision ». Cette partie pourrait éventuellement être renforcée en considérant par exemple d'autres approches de prise de décision (que la prise de décision multicritères) comme par exemple des approches fondées sur la théorie de jeux, l'argumentation etc.

- b) *Structuration et pilotage* : la structuration à deux équipes (Data Mining et Décision (DMD) et Systèmes d'Information Décisionnels (SID)) paraît bonne et justifiée. Même si on peut considérer qu'il puisse y avoir des chevauchements entre des sujets de recherche développés au sein de deux équipes ainsi que des interactions entre des thématiques de deux équipes (ce qui est souhaitable d'ailleurs), on peut distinguer des différences concernant les sujets sur lesquels sont focalisées les recherches développées au sein de chaque équipe. Néanmoins, on peut constater une (petite) différence au niveau des effectifs (12 vs 8) en faveur de l'équipe DMD ce qui demanderait éventuellement un rééquilibrage en faveur de l'équipe SID au moment de prochains recrutements (à moins que cette différence soit justifiée par des raisons qui n'apparaissent pas dans le rapport).

Concernant le pilotage, les rôles des différentes instances impliquées semblent bien distincts. On pourrait seulement se demander si vu la taille du laboratoire une structure de pilotage avec un niveau en moins ne serait pas plus simple et efficace. L'idée serait d'avoir un conseil de laboratoire avec un petit nombre de membres (à définir) où les responsables d'équipes seraient membres de droit et où le reste des membres seraient élus, représentant les membres permanents, les doctorants, post-doctorants, BIATSS, et membres associés. Ainsi on aurait la direction et le conseil pour prendre les décisions sur des sujets courants et il y aurait exceptionnellement et seulement pour certaines décisions plus importantes (p.ex. concernant une restructuration au niveau des équipes etc.) un vote de l'ensemble des membres du laboratoire (qui constituerait ainsi l'assemblée générale) pour valider ces décisions. Néanmoins cette proposition n'est qu'une simple suggestion et c'est aux instances du laboratoire d'examiner si une telle structuration de pilotage peut présenter quelques avantages supplémentaires par rapport au mode de fonctionnement actuel.

- c) Le laboratoire a une activité scientifique soutenue. La qualité aussi est présente puisque on peut noter plusieurs publications dans des revues et conférences internationales de très bonne réputation. Le laboratoire a eu aussi plusieurs projets (dont la majorité sont nationaux) et contrats ce qui montre une dynamique au niveau de la recherche des collaborations et de financements supplémentaires (c.a.d. autre que la dotation récurrente). J'aurais néanmoins une suggestion concernant les chercheurs invités. Je considère que le renforcement des liens avec des chercheurs étrangers et leurs institutions à travers un nombre plus important d'invitations de chercheurs confirmés et séjours bilatéraux, permettraient une plus grande encore visibilité du laboratoire au niveau international ainsi que des facilités au niveau du montage de projets européens.
- d) Le rapport de mi-parcours est bien structuré et bien écrit. J'aurais néanmoins certaines suggestions en vue de la préparation du rapport d'évaluation HCERES :

- Concernant la section 1.4.4 « Faits Marquants » il serait plus approprié (en accord aussi avec les documents type d'HCERES) de dédier cette section à des faits marquants concernant les résultats scientifiques du laboratoire (p.ex. prix reçus pour des papiers, des résultats théoriques et/ou expérimentaux importants, production des logiciels « open source » utilisés par la communauté, actions entreprises concernant le transfert technologique, etc.). D'autre part dans les documents types d'HCERES il est suggéré que les « faits marquants » soient présentés au niveau des équipes (dans le cas des laboratoires multi-équipes)
- Dans les différentes sections qui présentent les travaux des équipes selon différents axes, il serait plus approprié de se focaliser sur les travaux les plus importants/significatifs et qui ont été publiés dans les meilleures revues et conférences (p.ex. le top 20% des publications) des différents domaines. Il me semble que ce n'est pas actuellement le cas. Cela est aussi en relation avec la demande d'HCERES de présenter (et intégrer au dossier) seulement une liste de top 20% des publications et de préparer une liste avec toutes les publications seulement pour le cas où celle-ci serait demandée par le président du comité.



Prof. Pavlos MORAITIS

Directeur du Laboratoire
d'Informatique Paris Descartes
(LIPADE)

Université Paris Descartes

Commentaires sur le rapport d'activité à mi-parcours du laboratoire ERIC

Le laboratoire ERIC se caractérise par une activité de recherche très dynamique tant en science des données que dans le domaine de l'informatique décisionnelle. Les synergies entre ses différentes compétences lui permettent d'être un acteur reconnu dans le domaine du Big Data et tout particulièrement dans le domaine des humanités numériques.

Les deux équipes, *Data Mining et Décision* (DMD) et *Systèmes d'Information Décisionnels* (SID) sont équilibrées. La production scientifique est importante et de très bon niveau. La laboratoire peut se prévaloir d'un rayonnement national et international (publications, éditions, organisations de conférences et implication dans les sociétés savantes). La politique de recrutement est pertinente et les recommandations des précédentes évaluations ont été prises en compte.

L'équipe DMD est composée de 12 permanents structurée autour de quatre thèmes. Les thématiques sont pertinentes et offre de nombreuses possibilités de collaboration entre statisticiens et informaticiens. La largeur du spectre de compétence sur le *big data* est tout autant remarquable que rare dans le paysage universitaire français. Il est donc important de préserver et d'encourager cette multi-disciplinarité. Le résultat de ses interactions est cependant peu visible dans les contrats ou les publications. Une politique volontariste (bourse de master fléchée par exemple) permettrait de mieux profiter des synergies de compétences entre des chercheurs d'horizons différents, avec des systèmes d'évaluation indépendants mais avec des thématiques communes. Quelques chercheurs devraient être encouragés à publier leur travaux dans des conférences plus reconnues ou dans des journaux mieux indexés.

Le remarques suivantes sont destinées à aider la réflexion de l'équipe et ne constitue en rien des critiques :

- L'application aux humanités numériques (HN) gagnerait à être mieux explicitée : le nombre d'applications aux HN aboutit à un travail principalement valorisé dans des revues et conférences principalement informatiques. Les politiques de regroupement des universités et l'absence de label CNRS nuisent à la reconnaissance institutionnel de cette forte originalité du laboratoire. En ce sens l'intégration dans le labex IMU est un pas dans la bonne direction mais doit être complétée par des actions complémentaires. À titre d'exemples on peut penser à des co-tutelles de thèse, des accueils en délégation de chercheurs en

sciences humaines ou des contrats industriels vers les domaines des sciences humaines

- La présence de quatre axes par équipes fragmente fortement l'effort de recherche et disperse la recherche de financement. La possibilité de regroupement pourrait être imaginée
- La recherche de financement pourrait être diversifié afin d'éviter les à-coups inévitables avec l'ANR ou les projets européens. En particulier des collaborations de type chaire industriels permettraient une politique industrielle plus pérenne
- La laboratoire étant majoritairement rattaché à la section 27 il est important de préserver les particularités des chercheurs en science des données afin que ceux-ci puissent faire reconnaître leur travail au sein de la section 26 du CNU. Le financement de stage de master, le co-encadrement de thèse, l'accueil et le recrutement d'enseignant-chercheurs doivent être favorisés.
- Le recrutement d'un ingénieur support pour le traitement des données ou le développement d'applications industrielles seraient un vrai plus.
- Enfin il semble que les responsabilités administratives sont fortement concentrées sur un petit nombre de chercheurs.

En conclusion l'activité et le positionnement du laboratoire ERIC sont tout à fait remarquables mais doivent être consolidées pour se maintenir dans un environnement institutionnel en forte évolution .

Avner Bar-Hen
Conservatoire National des Arts et Métiers
2 rue Conté
75003 Paris, FRANCE



Bruxelles, le 4 février 2017

L'unité de recherche Eric se positionne dans un domaine extrêmement stratégique de nos jours, à savoir la science des données et l'informatique décisionnelle. L'unité est structurée en deux équipes de recherche, l'équipe Systèmes d'Information Décisionnels (SID), spécialisée dans le domaine des entrepôts de données, et l'équipe Data Mining et Décision (DMD), spécialisée dans le domaine des fouilles de données et décision. Une caractéristique originale des recherches de l'unité est de valoriser les mégadonnées (big data), notamment dans les domaines des sciences humaines et sociales (SHS). La gestion de l'unité est assurée par une structure efficace qui assure le bon fonctionnement de l'unité du point de vue administratif et scientifique.

Les objectifs du quinquennal en cours sont les suivants

- Conforter le positionnement leader du laboratoire dans le domaine de l'informatique décisionnelle au plan national, et renforcer sa visibilité et internationale.
- Poursuivre la politique de collaboration scientifique du laboratoire avec les SHS.
- Viser une reconnaissance CNRS à échéance du prochain contrat quinquennal via un projet pluridisciplinaire lié aux humanités numériques.
- Renforcer les liens existant avec les autres laboratoires d'informatique du site Lyon-Saint-Étienne.
- Affirmer un positionnement scientifique sur le thème big data analytics et notamment sur l'analyse des média sociaux.

Les activités du laboratoire Eric sont tout à fait en accord avec les objectifs quinquennaux visés. Au plan national, elles positionnent le laboratoire Eric comme leader des communautés scientifiques qui se sont constituées autour des entrepôts de données et de la fouille de données. Ces activités portent sur un large spectre qui couvre tant la recherche fondamentale et appliquée, tant dans le domaine de l'informatique que dans l'application de l'informatique dans les sciences humaines.

Au vu du rapport d'activités, quelques suggestions qui pourraient aider le laboratoire à mieux attendre les objectifs visés sont les suivantes.

- On remarque une base considérable des recettes du laboratoire dans l'année 2016. Si ce phénomène est tout à fait normal étant donné le caractère cyclique de la recherche qui est structurée autour de projets financés par des bailleurs de fonds externes, le rapport ne précise pas quelles sont les actions menées pour assurer la continuité de l'arrivée de fonds extérieurs pour alimenter la recherche. Une meilleure gestion prévisionnelle pour la soumission de projets de recherche en fonction des dates de clôture des projets en cours pourrait compenser ce phénomène.



- Il est souhaitable de définir une stratégie pour augmenter le rayonnement international du laboratoire Eric. Malgré que les activités du laboratoire contribuent à cet objectif, on peut constater une majorité d'activités à portée nationale. En particulier, il est souhaitable d'augmenter la soumission de projets européens, même en sachant la nature extrêmement compétitive de ceux-ci. Un projet européen, même s'il n'est accepté, est l'occasion de tisser des liens scientifiques avec d'autres équipes européennes travaillant dans le même domaine de recherche.
- De manière plus générale, le rapport ne montre pas un plan stratégique spécifique pour atteindre chacun des objectifs visés. Si toutes les activités de laboratoire vont dans le bon sens, il faut définir une stratégie claire qui permet de prioriser les activités qui auront un meilleur « retour sur investissement » dans l'atteinte des objectifs.

L'équipe Systèmes d'Information Décisionnels (SID) structure ses recherches autour des quatre axes suivants : (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. Dans chacune de ces quatre axes, l'équipe SID réalise des recherches tout à fait novatrices, plus particulièrement dans les deux premiers axes.

Les suggestions générales ci-dessus sont également valables pour l'équipe SID. D'autres suggestions qui pourraient aider l'équipe à mieux atteindre les objectifs généraux de l'unité sont les suivantes.

- Même si dans les deux derniers axes de recherche l'équipe SID aborde des thématiques très pertinentes dans le contexte actuel des entrepôts de données, la production de recherche est moins importante par rapport aux deux premiers axes. Ceci est peut-être dû au nombre de chercheurs qui travaillent dans les deux derniers axes. Une intégration des deux derniers axes de recherche dans l'un des deux premiers permettrait d'atteindre une masse critique supérieure. Par ailleurs, d'un point de vue scientifique, les deux derniers axes s'intègrent bien, par exemple, dans le premier.
- On remarque une diminution de la production scientifique de l'équipe et le rapport ne mentionne pas une analyse des raisons de cette diminution ni les mesures proposées pour pallier au problème. De plus, étant donné la qualité des résultats de recherche de l'équipe, c'est tout à fait regrettable que ces résultats ne soient pas valorisés par des publications dans des revues ou des conférences de premier rang.
- Etant donné que le rapport ne précise pas la composition des deux équipes SID et DMD, il n'est pas possible de déterminer combien les deux équipes collaborent. La complémentarité des deux équipes fait que des recherches tout à fait originales pourraient voir le jour si une réelle collaboration s'installe entre les deux équipes. Un domaine évident autour duquel les deux équipes pourraient articuler des recherches communes est celui des graphes, qui est abordé par les deux équipes mais avec des points de vue différents.

Bien à vous

Prof. Esteban Zimányi

Directeur du Département Computer and Decision Engineering
Université Libre de Bruxelles

Laboratoire des Sciences de L'Information et des Systèmes
LSIS – UMR CNRS 7296

Patrice Bellot – patrice.bellot@univ-amu.fr / (+33-0) 6 43 18 44 07 / (+33-0) 4 91 05 60 79
Aix-Marseille Université – LIS (CNRS UMR 7296) – Polytech' Marseille
Domaine universitaire de St Jérôme- Avenue Escadrille Normandie Niemen
F-13397, Marseille Cedex 20 (France)

Commentaires sur le rapport d'activités à mi-parcours du laboratoire ERIC

Le présent document est rédigé à partir du rapport d'activité du laboratoire ERIC pour la période 2012 à 2016. Ce rapport présente sur un peu plus de vingt pages les activités de ses membres et des deux équipes qui le constituent. Cette présentation est suivie par des Annexes d'une quarantaine de pages qui listent dans le détail les publications, les sujets de thèses, les projets, les séminaires et les principales responsabilités scientifiques et administratives de ses membres. Cette structuration du document permet de relever les résultats les plus remarquables et d'identifier clairement les lignes de force et la stratégie du laboratoire pour les années à venir.

L'unité, âgée de 22 ans, est équipe d'accueil sous la double tutelle des universités Lyon 1 et Lyon 2. Elle est membre de l'Institut des Sciences de l'Homme (ISH, unité CNRS) depuis 2012, Institut qui a lui-même été dirigé par l'un des Professeurs du laboratoire ERIC de 2011 à 2016. ERIC est partenaire du LabEx Intelligences des Mondes Urbains (IMU) et de l'Institut du Genre (groupement d'intérêt scientifique).

Les recherches d'ERIC s'inscrivent dans le cadre général de la science des données, plus particulièrement des mégadonnées, et de l'informatique décisionnelle. Le laboratoire positionne sans ambiguïté les Sciences Humaines et Sociales (les humanités numériques) comme un axe stratégique qui fait l'objet d'une attention particulière dans le rapport (Annexe A3). Un projet de laboratoire pluridisciplinaire lié aux humanités numériques est présenté comme la priorité en vue d'une reconnaissance CNRS pour le prochain contrat. Il s'agit d'une voie prometteuse, tout à fait pertinente en fonction du contexte local (dont la double-tutelle Lyon 1 – Lyon 2), national (peu voire pas de laboratoires positionnent cette voie de façon aussi nette) et international (se référer aux infrastructures DARIAH et CLARIN) mais aussi tout à fait en accord avec les efforts fournis durant la période écoulée par ERIC. Il ne s'agit donc pas d'un changement de cap pour le prochain contrat mais bel et bien d'une confirmation de la stratégie entamée. Elle est par ailleurs bien en accord avec certaines orientations du CNRS qui promeuvent des actions inter-instituts INS2I-INSHS, par exemple autour des données et des méga-données et des objets d'étude que sont, pour certaines disciplines SHS et pour l'informatique et les mathématiques, l'internet et les réseaux et média sociaux. Pour la

concrétiser, il pourrait s'avérer pertinent de flécher un prochain recrutement sur ce thème des humanités numériques en complément du recrutement d'un PR sur ce thème effectué en 2015.

ERIC a choisi d'orienter ses recherches autour des humanités numériques par l'intermédiaire de la fouille de données textuelles (grands corpus dont corpus du Web et des réseaux sociaux mais aussi corpus de spécialité) et le traitement automatique des langues. Il s'agit d'une activité bien complémentaire avec les thèmes du LIRIS qui considère le texte dans l'optique de la recherche d'information, la reconnaissance de caractères ou la structuration de documents mais pas dans celle du traitement automatique des langues (analyse de sentiment notamment).

La fouille de données textuelles est abordée dans le cadre de l'analyse de graphes, par exemple pour modéliser et prévoir la diffusion de l'information à travers des média et graphes sociaux. Elle est exploitée pour identifier les rôles des intervenants dans des forums de discussions, pour détecter des communautés, pour extraire et suivre des opinions (équipes SID et DMD) ou modéliser des controverses (équipe DMD), pour détecter des événements, ou plus récemment, pour l'analyse d'articles scientifiques (équipe SID) et la désambiguïsation des noms d'auteurs (axes 2 et 4 de l'équipe DMD). Notons que je n'indique dans la phrase précédente des noms d'équipe que pour souligner que l'axe Humanités numériques et étude de données textuelles est clairement suivi par les deux équipes. Ces différentes questions correspondent à des problématiques importantes des humanités numériques du point de vue par exemple de la linguistique, des sciences de l'information, de la bibliométrie ou de la sociologie. Les méthodologies proposées peuvent engendrer de nouveaux angles d'étude pour ces disciplines qui à leur tour doivent alimenter les recherches en informatique (l'un ne doit pas seulement être un 'service' de l'autre). Il s'agira de mon point de vue de bien cerner les apports réciproques afin que les approches et les recherches soient clairement transdisciplinaires. Cet aspect ne transparait pas dans le rapport mais j'ai bien conscience que la question est facile à poser et qu'il est difficile d'y répondre. Il s'agit d'une interrogation pour l'ensemble de la communauté scientifique des humanités numériques. ERIC peut y jouer un rôle notable, et la responsabilité du Master Humanités Numériques d'une de ses membres est un point fort pour cela (liaison études doctorales — j'aurais été intéressé par connaître les avancées dans ce domaine concernant le nombre d'étudiants et leur devenir mais aussi le positionnement vis-à-vis des écoles doctorales pour les étudiants continuant en thèse).

Dans ce qui suit, je reprends quelques éléments notables puis reviens sur chacune des équipes et enfin je présente ce qui constitue selon moi les points forts à maintenir et les points pouvant être améliorés.

Structuration :

Le laboratoire est structuré en deux équipes rassemblant des chercheurs des universités Lyon 1 et Lyon 2 alors qu'il était structuré en 3 axes (équipes ?) sur la période précédente. Leurs tailles sont équilibrées ainsi que le ratio PR / MCF avec ou sans HDR. Les noms des deux équipes font tous

deux apparaître le terme « Décision ». Cela est compréhensible mais je dois avouer avoir ressenti ensuite un certain décalage entre les noms et les thèmes suivis. Une plus grande différenciation dans les noms, entre les deux équipes, ou une meilleure proximité avec les termes définissant le nom du laboratoire pourrait donner une meilleure lisibilité : entrepôts, représentation et ingénierie des connaissances sont autant de termes non explicites et seulement une interprétation possible de « Data Mining » et de « SI décisionnels ». L'orientation SHS – DH ne transparait pas non plus. Cela n'est peut-être pas très important mais peut aussi signifier qu'une réorganisation des équipes serait pertinente.

Le conseil de laboratoire comprend tous les membres permanents (20 enseignants-chercheurs), un représentant élu du personnel administratif (1 seul BIATSS dans le laboratoire), un seul des doctorants et post-doctorants (parmi 27) et un seul des membres associés (parmi 6). Deux représentants de chacune de ces catégories de personnels pourrait permettre une meilleure représentativité.

Thèses :

Le nombre de doctorants est globalement stable sur la période. Du fait de l'augmentation du nombre de personnels (dont MCF HDR) on aurait pu s'attendre à une hausse. Cela s'explique par les variations du nombre de projets ANR / PIA obtenus et semble être compensé par une augmentation du nombre de contrats CIFRE.

Budget :

Une part importante du budget dépend des projets ANR et PIA obtenus par le laboratoire. Cela explique les fortes variations sur la période (budget 2016 pratiquement inférieur de moitié à celui de 2015). L'effort en faveur de projets collaboratifs est net. Par ailleurs, ce problème est général pour l'ensemble des laboratoires. La relative petite taille d'ERIC le rend plus sensible à ce problème qui peut remettre en cause la direction stratégique choisie pour répondre à des opportunités variées.

Projets :

Coordination scientifique d'un projet européen (2011-2014) et d'un projet ANR (2012-2015) par l'équipe DMD et de nombreux projets dans l'environnement lyonnais et CIFRE pour l'équipe SID. L'effort et la réussite d'ERIC est claire même s'il faut espérer plus de succès auprès de l'ANR dans un futur proche. Il s'agit là encore d'un problème récurrent et l'on ne peut qu'encourager le laboratoire à intensifier ses efforts autour des appels H2020 (dont ERC).

Publications :

La qualité des publications du point de vue des référentiels internationaux (préciser les versions de CORE utilisées) a augmenté par rapport à la période précédente surtout pour les revues de rang A. Ces référentiels, Scimago et CORE, sont pertinents mais ne sont pas les seules sources témoignant de la richesse d'une production scientifique. Dans les domaines correspondants aux activités du laboratoire, ils ne rendent pas toujours bien compte des publications pluridisciplinaires ou à destina-

tion de champs disciplinaires variés, des publications en accès ouverts, ni de l'impact de la littérature dite grise qui participe à la visibilité. Des indicateurs reflétant les usages mais aussi la visibilité sur les plateformes Web et les plateformes ouvertes peuvent fournir un regard complémentaire intéressant (HAL, Google Scholar, DBLP, métriques alternatives de type AltMetrics, consultations Web). Cela dit la liste des publications souligne bien que nombreuses sont celles présentes sur HAL et bénéficiant d'un DOI.

Si la qualité est au rendez-vous et la présence forte parmi les meilleures conférences du domaine de la fouille de données, il pourrait s'avérer pertinent d'intensifier l'effort de publications dans les conférences internationales des humanités numériques (conférence DH par exemple) ou comprenant des ateliers dédiés.

Éléments notables relatifs à chacune des deux équipes.

Equipe DMD :

- découpée en quatre axes thématiques : apprentissage automatique, fouille dans les graphes, aide à la décision multicritères, fouille d'opinions ;
- les données sont complexes et structurées, hétérogènes, dynamiques, imprécises, volumineuses. Parmi les données structurées : les données textuelles. Cela semble être une caractéristique encore plus forte que pour l'autre équipe ;
- les conférences non classées sont relevées dans le rapport comme constituant un point de vigilance. Cela n'est probablement pas inquiétant à partir du moment où les revues et conférences les plus qualitatives sont aussi en augmentation. Participer à des conférences non classées peut contribuer à la vitalité de communautés scientifiques ou à l'échange avec d'autres disciplines ;
- a organisé plusieurs ateliers tournés vers les SHS : fouille d'opinion dans le web social, étude du web politique (revue Réseaux) ;
- collaborations : de multiples collaborations sont évoquées, aussi bien nationales qu'internationales, mais il est difficile d'en percevoir la nature et l'impact. Il semble qu'il serait bien de les structurer et de les valoriser selon des axes stratégiques privilégiés ;
- forte implication sur différents projets collaboratifs (ANR, PIA, Europe).

Équipe SID :

- quatre axes : (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 ;
- animation forte : congrès inforsid 2014, action dans le GDR MaDICS, association EGC, journées d'étude... mais aussi mandatures au CNU ;

- de nombreuses collaborations sont évoquées mais il est difficile à la lecture du document d'en percevoir les objets et les résultats, surtout au niveau des collaborations nationales ;
- du point de vue des humanités numériques, l'équipe SID a notamment obtenu le financement de :
 - o un projet PEPS du CNRS visant à étudier les dispositifs numériques mis en place pour diffuser et échanger les informations relatives au patrimoine naturel et culturel du territoire,
 - o un projet LabEx IMU visant à décrire et caractériser des pratiques de documentation et d'archivage du territoire urbain par la photographie,
 - o un projet avec le laboratoire Environnement Ville Société (UMR EVS) sur la détection de métaphores dans le discours scientifique en géographie ;
- Soutenance de 15 thèses dont 3 en co-tutelles et 2 en do-encadrement avec l'équipe DMD.

Synthèse :

ERIC possède une identité forte autour de la fouille et des entrepôts de données qui en font l'un des laboratoires les plus visibles dans ces domaines (les laboratoires français qui ont le plus de personnels ont parfois des équipes dédiées aux mêmes thèmes qui ont la même taille qu'ERIC mais cela ne leur donne pas forcément une visibilité aussi forte). L'orientation stratégique vis-à-vis des humanités numériques pour la période à venir est à la fois pertinente et prometteuse. Les résultats déjà obtenus confirment cela. Il est important qu'ERIC se positionne clairement vis-à-vis des initiatives nationales et européennes autour des humanités numériques. Maintenir ce cap tout en préservant les recherches les plus disciplinaires autour des thèmes historiques du laboratoire paraît être une voie intéressante dans l'optique d'une demande d'association au CNRS pour le prochain contrat.

Points forts :

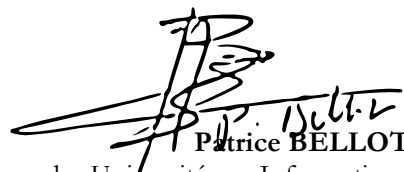
- Bonne croissance du nombre de membres du laboratoire : 2 PR et 2 MCF sur la période, dont 2 en section 27 et 2 en section 26
- La proximité des thèmes permet de réunir dans un seul laboratoire un grand nombre de membres dont les compétences sont complémentaires autour de familles de techniques assez similaires. Cela assure une originalité qui évite l'éparpillement et lui donne une identité forte associée à une bonne visibilité sur ses thèmes
- La moitié des doctorats correspondent à des cotutelles (17%) ou des bourses internationales (33%). Cela conforte la visibilité internationale du laboratoire.

- Des efforts de (re)structuration ont été accomplis durant le précédent contrat, aussi bien du point de vue du nombre d'équipes que de l'identification des axes au sein des 2 équipes retenues. Il pourrait s'avérer pertinent de définir des axes applicatifs stratégiques autour des lignes de force bien identifiées, notamment autour des humanités numériques (quelles communautés ou disciplines SHS prioritaires ?)
- Efforts ayant porté ses fruits durant la dernière période sur le nombre de publications en revues classées A
- Projets et publications communes entre certains membres des deux équipes
- Très forte implication des membres du laboratoire dans les missions des tutelles : direction de l'institut ISH, administration, conseils centraux (CA, commission recherche, école doctorale...), formations (responsabilités de plusieurs Masters et années de Licence, de DU et d'un Master Européen)... Cela présente certes des inconvénients du point de vue du temps passé à la stricte recherche mais permet au laboratoire une forte visibilité locale et certainement une bonne continuité entre Master et études doctorales.
- Animation scientifique dans l'environnement lyonnais (journées d'étude et notamment celles ayant donné lieu, autour des humanités numériques, à des publications ou un numéro spécial de la revue Réseaux) et local (nombreux séminaires dont beaucoup correspondent à des chercheurs étrangers invités). Ceci couplé à l'organisation de conférences (ALT, DS en 2012, Inforsid en 2014, EA 2015) et à une forte présence et implication au sein de la communauté EGC mais aussi du GDR MaDICS.
- Présence au sein de comités éditoriaux, comités de lecture et de programme, jurys de thèse et expertises

Points pouvant être améliorés :

- 1 seul personnel BIATSS pour 54 membres. Cela est très insuffisant pour permettre un bon appui à la recherche aussi bien du point de vue de l'ingénierie et du transfert que du suivi administratif et technique.
- Si la structuration en 2 grandes équipes paraît pertinente, la présentation des axes de chacune d'elles interroge parfois : les recoupements sont notables entre les deux équipes et par ailleurs la proximité des axes à l'intérieur d'une même équipe paraît parfois plus faible qu'entre ces mêmes axes et ceux de l'autre équipe. Cela est vrai pour les axes travaillant sur des données textuelles, sur les réseaux sociaux, sur les approches à base de graphes. Peut-être ne s'agit-il que d'un affichage à revoir en distinguant bien lignes de force et axes applicatifs ou données cibles. Par ailleurs, il serait intéressant d'estimer pour chaque ligne de force ou axe le nombre de personnes impliquées.

- Relatif déséquilibre entre les 2 équipes du point de vue des projets nationaux et européens d'un côté et des projets pluridisciplinaires plutôt locaux (Rhône Alpes) de l'autre. Cela peut peut-être s'expliquer par la façon dont les projets ont été présentés dans le rapport soulignant l'équipe coordinatrice et non pas les participations effectives.
- Le positionnement par rapport au LIRIS ou au centre Magellan (équipe SI) ainsi que par rapport aux autres laboratoires nationaux ou infrastructures nationales ou européennes qui abordent la fouille de textes n'est pas explicité dans le rapport. Les développements logiciels réalisés pourraient par exemple être présents et diffusés au sein de l'infrastructure européenne CLARIN. Le laboratoire pourrait se positionner vis-à-vis des (et avec les) infrastructures nationales de recherche que sont la TGIR Huma-Num, l'IR OpenEdition et les ERICs européens CLARIN et DARIAH.
- La forte implication des membres du laboratoire dans les missions et responsabilités des établissements doit aller de pair avec des responsabilités tournantes. Il n'est pas clair si ces responsabilités donnent lieu à des compensations en termes de service d'enseignement. Il s'agit d'un aspect sur lequel les tutelles doivent être vigilantes pour permettre qu'ERIC maintienne son activité de recherche sur la durée. Il n'est pas précisé dans le rapport les aménagements en faveur des nouveaux enseignants-chercheurs dans l'établissement.
- 2 HDR ont été soutenues sur la période. La production scientifique tend à montrer que d'autres maîtres de conférences pourraient s'engager dans cette voie.


Patrice BELLOT

Professeur des Universités en Informatique
Responsable équipe DIMAG
Chargé de mission Fouille de textes – INS2i CNRS
Aix-Marseille Université - CNRS (LSIS UMR 7296)