

SECTION 06

SCIENCES DE L'INFORMATION : FONDEMENTS DE L'INFORMATIQUE, CALCULS, ALGORITHMES, REPRÉSENTATIONS, EXPLOITATIONS

Composition de la section

Hubert COMON (président de section); Simon PERDRIX (secrétaire scientifique); Pierre ABOULKER (depuis septembre 2019); Pablo ARRIGHI (depuis septembre 2019); Sandrine BLAZY; Agnès BRAUD (jusque septembre 2018); Pierre CLAIRAMBAULT; Pascal DAYRE; Rémy DERNAT (depuis septembre 2019). Clarisse DHAENENS; Laurence DUCHIEN; Amélie GHEERBRANT (jusque septembre 2019); Nathalie GILLES; Jean-Marc LARRE (jusque janvier 2018); Dominique LAVENIER; Leo LIBERTI; Philippe OWEZARSKI; Celine SCORNAVACCA; Sylvain SENÉ (jusque septembre 2019); Pierre SENELLART; Pierre SENS; Gilles VILLARD; Igor WALUKIEWICZ; Laurent WEINHARD.

Introduction

Nous avons choisi dans ce rapport de faire porter l'essentiel de notre effort sur la collecte d'informations statistiques : effectifs, répartitions thématiques, genre, etc. Ces données ont été collectées par les membres de la section durant l'année 2018-2019. Nous proposons aussi des statistiques sur les concours de recrutement et les carrières des chercheurs de la section, à partir des observations effectuées lors des 3 années de mandature écoulées. Cette collecte d'informations peut comporter des erreurs, que nous espérons marginales. Par exemple, les chiffres évoluent en permanence et certaines incohérences peuvent être dues à des dates de collecte d'informations légèrement différentes.

La part de l'interprétation et de la description des thèmes de recherche et de leurs évolutions est de ce fait plus réduite.

Dans la partie I, nous découpons en thèmes

le périmètre couvert par la section 6. Nous comparons aussi les effectifs respectifs des chercheurs CNRS, des membres des UMR, des thèses soutenues, dans chacun des thèmes. Le tableau 1 résume ainsi l'un des principaux travaux effectués par la section pour ce rapport.

Dans la partie II, nous étudions la place des femmes dans la section (globalement faible) et la comparons à la place des femmes dans la discipline. Nous donnons aussi la répartition par genre dans les concours de recrutement (CR et DR). Les statistiques semblent indiquer que les femmes se censurent pour les candidatures CR, mais en revanche nous n'observons pas de « plafond de verre ».

Dans la partie III, nous détaillons, de manière plus traditionnelle, l'évolution des thèmes de recherche de la section.

Enfin, dans la partie IV nous nous penchons sur les recrutements et les carrières des chercheurs en mettant à disposition toutes les statistiques que nous avons pu collecter. Un point

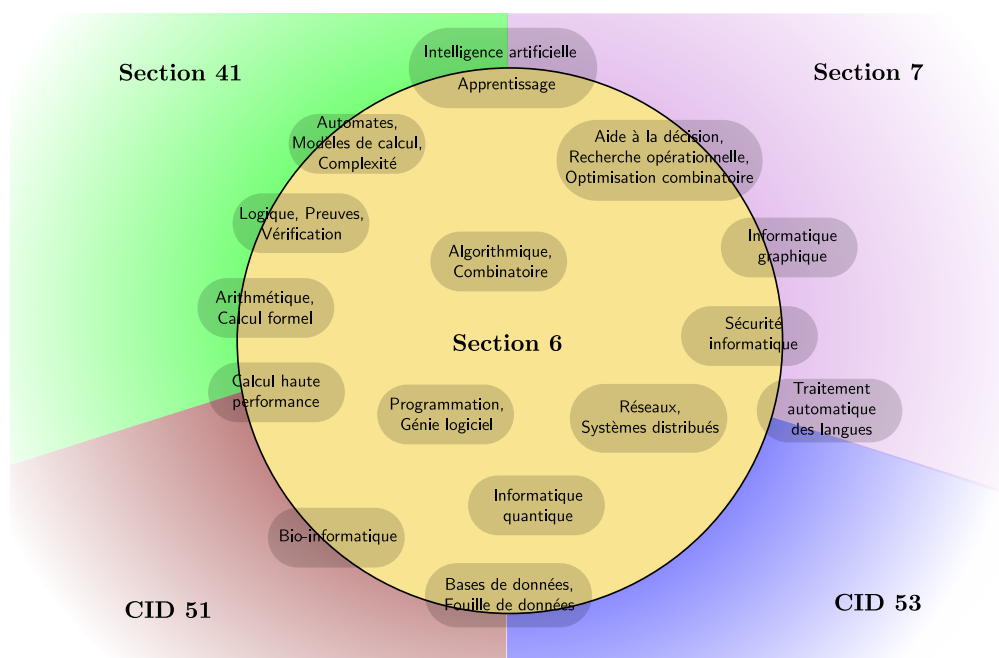


Figure 1 – Périmètre approximatif de la section

saillant est l'importante mobilité sortante dans la section. Nous n'avons hélas pas pu la quantifier précisément car il faudrait des statistiques sur la durée, ce qui est difficile à obtenir à cause de la réorganisation en sections 6 et 7, relativement récente.

Dans tout le document, nous n'avons pas adopté l'écriture inclusive pour faciliter la lecture mais, bien entendu, quand nous parlons de « chercheurs », il faut comprendre « chercheuses ou chercheurs », par exemple.

I. Périmètre thématique et évolutions

Le périmètre thématique de la section et de ses voisines est grossièrement représenté dans la figure 1. Bien entendu, les thèmes de recherche ne sont pas étanches ; il s'agit d'une approximation. Les thèmes « apprentissage » et « intelligence artificielle », particulièrement en vogue, sont revendiqués par de nombreuses sections. Ils font partie des mots-clefs de la section 6 depuis sa création, mais dépassent aujourd'hui largement le cadre de la section. Par exemple, le versant « statistique » peut relever de la section 41. L'apprentissage pour la classification d'images, ou le traitement automatique des langues, relèvera en revanche plutôt de la section 7, tandis que l'apprentissage statistique pour la mise au point de systèmes de recommandation relève de la section 6, etc.

Dans le tableau 1, nous reportons la répartition thématique des chercheurs, enseignants-chercheurs et thèses. Les chiffres donnés dans le tableau ont été obtenus de la manière suivante :

Chercheurs 6 : il s'agit de tous les chercheurs CNRS en activité (donc ne sont pas considérés ici les chercheurs émérites, les chercheurs en disponibilité ou en détachement...) dans la section 6 au 1^{er} janvier 2019. L'étiquetage thématique a été effectué manuellement par les membres de la section, à partir des rapports d'activité des chercheurs. Un chercheur peut émarquer à 1, 2 ou 3 thèmes (nous nous

sommes arbitrairement limités à 3). Dans les deux derniers cas il est compté respectivement $\frac{1}{2}$ et $\frac{1}{3}$ dans chacun des thèmes. Un chercheur de la section 6 peut aussi émarquer à des thèmes hors section.

UMR-non-CNRS : il s'agit des chercheurs permanents, membres des UMR (ainsi que des UPR, UMI, ERL) dépendant de l'INS2I et associées à la section 6, et qui ne sont pas rémunérés par le CNRS. Pour notre discipline, il s'agit très majoritairement d'enseignants-chercheurs, mais aussi par exemple de chercheurs Inria. Les données ont été extraites de l'annuaire CNRS des UMR.

Comme l'effectif total à considérer est très important (3575 personnes), nous avons tiré uniformément un sous-ensemble de 25% des personnels, puis étiqueté thématiquement leur activité de recherche, en nous appuyant sur leurs publications et leur page personnelle décrivant leurs travaux.

Comme ci-dessus, chaque chercheur peut émarquer à plusieurs thèmes et donc être compté pour une fraction dans plusieurs thèmes.

Thèses : les statistiques portent sur l'ensemble des thèses soutenues en 2017 en informatique. Nous avons utilisé l'API du site theses.fr pour récupérer l'ensemble des thèses soutenues en informatique et dans des disciplines voisines, et effectué une analyse manuelle des thèses soutenues durant cette année-là. 292 thèses étiquetées « informatique » (sur 874) ne relèvent pas de la section. Il y a bien sûr des thèses en traitement automatique des langues, en interaction homme machine... qui relèvent de la section 7. Mais, surtout, il y a beaucoup de ces thèses qui relèvent des mathématiques appliquées, voire d'autres disciplines très différentes (biologie, physique, sciences sociales...) pour lesquelles l'informatique n'est qu'un outil.

Dans tous les cas, les totaux peuvent différer légèrement de l'effectif total, à cause du cumul des erreurs d'arrondi. Soulignons aussi que l'étiquetage manuel requiert des interprétations qui peuvent parfois être contestables.

Tableau 1 – Répartition thématique des chercheurs de la section 6, des membres des UMR rattachées à la section et des thèses en informatique. Un échantillon d'un quart des personnels est considéré, comptage par fractions pour l'appartenance à plusieurs thèmes.

Thème	Chercheurs 6		UMR-non-CNRS	Thèses 2017	
	Nb	%	%	Nb	%
Algorithmique	31	10.2%	3.3%	18	3.1%
Combinatoire	37	12.2%	3.9%	24	4.1%
Arithmétique, calcul formel	11	3.6%	2.3%	5	0.9%
Calcul haute performance, calcul parallèle	9	3.0%	7.0%	32	5.5%
Automates, Modèles de calcul, Complexité	26	8.6%	2.6%	9	1.5%
Preuve, vérification, théorie de la programmation	46	15.6%	10.5%	31	5.3%
Sécurité informatique	16	5.3%	4.4%	47	8.1%
Génie de la programmation, génie logiciel	7	2.3%	10.0%	56	9.6%
Réseaux, systèmes distribués	15	5.0%	12.1%	90	15.5%
Données et connaissances	16	5.6%	16.8%	106	18.2%
Intelligence artificielle	28	8.9%	11.4%	77	13.3%
Recherche opérationnelle, aide à la décision	24	7.9%	10.2%	47	8.1%
Bio-informatique	27	8.9%	2.6%	21	3.6%
Informatique quantique	7	2.3%	0.7%	2	0.3%
Autres (section 6)	2	0.7%	2.3%	16	2.8%
Hors section 6	2			292	
Total	305			874	

A. Liens avec les autres disciplines

L'informatique est naturellement présente, sur le versant applicatif, dans de nombreux domaines frontières avec les autres disciplines. L'informatique entretient aussi des liens forts avec les mathématiques : souvent des candidats sont admissibles à la fois en section 6 et en section 41, typiquement, en cryptographie, en calcul formel, logique informatique, combinatoire... Jusque récemment, le CNRS proposait des postes croisés.

CID 51 : Modélisation, et analyse des données et des systèmes biologiques : approches informatiques, mathématiques et physiques. Des chercheurs à l'interface entre informatique et biologie, et dont les thématiques de recherche relèvent explicitement de la section 6, sont régulièrement recrutés par la CID 51. La production massive de données biologiques dans les domaines de l'agronomie, de l'environnement ou de la santé, par exemple, devrait encore renforcer les liens avec la section sur des thématiques telles que la science des données, le calcul haute performance et la sécurité.

CID 53 : Méthodes, pratiques et communications des sciences et des techniques. Les liens entre l'informatique et les sciences humaines et sociales alimentent des problématiques de recherches traitées en CID 53, dont le comité comporte deux chercheurs de la section 6. Actuellement, 3 chercheurs (2 CR et 1 DR) sont rattachés à la fois à la section 6 et à la CID 53. Parmi les candidats à cette CID, des chercheurs abordent des problématiques en lien avec l'informatique comme éthique, droit et traitements de données individuelles (de santé, de réseaux sociaux) ; sciences cognitives et formalisation du raisonnement ou des comportements individuels et collectifs, humanités numériques (traitements massifs de données de la recherche en SHS : textes, données, images...), étude des réseaux sociaux, philosophie des sciences,

science et société, enjeux sociaux des nouveaux systèmes informatiques (objets connectés, robotique et assistants intelligents, exploitation de données et surveillance) ou encore analyse du langage naturel (en lien avec la section 7).

Informatique quantique. L'informatique quantique est un domaine dans lequel des compétences en physique et en informatique sont nécessaires. Chaque année, la section considère des candidatures de physiciens. L'un d'eux a été recruté en 2017.

II. La place des femmes

Le tableau 2 donne les effectifs par grade et genre des chercheurs en activité dans la section 6, au 1^{er} janvier 2019. On observe une très faible proportion de femmes. En revanche, il ne semble pas y avoir de phénomène de « plafond de verre » dans la section puisque la proportion de femmes DR est supérieure à la proportion de femmes CR.

Cette analyse est confirmée par le tableau 3, qui montre la proportion de femmes recrutées sur les concours récents : le taux de femmes recrutées reste toujours supérieur au taux de femmes dans la section. Les deux tableaux précédents et le tableau 4 montrent ainsi que les carrières des femmes ont tendance à être légèrement plus rapides que celles des hommes en section 6.

Pour aider à analyser le faible nombre de femmes dans la section, le tableau 5 donne le ratio de femmes dans les UMR associées à la section (et dont les thèmes de recherche relèvent de la section 6), le ratio des thèses en informatique soutenues par des femmes (en 2017, sur des thèmes relevant de la section) et la proportion de candidatures de femmes aux concours CR ces trois dernières années (à nouveau restreinte aux thèmes de la section).

Nous observons donc d'une part que les femmes (en section 6) sont moins bien représentées au CNRS que dans l'enseignement su-

Tableau 2 – Effectifs par genre en section 6

Hommes	247				81.5%					
Femmes	56				18.5%					
	CR				DR					
Hommes	150		83.3%		97		78.9%			
Femmes	30		16.7%		26		21.1%			
	CRCN		CRHC		DR2		DR1		DRCE	
Hommes	146	83.9%	4	80%	63	80.8%	29	72.5%	5	100%
Femmes	29	16.1%	1	20%	15	19.2%	11	27.5%	0	0%

Tableau 3 – Recrutements CR récents en section 6, par genre

	Année de recrutement						
	2012	2013	2014	2015	2016	2017	2018
Femmes recrutées CR sec. 6 (nb)	3	2	2	3	3	2	3
Femmes recrutées CR sec. 6 (%)	25%	20%	25%	27%	21%	20%	30%

Tableau 4 – Proportion de femmes au concours DR

	2017	2018	2019
Candidates DR	9.5%	13.6%	10.8%
Recrutées DR	16.7%	28.6%	25%

Tableau 5 – Place des femmes dans la recherche sur les thèmes couverts par la section 6

	section 6	UMR-non-CNRS	Thèses (2017)	Candidatures CR		
				2017	2018	2019
Femmes	18.5%	23.1%	25.3%	17.1%	15.9%	10.7%

périeur. Nous observons aussi un net aggravement de la situation, puisque de moins en moins de femmes se portent candidates. Soulignons en particulier que seulement 10.7% des candidatures relevant de la section sont des candidatures de femmes en 2019, alors que 25.3% des thèses relevant de la section en 2017 sont soutenues par des femmes.

III. Évolution thème par thème

A. Algorithmique

Les algorithmes sont parmi les objets centraux de l'informatique. S'attachant à leur conception, leur analyse et leur étude, l'algorithmique se retrouve donc dans tous les domaines couverts par la section 6. La conception et l'optimisation des ressources utilisées (temps, nombre de requêtes, encombrement, énergie, etc.) comportent un volet de mise en évidence de nouveaux paradigmes algorithmiques. Les interactions de la discipline avec les mathématiques sont fondamentales en particulier par le biais de la géométrie, de l'algèbre, de la logique et de la topologie (p. ex., l'étude des propriétés structurelles des algorithmes eux-mêmes). Les interactions avec la biologie sont également marquantes, par exemple en assemblage d'ADN et pour la modélisation du vivant.

L'étude des modèles non classiques (c.-à-d. non discrets) ouvre actuellement un champ d'investigation important. L'algorithmique quantique devrait prendre plus d'ampleur. L'impact sur les autres sciences, la sécurité, la massification des données et les questions énergétiques donnent lieu à des enjeux formidables.

Au sein du GdR IM l'algorithmique est spécifiquement étudiée au sein du GT CoA dont la composition montre de vastes connexions avec d'autres domaines. On observe que la majorité de chercheurs répertoriés dans le thème (voir tableau 1) sont aussi associés à un autre thème et ne sont donc comptés que pour une fraction. Le

nombre de chercheurs ayant une composante algorithmique dans leurs recherches est donc très supérieur à ce qui est indiqué dans le tableau 1.

B. Combinatoire

Ce thème couvre l'étude des familles d'objets discrets et se décline en combinatoire algébrique, combinatoire énumérative, combinatoire des mots, et théorie des graphes. Les problèmes rencontrés proviennent en grande partie de l'informatique : structures de données, calculabilité, complexité et analyse d'algorithmes. Les interactions sont très riches en bio-informatique, avec la physique et avec de nombreuses parties des mathématiques (géométrie, algèbre, théorie des nombres). La thématique se regroupe principalement au sein du GdR IM.

Les sujets d'actualité concernent la convergence d'objets aléatoires, ainsi que la combinatoire des cartes planaires plongées sur des surfaces de genre supérieur. La théorie structurelle et algorithmique des graphes orientés doit se développer, les liens entre les méthodes paramétrées et les méthodes d'approximation représentent des enjeux importants. On peut penser que les outils d'aléa discret pourront à terme contribuer à diversifier les angles d'approches pour l'étude théorique des modèles de processus de renforcement ou d'apprentissage. Autour des systèmes dynamiques, des questions importantes émergent, concernant notamment la robustesse au bruit. En biologie systémique symbolique, la question qui va se poser dans les prochaines années est celle du changement d'échelle dans les approches de modélisation.

La combinatoire est un thème historique de l'informatique en France, avec une grande visibilité et reconnaissance internationale. Elle est bien représentée au CNRS et on constate l'excellent niveau des candidatures aux concours.

C. Arithmétique et calcul formel

Les recherches autour de la manipulation informatique des objets fondamentaux de l'arith-

métique, de l'algèbre et de l'analyse ont fortement évolué avec des rôles clefs en mathématiques expérimentales, cryptologie/codage et modélisation (systèmes dynamiques, géométrie & images, biologie...). Les percées théoriques et pratiques ont permis de renforcer les liens avec les autres disciplines. Une évolution récente est la réponse à des enjeux de garantie sur les calculs (codes critiques, sécurité, etc.) qui s'appuie sur des validations numériques ou formelles.

L'ancrage en informatique, au sein des GT Arithmétique et Calcul formel du GdR IM, permet de se confronter aux défis de : concevoir des solutions algorithmiques, logicielles et matérielles (vitesse, faible encombrement, basse consommation) de l'embarqué jusqu'aux supercalculateurs ; permettre l'amplification d'une vision hybride symbolique, certifiée et numérique du calcul scientifique, assurant robustesse des modèles et fiabilité des résultats ; tirer parti d'automatisations pour le développement et la preuve. Les liens sont nets avec le GdR Sécurité et en architecture avec le GdR SOC2. On s'attend à une accentuation des travaux en calcul haute performance et à une ouverture vers l'informatique quantique.

Au plan sociétal, en particulier avec les logiciels Maple ou SageMath, la discipline a un impact fort en formation scientifique. C'est sur l'équilibre entre l'informatique et les mathématiques dans l'enseignement, et sur les échanges entre ces deux domaines (des chercheurs sont membres d'UMR de mathématiques), que reposent la position internationale de tout premier plan de la communauté française et ses recrutements. Les chercheurs répertoriés en tableau 1 sont souvent associés à d'autres thèmes de la section et sont comptés pour fraction.

D. Calcul parallèle et distribué, calcul haute performance

Le calcul parallèle est un thème historique de l'informatique en France présent essentiellement dans les GdR Calcul, IM, RSD et GPL.

Cette thématique recouvre un large spectre allant de l'exploitation efficace d'une machine aux calculs sur des architectures massivement parallèles et hétérogènes, ainsi que distribuées telles que les clouds. On y retrouve des activités fortes autour du calcul numérique, des algorithmes parallèles, des modèles de programmation, et des supports d'exécution (OS et intergiciel).

La communauté a connu une forte évolution thématique liée aux architectures exascale, à la virtualisation avec les clouds, à la distribution avec le fog/edge computing, à la nécessité de tolérance aux fautes des grandes infrastructures, et à l'arrivée des grandes masses de données et de l'apprentissage automatique en tant que consommateur de ressources et fournisseur de nouvelles approches.

La question sociétale de l'efficacité énergétique fait l'objet d'une action spécifique du GdR RSD en lien avec le GDS Ecoinfo. La communauté porte aussi des plates-formes nationales telles que Grid'5000 et maintenant SI-LECS (en cours de construction) qui vise à resserrer et structurer les liens entre les chercheurs en systèmes, parallélisme et réseau. Depuis 2018, les Journées Calcul Données (JCAD) aident à créer des liens entre les différentes communautés liées au HPC (chercheurs, utilisateurs, et centres de calcul).

La politique de coloriage au concours a contribué à renforcer récemment ce thème qui reste encore peu représenté au CNRS.

E. Modèles de calcul et complexité

Les thématiques autour des modèles de calcul et complexité sont abordées par une large communauté de la section 6 et touchent à de nombreux autres domaines parmi lesquels l'algorithmique, la combinatoire, la logique ou encore l'informatique quantique. La discipline est transversale dans la section, en relation privilégiée avec les mathématiques au niveau des méthodes. On trouve des activités fortes sur les modèles de cal-

cul centraux de l'école d'informatique française (automates finis, machines de Turing) et également sur des modèles moins conventionnels (automates cellulaires, auto-assemblage, algorithmique distribuée...). Les développements sont toujours plus grands en direction de la modélisation et de la vérification des propriétés de systèmes de diverses natures. La thématique s'inscrit principalement au sein du GdR IM et partage des liens étroits avec les GdR IQFA et BIM.

Les défis sont nombreux ; sans vouloir être exhaustif, on peut citer :

- trouver des propriétés calculatoires des modèles et définir de nouveaux moyens d'appréhender la calculabilité adaptés aux modèles étudiés ;
- développer l'étude structurelle des classes de complexité en adoptant différents points de vue (algorithmique, algébrique, arithmétique) ;
- concevoir de nouveaux modèles de calcul tirant parti de la phénoménologie physique (calcul quantique) ou biologique (calcul moléculaire) et développer la connaissance de ceux existants.

Le domaine partage des liens privilégiés des disciplines telles que la physique et la biologie, à travers l'analyse des systèmes naturels complexes, et l'explication et la prédiction de certaines de leurs propriétés.

F. Preuve, vérification, et théorie de la programmation

La section 6 regroupe une large communauté autour de ces trois thématiques, qui entretiennent des liens forts. Les objectifs et méthodologies regroupent l'analyse et certification de systèmes et programmes (analyse statique, vérification de modèles...), la preuve formelle (théorie de la preuve, démonstration automatique, assistants de preuve...), et la théorie de la programmation (compilation, théorie des types, concurrence, sémantique...).

Les recherches sur ces thèmes s'inscrivent à titre principal dans le GdR IM. Elles ont de fortes interactions avec le GdR GPL (en particulier autour des langages de programmation, du typage, et de la preuve de programmes), et également avec le GdR Sécurité et son GT «Méthodes formelles pour la sécurité». Certains sujets sont à la frontière de la section 41, comme la théorie de la preuve et de la programmation (via le GT LHC nouvellement créé avec le renouvellement du GdR IM) ou l'homotopie et les catégories, avec le GdR Top.

On a vu ces dernières années, quelques importantes évolutions thématiques. Les aspects probabilistes prennent une importance croissante dans l'analyse des systèmes et des programmes. L'analyse de modèles et systèmes concurrents voit une ouverture vers l'algorithmique distribuée. On observe une convergence entre la théorie de la programmation (avec ses outils catégoriques, algébriques, topologiques, etc.) et celles de la vérification, des automates et de la complexité. Un intérêt émerge autour du développement de méthodes formelles pour l'apprentissage et en particulier l'apprentissage statistique. Les liens avec l'IA se renforcent en particulier via les travaux en représentation de bases de connaissances. Les assistants de preuve formelle (Coq) et outils de démonstration automatique (SAT, SMT) ont un impact croissant dans de nombreux domaines en informatique et en dehors.

Ce thème de recherche dynamique est très bien représenté, aussi bien dans les UMR (10,5% selon notre échantillonnage) que dans la section 6 (15,6%). Ce thème est un point fort, aussi bien historique qu'actuel, de la section. Le vivier de candidatures aux postes CNRS est grand, et les candidats sont d'excellent niveau.

G. Sécurité informatique

La principale évolution structurelle récente de la section concerne la création du GdR sécurité informatique dont le but est, entre autres, de

rassembler les chercheurs de ce domaine, auparavant éparpillés dans d'autres GdR : IM, GPL, SOC2, ISIS, RSD... Ce thème de recherche est globalement en expansion (plus de thèses soutenues en 2017 que de chercheurs+enseignants-chercheurs dans le tableau 1).¹ La section 6 du CNRS est surtout présente en cryptologie (9.8 chercheurs), peu dans les 5 autres groupes de travail du GdR (moins de 7 chercheurs dans l'ensemble des autres GT). Il faut noter cependant que certains groupes de travail relèvent aussi de la section 7 : sécurité des données multimedia, sécurité des systèmes matériels.

La sécurité informatique ne doit pas être vue comme du ressort exclusif des ingénieurs, même spécialisés. Ce serait une vision court terme dont les conséquences seraient néfastes.

Les recrutements récents (et la politique de coloriage au concours) tendent à renforcer la présence du CNRS en sécurité informatique. Elle reste cependant faible, notamment en sécurité système, domaine dans lequel il y a très peu (voire pas) de candidats.

H. Génie de la programmation et du logiciel

Le GdR Génie de la Programmation et du Logiciel (GPL) a été créé il y a maintenant 10 ans. Ce domaine de recherche permet de résoudre les problèmes posés par la complexité croissante des logiciels. Il rassemble les communautés qui travaillent sur la définition, la formalisation, l'évaluation des artefacts des langages de programmation, de la compilation, de la preuve, de la modélisation, du test, de la vérification et la validation, les approches empiriques et les outils associés pour la production logicielle actuelle et future. Quelques membres du GdR ont signé en juin 2019 un manifeste sur l'importance de la recherche dans le domaine pour la société.²

1. Voir https://www.allistene.fr/files/2018/03/VF_cartographie_2017-06-13.pdf dont les données sont collectées de manière très différente des nôtres.

2. <https://gl.frama.io/manifeste/>

Ce domaine est actif de par sa représentation en termes d'enseignants-chercheurs et de doctorants. Les enseignants-chercheurs sont très sollicités dans les départements d'enseignement des universités et des écoles d'ingénieurs pour former les futurs masters et ingénieurs en développement logiciel dont les entreprises ont tant besoin ; ceci explique le nombre d'enseignants-chercheurs recrutés sur les 20 dernières années. Par contre, le pourcentage de chercheurs dans la section est excessivement faible (2,3% – près de 5 fois moins que la proportion d'enseignants-chercheurs), malgré un coloriage Sciences du Logiciel proposé par l'INS2I depuis plusieurs années. Depuis 3 ans, 3 jeunes chargés de recherche ont été recrutés sur ce coloriage. De plus, l'année 2019 marque un tournant avec une augmentation significative du nombre de candidatures de qualité dans le domaine.

Il a fallu plusieurs années pour amorcer un vivier de candidatures excellentes ; ceci s'explique en partie par les nombreuses propositions de postes à haut potentiel faites aux jeunes docteurs du domaine par l'industrie en France ou par les universités à l'étranger, par une autocensure des candidats et des équipes, mais aussi par l'absence de recrutement au CNRS sur ce thème pendant de nombreuses années.

I. Réseaux et systèmes

Le GdR Réseaux et Systèmes Distribués (RSD) existe depuis plus de 25 ans, son intitulé, ses mots clés, son organisation thématique, etc., ayant été modifiés plusieurs fois avec l'évolution de ses champs de recherche. Le GdR RSD vise à contribuer à l'animation scientifique, à la structuration, à la dynamisation, à la promotion des savoirs et à la mise en synergie de ces deux pôles de recherche fondateurs des grandes avancées et innovations dans le domaine des STIC. Le GdR RSD couvre et intègre ainsi tout un continuum thématique constitutif des systèmes communicants à large échelle susceptibles de supporter des applications critiques.

La communauté RSD comporte majoritairement des enseignants-chercheurs, ces derniers étant largement présents dans les universités et écoles d'ingénieurs vu les besoins en formation de masters et ingénieurs dont l'industrie a besoin. Le flux de recrutements d'enseignants-chercheurs dans la thématique RSD dans les universités et écoles est donc soutenu pour répondre à ces besoins en formation. A contrario, le pourcentage de chercheurs de la thématique RSD dans la section fait partie des taux faibles, alors que dans l'enseignement supérieur cette discipline est la deuxième en nombre d'enseignants chercheurs, et la seconde également en nombre de doctorants (5% au CNRS, 12,5% dans l'enseignement supérieur et 15,5% des docteurs). Malgré un coloriage « Objets communicants, système, réseaux », proposé par l'institut INS2I depuis plusieurs années, seulement 2 jeunes chargés de recherche dans la thématique RSD ont été recrutés ces 3 dernières années (moins d'un par an). En effet, le nombre de candidatures de qualité dans ce domaine est toujours faible. Ceci s'explique en partie par la concurrence des entreprises en France ou à l'étranger pour les jeunes docteurs du domaine mais aussi par une autocensure des candidats et des équipes liée en partie au faible nombre de recrutements au CNRS pendant de nombreuses années sur ce domaine.

J. Données et connaissances

Ce thème recouvre l'acquisition, la gestion et l'exploitation de données et connaissances : les bases de données; la fouille de données; les systèmes de recommandation; l'ingénierie des connaissances et le Web sémantique; la recherche d'informations; l'analyse du Web et des réseaux complexes. Une partie de ce thème est représenté dans le GdR MaDICS sur l'interaction avec les communautés scientifiques productrices et utilisatrices de données.

Ce thème est (voir tableau 1) celui du plus grand nombre d'enseignants-chercheurs dans les UMR (16.8%) et du plus grand nombre

de thèses soutenues (18.2% en 2017) parmi les thèmes de la section 6. Une partie importante de ces thèses (non précisément quantifiée) est réalisée en collaboration avec des entreprises (p. ex., thèses CIFRE) ou à l'interaction d'autre disciplines scientifiques. Ce thème est sous-représenté parmi les chercheurs CNRS de la section (5.6%); il est également sous-représenté parmi les candidats au concours, malgré l'affichage du thème « Science des données » au concours, en partie par manque d'attractivité des carrières de recherche au regard des conditions offertes par les industriels (en particulier les géants du Web), en partie par une autocensure vis-à-vis des concours CNRS.

La frontière de ce thème avec le thème « Intelligence artificielle » est parfois floue, par exemple entre fouille de données et apprentissage, ou entre Web sémantique et raisonnement sur les données. On note également des interactions avec certains thèmes de la section 7 : le traitement automatique de la langue, par exemple, relève de la section 7, tandis que la fouille de texte et la recherche d'information, qui utilisent des outils voisins, relèvent de la section 6.

K. Intelligence artificielle

L'IA fait traditionnellement partie des mots-clés de la section 6, mais ce thème est aujourd'hui partagé par plusieurs autres sections du CNRS (7, 41...). Les recherches en IA au sein de la section 6 concernent de manière historique des approches symboliques telles que la représentation des connaissances et la formalisation du raisonnement, entre autres par des modèles de graphes, la résolution de problèmes par satisfaction de contraintes, la planification et la recherche heuristique, la gestion de l'incertitude, les réseaux bayésiens, les systèmes multi-agents et la décision collective, ainsi que certains aspects de l'apprentissage automatique, en particulier l'apprentissage symbolique. Domaine particulièrement dynamique au regard du nombre de thèses soutenues, l'IA représente une part importante des chercheurs de la section, avec

des chercheurs très visibles et reconnus internationalement. Une majorité de ces chercheurs se concentrent sur les aspects logiques de l'IA (15 sur 28) alors que la répartition entre les différentes thématiques est plus équilibrée parmi les thèses soutenues.

Globalement, ces recherches sont impactées fortement par le succès de l'analyse de grandes masses de données à l'aide d'algorithmes d'apprentissage automatique et particulièrement des réseaux de neurones artificiels. L'apprentissage a parfois supplanté des algorithmes exploitant des modèles logiques de connaissances, mais le défi actuel est d'utiliser au mieux leurs complémentarités, en particulier au service de l'explicabilité des systèmes et de leurs résultats. La maîtrise des analyses par apprentissage sur de grandes masses de données est un enjeu économique et stratégique de premier plan, que les États mettent en avant par des plans nationaux depuis 2017.³

Depuis 2016, la communauté est structurée grâce au GdR IA « Aspects formels et algorithmique de l'intelligence artificielle », lancé courant 2018 après avoir fonctionné comme pré-GdR durant 2 ans. Ce GdR a mis en place 6 groupes de travail propres dont trois s'intéressent justement au croisement entre apprentissage et raisonnement, apprentissage et résolution de contraintes et explicabilité. 4 autres groupes sont communs avec les GdR BIM (sur la Biologie Systémique Symbolique), IM (Informatique Mathématique), RO (Recherche Opérationnelle) et MaDICS (Masses de données).

L. Recherche opérationnelle et aide à la décision

La Recherche Opérationnelle (RO) et l'Aide à la Décision (AD) sont des disciplines hybrides, à la frontière entre mathématiques, informatique (algorithmique et intelligence

artificielle), et certains secteurs du génie industriel et de Droit-Économie-Gestion (DEG), e.g. l'économétrie. La RO/AD offre clairement des points d'intersection avec les thématiques générales identifiées sous les étiquettes « MOA(Optimisation)/Jeux » de l'INSMI, « Aide à la Décision » de l'INSHS et, au sein de l'INS2I et avec la Section 7, autour des thématiques liées à la robotique et aux systèmes automatisés. Le caractère pluridisciplinaire et appliqué de la RO/AD pénalise parfois sa reconnaissance par des instances académiques. Néanmoins, le principal point fort de la RO/AD reste son ancrage au sein du monde socio-économique : discipline hybride, la RO/AD est naturellement proche du monde des entreprises.

L'émergence des nouvelles technologies autour de la mobilité, des systèmes de production, des communications induisent des prises de décision de plus en plus réactives, implémentées dans un contexte d'incertitude au travers d'architectures de communications complexes et de nature souvent collaborative. Elles tendent à poser des questions théoriques, algorithmiques et logicielles sensiblement nouvelles, souvent entre le niveau décisionnel système et le niveau embarqué, et à susciter de nouvelles déclinaisons de la notion d'approximation, mettant en jeu des systèmes à base de règles ou d'apprentissage supervisé.

Le tableau 1 montre que, pour la RO, la proportion des chercheurs CNRS / enseignants-chercheurs des UMR est en accord avec la proportion globale. L'historique des recrutements au CNRS montre un recrutement régulier au cours de ces dernières années.

M. Bio-informatique

La bioinformatique rassemble différents domaines scientifiques (bio, math, info) et cherche à améliorer la compréhension du vivant à partir des données morphologiques et moléculaires. Le GdR BIM structure cette communauté. La section 6 est bien représentée et participe acti-

3. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

vement avec 8.9% des chercheurs répartis dans 20 UMR affiliées principalement à l'INS2I.

L'évolution très rapide des technologies de séquençage et l'acquisition massive et automatisée de données, place les STIC au cœur de cette discipline. Ces données ont de fortes spécificités : elles sont massives, hétérogènes, multi-échelles et interdépendantes. Les domaines concernés par la section couvrent principalement l'algorithmique, la combinatoire, les bases de données, la théorie des systèmes dynamiques, la représentation des connaissances, l'apprentissage, les méthodes formelles et la modélisation de systèmes biologiques complexes.

Les principaux enjeux des années futures sont (1) la santé numérique; (2) la génomique environnementale et la biodiversité; (3) la gestion et l'intégration des données massives. Ils devraient renforcer des interactions avec nos thématiques comme, par exemple, la cryptographie homomorphe pour la gestion sûre des données patient, l'algorithmique parallèle pour traiter les grands projets metagénomiques, ou l'IA et l'apprentissage automatique pour identifier des signaux biologiques d'intérêt dans des données massives et bruitées.

N. Informatique quantique

La communauté de recherche autour de l'informatique quantique est structurée au niveau national par le Groupe de Travail Informatique Quantique du GdR IM (GT-IQ) et le GdR Ingénierie Quantique (IQFA), et en région parisienne par la fédération de recherche PCQC.

Le GT-IQ est à l'intersection de l'informatique quantique et des thématiques de la section 6, ce qui concerne les recherches en algorithmique, cryptographie et méthodes formelles afin de comprendre les capacités, les limites et l'utilisation de ce modèle de calcul émergent. Le CNRS a su être précurseur dans ce domaine avec une implication forte (cf. tableau 1), à noter cependant un seul recrutement de chercheur dans ce domaine dans la section depuis 2009.

Le GdR IQFA est plus large et regroupe phy-

siciens et informaticiens du domaine des technologies quantiques, avec une large majorité de physiciens. L'informatique quantique est un domaine en plein développement avec notamment le *flagship* de l'UE sur les technologies quantiques et aussi l'implication d'acteurs industriels. L'arrivée des NISQ (*Noisy Intermediate Scale Quantum computers*) ouvre des perspectives immenses et aussi des problématiques nouvelles sur l'utilisation et la vérification de l'ordinateur quantique.

IV. Le recrutement et les carrières des chercheurs

A. Mobilité sortante

Nous n'avons pas de statistiques sur la durée pour la mobilité sortante, les détachements et disponibilités. Ceci est dû entre autres au fait que la section n'existe que depuis 8 ans. Néanmoins, nous constatons que, au 1^{er} janvier 2019, 41 chercheurs sont en détachement ou disponibilité. Ces 41 chercheurs n'ont pas été pris en compte dans les statistiques précédentes. Il s'agit d'une proportion importante de l'effectif : 12%. Nous donnons dans le tableau 6 une répartition de cet effectif par année de première demande de détachement ou disponibilité.

Par ailleurs, parmi les chercheurs examinés par la section durant son début de mandat (en principe, ceci concerne la totalité des chercheurs de la section puisqu'il y a une évaluation tous les 2,5 ans), nous avons observé très peu de cas de chercheurs revenus dans la section après un détachement ou disponibilité. On ne peut que conjecturer, au vu du tableau 6, qu'en moyenne 4 nouveaux détachements sont demandés chaque année et donc qu'en moyenne au moins 4 chercheurs quittent la section chaque année, en plus des départs à la retraite. Ceci doit être complété par les autres départs (définitifs), par exemple vers l'enseignement supérieur.

Tableau 6 – Année de première demande de détachement

< 1999	1999 - 2002	2003 - 2006	2007 - 2010	2011 - 2014	2015 - 2018
1	2	3	4	17	14

Tableau 7 – Chercheurs en activité en section 6, par année de recrutement

2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
4	20	17	6	9	18	14	19	14	15

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
9	12	12	10	8	11	14	10	10	8

B. Concours de recrutement CR

Le tableau 7 rapporte le nombre de chercheurs actuellement en activité dans la section 6, par année de recrutement. Bien que la section, dans son périmètre actuel, n'existe que depuis moins de 10 ans, ce tableau donne les recrutements de chercheurs dont l'activité relève aujourd'hui de la section 6 et donne donc un panorama fidèle de l'évolution des recrutements dans la section.

On observe une très nette diminution des recrutements durant la dernière décennie (environ moitié moins de recrutements). Ceci s'observe aussi sur la pyramide des âges (tableau 8) : on observe un pic autour de 36 - 40 ans, alors que, comme nous le voyons dans le tableau 9, l'âge moyen de recrutement est inférieur à 31 ans.

Âge/Ancienneté des chercheurs recrutés en section 6. L'âge n'est cependant pas une donnée pertinente (et n'est pas pris en compte dans le recrutement). L'*ancienneté en recherche* est une mesure plus pertinente : elle est appréciée comme le nombre d'années consacrées à la recherche, thèse et éventuels stages pré-doctoraux compris.⁴ Avec cette mesure, le tableau 10 rap-

porte l'ancienneté, évaluée au moment d'un recrutement éventuel, des candidats déclarés admissibles par le jury. Elle montre la volonté du jury de privilégier les recrutements de chercheurs dont l'ancienneté est inférieure à 7 ans. Le tableau montre aussi que la fusion des grades CR2 et CR1 a eu pour conséquence un abaissement de l'ancienneté moyenne en recherche des chercheurs recrutés.

Nombre de candidatures. Le tableau 11 rapporte le nombre de candidatures et le nombre de candidats ces 3 dernières années. Le nombre de candidatures totalise les candidatures CR1 et CR2 quand la distinction existait, mais n'inclut pas les concours spécifiques (car en général les candidats sur les concours spécifiques sont aussi candidats sur le concours général). Comme il peut y avoir des candidats à la fois sur un concours CR2 et un concours CR1, quand c'est possible, nous indiquons le nombre total de candidats (donc inférieur au nombre de candidatures). Le nombre de postes en revanche tient compte de tous les concours CR.

Candidatures d'étrangers. Le tableau 12 rapporte les lieux de thèse des candidatures ces trois dernières années. Les thèses en cotutelle peuvent entraîner des nombres fraction-

4. Les maternités sont prises en compte pour ce calcul.

Tableau 8 – Pyramide des âges en section 6

	< 31	31 - 35	36 - 40	41 - 45	46 - 50	51 - 55	56 - 60	> 60
Total	14	39	67	60	47	35	28	13
CR	14	39	59	32	14	11	7	4
DR	0	0	8	28	33	24	21	9

Tableau 9 – Âge des chercheurs lors de leur recrutement comme CR en section 6

Âge	< 28	28 - 31	32 - 35	> 35
Proportion (2012 - 2016)	9%	56%	26%	9%
Proportion (2017 - 2018)	10%	60%	30%	0%

Tableau 10 – Ancienneté en recherche des candidats admissibles aux concours CR, évaluée au moment de leur possible recrutement

Ancienneté en recherche (thèse comprise)	4	5	6	7	8	9	10	> 10
2017	2	7	4	5	1	0	2	1
2018	2	7	3	6	2	1	1	0
2019	2	5	3	2	1	1	0	0
Total	6	16	11	13	4	2	3	1

Tableau 11 – Nombres de candidats et candidatures aux concours CR en section 6

	2013	2014	2015	2016	2017	2018	2019
Nombre de candidats					139	119	115
Nombre de candidatures	213	220	159	154	158	119	115
Retenus pour audition		90	63		40	37	32
Admissibles	21	20	21	17	22	22	16
Postes aux concours	7	8	8	9	8	11	8

Tableau 12 – Origine des thèses des candidats CR

	2017	2018	2019	Total
Amérique du Nord	19	6	4	29
Amérique du Sud	3	5	4.5	12.5
Asie	0	4	4.5	8.5
Europe hors UE	11	3	2	16
Océanie	0	2	1	3
Union Européenne	39.5	32	15.5	87
Total hors France	72.5	52	31.5	156
France	126.5	80	82.5	289

Tableau 13 – Succès aux concours CR des candidats ayant soutenu une thèse à l'étranger

	Candidatures	Auditionnés	Admissibles	Recrutés (en 06)
2017	36.4%	45.0%	45.4%	37.5%
2018	39.4%	29.7%	31.8%	18.0%
2019	27.6%	28.1%	18.8%	12.5%

naires.⁵ Le tableau 13 donne les statistiques du concours pour les candidats ayant soutenu une thèse hors de France. Ces tableaux indiquent une forte baisse du nombre de candidats étrangers. Ceux-ci réussissent en revanche plutôt bien le concours, mais le nombre de démissions est plus important.

Répartition thématique des candidats et politique d'affichage des thèmes prioritaires.

L'institut a mené ces dernières années une politique volontariste d'affichage de thèmes prioritaires pour le recrutement. Le tableau 14 indique la répartition des candidats par thème prioritaire (ou hors de thèmes prioritaires).

C. Promotions

Les tableaux 15 et 16 donnent respectivement les délais pour les passages aux grades DR2 et DR1. Le délai 0 correspond à un recrutement directement dans le grade.

D. Répartition géographique des chercheurs de la section

Le tableau 17 rapporte l'effectif des chercheurs rattachés à la section 6 dans les laboratoires ayant au moins 4 chercheurs de la section, avec la localisation géographique de chaque unité (pour la région parisienne, nous distinguons Paris Centre, Paris Est, Paris Nord, Paris Sud).

5. Les candidatures multiples, soit sur plusieurs concours, soit plusieurs années consécutives sont comptées plusieurs fois. Par ailleurs les totaux peuvent être inférieurs au nombre de candidatures, à cause de certaines informations manquantes au moment de l'écriture de ce rapport.

Tableau 14 – Candidats sur des thèmes prioritaires aux concours CR généralistes

	2017	2018	2019
Réseaux et systèmes	9	12	8
Science des données	16	9	
Calcul parallèle et haute performance^a		6	5
Sciences du logiciel	7	8	7
Sécurité informatique^b	13	9	4
Intelligence artificielle^c			7
Tous les thèmes prioritaires	39	42	35
Hors thèmes prioritaires	81	77	80
Total candidats^d	120	119	115

^a sauf 2017^b hors cryptologie pour 2019^c seulement 2019^d Comme certains candidats émargent à plusieurs thèmes prioritaires, le total des candidats sur au moins un thème prioritaire est strictement inférieur à la somme des candidats sur chacun des thèmes prioritaires.

Tableau 15 – Ancienneté au CNRS lors du recrutement DR2

Ancienneté (années)	0	1 - 5	6 - 8	9 - 11	12 - 14	15 - 17	18 - 20	> 20
Effectif	11	4	16	36	26	12	7	6

Tableau 16 – Ancienneté dans le grade DR2 lors d'une promotion DR1

Ancienneté (années)	0	1 - 4	5 - 7	8 - 10	11 - 13	14 - 16	> 16
Effectif	2	0	13	13	5	5	1

Tableau 17 – Répartition des chercheurs de la section dans les unités ayant ≥ 4 chercheurs

Unité	IRIF	LaBRI	LAAS	LIP	IRIT	LIP6	LIRMM	LIX
Effectif	Paris C 24	Bordeaux 23	Toulouse 17	Lyon 17	Toulouse 16	Paris C 16	Montpellier 16	Paris S 16
Unité	LIG	IRISA	LORIA	LRI	G-SCOP	I3S	LIGM	LSV
Effectif	Grenoble 13	Rennes 12	Nancy 11	Paris S 10	Grenoble 8	Nice 8	Paris E 8	Paris S 8
Unité	CRISTAL	LAMSADE	LIPN	DI ENS	CRIL	LS2N	VERIMAG	
Effectif	Lille 7	Paris C 7	Paris N 7	Paris C 6	Lens 4	Nantes 4	Grenoble 4	

Conclusion

Les points saillants de ces statistiques sont d'une part la baisse alarmante des candidatures féminines et d'autre part le faible nombre de postes au concours : pour que la section 6 (informatique) conserve un effectif constant, il faudrait entre 11 et 14 recrutements par an, bien au delà des postes mis au concours.⁶

La discipline a beaucoup d'opportunités pluri-disciplinaires, qui doivent alimenter la recherche coeur de métier. Parmi les thèmes émergents qui alimentent les recherches dans la discipline : exploitation de masses de données, respect de la vie privée, prise en compte de la consommation d'énergie, distribution massive de données et de programmes, explicabilité de l'IA, ordinateurs quantiques... L'évolution rapide des technologies présente de nouveaux défis à la discipline, qui alimentent les recherches du périmètre actuel de la section, qui n'a sans doute pas vocation à évoluer significativement.

6. Cette estimation est obtenue en divisant le nombre de membres de la section par le temps moyen qu'un chercheur reste dans la section. Ce dernier chiffre ne peut pas être calculé de manière précise en l'absence de statistiques sur le long terme ; il est compris entre 23 ans et 28 ans.