

SECTION 06

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

Composition de la section

Nathalie APPEL; Jérémie BOURDON (jusque juin 2024); Nicolas BOUSQUET (jusque juin 2024); Johanne COHEN; Antoine GENITRINI; Patricia GEORGEON; Yves GRANDVALET; Katria JAFFRÈS-RUNSER; Arnaud LEGRAND; Damian MARKHAM (depuis février 2024); Anca MUSCHOLL; Anastasia PAPARRIZOU; Loïc PAULEVÉ (depuis septembre 2024); Michaël POSS; Maria POTOP-BUTUCARU (jusque juin 2023); Jean-Florent RAYMOND (depuis octobre 2024); Romain ROUVOY; Yolande SALLENT; Pierre SENELLART (président de section); Thomas SEILLER (secrétaire scientifique); Ye-Qiong SONG; Alain TCHANA; Hélène WAESELYNCK

Résumé

La section 6 du Comité national de la recherche scientifique est, avec la section 7, une des deux sections traitant de la science informatique, et plus précisément de l'algorithmique et de la combinatoire, du calcul, du logiciel, de la sécurité, des réseaux et systèmes distribués, des données et connaissances, de l'intelligence artificielle et de l'aide à la décision, ainsi que de la bio-informatique et de l'informatique quantique. Ce rapport présente le périmètre thématique de la section, discute de la place des femmes ainsi que des évolutions récentes des pratiques de recherche, présente la conjoncture des différents thèmes de recherche et enfin décrit les carrières au CNRS des chercheurs et chercheuses de la section.

Introduction

Nous avons choisi dans ce rapport, comme dans celui de la précédente mandature, 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 2024. Nous proposons aussi des statistiques sur les concours de recrutement et les carrières des personnels de la section, à partir des observations effectuées lors des trois années de mandature écoulées. Cette collecte d'informations peut comporter des erreurs, que nous espérons marginales. La part de l'interprétation et de la description des thèmes de recherche et de leurs évolutions

est de ce fait plus réduite. Quand cela est possible, nous affichons les mêmes données que dans le rapport de conjoncture de la section 6 établi en 2019 [CPA⁺21] pour faciliter les comparaisons.

Dans la partie I, nous découpons en thèmes le périmètre couvert par la section 6. Nous comparons aussi les effectifs respectifs des chercheurs et chercheuses 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

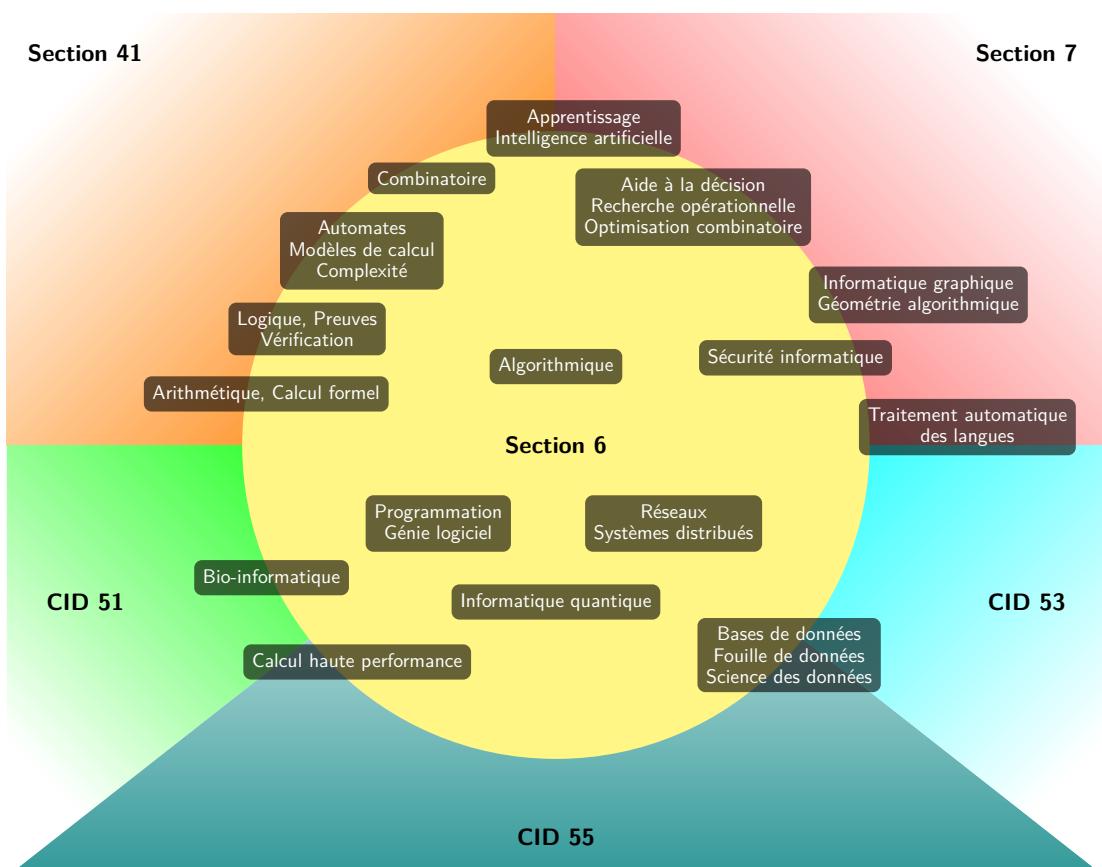


Figure 1 – Périmètre approximatif de la section

revanche nous n'observons pas de « plafond de verre ».

Dans la partie III, nous mentionnons des évolutions récentes des pratiques de la recherche dans la section, en terme de science ouverte et d'écoresponsabilité.

Dans la partie III, nous détaillons, de manière plus traditionnelle, l'évolution des thèmes de recherche de la section, en nous appuyant en particulier sur des éléments fournis par les GDR.

Enfin, dans la partie V 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 saillant est l'importante mobilité sortante dans la section.

I. Périmètre thématique

Le périmètre thématique de la section et les frontières avec 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 par exemple 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 ; les versants statistiques de l'apprentissage et de l'IA sont en premier lieu en section 7 (et certains aspects peuvent également relever de la section 41). Noter que dans la mandature à venir à compter de 2025, suite à une renumérotation des sections, le périmètre thématique de la section 6 devient celui de la section 2.

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 :

Section 6 : il s'agit de tous les chercheurs et chercheuses CNRS en activité (donc ne sont pas considérés ici les personnes émérites, en disponibilité ou en détachement, etc.) dans la section 6 au 1^{er} octobre 2024 (nous incluons également les chercheurs et chercheuses recrutés sur le concours 2024, même s'ils et elles n'avaient pas encore pris leur poste à cette date). L'étiquetage thématique a été effectué manuellement par les membres de la section, à partir des rapports d'activité. Un chercheur ou une chercheuse peut émerger 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. Quand nous donnons des statistiques sur les grades (par exemple dans le tableau 2), nous assimilons les personnes recrutées sur le concours handicap pendant leur année initiale de CDD à des CRCN, et les personnes recrutées comme DR en CDI à des DR1.

UMR-non-CNRS : il s'agit des chercheurs permanents, membres des UMR (ainsi que des UPR, IRL, UAR) rattachées à CNRS Sciences Informatiques 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 et d'enseignantes-chercheuses en section CNU 27, mais aussi par exemple de personnels Inria. Les données ont été extraites de l'annuaire CNRS des UMR mi-octobre 2024. Comme l'effectif total à considérer est très important (3 628 personnes), nous avons tiré uniformément au hasard un sous-ensemble de 25%, puis étiqueté thématiquement leur activité de recherche, en nous appuyant sur leurs publications et leur page personnelle décrivant leurs travaux. Chaque personne peut émerger à plusieurs thèmes et donc être comptée pour une fraction dans plusieurs thèmes.

Thèses : les statistiques portent sur l'ensemble des thèses soutenues en 2023 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. 465 thèses étiquetées « informatique » (sur 1 059) 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 il y a également des thèses qui relèvent des mathématiques appliquées, voire d'autres disciplines très différentes (biologie, physique, sciences sociales, etc.) 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 subjectives ou contestables.

Liens avec les autres disciplines

L'informatique est naturellement présente, sur le versant applicatif, dans de nombreux domaines

Tableau 1 – Répartition thématique des chercheurs et chercheuses de la section 6, des membres des UMR rattachées à la section et des thèses en informatique en 2023. Un échantillon d'un quart des personnels est considéré pour les membres des UMR non CNRS, comptage par fractions pour l'appartenance à plusieurs thèmes.

Thème	Chercheurs 6		UMR-non-CNRS		Thèses 2023	
	Nb	%		%	Nb	%
Algorithmique	34	10.9%		6.0%	25	4.2%
Combinatoire	35	11.3%		6.3%	19	3.2%
Arithmétique, calcul formel	12	3.7%		2.6%	10	1.6%
Calcul haute performance, calcul parallèle	10	3.2%		5.3%	21	3.5%
Automates, Modèles de calcul, Complexité	26	8.3%		6.3%	19	3.2%
Preuve, vérification, théorie de la programmation	47	15.1%		8.7%	36	6.0%
Sécurité informatique	18	5.7%		5.9%	53	9.0%
Génie de la programmation, génie logiciel	9	3.0%		6.5%	29	4.8%
Réseaux, systèmes distribués	19	6.3%		13.5%	119	20.0%
Données et connaissances	16	5.0%		11.4%	56	9.4%
Intelligence artificielle	25	8.1%		13.2%	94	15.8%
Recherche opérationnelle, aide à la décision	22	7.1%		8.5%	73	12.3%
Bio-informatique	23	7.5%		2.7%	29	4.9%
Informatique quantique	11	3.6%		0.8%	6	0.9%
Autres (section 6)	4	1.1%		2.2%	8	1.3%
Hors section 6	1				465	
Total	309				1059	

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, combinatoire, etc. À noter que chaque chercheur ou chercheuse peut choisir d'être co-évaluée par plusieurs sections et commissions interdisciplinaires (CID).

CID 51 : Modélisation, et analyse des données et des systèmes biologiques : approches informatiques, mathématiques et physiques. Des chercheuses et chercheurs à l'interface entre informatique et biologie, et dont les thématiques de recherche relèvent explicitement de la section 6, sont régulièrement recrutés et promus par la CID 51. La co-évaluation des dossiers des chercheurs par la section 6 et de la CID 51 donne une vision croisée et complémentaire des travaux réalisés. La section note que les chercheuses et chercheurs de la section qui sont clairement à cette interface ne profitent pas suffisamment de cette possibilité de co-évaluation.

CID 53 : Méthodes, pratiques et communications des sciences et des techniques. Cette commission interdisciplinaire a pour vocation le recrutement et l'évaluation de chercheurs et chercheuses travaillant à l'interface entre sciences de la nature et sciences humaines et sociales. L'impact croissant des technologies du numérique sur la société contribue à l'émergence de nouvelles problématiques dont l'étude et la compréhension nécessitent un dialogue interdisciplinaires entre les sciences humaines et sociales et les sciences informatiques. Le nombre de chercheurs travaillant à cette interface est naturellement amené à croître, et le rôle de la CID 53 pour le recrutement et l'évaluation de ces profils est essentiel.

CID 55 : Sciences et données Cette commission interdisciplinaire, créée en 2021, est principalement rattachée à CNRS Sciences Informatiques, mais vise en premier lieu à recruter des chercheurs et des chercheuses émergeant aux instituts secondaires qui la composent (CNRS Nucléaire & Particules, CNRS Terre & Univers, CNRS Sciences Humaines & Sociales, CNRS Chimie), et dont l'activité de recherche a une forte composante numérique, associant les données aux autres domaines scientifiques. Le comité de la CID 55 comporte

ainsi 2 membres de la section 6, et 2 membres de la section ayant une forte activité interdisciplinaire sont co-évalués par la CID 55.

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} octobre 2024. On observe une faible proportion de femmes, cependant en augmentation par rapport à 2019 (où la proportion était de 18,5%). 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 est en général et en moyenne 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 2023, sur des thèmes relevant de la section) et la proportion de candidatures de femmes aux concours CR ces trois dernières années (à nouveau restreintes aux thèmes de la section). Nous observons que les femmes en section 6 sont maintenant représentées à proportion similaire au CNRS et dans le reste de l'enseignement supérieur (ce n'était pas le cas en 2019, donc c'est une avancée notable). Cependant, nous nous inquiétons du faible nombre de candidatures de femmes au CNRS (12,3% en 2022, 12,0% en 2023, 20,8% en 2024 – cette dernière année était particulière avec un calendrier de concours décalé, ce qui a peut-être eu un impact), très en dessous de la proportion des thèses de doctorat soutenues par des femmes sur des thématiques relevant de la section en 2023 (24,9%, stable par rapport à 2017).

III. Évolution des Pratiques de Recherche

Tableau 2 – Effectifs par genre en section 6

Hommes	241	78,0%
Femmes	68	22,0%
CR		
Hommes	128	80,0%
Femmes	32	20,0%
DR		
Hommes	113	75,8%
Femmes	36	24,2%
CRCN		CRHC
Hommes	101	78,3%
Femmes	28	21,7%
DR2		DR1
Hommes	64	76,2%
Femmes	20	23,8%
DRCE		
Hommes	39	78,0%
Femmes	11	22,0%
	10	66,7%
	5	33,3%

Tableau 3 – Recrutements CR récents en section 6 : nombre de recrutés, nombre de personnes toujours en activité au CNRS actuellement, femmes recrutées

	Année de recrutement							
	2017	2018	2019	2020	2021	2022	2023	2024
Recrutements CR (nb)	10	11	8	11	11	7	10	10
Toujours au CNRS (nb)	8	10	6	11	10	7	10	10
Femmes recrutées CR (nb)	2	3	1	4	3	2	2	4
Femmes recrutées CR (%)	20%	27%	13%	36%	27%	29%	20%	40%

Tableau 4 – Proportion de femmes au concours DR

	2020	2021	2022	2023	2024
Candidates DR	7,1%	17,4%	21,1%	18,8%	12,6%
Recrutées DR	22,2%	12,5%	40,0%	22,2%	22,2%

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

	section 6	UMR-non-CNRS	Thèses (2023)	Candidatures CR 06/02		
				2022	2023	2024
Femmes	22.0%		22.2%	24.9%	12.3%	12.0%
						20.8%

Dans cette partie, nous résumons la perception de certaines évolutions au sein des pratiques de recherche, en s'appuyant sur une consultation des GDR sur les problématiques en lien avec la science ouverte et avec l'éco-responsabilité.

Les articles et prépublications présentant des contributions numériques sont maintenant assez fréquemment complétés de logiciels mis à disposition librement, plus souvent avec le code source que sous forme exécutable. La plate-forme privilégiée pour ce partage est GitHub, suivie de loin par GitLab.com, les GitLabs institutionnels et Zenodo. De façon moins systématique, les chercheurs et chercheuses mettent également à disposition des jeux de données, plus fréquemment brutes que post-traitées, sous des formes et sur des plates-formes plus diverses que dans le cas des logiciels. Dans certaines communautés, quand cela s'y prête, des jeux de données de référence sont également exploités.

Du côté des supports de publications, la grande majorité des communautés mettent à disposition leurs articles sur des serveurs d'archivage tels que Hal. Alors que des revues reconnues en accès ouvert diamant (sans paiement par les auteurs) existent et sont reconnues dans de nombreux domaines, elles restent marginales ou inexistantes dans d'autres. La plupart des actes de conférences sont édités par Springer et ACM, suivis de LIPics et IEEE. Enfin, l'utilisation de rapports de relecture transparents reste très rare, avec cependant quelques communautés où cela semble être devenu la norme. On peut également noter l'émergence d'initiatives originales en science ouverte, telle que les *Peer Community In*, qui valorisent le travail éditorial en lui attribuant un DOI et permettent une transparence totale sur le processus de relecture (par exemple *PCI Mathematical and Computational Biology*). Les pratiques des informaticiens et informaticiennes de la section 6 semblent donc dans l'ensemble suivre une trajectoire en accord avec les politiques nationales de science ouverte.

CNRS Sciences Informatiques ayant colorié plusieurs années de suite des postes en lien avec l'informatique éco-responsable, il nous a semblé légitime d'également sonder les GDR sur leur perception de l'évolution de cette problématique. Si nous avons reçu moins de retours que pour l'ouverture de la science, il apparaît assez clairement que la crise environnementale en cours a com-

mencé à affecter les pratiques de recherche, en particulier sur la façon de collaborer et de publier dont certaines communautés commencent à s'emparer. Sur le plan des expériences ou des moyens de calculs nécessaires à la recherche, certaines communautés considèrent la problématique comme pertinente mais non investie. Dans l'ensemble, l'évolution des pratiques semble freinée par les injonctions contradictoires (budgets à durée limitée, survalorisation des collaborations internationales et des conférences) et l'acceptabilité des contraintes qui en découlent (fatigue, surcoût, qualité des échanges amoindrie) et se rajoutent à l'existant (lourdeurs administratives, manque de moyens) en l'absence d'alternatives acceptables (transport, importance des conférences), en particulier dans un contexte où les pratiques des autres pays n'évoluent pas.

Sur le plan de l'évolution des thématiques de recherche, la consultation indique que les problématiques, telles qu'identifiées par le GDS EcoInfo, d'*efficacité* (optimisation de l'existant) et d'*évaluation* (mesure et modélisation des impacts) sont relativement présentes dans les GDR RO et RSD et commencent à émerger dans certaines autres communautés, alors que celles liées à la *résilience* (situations de rupture) et la *sobriété* (redéfinition des besoins, réorientation des usages) semblent dans l'ensemble considérées comme pertinentes mais non investies (à l'exception du GDR RO). Notons que les chercheurs et chercheuses réorientant leurs thématiques de recherche et ne trouvant pas leur compte dans leur communauté d'origine ont pu trouver un environnement plus favorable dans le GDS EcoInfo, qui est d'ailleurs amené à évoluer en GDRS. Si le développement d'une communauté plus à même de mener des recherches souvent pluri-disciplinaires est salué, il serait dommage que les communautés des disciplines historiques de l'informatique ne s'emparent pas de ces sujets.

Quoi qu'il en soit, l'ensemble de ces problématiques (pratiques et thématiques de recherche) semble induire des tensions (parfois générationnelles) ainsi qu'un certain désarroi.

IV. Conjoncture par thème

A. Algorithmique

Les algorithmes, au cœur des thématiques abordées par la section 6, sont analysés sous divers angles dans la majorité des GDR de cette section. Historiquement ancrés au sein des mathématiques — pour une analyse rigoureuse de leurs bornes inférieures et complexités, ou pour des applications à la géométrie et l’algèbre — leurs champs d’application s’étendent désormais à des disciplines variées, comme la chimie ou l’économie.

Compte tenu de la diversité des domaines d’application, l’algorithmique est étudiée sous différentes perspectives au sein de la section 6, interagissant avec d’autres thématiques comme la théorie des graphes, les sciences des données, l’intelligence artificielle (IA), la bio-informatique, les réseaux et le traitement d’images. Par ailleurs, les motivations pour concevoir et étudier des algorithmes peuvent aussi bien être pratiques que théoriques. Cette diversité impose une adaptation continue des approches algorithmiques, en fonction de contraintes telles que la connaissance partielle ou complète des instances, l’existence ou non d’aléa, les ressources limitées, ou les acteurs potentiellement impliqués dans l’exécution des algorithmes. Cela a conduit au développement de nouvelles catégories d’algorithmes, incluant les algorithmes d’approximation, de streaming, paramétrés, online, ainsi que ceux intégrant des mécanismes de prédiction.

Enfin, l’algorithmique constitue également un objet d’étude en soi, en particulier à travers des recherches consacrées à l’évaluation des impacts environnementaux et aux interactions entre les algorithmes et la société.

L’algorithmique est un des thèmes les mieux représentés au CNRS, comme en témoigne le tableau 1.

B. Combinatoire

La combinatoire est une thématique à l’interface entre mathématiques et informatique et a pour objectif l’étude d’objets, de modèles et de méthodes issus d’un contexte discret. La thématique se partage entre plusieurs directions autour des combinatoires algébrique, énumérative, bijective ou analytique, mais elle se déploie également, plus spécifiquement, dans le contexte de la théorie des graphes. La combinatoire entretient de plus des

liens étroits avec le calcul formel. Les problèmes d’application centraux proviennent de l’informatique. On peut citer par exemple le développement et l’étude de structures de données, l’analyse d’algorithmes ou les théories de la calculabilité ou de la complexité. Les interactions sont riches et variées, notamment avec la bioinformatique et la physique (statistique), mais aussi avec de multiples aspects des mathématiques (probabilités, géométrie, algèbre, théorie des nombres, systèmes dynamiques). La thématique est principalement représentée par des chercheuses et chercheurs membres du nouveau GDR IFM.

Les sujets actuels concernent entre autres la convergence d’objets aléatoires vers des modèles probabilistes continus et la combinatoire des cartes planaires. Par ailleurs, l’émergence de nouvelles largeurs de graphes en lien avec les métathéorèmes algorithmiques (à l’interface avec la logique) font émerger de nouvelles directions de recherche en théorie structurelle et algorithmique des graphes. Les approches d’algorithmique paramétrées et d’approximation sont également en effervescence. Enfin, les outils d’aléa discret et d’analyse se diversifient et proposent désormais des analyses pour l’analyse de systèmes complexes. Au cœur des systèmes dynamiques, la combinatoire des mots entretient des liens étroits avec l’étude des systèmes dynamiques symboliques, des pavages et des structures quasi-périodiques. La combinatoire est une thématique historique de l’informatique en France, avec une grande visibilité et reconnaissance internationale. Elle est bien représentée au CNRS et l’on constate un excellent niveau des candidatures aux concours.

C. Arithmétique et calcul formel

Le calcul formel et l’arithmétique des ordinateurs s’intéressent au développement, à l’algorithmique et à l’implémentation d’opérations élémentaires sur les objets mathématiques de base, comme les nombres (entiers, algébriques, réels, flottants), les polynômes, les matrices, etc. Les systèmes de calcul formel permettent de manipuler des formules mathématiques de façon symbolique ; les résultats sont exacts, mais souvent aussi plus coûteux à calculer et l’efficacité est un souci permanent. Dans le cadre de l’arithmétique des ordinateurs, les garanties sur les arrondis ou sur la vitesse d’exécution sont à prendre en compte. Les

principaux enjeux actuels relèvent de la conception de solutions (algorithmiques, logicielles et matérielles) allant de l'embarqué jusqu'aux supercalculateurs, du développement d'une vision hybride symbolique, certifiée et numérique, de la question de la reproductibilité des résultats, ainsi que de l'implication de méthodes de preuve formelle. L'efficacité, la précision, la fiabilité, la certification, et la sécurité sont ainsi des questions clé dans ce contexte. On a observé récemment un foisonnement de résultats autour de la multiplication des entiers en temps quasi-linéaire, ou de l'algorithmique des polynômes (résultants bivariés, composition modulaire), le tout accompagné d'un investissement logiciel majeur. Les domaines d'application sont variés et vont des mathématiques effectives, de la visualisation d'objets mathématiques complexes au spatial, en passant par la cryptographie.

D. Calcul parallèle et distribué, calcul haute performance

À des fins d'efficacité énergétiques, les plateformes de calcul et de traitement de données (supercalculateurs, *data centers*, *clouds*) sont devenues largement hétérogènes (plusieurs CPUs, GPUs, TPUs au sein d'un même nœud de calcul, plusieurs types de nœuds dans une même infrastructure) et distribuées (hiérarchies mémoire, réseaux d'interconnexion réguliers ou pas). Cette thématique recouvre un large spectre de contextes et de problèmes allant de l'exploitation efficace d'une machine à des calculs complexes sur des architectures massivement parallèles et hétérogènes. À une telle échelle, l'hétérogénéité et la diversité des ressources induit une tension entre les besoins de virtualisation (pour cacher cette complexité) et d'exposition (pour permettre des optimisations spécifiques et adaptées au contexte) qui se traduit par une complexité croissante des applications et des intergiciels permettant de les exploiter, et ce autant sur le plan algorithmique que logiciel et système. Le calcul parallèle, thème historique de l'informatique en France, est donc présent essentiellement dans les GDR Calcul, IM, RSD et GPL.

Dans un contexte de demande croissante de puissance de traitement, que ça soit de la part des scientifiques, de l'industrie, ou de la société civile, la communauté a connu une forte évolution liée aux architectures *exascale* et *hyperscale* et à l'arri-

vée des grandes masses de données et de l'apprentissage automatique. Si à ses débuts, le HPC était souvent associé aux grandes simulations pour la physique ou la chimie, l'enjeu est maintenant de nourrir ces simulations et codes de prédictions à l'aide de données issues de grands instruments ou de capteurs distribués à large échelle tout en maîtrisant les coûts opérationnels. Un point notable au sujet de ces nouvelles charges applicatives en lien avec l'IA, est qu'au delà de leurs différences structurelles, les besoins en terme de précision et de stabilité numérique sont moindres, ce qui conduit à un regain d'intérêt pour les calculs en précisions mixtes et donne également plus de souplesse sur le plan algorithmique (asynchronisme, structures plus maléables, etc.). Dans ce contexte, les paradigmes de programmation modernes, comme la programmation par tâches, apportent une réponse pertinente aux problèmes d'exploitation des ressources tout en simplifiant la conception des applications.

Le PEPR *Numérique pour l'Exascale* alimente et structure les recherche autour de ces enjeux (mouvements de données, gestion des entrées/sorties, algorithmes et schémas numériques adaptés, auto-paramétrage, modèles de programmation, composabilité, adaptabilité, etc.) avec une volonté de développement de piles logicielles qui puissent à terme s'intégrer dans les supercalculateurs exaflopique français et européens ou dans des infrastructures liées à des projets d'astronomie comme SKA. La communauté se structure aussi autour des plates-formes nationales telles que Jean Zay ou Grid'5000 (dont les 20 ans ont été fêtés en mai 2023) et maintenant SILECS, qui construit sur Grid'5000 et sur FIT IoT-Lab, et qui s'intégrera dans la plate-forme européenne SLICES-RI et permettra de renforcer les liens entre les chercheur(euse)s 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). Des réflexions autour de la création d'un GDR HPC sont en cours et permettraient de continuer à structurer une communauté qui rassemble aussi bien informaticien·ne·s que mathématicien·ne·s appliqués, physicien·ne·s, chimistes, biologistes, etc. Enfin, la question de l'efficacité énergétique continue de faire l'objet d'une action spécifique du GDR RSD en lien avec le GdS Ecoinf.

Malgré la politique volontariste du CNRS pour renforcer ce thème qui reste assez peu représenté

au CNRS, le nombre de candidat·e·s au concours CR reste très limité.

E. Modèles de calcul, complexité, automates

Les thématiques autour des modèles de calcul et la complexité sont abordées par une large communauté de la section 6 et touchent à de nombreux autres domaines parmi lesquels l’algorithmique, la logique, les fondements des bases de données, les fondements de la programmation, l’informatique quantique, ou les systèmes dynamiques. La discipline est transversale dans la section, en relation privilégiée avec les mathématiques au niveau des méthodes et des pratiques. Les thématiques s’inscrivent principalement au sein du GDR IFM et partagent des liens étroits avec les GDR RADIA et BIMM.

Le paysage français des activités de recherche en calculabilité est vaste : il va de l’étude de modèles de calcul moins conventionnels (automates cellulaires, équations différentielles, etc.) à des problèmes d’axiomatisation (mathématiques à rebours pour des théorèmes combinatoires de style Ramsey). Les probabilités jouent également un rôle fondamental, par exemple dans les liens entre la complexité de Kolmogorov et la théorie de l’information (protocoles de partage de clé secrète), ou encore à travers des notions d’aléatoire plus fortes, basées sur des stratégies (martingales) calculables de manière probabiliste.

La complexité est abordé par les chercheurs de la section également sous des angles très variés. Une avancée majeure récente a été obtenue dans la domaine de la complexité algébrique, en obtenant des bornes inférieures superpolynomiales pour les circuits algébriques de profondeur constante. Dans le domaine de la complexité implicite, où on s’intéresse à des ressources de programmes (temps, espace), on veut obtenir des systèmes de types avec des garanties sur la complexité des ressources. Un résultat récent a été une caractérisation des fonctions des second ordre calculables en temps polynomial par une classe de programmes combinant des aspects fonctionnels et impératifs de manière stratifiée. D’autres avancées récentes de chercheurs de la section sont liées à la théorie des bases de données au sens large, en lien avec la représentation de connaissances, ou le problème de satisfaction de contraintes, ou encore les problèmes d’énun-

mération pour les requêtes. La grande diversité des recherches dans le domaine de la complexité a été bien résumée par les *Complexity Days* qui ont eu lieu en 2023 sous l’égide du GDR IFM.

Le domaine partage aussi des liens privilégiés avec des disciplines telles que la physique et la biologie, à travers l’analyse des systèmes naturels complexes, et l’explication et la prédition 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 traditionnellement des liens forts. Les objectifs et méthodologies regroupent l’analyse et la certification de systèmes et programmes (analyse statique, vérification de modèles...), la preuve formelle (théorie de la preuve, démonstration automatique, assistants de preuve, etc.), et la théorie de la programmation (théorie des types, langages concurrents, probabilistes, quantitatifs, etc.).

Les recherches sur ces thèmes s’inscrivent à titre principal dans le GDR IFM (GT Daal, LHC, Scalp, Verif). Elles ont de fortes interactions avec le GDR GPL (en particulier autour des langages de programmation, de la compilation certifiée, du test, etc.), et également avec le GDR Sécurité (autour des 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, et le noyau commun à de nombreuses interactions est la logique.

On a vu ces dernières années quelques importantes évolutions thématiques. Les aspects probabilistes prennent une importance croissante dans la sémantique et l’analyse de programmes, ainsi que dans leur synthèse. Les interactions entre vérification et apprentissage automatique émergent sous la forme de « apprentissage par renforcement », un concept qui est étroitement lié aux techniques de synthèse basées sur les jeux sous forme de processus de décision markoviens. L’analyse de systèmes concurrents voit une ouverture vers l’algorithmique distribuée. Sur le plan théorique on a vu des avancées fondamentales sur la complexité des réseaux de Petri. On a vu également des interactions importantes entre la sémantique des langages de programmation (avec ses outils catégoriques, algébriques, topologiques, etc.) et la vérification

de modèles : analyse des programmes d'ordre supérieur, preuves circulaires, etc. Les assistants de preuve formelle (Coq) et les outils de démonstration automatique (SAT, SMT) ont également un impact croissant dans tous ces domaines.

Ce thème de recherche dynamique est très bien représenté, aussi bien dans les UMR (8,7% selon notre échantillonnage) que dans la section 6 (15,1%). 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 sécurité informatique française a connu un développement récent majeur de par sa structuration dans le GDR *Sécurité Informatique*. Le GDR SI a été créé en tant que pré-GDR en 2016, confirmé en 2019, et renouvelé 2024. Le GDR SI rassemble des chercheurs des sections 6 et 7 sur des thèmes qui étaient auparavant éparsillés dans d'autres GDR (IFM, GPL, SOC2, ISIS, RSD) : codes et cryptographie (GT commun avec le GDR IFM), méthodes formelles pour la sécurité, protection de la vie privée, sécurité et données multimédia, sécurité système, logiciel et réseaux, et sécurité des systèmes matériels. La section 6 y est majoritaire et surtout présente en cryptologie et méthodes formelles pour la sécurité. Les recrutements récents, et la politique de coloriage au concours, ont renforcé la présence du CNRS, en particulier sur les thèmes autour de la cryptologie ; ce renforcement a été moins efficace sur les autres aspects de la cybersécurité.

Spécifiquement en cryptologie, la période est marquée par l'effervescence et l'émulation suscitées par la compétition du *National Institute of Standards and Technology* (NIST) qui a lancé en 2017 un concours de cryptographie post-quantique qui a sélectionné en 2022 quatre propositions pour standardisation, et qui a complété l'appel en lançant en 2022 une nouvelle compétition uniquement sur les signatures. Les cryptosystèmes post-quantiques doivent pouvoir être implémentés sur de l'électronique classique tout en résistant à la puissance de calcul des ordinateurs quantiques. Ces appels à propositions doivent aboutir à la standardisation des mécanismes d'échange de clés et de signatures numériques appelés à être utilisés dans les décennies à venir. Les équipes françaises

sont particulièrement bien placées parmi les finalistes ; la communauté française participe activement à la conception et à l'analyse de sécurité des primitives mais aussi à leur cryptanalyse. Dans ce cadre s'est développée une cryptographie à base de codes correcteurs d'erreurs, de réseaux euclidiens (avec par exemple des réseaux définis par des idéaux de corps de nombres), d'isogénies, de systèmes polynomiaux multivariés, ou encore de fonctions de hachages. La cryptographie légère est également un thème en plein développement avec le développement d'outils pensés pour des systèmes sous contraintes de ressources (énergétiques, calculatoires) comme l'internet des objets. La recherche de générateurs pseudo-aléatoires pour la cryptographie, également cruciale pour l'implémentation de cryptosystèmes, suscite également une recherche féconde.

Le domaine de la sécurité informatique est très visible en France, et très bien soutenu. Le PEPR Cybersécurité est un outil financier très puissant qui a aidé ce domaine à acquérir une dynamique importante.

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 16 ans (2008). 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.

Malgré une baisse du nombre de soutenances de thèses obversées sur la période passée (comparer le tableau 1 avec celui de [CPA⁺²¹]), 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. En revanche, le pourcentage de chercheurs dans la section demeure excessive-

ment faible (3%), malgré un coloriage Sciences du Logiciel proposé par CNRS Sciences Informatique depuis plusieurs années. Depuis 3 ans, 1 chargé de recherche et 1 directeur de recherche ont été recrutés sur ce thème. Néanmoins, la question du vivier de candidatures demeure une préoccupation pour la section avec un nombre très faible de candidatures reçues (voir tableau 13).

I. Réseaux et systèmes distribués

Adapter nos systèmes et infrastructures aux défis d'une demande numérique massive, mais contrainte en ressources et ceci pour un éventail de contextes applicatifs très large (systèmes critiques, à haute disponibilité, à qualité de service garantie, non-terrestre, ubiquitaire) anime les recherches de la communauté en réseaux et systèmes distribués. Ce thème est marqué par une évolution importante vers des infrastructures de plus en plus programmables via des technologies de virtualisation (machine virtuelle, conteneurisation, réseaux programmables SDN/NFV). Ces systèmes offrent ainsi plus de flexibilité dans l'allocation des ressources, que ce soit dans les infrastructures hébergées dans des centres de calcul (services cloud, réseaux opérés) que dans les unités réparties en bordure (*edge computing*) ou dans les systèmes mobiles (*fog computing*). Les recherches s'attachent à offrir une palette de services plus réactifs, capables d'intégrer de plus en plus d'objets connectés (IoT) et de faire coopérer des réseaux et systèmes possiblement hétérogènes, tout en rationalisant l'usage des ressources généralement limitées. Les défis soulevés correspondent à l'allocation et ordonnancement efficace et optimale des ressources et services des réseaux et infrastructures distribués, la conception des nouveaux protocoles et services, ainsi que des algorithmes distribués capables d'offrir les garanties en termes de qualité de service, résilience aux fautes et aux cyber-attaques. L'efficacité énergétique dans ces réseaux et systèmes constitue aussi une thématique d'actualité majeure. Pour relever ces défis, la communauté développe des méthodes et outils en algorithme distribuée, ordonnancement, optimisation et évaluation de performances. La section observe un recours de plus en plus fréquent à des outils d'intelligence artificielle (apprentissage profond, apprentissage par renforcement et fédéré), que ce soit dans les recherches en réseaux, en cy-

bersécurité ou en optimisation des systèmes. On souligne des sujets émergents en réseaux autour de la 6G, des réseaux de drones et de robots, du wireless sensing, des réseaux quantiques, réseaux non-terrestres, les métasurfaces.

La thématique est principalement représentée par des chercheurs et chercheuses du GDR RSD (Réseaux et Systèmes Distribués). Les recherches aux interfaces avec les autres GDR existent avec le GDR SoC2, le GdRI EcoInfo et GDR SI (sécurité informatique). La communauté RSD se compose en grande majorité d'enseignants-chercheurs, ces derniers contribuant fortement aux besoins de formation de spécialistes en réseaux et systèmes dont l'industrie a particulièrement besoin. Le flux de recrutements d'enseignants-chercheurs dans la thématique RSD dans les universités et écoles est donc soutenu dans les universités et écoles d'ingénieurs. *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 première en nombre d'enseignants-chercheurs et de doctorants (6,3% au CNRS, 13,5% dans l'enseignement supérieur et 20% des docteurs). Grâce à un coloriage « Réseaux et systèmes distribués » proposé par CNRS Sciences Informatiques, 3 jeunes chargé(e)s de recherche dans la thématique RSD ont été recrutés ces 3 dernières années. En revanche, le comité constate que 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 liées 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.

Les défis associés à ce thème concernent le vo-

lume des données, leur hétérogénéité (en terme de formats, de contenus, de schémas, de granularité), l'estimation de leur qualité, le fait qu'elles puissent être difficile d'acquisitions. Beaucoup des recherches sur ces thèmes s'appliquent directement à des données du monde réelle ou sont motivées par de telles applications.

Une évolution majeure et très récente dans ce domaine a été l'avènement de *modèles de fondation*, des modèles génériques obtenus par apprentissage profond sur un très grand volume de données, tels que les grands modèles de langue. Ces modèles de fondation peuvent être directement appliqués ou adaptés pour un grand nombre de tâches d'extraction, de fouille ou de recherche d'information qui nécessitaient auparavant le développement de modèles ad-hoc.

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.

Ce thème est celui (voir tableau 1) où la proportion des personnels CNRS rapportés à la proportion des personnels des UMR non-CNRS est la plus faible (5.0% vs 11.4%).

K. Intelligence artificielle

Le terme d'intelligence artificielle est aujourd'hui rentré dans le vocabulaire courant pour faire référence à des thèmes scientifiques variés de l'informatique et des mathématiques. Le thème « intelligence artificielle » de la section 6 fait référence aux approches centrées sur :

- la représentation des connaissances, le raisonnement logique et le raisonnement en présence d'incertitude,
- la résolution de problèmes combinatoires par des méthodes génériques pour la satisfaction de contraintes, la planification ou la recherche heuristique,
- les systèmes multi-agents, l'interaction humain-IA et le choix social,
- l'apprentissage symbolique.

Les approches centrées sur l'apprentissage statistique relèvent principalement de la section 7, sauf quand elles rejoignent d'autres thématiques de la section, en étant par exemple un objet d'étude pertinent pour la thématique « algorithmique », en étant formulées comme des problèmes d'optimisation mathématique, ou en manipulant les objets d'étude d'autres thématiques de la section 6.

Les succès de l'apprentissage statistique, et en particulier de l'apprentissage profond, ont fortement impacté la recherche en IA sur les approches de la section 6, en motivant des recherches sur l'explicabilité et la confiance des systèmes d'IA basées sur le raisonnement logique, la modélisation des incertitudes ou l'argumentation. Les approches neuro-symboliques, qui combinent apprentissage et raisonnement, s'appuient sur des techniques d'IA symbolique pour la résolution de problèmes complexes, en ajoutant des capacités de raisonnement, en rendant les décisions plus explicables, et en intégrant des connaissances dans les approches d'apprentissage. Par ailleurs, l'impact croissant de l'IA sur la société amène à aborder des questions éthiques avec les formalismes développés dans la section 6.

Le GDR RADIA *Raisonnement, Apprentissage, et Décision en Intelligence Artificielle*, qui a pris la succession du GDR IA depuis 2022, est structurant pour la communauté. Ce GDR a mis en place 6 groupes de travail propres dont 1 spécifiquement sur les modèles hybrides (symbolique-numérique ou connaissances-raisonnement-apprentissage) et 3 autres qui mentionnent ce croisement entre apprentissage et raisonnement (apprentissage et programmation par contraintes, explicabilité et confiance, IA et jeux). Les 3 autres groupes sont communs avec les GDR BIM (sur la Biologie Systémique Symbolique) et ROD (sur Contraintes, RO, et IA, ainsi que sur la Théorie Algorithmique de la Décision et des Jeux).

Le tableau 1 montre que ce thème est aujourd'hui proportionnellement mieux représenté parmi les membres permanents des UMR que dans la population des chercheurs et chercheuses CNRS, et la proportion des thèses soutenues en IA est encore supérieure. La comparaison avec les données du précédent rapport de conjoncture révèle que cet écart, qui existait déjà en 2017, s'est amplifié depuis.

L. Recherche opérationnelle et aide à la décision

La Recherche Opérationnelle (RO) et l'Aide à la Décision (AD) sont des disciplines hybrides, regroupant des méthodes et outils quantitatifs orientés vers la recherche du meilleur choix dans la façon d'opérer en vue d'aboutir au résultat visé ou au(x) meilleur(s) résultat(s) possible(s) ou encore au résultat optimal. Ce domaine fait appel à la modélisation des processus, au raisonnement mathématique (incluant la combinatoire, l'algèbre linéaire, la logique, les probabilités, l'analyse des données), à l'algorithme et au développement logiciel.

La RO peut assister un décideur ou une décideuse confrontée à un problème d'optimisation combinatoire, aléatoire ou concurrentiel. Un problème est dit combinatoire s'il contient un grand nombre de solutions admissibles parmi lesquelles il s'agit de chercher une solution optimale ou quasi-optimale. La RO/AD traite ces problèmes grâce à l'algorithme, et en particulier, grâce à l'optimisation mathématique. Les problèmes aléatoires, quant à eux, cherchent une solution optimale à un problème dans un contexte d'incertitude, impliquant par exemple des variables aléatoires. Enfin, un problème est dit concurrentiel s'il consiste à trouver une solution optimale dans un cadre où les termes dépendent de l'interrelation entre ses propres agissements et ceux d'autres décideurs, en s'appuyant sur la théorie des jeux ou le choix social computationnel.

La RO/AD partage des liens thématiques de la section 7, autour des thématiques liées à la robotique, aux systèmes automatisés et aussi l'apprentissage. Elle est également proche du mot-clé « Contrôle et optimisation » de CNRS Mathématiques, ainsi que de la section 27, traitant d'Économie et gestion. Elle est structurée autour du GDR ROD.

Le tableau 1 montre que la recherche opérationnelle est également un thème où les chercheurs CNRS sont sous-représentés par rapport aux personnels des UMR et aux thèses soutenues.

M. Bio-informatique

La bio-informatique vise à améliorer la compréhension du vivant à travers la modélisation, l'analyse et l'intégration de données d'observa-

tions à plusieurs échelles. Elle rassemble différentes disciplines scientifiques expérimentales et fondamentales. Le GDR BIMMM structure cette communauté.

L'évolution très rapide des technologies de séquençage et l'acquisition massive et automatisée de données place les STIC au cœur de cette discipline. De plus, la compréhension et la pré-diction des processus biologiques nécessite des cadres de modélisation et d'analyse complexes tenant compte des dimensions spatiales, temporelles, et multi-échelles. Les domaines concernés par la section couvrent principalement l'algorithme, 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'algorithme parallèle pour traiter les grands projets metagénomiques, ou l'IA symbolique et numérique pour l'apprentissage automatique de modèles pour la découverte de stratégies thérapeutiques personnalisées.

N. Informatique quantique

L'information quantique est un domaine interdisciplinaire exploitant les comportements physiques quantiques pour de nouveaux avantages et applications dans divers domaines de la théorie de l'information, notamment le calcul, la communication et la métrologie. Avec une intense activité au niveau mondial, le domaine a connu une augmentation massive du volume de recherche et également de la diversité des thématiques. La France a traditionnellement été visible sur les sujets des algorithmes quantiques, de la complexité et de la communication, ce qui s'est étendu à de nombreux domaines de l'information quantique, notamment l'IA quantique, les modèles formels (tels que le calcul ZX), les fondements, l'informatique quantique d'ordre supérieur, les réseaux quantiques et la tolérance aux pannes.

Il y a eu un changement marqué dans l'avant-

Tableau 6 – Chercheurs et chercheuses en activité en section 6, par année de premier recrutement

1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
4	1	1	3	2	8	7	4	6	2
1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
2	4	2	3	4	3	18	13	6	7
2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
17	13	16	12	11	10	11	9	7	8
2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
9	12	7	10	6	12	11	8	10	10

cement de la technologie quantique ces dernières années, ce qui a fait évoluer le domaine à la fois en termes d'étendue des thématiques et de communauté. Il y a un nombre toujours croissant de chercheurs de la section 6) et un intérêt croissant de l'industrie, ce qui conduit à beaucoup plus d'interaction avec le monde universitaire, à travers des projets de recherche et des bourses CIFRE, ainsi qu'à des structures (par exemple via PCQT et divers DIM). Les sujets à motivations technologiques, tels que les algorithmes à court terme, dits « NISQ » (par exemple l'IA quantique et le calcul variationnel), la tolérance aux pannes, et la compilation sont de plus en plus représentés, tout comme les réseaux et la recherche d'applications cryptographiques plus larges.

La communauté de recherche française reste très visible et active, impliquée dans de nombreuses activités internationales (dans le cadre de projets phares et d'infrastructures européens, et de collaborations internationales, par exemple l'IRN CAFQA avec le Canada et plusieurs IRL à composantes quantiques). La communauté interdisciplinaire est structurée autour du GDR TEQ, qui est dominé par la physique, et comprend des sections mathématiques et ingénierie, ainsi qu'un GT IQ (du GDR IFM). Ces dernières années, il y a eu de nombreuses embauches au CNRS et à l'extérieur, et l'information quantique est un sujet prioritaire depuis 2020, ce qui a donné un flux constant de recrutements de haut niveau, et de nombreux candidats de haut niveau aux concours.

V. Carrières au CNRS

A. Pyramide des âges

Le tableau 6 rapporte le nombre de chercheurs et chercheuses 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 15 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. Sur la pyramide des âges (tableau 7), on observe un pic entre 41 et 50 ans, logiquement décalé de 5 ans par rapport au pic observé dans le rapport de conjoncture de 2019. Le volume actuel des recrutements (dont l'âge médian, voir tableau 8, est inférieur à 31 ans) n'assure pas le renouvellement de cette génération.

B. Concours CRCN

Nous discutons maintenant des concours CRCN en section 6. Noter que les CR en section 6 ne sont pas tous issus de ces concours : ils et elles peuvent également être recrutés en CID ou sur les concours handicap régulièrement ouverts ; enfin, les CR peuvent également demander à changer de section quand leurs thématiques de recherche évoluent.

Ancienneté des chercheurs recrutés en section 6. L'âge des candidats et candidates au concours n'est 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 (les maternités sont

Tableau 7 – Pyramide des âges en section 6

	< 31	31 - 35	36 - 40	41 - 45	46 - 50	51 - 55	56 - 60	61 - 65	> 65
Total	11	43	32	57	64	41	31	24	7
CR	11	43	26	28	28	13	5	5	2
DR	0	0	6	29	36	28	26	19	5

Tableau 8 – Âge 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 - 2021)	12%	63%	20%	6%
Proportion (2022 - 2024)	4%	63%	22%	11%

prises en compte pour ce calcul). Avec cette mesure, le tableau 9 rapporte l'ancienneté, évaluée au moment d'un recrutement éventuel, des candidates et 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, et qu'aucun chercheur ou chercheuse n'a été recruté avec une ancienneté supérieure à 10 ans.

Nombre de candidatures. Le tableau 10 rapporte le nombre de candidatures et le nombre de candidates et candidats ces cinq dernières années. Ces deux nombres sont différents dans les cas où plusieurs concours CR sont ouverts en section 6, avec des personnes postulant sur plusieurs concours, comme ça a été le cas en 2020 et 2021. Le nombre de candidatures a beaucoup baissé en 2022 et 2023, ce qui est source d'inquiétudes sur l'attractivité du concours ; la tendance s'est inversée en 2024, mais c'est peut-être dû au fait que le concours 2024 a été décalé dans le temps à une date (début février) plus propice aux candidatures que la date habituelle de début janvier.

Profil des CR recrutés. Au-delà de l'âge (tableau 8) et de l'expérience en recherche (tableau 9), nous nous intéressons au profil et en particulier à l'origine des personnes recrutées sur le concours CR. Le tableau 11 indique l'origine des thèses de doctorat : 35% de ces thèses ont été soutenues à l'étranger, principalement en Europe. Nous notons que, sur cette période, plusieurs autres candidats ayant soutenu leur thèse à l'étran-

ger (en particulier aux États-Unis) ont été classés admissibles mais se sont désistés, en partie en raison du manque d'attractivité des carrières académiques en France. Une autre manière de caractériser les CR recrutés est par leur profil pré-thèse comme reporté dans le tableau 12 ; à nouveau, 35% ont fait leurs études à l'étranger (en grande partie, mais pas seulement, les mêmes qui ont obtenu leur doctorat à l'étranger). Parmi les chercheurs et chercheuses ayant fait leurs études en France, c'est principalement au sein d'une École normale supérieure (39% du total), puis d'une université (18%), les diplômé·e·s d'école d'ingénieur étant minoritaire (8%).

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 13 indique la répartition des candidats par thème prioritaire (ou hors de thèmes prioritaires). Certains thèmes affichés comme prioritaires (comme Informatique quantique) n'ont aucun mal à attirer des candidatures, ce qui n'est pas le cas pour d'autres (en particulier, Sciences du logiciel et Éco-responsabilité).

C. Promotions et recrutement DR2

Les tableaux 14, 15 et 16 donnent respectivement les délais pour les passages aux grades CRHC, DR2 et DR1 pour les personnes ayant at-

Tableau 9 – Ancienneté en recherche des candidats admissibles au concours CR (06/02), évaluée au moment de leur possible recrutement

Ancienneté en recherche (thèse comprise)	4	5	6	7	8	9	10	> 10
2022	1	4	2	5	1	2	0	0
2023	0	6	5	4	0	2	1	0
2024	2	6	3	3	0	2	2	0
Total	3	16	10	12	1	6	3	0

Tableau 10 – Candidatures aux concours CR en section 6

	2020	2021	2022	2023	2024
Nombre de candidat·e·s	121	144	106	83	130
Nombre de candidatures	134	177	106	83	130
Retenus pour audition	31	38	32	32	33
Admissibles	21	17	15	18	18
Postes aux concours	9	10	7	8	8

Tableau 11 – Origine des thèses des CR recrutés récemment en section 6

	2020	2021	2022	2023	2024	Total
Amérique du Nord	1	0	0	0	1	2
Moyen-Orient	0	2	0	0	0	2
Europe hors UE	2	1	0	0	1	4
Union Européenne	1	1	2	2	3	9
Total hors France	4	4	2	2	5	17
France	9	7	5	7	4	32

Tableau 12 – Études avant thèse des CR recrutés récemment en section 6

	2020	2021	2022	2023	2024	Total
Étranger	4	4	2	2	5	17
Université française	3	1	3	1	1	9
École d'ingénieur française	0	2	0	2	0	4
École normale supérieure	6	4	2	4	3	19

Tableau 13 – Candidates et candidats sur des thèmes prioritaires au concours CR

	2022	2023	2024
Cybersécurité	6	6	
Éco-responsabilité		3	1
Informatique quantique	7	10	12
Intelligence artificielle	8		
Réseaux et systèmes			10
Sciences du logiciel	4	2	1
Tous les thèmes prioritaires^a	24	19	23
Hors thèmes prioritaires	82	64	107
Total candidats	106	83	130

^a Comme certains candidates et candidats émargent à plusieurs thèmes prioritaires, le total sur au moins un thème prioritaire peut être inférieur à la somme sur chacun des thèmes prioritaires.

Tableau 14 – Ancienneté au CNRS lors d'une promotion CRHC parmi les CRHC en activité

Ancienneté (années)	0	1 - 5	6 - 8	9 - 11	12 - 14	15 - 17	18 - 20	> 20
Effectif	0	0	1	2	2	8	11	7

Tableau 15 – Ancienneté au CNRS lors du recrutement DR2 parmi les DR en activité (hors DR recrutés au grade DR1)

Ancienneté (années)	0	1 - 5	6 - 8	9 - 11	12 - 14	15 - 17	18 - 20	> 20
Effectif	9	3	19	45	37	17	8	8

Tableau 16 – Ancienneté dans le grade DR2 lors d'une promotion DR1 parmi les DR1 et DRCE en activité

Ancienneté (années)	0	1 - 4	5 - 7	8 - 10	11 - 13	14 - 16	> 16
Effectif	3	2	23	20	13	4	0

Tableau 17 – Répartition des chercheurs et chercheuses de la section dans les unités en ayant au moins 3

Unité	IRIF	LaBRI	LIP	LAAS	LIRMM	IRIT	LIP6
Effectif	Paris C 31	Bordeaux 22	Lyon 20	Toulouse 19	Montpellier 19	Toulouse 16	Paris C 13
Unité	LIG	CRISTAL	LMF	LORIA	LIX	IRISA	I3S
Effectif	Grenoble 12	Lille 11	Paris S 11	Nancy 11	Paris S 10	Rennes 9	Nice 7
Unité	LAMSADE	LIGM	LIPN	LIS	CRIL	DI ENS	G-SCOP
Effectif	Paris C 7	Paris E 7	Paris N 7	Marseille 6	Lens 4	Paris C 4	Grenoble 4
Unité	LIRIS	LISN	LS2N	VERIMAG	ICube	I2M	
Effectif	Lyon 4	Paris S 4	Nantes 4	Grenoble 4	Strasbourg 3	Marseille 3	

teint ces grades, en poste actuellement. Le délai 0 correspond à un recrutement directement dans le grade : 9 sur le concours DR2 et 3 sur des concours DR1 (ou recrutés en CDI). Notons que, pour les promotions CRHC, ces chiffres sont biaisés par le fait que ce grade n'existe que depuis 2018 et donc que les premières attributions de ce grade se sont concentrées sur des chercheurs avec une ancéneté importante au CNRS.

D. Répartition géographique

Le tableau 17 rapporte l'effectif des chercheurs rattachés à la section 6 dans les laboratoires ayant au moins 3 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).

E. Mobilité sortante

Nous n'avons pas de statistiques sur la durée pour la mobilité sortante, les détachements et disponibilités. Il est par ailleurs difficile de savoir, au moment d'un détachement ou d'une disponibilité, si la personne en question compte reprendre son poste au CNRS. Informellement, et comme la mandature précédente, nous observons cependant peu de retour au CNRS de chercheurs et chercheuses parties en disponibilité ou détachement. Nous notons que les départs peuvent avoir lieu peu après le recrutement : ainsi (cf. tableau 3), sur les 51 CR recrutés lors de la mandature 2016 - 2021, 6 ne sont plus actuellement en poste au CNRS (soit 12%).

L'effectif de la section reste cependant pour l'instant stable : la section est passée de 303 membres en poste en 2019 (voir [CPA+21]) à 309 en 2024 (voir tableau 2).

Conclusion

Les conclusions du rapport précédent pointaient deux éléments alarmants : la baisse des candidatures féminines et le nombre de postes au concours insuffisant pour maintenir l'effectif de la section. Le premier point reste inquiétant : même si la proportion de femmes dans la section a augmenté, la proportion de candidates au concours reste anormalement faible, y compris en tenant compte de la proportion de femmes soutenant des thèses en informatique. Concernant l'effectif de la section, il reste stable pour l'instant. Le nombre relativement important de départs du CNRS peu après le recrutement est en revanche un point à surveiller, qui peut être un indicateur de l'attractivité des carrières scientifiques.

Les thématiques de la science informatique de la section 6 recouvrent un champ disciplinaire très dynamique, avec des résultats majeurs disciplinaires et en interface avec d'autres sciences (notamment en ce qui concerne l'informatique quantique ou le calcul naturel). Les progrès spectaculaires de l'apprentissage profond, des modèles de fondation et de l'intelligence artificielle générative, bouleversent certains domaines de l'informatique. On note enfin deux évolutions majeures des pratiques de la recherche : une plus grande ouverture de la science via les pratiques de publication

et de partage des données et programmes ; et une prise de conscience des enjeux écologiques de l'informatique.

Références

- [CPA⁺21] Hubert Comon, Simon Perdrix, Pierre Aboulker, Pablo Arrighi, Sandrine Blazy, Agnès Braud, Pierre Clairambault, Pascal Dayre, Rémy Dernat, Clarisse Dhaenens, Laurence Duchien, Amélie Gheerbrant, Nathalie Gilles, Jean-Marc Larre, Dominique Lavernier, Leo Liberti, Philippe Owezarski, Celine Scornavacca, Sylvain Sené, Pierre Senellart, Pierre Sens, Gilles Villard, Igor Walukiewicz et Laurent Weinhard : Section 06. sciences de l'information : fondements de l'informatique, calculs, algorithmes, représentations, exploitations. *In Rapport de conjoncture 2019*. CNRS Éditions, 2021. <https://hal.science/hal-04884956>.