

CID 55

SCIENCES ET DONNÉES

Composition de la commission interdisciplinaire – CID

Yves GRANDVALET (président de la commission) ; Laurence TRESSE (secrétaire scientifique) ; Nabila AGHANIM ; Sihem AMER YAHIA (référente diversité, équité et inclusion) ; Éric AUBOURG (membre du bureau) ; Isabelle BLOCH ; Christophe CASSEN ; Sandrine CODIS (de 2021 à 2023) ; Daniel CORDIER (de 2021 à 2024) ; Gaëlle DUFOUR ; Bernard DUSSOUBS ; Aurélien GARIVIER (référent diversité, équité et inclusion) ; Damien LAAGE (depuis 2024) ; Arnaud LEGRAND (depuis 2023) ; Antonio MONARI (membre du bureau) ; Nicolas POIRIER ; Tania REGIMBAU (de 2021 à 2023) ; Glenn ROE (membre du bureau) ; David ROUSSEAU ; Cyril RUCKEBUSCH (de 2021 à 2023) ; Dominique SIMON ; Efstathia SOROLI ; Steeve ZOZOR.

Résumé

La CID 55 a été créée en 2021. Cette nouvelle commission interdisciplinaire a une organisation singulière, en ce sens qu'elle est pilotée par l'institut CNRS Sciences informatiques mais qu'elle vise en premier lieu à recruter des chercheuses et des chercheurs qui relèvent des quatre autres instituts concernés pour cette mandature : CNRS Nucléaire & particules, CNRS Terre & Univers, CNRS Sciences humaines & sociales, CNRS Chimie. La création de cette CID témoigne de l'ubiquité des sciences informatiques et de la nécessité de promouvoir une recherche interdisciplinaire associant le numérique aux autres domaines scientifiques pour répondre aux nouveaux défis de recherche associés aux données. Notons que la CID 55 ne couvre pas toutes

les activités interdisciplinaires liées aux données dans les sciences. En particulier, les questions de recherche liées aux données en sciences du vivant ainsi que celles sur les données issues des sciences et technologies du numérique ne relèvent pas du périmètre de la CID 55.

Introduction

Les sciences du numérique jouent un rôle central dans la production, la collecte et l'exploitation des données dans tous les domaines scientifiques. Elles contribuent directement à la production de données non seulement par la

numérisation et la collecte, mais aussi par la génération de données à l'aide de modèles numériques. Le calcul à haute performance (HPC) joue également un rôle clé en facilitant ces processus, en rendant possibles le traitement et l'analyse de vastes volumes de données avec une précision et une rapidité accrues. Ces possibilités de traitement ont par ailleurs un effet indirect, en motivant le développement d'instruments scientifiques dont le bon usage repose sur le numérique. Tirer pleinement parti de ces données nécessite de répondre aux différents besoins de gestion et d'analyse efficace des données produites.

La création de la CID 55 vise à développer et maintenir les recherches interdisciplinaires nécessaires pour réaliser des innovations technologiques et méthodologiques inédites, spécifiquement adaptées aux besoins des disciplines scientifiques couvertes par les instituts CNRS Nucléaire & particules, CNRS Terre & Univers, CNRS Sciences humaines & sociales et CNRS Chimie. Dans ces recherches, le numérique n'est pas simplement au service des autres disciplines ; il participe activement à leur évolution en soulevant de nouvelles questions de recherche et en ouvrant de nouvelles pistes de résolution. La CID 55 ne recrute donc pas des personnels d'appui à la recherche, mais des chercheurs et chercheuses proposant des approches innovantes, visant à l'avancement de leur domaine scientifique, dans lesquelles les données jouent un rôle clé. Ces innovations nourrissent en retour la science des données.

Les démarches de recherche originales fortement axées sur les données, qui peuvent être perçues comme insuffisamment disciplinaires en section, trouvent leur place dans cette CID. Ces recherches, interdisciplinaires aujourd'hui, visent à s'intégrer pleinement aux disciplines, en enrichissant leurs approches et leurs méthodologies. La définition de l'interdisciplinarité pratiquée dans la CID 55 a donc vocation à être adaptée à chaque domaine scientifique et à être continûment réajustée à mesure que chaque champ disciplinaire intègre et assimile de nouvelles techniques et méthodes issues des sciences du numérique.

I. Thématisques

La CID 55 s'intéresse à la formalisation et au déploiement de toutes les étapes du cycle de vie des données, depuis leur production et leur collecte jusqu'à leur visualisation, leur analyse exploratoire et l'extraction de valeur, en passant par le stockage, l'interrogation, la modélisation et l'optimisation. Ces étapes nécessitent une forte intégration de deux expertises : celle qui relève des données et de leur traitement, et la connaissance approfondie d'un domaine scientifique, qu'il s'agisse d'astronomie, de chimie, d'archéologie, de traitement automatique des langues ou de physique. Le développement et l'utilisation de nouvelles approches d'intelligence artificielle adaptées aux besoins spécifiques de chaque discipline scientifique constituent une compétence clé des chercheuses et chercheurs de la CID 55. Cette double compétence est essentielle pour assurer une visibilité et une productivité durables dans les communautés de recherche situées à l'intersection d'une discipline scientifique et de la science des données. Elle est également cruciale pour la pérennité de nombreuses disciplines où les données et l'intelligence artificielle jouent un rôle de plus en plus central.

A. Production, collecte et gestion de données

Le numérique permet de recueillir des données en grandes quantités grâce à des dispositifs automatisés, que ce soit sur de grands instruments comme les accélérateurs de particules ou les télescopes, ou sur des détecteurs distribués comme des capteurs océanographiques, les réseaux de radiotélescopes ou de sismomètres. Le traitement des données produites par ces instruments requiert des compétences spécifiques et des collaborations établies en traitement du signal et de l'information, ainsi que des projets de recherche dédiant une part significative, voire prépondérante, au

développement de méthodes et d'outils spécialisés. Le numérique permet également de convertir des documents ou des informations existantes sur des supports hétérogènes (vidéos, audio, images, archives, manuscrits, supports épigraphiques, numismatiques, etc.) sous une forme plus facilement exploitable pour leur analyse, ouvrant ainsi la voie à des approches quantitatives considérant des échelles inédites.

Par ailleurs, l'augmentation des moyens de calcul contribue au développement de modèles numériques complexes. Certains modèles génèrent des reproductions virtuelles de phénomènes ou d'artefacts, qui participent à leur interprétation ; d'autres produisent des données synthétiques de simulation, qui doivent ensuite être analysées. La production de ces données requiert des compétences interdisciplinaires pour la production elle-même, en assurant la qualité des données produites, leur interopérabilité, mais aussi pour que les capacités de traitement et de stockage puissent suivre le rythme de leur génération. Les données produites permettent d'initier ou de renforcer des analyses quantitatives, en particulier au travers de leur structuration, qui vise à faciliter leur analyse, qui peut être automatique, en particulier par des méthodes issues de l'intelligence artificielle, mais aussi à garantir leur accessibilité, leur traçabilité et leur conservation à long terme.

B. Infrastructures, stockage, calcul

Les besoins en calcul intensif pour le développement des simulations numériques et le passage à l'exascale sont bien identifiés des communautés qui opèrent les grands instruments. Les problèmes d'infrastructure informatique, de stockage, de calcul varient significativement selon les domaines scientifiques, en raison des spécificités des données produites. Il faut par exemple pouvoir intégrer des données hétérogènes et géographiquement distribuées, croiser des données multimodales

(par exemple orales, visuelles, écrites), gérer des pétaoctets ou exaoctets de données produites par les grands instruments et assurer leur accessibilité, pour les analyses en temps réel ou différé, transférer les données vers les centres de calcul, ou encore numériser et garantir l'accessibilité des données et de leurs annotations. Par ailleurs, ces infrastructures consommant d'importantes ressources, il est crucial de développer des approches plus durables.

Sur toutes ces questions, les interactions interdisciplinaires sont nécessaires pour assurer la meilleure adaptation des infrastructures et de leurs utilisations aux besoins en amont et en aval, pour la production et le traitement des données.

C. Analyse exploratoire, visualisation de données

La nature et la quantité des données acquises ouvrent de nouvelles pistes de recherche qui nécessitent l'utilisation d'outils d'analyse des données adaptés. Pour les domaines ou sous-domaines où l'usage du numérique est moins développé, l'intégration de nouvelles méthodes peut relever de l'interdisciplinarité. Par exemple, l'étude historique à grande échelle des réseaux prosopographiques a motivé l'introduction dans ce domaine de modèles de graphes. Pour les domaines plus familiarisés avec l'usage du numérique, l'interdisciplinarité est typiquement marquée par des développements méthodologiques importants, nécessités par la spécificité des données.

De manière générale, concevoir des méthodes de représentation et d'exploration de données sur plusieurs dimensions (visualisation multi-D de positions, vitesses, températures, densités, intensités, etc.) dans un espace visuel ou interactif nécessite des compétences interdisciplinaires. Bien que ces approches soient essentielles pour comprendre et interpréter des phénomènes complexes au travers de données, elles restent encore insuffisamment développées dans l'ensemble des domaines scientifiques couverts par la CID.

D. Modélisation, simulation, criblage, optimisation

La modélisation numérique est un élément de validation de théories, voire d'élaboration de théories en mettant à jour des phénomènes ou des interactions entre phénomènes à partir de l'analyse de ces données de simulation. Les codes de simulation reposent sur des modèles de plus en plus complexes, qui permettent d'analyser des phénomènes non-observables ou difficiles à observer, ou encore de les simuler à grande échelle, sans contraintes expérimentales. Ces outils permettent également de contrôler et d'évaluer ainsi l'impact d'événements divers et de tester des situations encore inédites. Les modèles peuvent être multi-physiques, multi-échelles, multisources. La nature des objets modélisés est ainsi très large, que ce soit en échelles de temps ou d'espace, ou en nombre d'entités participant au phénomène global.

La modélisation peut également donner une place importante aux observations avec l'analyse statistique des signaux et des données. Les projets qui couvrent le développement d'estimateurs statistiques (y compris statistiques d'ordre supérieur), de méthodes d'inférence, etc., se distinguent par des aspects méthodologiques importants, sont pertinents pour la CID 55.

Outre les modèles de simulation de phénomènes physiques, certains jumeaux numériques peuvent avoir comme objectif de réaliser un rendu 3D de sites ou d'artefacts à des fins d'analyse ou de restitution historique tenant compte des hypothèses de travail, et plus généralement des incertitudes et de leur propagation dans les traitements successifs.

E. IA, Apprentissage automatique

Les méthodes et outils de l'intelligence artificielle, en particulier ceux de l'apprentissage automatique, sont mobilisés dans l'ensemble des domaines scientifiques couverts par la CID 55. Différentes techniques peuvent venir en appui des opérations de traitement de grands volumes de données. Ces volumes peuvent couvrir des échelles très différentes entre domaines scientifiques, mais elles ont en commun de ne plus permettre l'analyse systématique de chaque observation, ce qui rend critique le contrôle rigoureux des incertitudes et des biais. Comme pour les thématiques précédentes, les projets de mise en œuvre de ces méthodes sans développement de nouveaux modèles ne sont considérés comme relevant de la CID 55 que dans les sous-domaines où ces méthodes sont inédites.

Les méthodes d'apprentissage automatique sont utilisées avec des objectifs divers. Elles sont bien sûr utilisées dans un but prédictif, par exemple pour catégoriser les observations, mais de nombreux projets en font usage pour accélérer des calculs, pouvant être réalisés par d'autres méthodes plus coûteuses en calcul. Dans ce cadre, l'efficacité relative des méthodes d'apprentissage peut permettre de traiter plus de données ou de passer à des échelles de temps ou d'espace bien supérieures à ce que les méthodes paramétriques permettent. Les projets scientifiques soumis à la CID 55 visent souvent à assurer que les modèles appris satisfassent certaines propriétés connues des phénomènes modélisés, ce qui peut passer par des contraintes appliquées de manière *ad hoc* à des modèles flexibles génériques, ou par la proposition de modèles satisfaisant nécessairement ces contraintes, par exemple en passant par la définition de noyaux adaptés.

Les systèmes d'intelligence artificielle générative sont de plus en plus utilisés pour simuler des phénomènes ou leurs environnements, des artefacts, ou pour analyser ces phénomènes

complexes. Cependant, leur fiabilité soulève un défi majeur, notamment pour ce qui est de la pertinence des résultats produits. Les questions sur le niveau de confiance à accorder à ce type d'étude restent, à ce jour, largement ouvertes, ce qui constitue un frein majeur à leur développement. Il est donc essentiel de développer des approches théoriques et pratiques rigoureuses pour évaluer ces systèmes. En effet, les grands modèles de langage (LLM) peuvent produire des « hallucinations », c'est-à-dire des réponses plausibles mais factuellement incorrectes ou en décalage avec l'intention de l'utilisateur. Les méthodologies de validation, intégrant des mesures de contrôle de qualité, d'explicabilité et de raisonnement auditabile, doivent être abordées de manière interdisciplinaires afin d'intégrer la connaissance métier spécifique à chaque domaine d'application. L'enjeu consiste à exploiter pleinement le potentiel de l'IA générative tout en garantissant un cadre approprié à la prise de décision rationnelle et éclairée. Ces défis sont au cœur de nouvelles thématiques de recherche en « Sciences pour l'IA, l'IA pour les sciences » : d'une part, l'IA (y compris l'IA générative et conversationnelle) apporte de nouvelles fonctionnalités aux disciplines scientifiques, et d'autre part, ces disciplines enrichissent et étendent les fondements de la science des données et de l'intelligence artificielle. Ainsi, pour maximiser l'intérêt et l'utilité de l'IA générative tout en minimisant ses risques, il est impératif d'adopter une démarche intégrative combinant expertise technique, validation scientifique et compréhension des besoins propres à chaque domaine d'application.

F. Science ouverte

En CID 55 comme ailleurs, la science ouverte permet d'accéder aux informations critiques, que ce soit les données, leur traitement ou les résultats issus de ces traitements. Dans un contexte interdisciplinaire, elle favorise plus encore l'acquisition des nouvelles compétences nécessaires pour travailler aux

interfaces ; elle stimule également l'innovation en permettant d'exploiter plus facilement des découvertes et des outils issus d'autres disciplines pour répondre à des questions spécifiques. Par ailleurs, la science ouverte encourage l'adoption de standards ouverts pour faciliter l'échange et l'utilisation des données entre disciplines (standardisation, « fairisation », innovations pour l'interopérabilité, etc.).

II. Positionnement vis-à-vis des sections et CID

Les thématiques de la CID 55 sont pertinentes pour l'ensemble des sections des instituts concernés, et au-delà. Nous détaillons les positionnements qui nous ont paru informatifs pour illustrer par des exemples concrets notre approche pour définir le périmètre de la CID.

De manière générale, les profils des chercheurs et chercheuses recrutés en CID 55 se distinguent par une singularité sur l'interdisciplinarité entre une discipline scientifique et les sciences du numérique. Cette singularité se réalise de manière très différenciée entre les disciplines, ne serait-ce que de par la proximité de ces disciplines au numérique, typiquement liée à la formation initiale de ces chercheuses et chercheurs, qui peut être plus littéraire ou plus scientifique, et également par le degré de maturité des méthodes numériques au sein de chaque sous-discipline. Certains profils pourront donc s'apparenter à des « super-utilisateurs », pouvant transférer et adapter des concepts, méthodes et outils à leurs questionnements de recherche, alors que d'autres seront beaucoup plus avancés sur les aspects numériques, permettant ainsi de contribuer au développement d'algorithmes ou de matériels de production de données innovants, dédiés à la réduction de données nécessaire pour répondre à leurs questionnements. Toutefois, les profils purement techniques, ne mettant pas en avant de questionnements scientifiques, ne relèvent pas du recrutement en CID.

A. Usages disciplinaires et interdisciplinaires

Dans certains domaines couverts par la CID 55, l'usage du numérique est particulièrement bien intégré. Néanmoins, il existe des variations importantes entre sous-disciplines : celles où le numérique est consubstantiel du domaine de recherche (instrumentation du grand collisionneur de hadrons (LHC) ou en haute résolution angulaire pour les télescopes spatiaux, sismologie, prosopographie, etc.); et celles où l'usage du numérique est moins prégnant, en croissance (physique des ions lourds, planétologie, géodynamique, méthodes qualitatives, économie, etc.).

Dans ce contexte, les sections évaluent et recrutent des scientifiques dont le profil a une composante « numérique et traitement des données » importante, quand les projets de recherche sont fortement ancrés dans les problématiques du domaine. La CID 55 est complémentaire des sections en évaluant et recrutant des scientifiques fortement interdisciplinaires, qui apportent (par leur formation initiale, leur recherche, leurs collaborations, leur projet) des compétences numériques nouvelles pour leur sous-discipline, et qui portent un projet donnant une grande place au numérique.

B. Linguistique informatique et traitement automatique des langues

Pour la CID 55, le traitement automatique des langues (TAL) présente une spécificité notable puisqu'il est abordé sous deux aspects complémentaires dans les sections du comité national : sous un aspect essentiellement informatique et algorithmique, par la section 7, avec des questions qui concernent la génération, la compréhension et la traduction automatique des langues ; et sous un aspect essentiellement linguistique, centré sur l'humain par la

section 34, avec des questions sur la formalisation des structures linguistiques, les modèles de traitement et d'annotation, basés sur des méthodologies qui prennent en compte la richesse et la complexité des langues naturelles. Les projets combinant ces deux aspects sont particulièrement pertinents pour la CID 55.

Il peut s'agir de questions liées à la mise en œuvre d'infrastructures numériques, avec le développement de chaînes de traitement de données appliqués à la linguistique informatique et au TAL, intégrant le nettoyage et l'exploration des données, la construction et l'entraînement de modèles, ou la méthodologie de validation. Par exemple, les équilibres entre l'expertise humaine et l'exploitation des grands modèles de langage nécessitent de revisiter les étapes d'annotation, d'enrichissement des données et d'évaluation des résultats, en priorisant la traçabilité, l'interopérabilité et l'accèsibilité des données.

L'essor des modèles génératifs d'apprentissage et des technologies d'intelligence artificielle appliquées au TAL ouvre ainsi de nouveaux enjeux interdisciplinaires, nécessitant une collaboration étroite entre linguistes, informaticiens et experts en sciences des données.

C. CID 51. Modélisation mathématique, informatique et physique pour les sciences du vivant

Le recouvrement thématique de la CID 55 avec la CID 51 « Modélisation mathématique, informatique et physique pour les sciences du vivant » est important pour ce qui est des méthodes numériques, mais les périmètres des deux commissions sont bien distincts de par leurs domaines scientifiques. La CID 51 est centrée sur le vivant, pilotée par l'institut CNRS biologie qui n'est pas concerné par la CID 55. La CID 51 conserve donc dans son périmètre les aspects « sciences et données »

qui relèvent de la biologie, des neurosciences, de l'écologie ou de l'épidémiologie.

Il est bien sûr possible qu'une activité de recherche relève des deux commissions, en particulier dans le domaine de la biochimie. La chimie des organismes vivants, répondant à des questions de biologie, relève du périmètre de la CID 51, même quand elle fait appel aux méthodes du numérique pertinentes pour la CID 55. En revanche, l'application des mêmes méthodes à la chimie inorganique ou aux aspects moléculaires de la chimie organique, même dans un contexte de biochimie, relève pleinement de la compétence de la CID 55.

D. CID 52. Environnements sociétés : du savoir à l'action

Les questionnements scientifiques de la CID 52 s'articulent autour de l'exploration, l'analyse et la compréhension des systèmes écologiques et de leur fonctionnement, en y intégrant les questions socio-politiques sous-jacentes aux questions environnementales et la proposition de solutions, d'innovations conceptuelles, technologiques, ou socio-politiques pour la gestion, la remédiation et la préservation de l'environnement, tant d'un point de vue fondamental, théorique que d'un point de vue opérationnel. Les recouvrements entre la CID 55 et la CID 52 résultent du besoin de gestion et d'analyse de données, de plus en plus nombreuses, issues de sources multiples et hétérogènes, utilisées pour décrire la complexité environnementale et pour réaliser des outils d'aide à la décision. Dans ce mandat, la CID 55 a ainsi pu évaluer des candidatures à la croisée de la géographie, l'urbanisme et l'aménagement du territoire, en lien avec la pollution et la qualité de l'air ou de l'eau, ainsi que des candidatures visant l'attribution du rôle du changement climatique dans les événements météorologiques extrêmes. Certaines de ces candidatures ont été jugées admissibles par les deux commissions.

E. CID 53. Sciences en société : production, circulation et usages des savoirs et des technologies

Dans le vaste périmètre scientifique de la CID 53, certaines questions scientifiques font intrinsèquement appel à des approches quantitatives reposant sur l'analyse de données. C'est par exemple le cas de l'étude des enjeux sociaux des technologies numériques. Cette interdisciplinarité entre sciences humaines et sciences informatiques n'est donc pas du ressort de la CID 55. Ainsi, jusqu'ici, aucune des candidatures communes entre CID 55 et CID 53 n'a été jugée admissible par les deux commissions.

III. Recrutements 2022-2024

Au total, sur les 3 années de la période 2022-2024, 13 CRCN et 8 DR2 ont été recrutés. Ce nombre est donc relativement faible, et les résumés statistiques qui suivent doivent donc être interprétés avec circonspection : 1 personne recrutée CRCN ou DR2 correspond respectivement à approximativement 8% et 12% de la population recrutée. Les taux de pression sont beaucoup plus importants sur le concours CRCN, avec 285 candidatures pour 13 recrutements, que sur le concours DR2, avec 62 candidatures pour 8 recrutements.

A. Effectifs par institut de rattachement

Les postes ouverts au concours CRCN dans les CID émanent des instituts, qui ont des politiques de coloriage qui leurs sont propres. Les postes DR2 sont directement distribués par la direction de l'organisme, sans aucun coloriage.

Comité national de la recherche scientifique

Sur la période, les 13 postes ouverts ont été distribués par : CNRS Nucléaire & particules (IN2P3, 2 postes), CNRS Terre & Univers (INSU, 6 postes), CNRS Sciences humaines & sociales (INSHS, 2 postes), CNRS Chimie (INC, 2 postes) et CNRS Ingénierie (INSIS, 1 poste). Aucun poste n'a été mis au concours par l'institut pilote CNRS Sciences informatiques (INS2I).

Le tableau 1 indique le nombre de chercheuses et chercheurs recrutés par institut.

Tableau 1 : Nombre de recrutements aux concours CRCN et DR2 par institut d'accueil.

	CRCN	DR2
IN2P3	1	0
INC	2	1
INEE	1	0
INP	0	0
INS2I	0	1
INSB	0	0
INSHS	3	0
INSIS	0	0
INSMI	0	0
INSU	6	6

Tous les instituts de rattachement de la CID (hors institut pilote) ont recruté au moins un ou une CRCN. Un recrutement en dehors des instituts porteurs a également été réalisé dans un laboratoire rattaché à CNRS Écologie & environnement (INEE).

B. Effectifs par genre

Le tableau 2 indique la proportion de femmes candidates et recrutées sur la période. Étant donné la complexité à estimer les viviers des candidatures pertinentes pour la CID 55, il est difficile d'y estimer la proportion de femmes.

Tableau 2 : Proportion de femmes aux concours CRCN et DR2 (%).

	CRCN	DR2
Candidates	25	18
Admissibles	36	33
Recrutées	38	25

Par rapport aux chiffres du rapport social unique de 2022, le déséquilibre de genre au niveau des candidatures est plus important que celui de la moyenne au CNRS (35 % en CRCN et 32 % en DR2). Si le taux de femmes recrutées est supérieur à celui des candidatures, il reste inférieur aux taux moyens du CNRS (42 % en CRCN et 44 % en DR2 sur l'année 2022).

Les chiffres par institut ou section de rattachement sont trop faibles pour être significatifs et ne sont donc pas détaillés ici, mais sur le plan des recrutements, nous n'avons pas noté d'effet croisé de la discipline et du genre.

Conclusion

Les collaborations interdisciplinaires sont nécessaires pour répondre aux enjeux scientifiques actuels associés aux données. Chaque domaine scientifique pose des défis spécifiques, dont les solutions se construisent autour de problématiques méthodologiques partagées, nécessitant une synergie entre experts disciplinaires et spécialistes des données et de leur traitement. Dans ce contexte, l'ensemble des pratiques de science ouverte, visant à rendre la recherche scientifique, ses outils méthodologiques, ses données et ses résultats accessibles à tous, est un élément essentiel, qui favorise les échanges entre experts des domaines scientifiques et spécialistes en calcul intensif, IA ou gestion de données.

Les formations mises en place ces dernières années ont contribué à l'émergence et au développement de profils interdisciplinaires parmi les jeunes chercheuses et chercheurs, qui sont capables d'innover dans les méthodologies numériques tout en répondant à des questions disciplinaires spécifiques. Cependant, que ce soit à l'échelle nationale ou internationale, les opportunités d'échange entre spécialistes en sciences des données et experts d'autres disciplines scientifiques sont encore insuffisantes. Par ailleurs, les recherches ayant une pertinence transdisciplinaire rencontrent encore davantage de difficultés à diffuser au-delà des frontières de leur domaine scientifique d'origine.

La diffusion de la science des données à travers diverses disciplines scientifiques est

un fait incontestable. Les profils de candidats maîtrisant l'intégralité de la chaîne de traitement et analyse des données continueront à évoluer et à se diversifier, et les progrès des sciences basées sur l'exploitation de données massives dépendra en grande partie de l'intégration de ce type de profil. Le CNRS se doit d'être attractif pour ces scientifiques, et la CID 55 est l'instrument idéal pour les accueillir. Nous invitons tous les chercheurs et chercheuses alliant une expertise en science des données avec une compétence forte dans un autre domaine scientifique à demander à être co-évalués par cette CID, dont les avis, réalisés sous le prisme particulier de l'interdisciplinarité, complètent ceux des sections disciplinaires.

