
Réseau Calcul

Évaluation

PLATEFORME DES RÉSEAUX POUR LA MISSION POUR
L'INTERDISCIPLINARITÉ

2017

1 Présentation du réseau

Le calcul, qu'il soit scientifique ou intensif, s'est largement développé au sein de la communauté scientifique ces dernières décennies aussi bien dans la recherche publique que dans le monde du privé. Les problèmes étudiés sont de plus en plus complexes tant au niveau des méthodes numériques qu'au niveau des outils informatiques utilisés. Il est maintenant nécessaire de réunir différentes communautés (mathématiciens, informaticiens, biologistes, physiciens, etc.) pour résoudre des problèmes ayant un impact direct dans notre quotidien (climat, énergie, environnement, etc.). Le calcul est donc par nature interdisciplinaire.

D'un point de vue technique, le calcul évolue également en fonction des avancées considérables des architectures informatiques. Dans les années 90, les codes de calcul étaient exécutés sur une seule machine et la taille des problèmes était limitée. Aujourd'hui, les codes de calcul nécessitent un fort parallélisme du fait que nos machines sont multi-cœurs, qu'elles ont des cartes accélératrices de type GPU ou Xeon Phi et qu'elles sont interconnectées par des réseaux rapides. L'utilisation de ces architectures est donc de plus en plus complexe et nécessite une expertise très forte des agents du CNRS pour exploiter pleinement la puissance de calcul disponible.

Les langages de programmation ainsi que tous les outils du développement logiciel (système de gestion de version, tests, documentation, intégration continue, reproductibilité des calculs) ont également fortement évolué. Il n'est plus possible aujourd'hui de ne connaître qu'un petit nombre d'outils indépendants pour développer efficacement et dans un contexte collaboratif. Là encore, la complexité est grande car si le nombre d'outils croît, il faut également comprendre comment les agencer entre eux. Les combinaisons sont multiples et dépendent du problème étudié.

Le réseau Calcul s'inscrit dans ce contexte.

1.1 Historique

Il y a une quinzaine d'année, les discussions autour de problématiques de calcul devenant de plus en plus fréquentes dans la liste de discussion du réseau Mathrice (réseau des *administrateurs systèmes et réseaux* des laboratoires de mathématiques du CNRS), un petit groupe s'est constitué en 2003 et a commencé à créer des événements pour réunir la communauté du calcul en France. Dans le but de consolider cette initiative, le Groupe Calcul a été formalisé en 2009 en créant :

- un réseau métier porté par la MRCT (ancienne structure du CNRS remplacée par la suite par la plateforme des réseaux de la Mission pour l'Interdisciplinarité) ;
- un GdR porté par l'INSMI, en partenariat avec l'INS2I.

Bien que le réseau et le GdR aient été créés ensemble, ils correspondent chacun à un projet différent d'animation de la communauté du calcul scientifique : le GdR Calcul traite de sujets de recherche alors que le réseau discute de sujets techniques et technologiques mis en oeuvre pour le soutien aux projets de recherche.

- Le GdR Calcul s'intéresse aux nouvelles méthodes numériques et aux nouveaux algorithmes qui sont à la pointe de la recherche en calcul scientifique. Ces thématiques ont une forte composante en mathématiques appliquées et en informatique mais sont également pluri-disciplinaires (physique, chimie, biologie, ingénierie, etc.). En règle générale, la validation de ces nouvelles méthodes passe par l'élaboration de logiciels de calcul qui sont développés au sein des laboratoires et sont distribués à la communauté du calcul en France et à l'international. Le GdR participe à la promotion de ces logiciels et à la formation de la communauté d'utilisateurs.
- Le réseau Calcul s'intéresse aux nouvelles architectures matérielles et aux logiciels, en partageant son expertise sur la façon d'utiliser au mieux les ressources de calcul dans le cadre d'un projet de recherche donné. Cela va de la connaissance et de l'analyse détaillée du fonctionnement des machines, à l'apprentissage des langages de programmation, en passant par le développement et l'utilisation de bibliothèques adaptées à un problème de recherche donné, permettant ainsi d'exploiter de manière optimale ces architectures. Le réseau Calcul fait le lien avec les nouvelles méthodes numériques dans leurs mises en oeuvre informatique, leur validation et leur diffusion dans des codes et logiciels.

Les thématiques traitées par chacune des structures ne sont pas totalement sans recouvrement (cf. figure 1), mais elles y sont abordées de façon différente. Par exemple, le GdR discute de nouvelles méthodes numériques et nouveaux algorithmes qui peuvent être implantés dans des bibliothèques scientifiques largement utilisées par la communauté. Ces bibliothèques peuvent alors faire l'objet d'une présentation par le réseau. Le GdR sera plus concerné par les discussions autour des propriétés de ces méthodes, de leur analyse, alors que le réseau sera plus intéressé par synthétiser celles qui sont exploitables pour un problème donné et par leur utilisation éventuellement dans le cadre de bibliothèques et outils de calcul numérique.

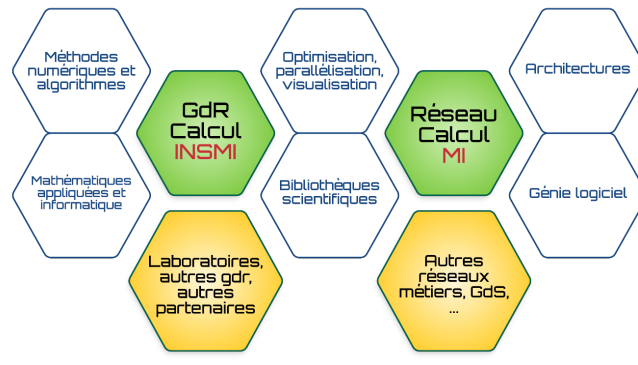


Figure 1: *gdr et réseau*

La finalité n'étant pas la même, les actions et les participations à des financements sont toujours disjointes. Côté financement, le GdR a un soutien de la part de l'INSMI du CNRS alors que le réseau dépend de la Mission pour l'Interdisciplinarité (MI) du CNRS et à ce titre est transversal à l'ensemble des Instituts. Les demandes de financement des actions du réseau sont faites de façon principale auprès de la MI. Pour les mener à bien, elles reçoivent parfois un financement complémentaire venant d'autres entités (laboratoires, autres réseaux métiers, partenaires, etc.).

1.2 Objet du réseau

Le réseau Calcul a pour objectif de proposer un ensemble d'actions complémentaires destinées à favoriser les échanges et contribuer à l'animation de la communauté du calcul scientifique en France. Il s'agit principalement de :

- Diffuser de l'information via la liste de discussion du réseau (1837 abonnés) pour encourager les échanges, fédérer les compétences et favoriser la circulation des idées :
 - échanges techniques (méthodes numériques, programmation, matériel, etc.) ;
 - retours et partages d'expériences ;
 - diffusion d'informations de conférences, workshops, formations, etc., mais aussi d'offres de thèses, post-doctorats et d'emploi en général.
- Communiquer sur le site internet les annonces (écoles, formations, etc.), les nouvelles en lien avec le calcul, et mettre à disposition des documents et des supports de formation.
- Organiser des formations afin de permettre à chacun de développer ses compétences :
 - journées scientifiques et techniques ;
 - ateliers et écoles.
- Organiser des actions communes :
 - avec la communauté des développeurs, notamment avec le réseau DevLOG, sur l'activité de développement logiciel en lien avec le calcul ;
 - avec le réseau RESINFO sur les infrastructures de calcul, en lien notamment avec les mésocentres ;
 - avec le GIS France Grilles et GENCI sur l'utilisation des différentes infrastructures (grille, cloud, mésocentres) pour le calcul.

Le terme *calcul* ne désigne pas uniquement les activités liées au calcul intensif et à la simulation numérique, mais le calcul au sens large, ce qui inclut le calcul formel, les statistiques, la théorie des nombres, etc. Les thèmes comme les pré/post-traitements, la visualisation, les maillages, le parallélisme, l'architecture des machines, l'algorithmique et la programmation y ont aussi leur place. Voici une liste des principaux sujets abordés dans les discussions ayant lieu sur la liste de diffusion ou lors des événements organisés par le réseau :

- architecture matérielle ;
- logiciels de calcul ;
- compilateurs et compilation ;
- outils de développement ;
- outils de visualisation ;
- bibliothèques scientifiques ;
- langages de programmation ;
- méthodologie de programmation ;
- méthodes numériques.

1.3 Organisation interne

Le réseau Calcul est géré et animé par un bureau (comité exécutif, appelé CoPil par la MI) chargé de mettre en place une organisation et des actions structurantes pour la communauté. Il s'est doté en 2015 d'une charte, sur la base des éléments proposés dans la charte des réseaux du CNRS et a été communiquée à la MI. Il y a été spécifié de manière collective la finalité du bureau et le rôle de ses membres, auquel chacun souscrit en s'engageant dans l'animation du réseau. Elle est révisée tous les ans en fin d'année. Au sein du bureau, trois personnes se chargent des rôles définis par la MI pour correspondre avec elle (appelé bureau par la MI), conformément à la charte proposée. Ces charges d'interlocuteur avec la MI sont assurées à tour de rôle et révisées périodiquement. Ces correspondants sont tous inscrits dans la liste `coord_mireseaux@services.cnrs.fr`. Le bureau du réseau siège en même temps que le bureau du GdR Calcul, afin de passer l'information le plus facilement et le plus rapidement possible entre les deux structures. Ceci permet également de maintenir une certaine cohérence entre les actions proposées par chacune d'elles, au bénéfice de la communauté. Lorsque c'est possible, un scénario enrichissant consiste à réaliser des actions successivement dans les deux structures, en abordant respectivement les concepts et leur mise en application pratique, à travers des méthodes et outils. C'est le cas par exemple sur la reproductibilité en calcul scientifique, pour laquelle le GdR a consacré des journées, déclinées au niveau du réseau sur une de ses applications pratiques, avec la formation aux outils d'expérimentation numérique comme Execo ou OpenMOLE. Chacun participe ainsi à la diffusion de bonnes pratiques, bénéfiques à tous les aspects du métier.

Le bureau se réserve par ailleurs la possibilité d'inviter des membres extérieurs, partenaires du réseau (Instituts du CNRS, Mission pour l'Interdisciplinarité, etc.) lorsque cela est jugé nécessaire.

Constitution du bureau au 1er septembre 2017 :

• Réseau métier

- Matthieu Boileau* IRMA, Strasbourg
- Anne Cadiou*, LMFA, Lyon
- Bernard Dussoubs, Mines de Nancy
- Loïc Gouarin*, Université Paris Sud
- Matthieu Haefele*, Maison de la simulation, Saclay
- Florent Langrognet, Université Besançon
- Pierre Navaro, IRMAR, Rennes

* *correspondant pour la MI* (avec un effet de tuilage sur 2017, ce qui explique le nombre de quatre)

• GdR

- Matthieu Boileau, IRMA, Strasbourg
- Roland Denis, ICJ, Lyon

- Benoît Fabrèges, ICJ, Lyon
- Sylvain Faure, Université Paris Sud
- Alain Franc, INRA, Bordeaux
- Loïc Gouarin, Université Paris Sud
- Violaine Louvet, GRICAD, Grenoble
- Marco Mancini, LUTH, Observatoire de Paris
- Fabrice Roy, LUTH, Observatoire de Paris
- Laurent Series, Centrale Supélec

Le bureau a mis en place des groupes de travail internes, coordonnés par deux membres, afin de faire émerger des idées et de faire avancer des projets. Quatre groupes de travail sont identifiés à l'heure actuelle :

- formation
- communication
- veille technologique
- partenariats

Le bureau échange et coordonne ses actions à l'aide d'une liste de diffusion et d'un projet hébergé sur le service gitlab de l'Institut de Recherche Mathématique Avancée (IRMA) à Strasbourg. Gitlab permet aux membres du bureau d'éditer un wiki interne, de définir des tâches, de les attribuer aux membres. Chacun peut ainsi suivre et participer aux fils de discussion en cours.

Deux réunions du bureau sont organisées annuellement en présentiel, auxquelles tous les membres du bureau sont conviés. Des visio-conférences permettent de discuter des questions urgentes qui ne peuvent attendre la prochaine réunion.

Le bureau est renouvelé régulièrement par cooptation ou à la suite d'un appel à engagement, en assurant une rotation annuelle d'un de ses membres. Ceci permet d'avoir constamment de nouvelles personnes tout en gardant un historique favorisant leur intégration. Ceci permet de maintenir le dynamisme et l'arrivée de nouvelles idées, de nouvelles disciplines et de nouveaux points de vue. Le bureau se doit également de conserver une pluralité dans les instituts et les communautés scientifiques auxquelles appartiennent les membres du bureau, avec un équilibre aussi sur les sensibilités qui relèvent du calcul intensif et ceux qui n'en relèvent pas.

1.4 Une communauté ouverte

Le réseau a pour vocation d'être largement ouvert à tous les acteurs du calcul quel que soit leur statut (ingénieurs, chercheurs, doctorants, etc.), quelle que soit leur discipline (mathématiques, physique, biologie, etc.) et qu'ils exercent dans un établissement public ou privé. Ces acteurs sont, pour certains, fournisseurs de moyens ou d'outils de calcul (administrateurs d'infrastructures, développeurs de codes ou de bibliothèques), pour d'autres, utilisateurs de ces moyens et outils.

Cette large ouverture permet d'avoir des échanges d'expériences, de susciter de nouvelles collaborations et de faciliter le travail multi-disciplinaire. La liste de diffusion est par exemple un lieu d'échanges réguliers d'informations entre communautés scientifiques différentes sur l'utilisation de méthodes numériques, mais aussi entre des constructeurs de matériels ou développeurs de logiciels et leurs utilisateurs.

Aux disciplines scientifiques traditionnellement utilisatrices du calcul (astrophysique, climatologie, mécanique des fluides, biologie, chimie, etc.) s'ajoutent depuis quelques années de nouvelles communautés travaillant à l'analyse et la visualisation de grands volumes de données, notamment grâce à des outils informatiques et statistiques. Avec l'augmentation du volume des données à traiter, ces communautés se voient confrontées à de nouveaux défis qui relèvent en partie du calcul scientifique. Le réseau Calcul voit donc son périmètre s'élargir, en particulier vers les laboratoires de lettres ou de sciences sociales qui utilisent de telles méthodes d'analyse.

2 Bilan des actions du réseau Calcul

Le réseau Calcul est très largement ouvert à tous les acteurs du calcul : il concerne toutes les disciplines du monde académique ayant des problématiques de calcul scientifique, mais aussi celles des entreprises privées.

L'espace d'échanges le plus utilisé est la liste `calcul@listes.math.cnrs.fr`. Multidisciplinaire et ouverte à tous, elle compte aujourd'hui 1837 abonnés. Sur cette liste sont postés :

- des offres d'emploi ;
- des annonces de rencontres, conférences et formations ;
- les sorties de nouvelles versions de logiciels liés au calcul scientifique ;
- des demandes d'aides techniques pour le développement ou l'achat de matériel.

La vie du réseau Calcul est animée par des actions qui se déclinent actuellement sous trois formes :

- écoles ;
- ateliers ;
- journées thématiques.

Ces évènements ont lieu dans toute la France, sur des durées variant de 1 à 5 jours.

2.1 Écoles

Une fois par an, le réseau Calcul organise une formation d'une durée de 4 à 5 jours où les participants sont logés. Certaines d'entre elles ont le label A.N.F (Action Nationale de Formation) du CNRS.

Dans le cadre d'une thématique donnée, chaque formation porte sur différents outils, présentés par des spécialistes, avec des travaux pratiques approfondis qui permettent aux participants d'apprendre à les utiliser sur des exemples concrets.

Les premières écoles étaient centrées autour des bibliothèques de calcul scientifique, des outils de développement sur les architectures parallèles et des langages, plus particulièrement Python. Progressivement, les sujets sont devenus plus ciblés en suivant les évolutions des métiers liés au calcul. Plusieurs écoles ont eu pour sujet "l'optimisation" qui reste un aspect central.

Ces dernières années, le réseau s'est ouvert aux "sciences des données", les techniques autour du "big data", du "deep learning" et l'analyse de données massives étant au cœur des préoccupations actuelles des acteurs du calcul. Le réseau Calcul continue également de proposer des formations sur des aspects fondamentaux comme la visualisation, la reproductibilité et la précision.

Voici la liste des formations proposées au cours des 4 dernières années de la plus récente à la plus ancienne :

- *Réduction de la dimension dans la fouille de données massives : enjeux, méthodes et outils pour le calcul* du 25 au 29 septembre 2017 sur l'île d'Oléron.
- *Précision, Reproductibilité en Calcul et Informatique Scientifique* du 15 au 19 mai 2017 à Fréjus.
- *Boîtes à outil éléments finis open source* du 24 au 26 janvier 2017 à Jussieu, Paris.
- *Visualisation et données* du 28 novembre au 2 décembre 2016 à Orsay.
- *Installation et exploitation d'un cluster de calcul*, coorganisée avec l'Inra en janvier 2016 à Lyon.
- *R avancé et performances* du 4 au 9 octobre 2015 à Aussois.
- *Informatique pour le calcul scientifique, bases et outils*, du 22 au 26 septembre 2014 à Saint-Dié (Vosges).
- *Optimisation* dans le cadre du projet Equip@Meso avec la Maison de la Simulation, du 7 au 11 juillet 2014 à l'université de Strasbourg.
- *Python avancé en calcul scientifique* du 2 au 6 décembre 2013 à Biarritz.
- *Précision et reproductibilité en calcul numérique* du 25 au 29 mars à Fréjus.
- *Green HPC - Energie et calcul haute performance* du 21 au 25 septembre 2015 à Inria Rocquencourt.
- *Optimisation* du 7 au 11 octobre 2013 à la Maison de la Simulation à Saclay.

2.2 Ateliers

En plus des formations sur une semaine, nous avons proposé des ateliers de 2 à 3 jours au rythme moyen de 2 par année. Ces ateliers sont centrés sur un outil spécifique lié au calcul, mais traitent aussi des méthodologies pertinentes du génie logiciel (gestion de version ou intégration continue par exemple) ou du choix des licences. Ils permettent aussi de faire découvrir les nouveautés comme les outils python Numba et Pythran, ou un langage récent et prometteur comme Julia, sous l'impulsion de spécialistes reconnus. Les langages historiques évoluent également : ainsi, les nouvelles normes 11 et 14 du C++ ont pu faire l'objet d'ateliers, vivement appréciés aussi bien par les membres déjà expérimentés que par ceux qui découvraient le langage. Le développement et la réalisation de calculs performants et efficaces sont un dénominateur commun à tous ces ateliers.

- *Python avancé en calcul scientifique*, du 4 au 6 juillet 2017 à Paris.
- *Intégration continue*, du 29 au 31 mai 2017 à Paris.
- *C++ : les bases du 11 et du 14* du 29 novembre au 1er décembre 2016 à Rennes.
- *C++ : niveau intermédiaire*, du 2 au 4 mai 2016 à Lyon.
- *C++ : les bases du 11 et du 14*, du 8 au 10 septembre 2015 à Besançon.
- *Julia* les 16 et 17 juin au 2015 au CNAM Paris.
- *Git et Gitlab au quotidien : commandes principales et workflows* le 4 juin 2015 à CentraleSupélec sur le campus de Châtenay-Malabry.
- *C++ : les bases du 11 et du 14* du 28 au 30 avril 2015 à l'IHP, Paris.
- *C++ : vers 2011 et au-delà* du 29 au 31 octobre 2014 à l'IHP, Paris.
- *Profilage de codes de calcul* avec le mésocentre de Centrale Paris, le 11 juin 2014, Châtenay-Malabry.
- *PETSc avancé*, du mardi 11 juin au jeudi 13 juin 2013 à la Maison de la Simulation, Saclay.
- *Initiation à PETSc* du lundi 18 mars au mercredi 20 mars 2013 à la Maison de la Simulation et du 13 mai au 15 mai 2013 à Lyon.

2.3 Journées thématiques

Les journées thématiques sont l'occasion de suivre une série de présentations sur l'utilisation du calcul dans des domaines variés (chimie, biologie, mécanique, physique, etc.). Elles rassemblent un public d'horizons différents favorisant les discussions et retours d'expérience. Elles permettent de faire émerger des collaborations. Il faut noter qu'à certaines de ces journées sont présents des fournisseurs de grandes marques de matériel (Intel, NVIDIA).

Le réseau Calcul a joué un rôle majeur dans l'animation des mésocentres et sa coordination au niveau national depuis dix ans, en s'engageant à ce que les acteurs institutionnels rencontrent chaque année les utilisateurs du calcul intensif. Nous avons conscience que cette animation devrait être portée par une coordination des mésocentres mais celle-ci n'existe pas à ce jour. Or, pour nous, il nous semble très important que ces mésocentres aient une forte visibilité dans l'univers du calcul français. Les journées mésocentres leur donnent donc une large place et permettent aux personnes désireuses d'accéder à des ressources de calcul régionales d'identifier rapidement les acteurs locaux. Ces journées permettent également à ces acteurs de se rencontrer une fois par an et d'échanger sur leurs pratiques. Elles traitent d'aspects technologiques et s'efforcent de faire le lien avec nos institutions.

- *Journée Python et Data Science*, le 19 décembre 2017, Rennes.
- *Journées SUCCES*, les 16 et 17 octobre 2017 à GRICAD, Grenoble.
- *10es journées mésocentres*, les 26 et 27 septembre 2017 à l'IHP, Paris.
- *Journée Runtime*, le 20 janvier 2017 à l'Inria, Paris.
- *Journées SUCCES*, les 23 et 24 novembre 2016 à l'IPGP, Paris.
- *9es journées mésocentres*, les 11 et 12 octobre 2016 à l'IHP, Paris.
- *Journées SUCCES*, les 5 et 6 novembre 2015 à l'IPGP, Paris.
- *8e journée mésocentres*, le 6 octobre 2015 à l'IHP, Paris.
- *7e journée mésocentres*, le 7 octobre 2014 à l'IHP, Paris.
- *Journées SUCCES*, les 13 et 14 novembre 2013 à l'IHP, Paris.

- *6e journée mésocentres*, le 19 septembre 2013 à l'IHP, Paris.
- *10 ans de Calcul : Histoire du Calcul*, le 9 avril 2013 à l'IHP, Paris.

2.4 Quelques chiffres

On constate que ces différentes actions autour du calcul sont bien suivies et remplissent un rôle de lien dans la communauté : les formations regroupent 30 à 40 participants et les journées thématiques 60 à 80 personnes.

L'organisation d'une journée de formation nécessite un financement variable selon que cette formation est en résidentiel ou non. Le coût moyen d'une formation en résidentiel est de l'ordre de 180 euros par jour et par participant. Les écoles, ainsi réalisées sur 5 jours avec 30 participants, ont nécessité en moyenne des budgets de l'ordre de 20 000 euros pour couvrir l'ensemble des frais. Un atelier pour 30 personnes nécessite quant à lui un budget de l'ordre de 2000 euros pour couvrir les frais de pause café, repas et réservation de salle. Cela ne couvre pas les frais de déplacement des intervenants.

Les candidats aux différents concours de la fonction publique sur les postes d'ingénieurs calcul ont souvent déjà participé à au moins une de nos formations. Les nouveaux arrivants déclarent s'en servir pour compléter leur formation et élargir leur domaine de compétences. L'ensemble des disciplines (physique, chimie, ingénierie, biologie, etc.) bénéficie à travers le réseau Calcul d'une mise en lien avec les mathématiques appliquées et l'informatique pour le calcul ainsi que d'un regard vers les autres disciplines.

2.5 Communications et interactions

Le réseau Calcul développe et gère son site web <http://calcul.math.cnrs.fr>, une liste de discussion (tous deux hébergés par Mathrice) ainsi qu'un fil Twitter. Parmi les actions de communication, notons :

- la diffusion de sondages de la communauté sur les pratiques métier (par exemple sur les données en 2016) ;
- depuis 2016, un appel à contribution et expression des besoins est réalisé chaque année via la liste de discussion (en novembre) et sur Twitter ;
- la participation aux réflexions sur l'évolution des métiers.

Le réseau Calcul est partenaire des réseaux DevLOG et RESINFO, avec lesquels il collabore régulièrement sur des problématiques logicielles et informatiques (Plume et Envol, par exemple). Il a par ailleurs des correspondants dans ces réseaux. Il participe aux journées réseaux de la Mission pour l'Interdisciplinarité (poster, vidéo et stand en janvier 2016) ainsi qu'à des groupes de travail inter-réseaux (e.g. sur le cycle de vie des données depuis 2016). Enfin, il participe à la cellule de veille technologique de GENCI, créée en 2015 pour préparer l'arrivée de l'exascale.

Lorsque le bureau du réseau discute des actions à mener, en particulier des actions de formation, il les priorise en fonction des autres actions de formation proposées par ailleurs, notamment par l'IDRIS ou le CINES pour ce qui concerne le Calcul Haute Performance. Les membres du bureau ont constaté qu'une partie des demandes qui remontent des sondages effectués sur la liste de discussion du réseau correspondent à des formations déjà existantes. C'est le cas par exemple de demandes concernant les bases des langages de programmation (Fortran/C++) ou la parallélisation (MPI/OpenMP). L'IDRIS et le CINES réalisent ces formations depuis de nombreuses années et sont experts dans le domaine. Il ne nous semble donc pas nécessaire de les monter au niveau du réseau mais bien de les relayer pour mieux les faire connaître. Néanmoins, sur certaines actions, il nous semble intéressant d'aller chercher les intervenants de GENCI - et de l'IDRIS en particulier, au titre de leur expertise. C'est le cas par exemple de l'ANF proposée en 2018 sur les outils HPC de support au calcul dans laquelle est impliquée une ingénieure de l'IDRIS. Le bureau du réseau reste persuadé que son rôle dans le choix des actions de formation consiste à mettre en avant des formations nouvelles, en avance de phase par rapport aux formations initiales ou avancées, telle que celles proposées par exemple par la formation permanente du CNRS via les Délégations Régionales ou l'IDRIS.

Afin d'avoir une bonne visibilité sur les formations proposées en France en lien avec le calcul, le réseau a contribué à constituer le portail de formations sur le calcul au niveau national (<http://formation-calcul.fr/>) avec AMIES et la Maison de la Simulation, et ce avec le support de MaiMoSiNE et de Mathrice. Focal est destiné à recenser les formations autour du calcul

au niveau national, provenant de différents instituts. Le but étant d'éviter les doublons et de travailler ensemble sur les besoins de la communauté.

3 Projet du réseau

3.1 Analyse de l'existant et prospective de l'objet du réseau

3.1.1 État de l'art

Concernant le métier, les éléments qui ont motivé la création du réseau en 2009, présentés en première partie de ce rapport, sont toujours d'actualité. Les techniques ont évolué, le métier aussi, mais la finalité du réseau reste la même.

L'organisation actuelle du réseau Calcul et son interaction avec les différents partenaires du calcul scientifique tant académiques qu'industriels lui permettent :

- de favoriser les échanges au sein de la communauté du calcul sur les différents aspects du métier ;
- de croiser les pratiques émanant de différentes disciplines ;
- de partager le savoir-faire et l'expérience ;
- d'accompagner l'évolution des méthodes et des outils en diffusant les connaissances et les techniques liées au calcul scientifique ;
- de faire remonter les réalités du terrain vers les tutelles, par le lien avec les structures institutionnelles.

Ces actions répondent au besoin actuel d'entre-aide de la communauté, dans sa mise en œuvre du calcul scientifique et sa prospective de veille technologique.

En pratique, au niveau du réseau, les échanges prennent les formes décrites dans la section sur le bilan des actions. La formation y tient une large part car elle est basée sur des mises en application et des retours d'expérience sur des situations réelles.

Le spectre des expertises actuelles des acteurs du calcul, trouvées au sein du réseau Calcul couvre :

- l'usage de différents langages de programmation pour le calcul scientifique ;
- le développement, la diffusion et l'utilisation de bibliothèques et de codes de calcul ;
- le déploiement, l'administration et l'exploitation d'architectures adaptées au calcul scientifique ;
- l'implémentation et la validation de méthodes numériques et mathématiques appliquées adaptées au problème posé ;
- l'optimisation de codes de calcul, l'usage expert de différentes techniques de parallélisation ;
- l'utilisation d'outils de génie logiciel répondant aux problématiques du calcul scientifique que ce soit pour du développement de méthodes ou pour leur mise en œuvre dans des campagnes de simulation ;
- le développement et la mise en œuvre de calculs sur les données, couplés aux simulations ou à l'analyse de leurs résultats, avec un usage expert d'un certain nombre d'outils logiciels associés.

3.1.2 Développer des compétences

Lors de la création du réseau métier en 2009, il n'existait pas encore en France de formation initiale intégrant pleinement le calcul haute performance dans un paysage où la modélisation et la simulation numérique étaient, elles, largement enseignées car répandues depuis plusieurs décennies dans tous les domaines scientifiques. Depuis, des initiatives ont vu le jour et laissent espérer le déploiement d'une culture initiale du calcul haute performance, favorable à l'émergence de solutions innovantes dans les travaux de recherche. Cependant, le recours au calcul intensif et l'augmentation des ressources en calcul s'accompagnent aussi d'une complexité croissante des problèmes de recherche explorés (couplés, multiphysiques et multiéchelles) qui nécessite de maîtriser un spectre de techniques de plus en plus large (exploitation d'architectures hybrides, couplages du calcul et données, interfaçage de plusieurs langages et de bibliothèques logicielles,

etc.) tout en croisant souvent différentes disciplines scientifiques et méthodes numériques associées. La formation initiale ne suffira donc pas à elle seule à suivre les réalités du métier. Le réseau prend donc le relais de ces formations pour continuer à développer les compétences.

Par exemple, le recours au calcul intensif dans toutes les disciplines de la recherche et l'exploration par la simulation de modèles numériques toujours plus fidèles, ont favorisé l'émergence de grands codes communautaires qui rendent possibles ces explorations, au prix d'une forte expertise dans leur utilisation. Toutefois, optimiser continuellement ces grands codes pour les rendre compatibles avec les nouvelles architectures et techniques de parallélisme nécessite des moyens importants et des compétences bien spécifiques, parfois distinctes de celles de leurs utilisateurs. Par ailleurs, il est également nécessaire de réaliser des prototypes pour aborder des problèmes nouveaux, à la croisée des mathématiques appliquées, de l'informatique et des sciences concernées (physique, biologie, chimie, etc.).

Ainsi, les compétences que le réseau métier doit continuer à soutenir par ses actions concernent :

- l'utilisation efficace des architectures de calcul par les applications scientifiques et les bibliothèques logicielles pour exploiter au mieux les ressources et répondre aux besoins de la recherche ;
- la connaissance et l'usage expert des langages et bibliothèques logicielles pour développer ou sélectionner les outils adaptés à un problème donné ;
- la culture approfondie des méthodes de calcul, de leur implémentation informatique et de leur validation pour susciter de nouveaux développements au profit de la simulation numérique et de l'analyse des données de la recherche.

Continuer à contribuer à l'articulation de ces compétences est une des vocations du réseau Calcul. Cela nécessite un dialogue entre disciplines : les mathématiques appliquées, l'informatique et les sciences applicatives (physique, biologie, etc.).

3.1.3 Défis scientifiques et technologiques

Dans un avenir proche, dans le cadre du prochain quinquennal, on peut identifier deux verrous technologiques principaux : d'une part l'évolution nécessaire des techniques de parallélisme — comment aller vers le calcul exaflopique — et d'autre part à l'explosion quantitative des données — comment analyser les mégadonnées issues des simulations. Les défis scientifiques consistent à répondre efficacement à l'aspect de plus en plus multidisciplinaire des applications du calcul.

Les supercalculateurs ont commencé une profonde mutation liée aux limites atteintes dans la miniaturisation des micro-processeurs et aux contraintes de dissipation thermique et de consommation électrique. On tend vers des architectures fortement distribuées, de plus en plus hétérogènes, et généralement hybrides. Cette mutation s'accompagne d'un passage à l'échelle dans le transfert et le traitement des données qui entraîne une révolution des pratiques logicielles.

On peut identifier les défis suivants :

- Accompagner la communauté dans l'exploitation de ces nouvelles architectures parallèles, hybrides et hétérogènes. Cela passe par de nouvelles méthodes de parallélisation et l'usage de bibliothèques associées.
Mots-clefs : exascale, passage à l'échelle, performance, optimisation, parallélisme massif, hiérarchie mémoire profonde, architecture hétérogène et hybride, accélérateurs, CPU, GPU, FPGA, modèles de programmation (MPI, OpenMP, threads, tâches, Domain Specific Languages), Middleware (systèmes runtime), etc.
- Renforcer la connaissance des méthodes mathématiques pour le traitement et l'analyse des données. Les solutions se trouvent à l'interface entre les statistiques, l'algèbre linéaire, la modélisation physique et l'outil informatique. Mots-clefs : données massives, modélisation et simulation, workflows, transfert, efficacité énergétique, calcul statistique, cycle de vie des données, etc.
- Traiter de manière efficace la tolérance aux fautes. La probabilité d'une panne augmentant avec le nombre d'unités de calcul, l'utilisation de machines exaflopiques nécessite de revoir la programmation des codes pour en gérer l'occurrence. Mots-clefs : reproductibilité, quantification des incertitudes, etc.
- Maîtriser la précision des calculs. Cet aspect est une constante du calcul numérique, mais dans certains domaines d'application, il est particulièrement critique (notamment dans le cas de problèmes multiéchelles). La maîtrise de

la propagation des erreurs est un défi, notamment dans les problèmes de grande taille, bien que sa mise en œuvre puisse être popularisée partout. Dans certains cas, la précision informatique peut être dégradée lorsque la précision numérique est suffisamment assurée, pour une meilleure efficacité de calcul. Mots-clefs : précision, propagation des erreurs, etc.

- Assurer la reproductibilité des expérimentations numériques — comment réexécuter un code dans un environnement logiciel et matériel évoluant constamment et de manière irréversible ? Cela touche tous les aspects du calcul, quelle que soit son échelle.
- Permettre la réutilisabilité des données et leur exploitation par le calcul. Le calcul produit des données mais est aussi utilisé pour qu'elles fassent sens. Une perspective à plus long terme consiste à coupler données et calculs, ce qui nécessitera aussi de revisiter les usages actuels.
- Renforcer le développement et l'utilisation de bibliothèques de calcul. Les projets de recherche de plus en plus multidisciplinaires posent de nouveaux défis de calcul, qui nécessitent le développement de bibliothèques adaptées et leur diffusion dans la communauté. Leur usage contribue à propager de bonnes pratiques dans la réalisation de calculs scientifiques.

3.2 Plan d'action

3.2.1 Pour tenter de lever les verrous technologiques ou aider à l'amorçage de projets visant à lever ces verrous

Les solutions pour répondre à ces défis sont probablement à l'interface entre différentes disciplines applicatives et différentes communautés de recherche.

Dans l'optimisation et l'exploitation efficace des codes de calcul, les disciplines qui ont depuis longtemps recours à la simulation intensive sont généralement à la pointe des dernières technologies. Elles les mettent en œuvre dans leurs calculs intensifs (physique nucléaire, mécanique des fluides, climat, etc.) avec l'aide des grands centres de calcul nationaux ou européens. Ainsi, la participation du réseau Calcul aux veilles technologiques menées par ces communautés devrait accélérer l'amorçage de projets avec des retombées sous forme de retour d'expérience vers l'ensemble du réseau, par exemple avec la cellule de veille technologique de GENCI sur les prototypes Frioul (architecture Intel Xeon Phi KNL) et Ouessant (architecture OpenPower, CPU Power8+ et GPU NVidia Pascal).

Dans l'analyse des données massives, les communautés comme l'astrophysique, la sismologie ou la biomécanique ont développé des solutions qu'il est intéressant de faire connaître aux autres disciplines. L'ANF sur la fouille de données massive avait été discutée l'année passée avec la communauté de BioInformatique de l'INRA et une équipe de l'Inria sur des aspects d'algèbre linéaire. Sur cette thématique, les aspects informatiques et statistiques développés par la communauté du GdR Madics pourraient être intéressants à faire connaître au niveau du réseau.

La méthodologie adoptée par le bureau pour faire émerger des projets consiste donc à :

- continuer à relayer les actions des autres communautés sur la liste de discussion de Calcul ;
- croiser les pratiques d'autres communautés (Centres de Calcul nationaux, européens, GdR, etc.) ;

en assistant ou participant à leurs manifestations (e.g. forum ORAP, participation au Workshop TDMF, etc.).

Par ailleurs, dans son travail de structuration de la communauté, le bureau du réseau Calcul travaillera à :

- renforcer le partenariat avec les réseaux experts de différents instituts et EPST partenaires (Inria, INRA, Ifremer, CEA, etc.) ;
- articuler ses actions avec les réseaux métiers voisins (DevLOG, RESINFO, QeR, Medici, etc.) ;

par le biais d'actions communes (de type journées thématiques dans lesquelles seront fait état de retours d'expériences ou proposées des formations s'appuyant sur des experts des deux communautés).

3.2.2 Pour accompagner la communauté et acquérir les compétences nécessaires à l'évolution du métier et/ou de la technologie

Le réseau Calcul continuera dans ses actions structurantes à explorer des pistes de collaboration avec les Instituts, les Délégations Régionales et CNRS Formation Entreprise pour permettre aux experts du réseau de :

- contribuer à la réflexion sur l'évolution du calcul, ses enjeux et ses réalisations ;
- contribuer à la formation permanente sur demande des Délégations Régionales ;
- travailler avec CFE sur des actions totalement externalisées ;
- concentrer les formations sur la popularisation de nouvelles pratiques, à la pointe des développements récents.

De nombreux Instituts du CNRS sont concernés par le calcul scientifique et intensif, avec qui le réseau Calcul espère poursuivre le dialogue : des contacts ont été pris en 2016 avec les correspondants formation de ces Instituts pour initier une discussion. L'année prochaine, ce contact sera renouvelé et étendu aux Chargés de mission Calcul, s'ils existent dans ces Instituts. L'objectif est dans un premier temps de faire connaître le réseau et de dialoguer de manière informelle des problématiques associées au Calcul. Ces prises de contact explorent des pistes de collaboration, discutées au bureau. Dans le même esprit, les membres du bureau s'informent et échangent sur les évolutions du calcul. Cette activité, à la frontière entre le réseautage et la veille technologique conduit à entretenir des liens avec les différents acteurs du calcul (MICADO, COCIN, GENCI, AMIES, INRIA, etc.)

Concernant les aspects de formation, le relais vers les Délégations Régionales a été exploré, partant du constat suivant. Depuis plusieurs années, le réseau Calcul a proposé des formations sur différents aspects du calcul. Certaines de ces formations se déclinent à deux niveaux : débutant et avancé. Les formations de niveau débutant sont toujours très recherchées par les membres du réseau et servent de base aux formations plus expertes. Cependant, afin de concentrer la demande de moyen sur les aspects les plus nouveaux, le bureau du réseau souhaite inciter les partenaires institutionnels à s'appuyer sur le vivier d'experts lorsqu'il définit ses offres de formation. Il a donc été décidé de prendre contact régulièrement avec les Délégations Régionales pour favoriser cette démarche et recueillir auprès d'elles l'expression de besoins autres que ceux exprimés habituellement au sein du réseau (via la liste de diffusion ou lors de journées thématiques et ateliers à la suite desquels un sondage sur les besoins futurs est réalisé).

Via la Mission pour l'Interdisciplinarité, le bureau est en contact avec CNRS Formation Entreprise, suivant les procédures de la formation permanente du CNRS. Suite à des discussions avec CFE, le bureau du réseau a pour projet de recueillir également des idées de formation à mettre au catalogue de CFE sur des actions externes.

Enfin, le réseau Calcul et son bureau continueront d'alimenter le site FoCal, catalogue partagé des formations en calcul <http://formation-calcul.fr/> et d'en faire la promotion auprès des différents partenaires.

Une réflexion est menée au sein du bureau pour accompagner encore plus la communauté en renforçant, dans les formations et les ateliers, la mise en pratique sur des cas applicatifs provenant des participants. Cela est un exercice délicat lorsque les participants proviennent de milieux disciplinaires hétérogènes. Les ANF sur des outils transverses à plusieurs communautés (visualisation, fouille de données...) s'y sont frottées, sous forme de démonstrateurs. Pour fonctionner sur un mode de type *sprint*, il est sans doute nécessaire de découpler la découverte des outils et la mise en pratique sur les cas des participants, à un intervalle de 6 mois à 1 an. Cette solution est explorée dans l'ANF proposée pour 2018 sur les moteurs d'expérimentation numérique.

3.2.3 Pour organiser le partage et la diffusion des connaissances, savoirs et savoir-faire

Voici la liste des actions qui sont prévues.

- Continuer à relayer l'information via la liste de discussion, le site web et le compte Twitter de Calcul. Le site web est actuellement en refonte pour permettre un accès plus largement collaboratif.
- Continuer à soutenir des actions de formation, des journées thématiques, ateliers et mini-workshops. Sont prévues à court terme :

- une action d'état de l'art et retour d'expérience autour du GPU, suite à une interaction avec l'INS2I, en lien avec le traitement des données massives ;
 - une formation à l'utilisation des moteurs d'expérimentation numérique au service de la reproductibilité des campagnes de simulation afin de populariser ces pratiques ;
 - une formation à l'administration de modules et bibliothèques scientifiques sur les clusters de calcul (en partenariat avec RESINFO).
- Continuer à lancer des appels à contribution et à expression de besoins, en suivant l'ouverture d'une boîte à idées proposée en novembre 2016. Cela sera reconduit à la même période chaque année pour les actions arrêtées en fin d'année. Un appel à contribution similaire sera lancé pour la définition des ANF, suivant un calendrier qui sera calé sur les Écoles Thématiques à partir de l'année prochaine.
 - Favoriser la reproductibilité et le partage des connaissances au travers de documents interactifs pour la résolution de problèmes numériques. Aujourd'hui, on voit de plus en plus de formats numériques permettant d'avoir des articles, des livres scientifiques ou technologiques interactifs. L'interactivité signifie qu'un lecteur peut aisément réexécuter les lignes de codes intégrées aux documents, de même que modifier les paramètres (voir les outils autour de Jupyter). Le réseau Calcul s'appuie déjà sur ces technologies pour ses formations (voir "Python avancé pour le calcul scientifique"). Un autre avantage, lorsque l'on a une infrastructure de type JupyterHub, est d'avoir des espaces de formations conviviaux avec tous les outils déjà installés. Les participants n'ont plus qu'à aller, via leur navigateur web, sur cet environnement. Le réseau Calcul pourra également apporter son savoir-faire pour enrichir l'expérience des acteurs du calcul dans la rédaction de documents interactifs et la promotion des savoir-faire.
 - Améliorer les pratiques autour du packaging d'applications permettant aux acteurs du calcul de diffuser facilement leur travail et en faire bénéficier la communauté. On pourra s'appuyer par exemple sur des outils tels que Docker, cookie-cutter, conda-forge, ... et des outils d'intégration continue disponibles sur github ou sur des gitlab locaux permettant d'automatiser ce packaging. Il est également indispensable de travailler avec les services de valorisation notamment pour tout ce qui est problème de licences. Le réseau Calcul a déjà mené des actions de sensibilisation autour de ce sujet mais c'est un travail qui doit se faire de manière régulière.
 - Renforcer le réseau en accentuant le développement de journées thématiques faisant un état de l'art des techniques et pratiques provenant de différents instituts de recherche en France et si possible en Europe. (e.g. le réseau a été contacté par le CNES l'année dernière pour un échange de pratiques autour du calcul et des données; les actions de formations menées par PRACE et pilotées en France par la Maison de la Simulation sont actuellement communiquées au réseau via sa liste de diffusion).