

LABORATOIRE DE L'ACCÉLÉRATEUR LINÉAIRE

IN2P3-CNRS et Université PARIS-SUD

Centre Scientifique d'Orsay - Bât 200 - B.P. 34

91898 ORSAY Cedex (France)



Guy WORMSER

Directeur

Tél. : 01 64 46 83 01

Fax : 01 69 86 98 63

Email : wormser@lal.in2p3.fr

Le 12 janvier 2006

Proposition d'un Nouveau Programme Interdisciplinaire ICAD Infrastructure de Calcul Distribué

Guy Wormser

Responsable scientifique d'EGEE-CNRS et EGEE-France

au nom des participants du projet EGEE

Objectifs et attendus

1. Les grilles de calcul

La très forte croissance des besoins en calcul scientifique et en capacité de stockage de données liée à l'exploitation de nouveaux instruments (citons par exemple les accélérateurs de très haute énergie et/ou intensité en physique subatomique, les télescopes et satellites en astrophysique, les bases de données liées au décodage du génome ou à l'imagerie en biologie et en médecine, les nombreux capteurs en science de la Terre) nécessite la mise en place de nouvelles infrastructures de calcul complémentaires des grands centres nationaux. En effet, il devient impossible par exemple de stocker l'ensemble des données issues d'un même instrument dans un seul site informatique. : la mise en place d'un réseau de sites de stockage s'impose donc.

Une des voies les plus prometteuses, vigoureusement poursuivie actuellement au CNRS, est basée sur le concept de grilles de calcul, vastes infrastructures informatiques distribuées permettant un accès facile et transparent à des ressources de calcul et de stockage hétérogènes. Les très hautes performances des réseaux informatiques, atteignant maintenant de façon courante plusieurs dizaines de Gigabits par seconde, soit des débits comparables à ceux d'une unité centrale, rendent désormais réaliste la distribution géographique des données sur de grandes distances.

Le caractère distribué d'une telle infrastructure reflète de plus en plus celui de ses utilisateurs tant à l'intérieur d'une même communauté, ou organisation virtuelle, que globalement.

La très grande flexibilité et le caractère dynamique de l'allocation des ressources tant à l'intérieur d'une même communauté qu'entre communautés contribuent à le rendre très pertinent sur le plan économique et le mieux adapté à répondre à des situations où les délais d'obtention de résultats sont des enjeux essentiels : simulations liées à une catastrophe naturelle, à une opportunité ou une urgence scientifique. Des exemples récents en sont fournis plus bas.

Les grilles permettent de fédérer et de mutualiser les ressources des différentes communautés scientifiques mais aussi de développer leur utilisation des centres de calculs nationaux (IDRIS, CINES) par l'établissement des passerelles adéquates. En effet, dans la mesure où les besoins

informatiques de chaque communauté sont mieux définis, les tâches nécessitant les performances de supercalculateurs peuvent être plus efficacement aiguillés vers les centres nationaux.

2. L'état de l'art

Une telle infrastructure est maintenant en place. [EGEE \(Enabling Grid for European E-science\)](#) est la plus grande infrastructure de calcul distribué du monde actuellement déployée et en cours de production. La grille de calcul EGEE regroupe plus de 15000 processeurs et 5 PétaOctets de stockage répartis sur 150 nœuds en Europe, en Amérique et en Asie. EGEE compte plusieurs milliers d'utilisateurs provenant d'une dizaine de communautés scientifiques internationales les plus diverses, allant de la physique des hautes énergies à la finance en passant par la biologie, la médecine, la pharmacologie, l'astrophysique, la cosmologie, la chimie, sciences de la Terre, les bibliothèques numériques, etc... Ces utilisateurs soumettent plusieurs milliers de travaux par jour à l'infrastructure qui fonctionne 24 heures sur 24, 7 jours sur 7.

Cette infrastructure de production se distingue et intervient en complément des autres grilles qui constituent la grille de recherche GRID5000 et le réseau de supercalculateurs DEISA. Le programme proposé permettra d'augmenter considérablement les relations entre ces différents outils et entre les communautés correspondantes.

3. Les objectifs du programme ICAD

Le programme interdisciplinaire ICAD poursuit quatre objectifs :

- Maintenir, amplifier, *pérenniser* l'infrastructure de calcul distribuée construite autour d'EGEE
- Élargir le cercle des utilisateurs de cette infrastructure tant à l'intérieur des communautés déjà utilisatrices que vers de nouvelles communautés
- Favoriser la recherche sur les grilles en tant qu'objets informatiques du futur
- Développer une collaboration avec les centres de calcul nationaux (IDRIS, CINES) pour la prise en charge des besoins spécifiques en calcul intensif

La création de ce programme permettra donc de beaucoup mieux intégrer les ressources de calcul offertes par EGEE dans la panoplie des moyens de calcul disponibles au CNRS, panoplie dont la diversité est rendue nécessaire par la très grande palette des besoins et permet de la satisfaire de façon optimale.

Contexte national, européen et international

Le projet de grille nationale

Une des conditions fondamentales du succès d'EGEE-II est de pouvoir proposer dès 2007 une organisation permettant sa pérennisation à l'Union Européenne et à l'ensemble des pays partenaires. La vision de cette structure pérenne est pour l'instant très similaire à celle des réseaux informatiques : des agences nationales de statut pérenne (RENATER), organisées en réseau au niveau européen (DANTE), des projets cofinancés par l'UE et les pays membres (GEANT), avec un support opérationnel sous traité à l'industrie (Communications et Systèmes en France). Notons [la proposition du CERN](#) de servir d'hôte à cette structure pérenne européenne détaillée dans la référence [EGO](#).

Certains pays européens (Angleterre, Italie, Allemagne) se sont d'ores et déjà dotés d'une structure nationale de grille ([UK E-science](#), [Grid.it](#), [D-GRID](#)). **Une structure équivalente manque cruellement en France et c'est l'objectif majeur d'ICAD de constituer ce projet de grille nationale.**

Il existe en France d'autres grilles et en particulier GRID5000. Cet outil de recherche ne vise pas, au contraire d'EGEE, à fournir des ressources de calcul soutenues à une très vaste communauté d'utilisateurs mais à étudier le fonctionnement optimum d'une grille de calcul et à mettre au point les logiciels du futur. On conçoit donc facilement l'aide mutuelle que peuvent s'apporter ces deux systèmes : on pourra tester à pleine échelle sur l'infrastructure de production les nouvelles méthodes mises au point sur GRID5000 et fournir via l'étude détaillée du fonctionnement de la très grande infrastructure EGEE avec ses milliers d'utilisateurs les données nécessaires aux chercheurs en informatique pour la meilleure compréhension de l'objet grille. Le programme ICAD fournira un cadre très efficace pour mener ces travaux.

Le développement de la grille nationale s'accompagne de nombreuses initiatives locales très structurantes autour des nœuds importants de la grille. Ainsi, la région Auvergne (projet INSTRUIRE) a réussi à créer des synergies avec de nombreux partenaires en provenance des campus de recherche ou des industriels locaux, et avec le soutien actif des institutions régionales. L'Ile de France ou Rhône-Alpes suivent de près ce modèle.

La valorisation industrielle de ces développements informatiques fait également l'objet d'intenses travaux en France. Plusieurs entreprises importantes sont partenaires de ce projet : Communication et Systèmes, la Compagnie Générale de Géophysique sont des acteurs très entreprenants dans EGEE, IBM a signé un accord de partenariat avec l'IN2P3 dans le cadre duquel des développements logiciels importants ont été effectués, plusieurs projets mêlant laboratoires publics et entreprises privées ont été soutenus par diverses instances, citons en particulier les projets RUGBI et OPENPLAST dans la région Rhône Alpes. Le forum industriel d'EGEE sous la cotutelle du CNRS et de l'Ecole Centrale de Paris joue un rôle clé pour faciliter l'appropriation dans l'industrie des outils de grille développés par EGEE et rassemble plus d'une centaine de partenaires industriels de tous horizons et de toutes tailles (<http://public.eu-egee.org/industry/>), dont beaucoup en France.

Contexte européen et international

L'Union Européenne, dans le cadre de son 6ème PCRD, est le principal bailleur de fonds d'EGEE. EGEE est un projet d'infrastructure I3 dépendant de la DG Recherche et financée à hauteur de 32 M€ pour la période 1/4/2004-31/03/2006. Les soixante-dix partenaires d'EGEE provenant de vingt-sept pays différents —de tailles très diverses puisqu'on trouve aussi bien des grands organismes de recherche comme le CNRS ou le CEA ou des petits laboratoires, instituts ou sociétés privées— sont coordonnés par le CERN et regroupées en 11 fédérations qui rassemblent les différents partenaires d'un même pays pour les grands pays et des groupements de pays pour les plus petits. Le point commun de tous ces partenaires est d'être les leaders de l'effort de grilles dans leur région respective. Les 11 fédérations d'EGEE sont :

- Angleterre-Irlande
- Allemagne-Suisse
- France
- Italie
- Europe Centrale
- Europe du Nord
- Europe du Sud-Ouest
- Europe du Sud –Est
- Russie
- USA
- CERN

Une nouvelle fédération Asie (Japon, Taiwan, Corée) va rejoindre EGEE-II et de nombreux partenariats sont noués entre EGEE et des pays émergents comme ceux de l'Amérique latine ou du Maghreb.

Il existe aux Etats-Unis une grille aux ambitions à peu près comparables à EGEE, [Open Science Grid](#), quoique de taille nettement plus petite. Au Japon, la grille NAREGI poursuit également des objectifs similaires. Ces grilles sont interopérables car elles font toutes au moins en partie membres du même projet WLCG dévolu à l'analyse des données du [LHC](#), prochain instrument majeur de la recherche en physique des particules, où les communautés américaine et japonaise jouent un rôle important. La réunion d'EGEE et d'OSG constitue une des grilles universelles pluridisciplinaires les plus avancées actuellement sur les plans techniques et opérationnels.

Communautés visées

Le programme ICAD vise toutes les communautés d'utilisateurs potentiellement intéressés et la communauté des chercheurs en informatique. Les six départements du CNRS et ses deux instituts sont donc concernés. Tous ces départements et instituts sont en fait déjà, à l'exception notable des sciences de l'homme et de la chimie, utilisateurs d'EGEE, à des degrés divers. Notons qu'EGEE accueille cependant plusieurs applications relatives aux sciences humaines et à la chimie dans les domaines de l'archéologie, des bibliothèques digitales, de la finance, de la chimie quantique, mais sans participation institutionnelle pour l'instant d'équipes du CNRS. Le caractère multidisciplinaire d'EGEE est donc tout à fait à l'image de celui du CNRS, ce qui justifie pleinement à la fois le rôle moteur joué par le CNRS dans le secteur applicatif d'EGEE et le rôle que le CNRS doit prendre grâce à ICAD dans la structuration de l'infrastructure de grille nationale.

Il faut bien garder à l'esprit que toutes les applications ne sont pas adaptées à l'infrastructure de grille proposée. En effet, bien qu'il soit possible de soumettre des travaux nécessitant la communication inter-processeurs (via par exemple le protocole MPI), les travaux nécessitant des calculs couplés intenses sont traités beaucoup plus efficacement sur des supercalculateurs comme ceux de l'IDRIS ou du CINES. En revanche, certains des travaux soumis à ces centres pourraient très bien s'effectuer sur la grille ICAD et libérer ainsi de précieuses ressources pour ceux qui en ont le plus besoin. On voit donc la nécessité de proposer aux utilisateurs une offre complète et complémentaire de moyens de calcul pour satisfaire les besoins de la façon la plus économique et la plus efficace possible.

Les quelque cinq mille travaux soumis quotidiennement sur la grille EGEE concernent actuellement majoritairement la physique des particules, la biologie et la médecine, les sciences de la terre et l'astrophysique. Les exemples les plus significatifs de ces applications sont indiqués ci-dessous.

Le premier succès d'EGEE/WLCG est la satisfaction des immenses besoins de calcul de la seule communauté qui a fait le pari de la grille comme outil unique pour l'ensemble de ses sept mille utilisateurs répartis dans le monde entier et désireux d'analyser les données du LHC. Cette infrastructure, aujourd'hui déployée opérationnellement avec une taille conforme aux prévisions, est maintenant utilisée par tous et a fait la preuve d'une fiabilité et d'une flexibilité satisfaisantes.

D'autres succès d'EGEE concernent son utilisation dans d'autres domaines scientifiques, ouvrant cette fois des champs complètement nouveaux. La grille permet des avancées scientifiques spectaculaires illustrées ci-dessous grâce à deux caractéristiques essentielles :

- la possibilité de mobiliser des ressources informatiques très importantes pour un temps court pour certaines applications devenues critiques avec une flexibilité maximale
- la possibilité d'accéder de façon transparente à des bases de données distribuées géographiquement et dans des systèmes informatiques, administratifs et institutionnels très hétérogènes et souvent inaccessibles de façon commune autrement.

Notons par ailleurs la très forte participation des équipes françaises et du CNRS dans tous ces « success stories ».

En *bioinformatique* (C. Blanchet/IBCP), le portail [GPS@](#) est maintenant ouvert. Il permet d'offrir une puissance considérablement accrue à des demandes de séquençage génomique de toute provenance en s'appuyant sur un réseau de bases de données également très élargi. Différentes initiatives très prometteuses lancées par le groupe GIGn (<http://ibcp.gign.fr>) ont jeté les bases en France d'une véritable communauté d'utilisateurs de la grille pour la recherche en génomique.

En *médecine* (V. Breton/LPC Clermont), le projet [WISDOM](#) a permis de mobiliser en quelques semaines l'ensemble de la grille pour sélectionner, à partir de plusieurs dizaines de millions de structures candidates, la centaine la plus prometteuse pour lutter efficacement contre la malaria. Ce résultat sans précédent ouvre de nouvelles perspectives pour traiter une maladie tuant des millions de personnes dans le monde et qui pourtant ne bénéficie pas par ailleurs de l'attention nécessaire. Le très faible coût marginal de ce genre d'action pour la grille est donc aussi essentiel dans ce cas.

En *sciences de la Terre* (M. Petitdidier/IPSL), c'est d'abord la création d'une communauté européenne des sciences de la Terre qu'il faut saluer. La grille a permis de rapprocher des équipes de très nombreux pays européens et du Sud de la Méditerranée dans des domaines aussi variés que l'observation de la couche d'ozone, la sismologie, l'hydrologie, la dynamique terrestre, la climatologie, etc. Les résultats les plus spectaculaires concernent la gestion des catastrophes : traitement complet d'un tremblement de terre moins de vingt-quatre heures après son déclenchement (essentiel pour en connaître les causes, modalités, probabilités de répliques, etc.), gestion des inondations, scénarios de développement durable, etc.

Concernant les *applications industrielles* (D. Thomas/CGG), [EGEODE](#) est un logiciel propriétaire de la compagnie CGG permettant la reconstruction 3D de nappes pétrolières à partir de campagnes de mesures sismiques. Grâce à un système innovant de gestion de licences, ce logiciel très performant a pu être mis à la disposition d'équipes académiques.

Le [partenariat scientifique entre les pays du Sud et du Nord](#) (G. Wormser/LAL Orsay) bénéficie beaucoup de la facilité d'accès et de mobilisation des ressources apportées par la grille. Ceci a été récemment démontré par la conférence sur le partage du savoir autour de la Méditerranée qui comporte une session permanente sur la Grille et ses applications dans ce contexte. On constate avec satisfaction un double flux très prometteur: l'accès par des scientifiques du Sud aux données scientifiques du Nord (physiciens marocains au LHC par exemple) et le déploiement d'applications issues du Sud sur les grilles du Nord (gestion de l'eau en Tunisie par exemple). Signalons que ce partenariat sera renforcé dans le futur par le projet EUMEDGRID soutenu par l'Union Européenne et dont EGEE est l'interlocuteur principal.

Descriptif du programme et modes d'action

Les trois actions du programme ICAD sont donc :

a) le soutien, le développement et la pérennisation de l'infrastructure existante

Les actions de soutien actuelles à l'infrastructure comportent les frais de fonctionnement du personnel permanent participant soit environ 150 k€ par an, le soutien aux nœuds existants, pour l'instant de 100 k€ en 2005 répartis entre les différents nœuds partenaires. Le développement absolument nécessaire de l'infrastructure dans le cadre du programme par répartition des moyens disponibles pour l'achat du matériel se fera par le comité scientifique du programme en concertation avec la cellule TGE/TGI responsable des moyens de calcul. Le tableau financier ci-dessous indique

la montée en puissance prévue de ce soutien à l'infrastructure, de 100 k€ en 2005 à 300 k€ en fin de programme.

Il faut noter que l'infrastructure ne se limite pas aux moyens de calcul et de stockage. Il faut également soutenir les centres de formation et de dissémination.

b) le renforcement de la communauté des utilisateurs.

Deux axes seront à poursuivre : un axe « horizontal » de recrutement de nouvelles communautés et un axe « vertical » pour amplifier l'usage et le retour scientifique des utilisateurs déjà présents.

Pour l'axe horizontal, nous proposons de reproduire au niveau national la procédure qui a bien fonctionné au niveau international dans EGEE : un conseil de recrutement et de suivi des applications (CRSA), formé d'experts indépendants, lance deux fois par an des appels à projets en direction de la communauté scientifique de la façon la plus large possible. L'objectif n'est pas de sélectionner les meilleures applications mais de définir les moyens à mettre en œuvre pour les intégrer toutes de façon maîtrisée. Le comité s'attache à déterminer

- L'impact scientifique potentiel apporté par la grille à cette nouvelle communauté
- le degré de maturité de la collaboration
- le degré d'expertise logicielle dans l'informatique
- la main d'œuvre nécessaire au déploiement de l'application

Ce même comité est en charge du suivi régulier des applications labellisées.

Un contingent de main d'œuvre spécialisée sera mis à la disposition des nouveaux arrivants pour faciliter leur intégration.

c) la recherche en informatique associée

Cette action du programme vise à favoriser l'échange scientifique entre les différents acteurs de la recherche informatique autour du calcul distribué. Un certain nombre de forums existent déjà tant au niveau national qu'international. Le rôle spécifique de celui-ci sera d'encourager la recherche pertinente pour l'infrastructure mise en place, visant tant à son étude qu'à son amélioration.

En particulier, il faudra veiller à un dialogue beaucoup plus actif qu'aujourd'hui entre l'infrastructure de grille construite autour d'EGEE, les travaux menés autour de DEISA, réseau de supercalculateurs européens coordonné par l'IDRIS, la grille de recherche française GRID5000, les réseaux d'excellence européens de recherche amont et en particulier COREGRID.

Cette action aura également pour rôle de coordonner les réponses aux différents appels d'offre émanant de l'ANR, du RNTL, de l'Union Européenne.

Un des avantages importants de la création de ce programme est celui de créer la structure ad hoc permettant son évaluation. En effet, pour l'instant, aucune instance française n'évalue EGEE. Un conseil scientifique de haut niveau attaché au programme permettra de combler cette grave lacune.

Enfin ce programme permettra de soutenir l'enseignement en France des techniques de grilles par le biais d'écoles, de cours, etc...

Partenaires

Le programme ICAD ayant vocation à devenir le cœur de l'infrastructure nationale de grille insérée au sein de l'entité européenne, les partenariats devront se nouer tant au niveau national parmi les différents organismes potentiellement intéressés par l'utilisation de la grille qu'avec les interlocuteurs naturels au niveau européen et international que sont les membres d'EGEE.

• Nationaux

Les partenaires nationaux sont bien sûr à chercher parmi les membres d'EGEE, soit le CEA, l'Ecole Centrale de Paris, le CNES ainsi que les deux sociétés CGG et CS. Ce partenariat pourra s'enrichir avec RENATER et l'INRIA. Certaines universités très actives pourraient également rejoindre le programme. Il serait souhaitable d'avoir une implication forte du ministère de la recherche.

Comme indiqué plus haut, ICAD nouera des partenariats scientifiques beaucoup plus étroits qu'aujourd'hui avec les autres acteurs des différentes grilles de calcul et en particulier GRID5000 la grille de recherche française, DEISA le réseau européen de supercalculateurs.

• Internationaux

Les partenaires internationaux naturels du projet sont bien sûr tous les participants du projet EGEE et EGEE-II qui regroupe la quasi totalité des pays européens, les Etats-Unis, la Russie et les acteurs principaux en Asie, ainsi que certains pays émergents associés. Ce lien permet également le partenariat avec l'Union Européenne, principal bailleur de fonds d'EGEE et le CERN, organisme intergouvernemental actuel coordinateur d'EGEE et qui se propose d'accueillir la future structure pérenne européenne.

Horizons futurs,

Le futur d'ICAD est une infrastructure de grille nationale qui doit comporter une cellule opérationnelle chargée de l'opération et de la maintenance de la grille, une interface avec la structure européenne dans laquelle elle s'intègre, et un partenariat avec RENATER, la structure nationale en charge des réseaux informatiques pour la recherche. Un modèle à explorer consiste à élargir les missions de RENATER pour lui confier l'ensemble de ces tâches. L'Union Européenne souhaite très vivement définir rapidement les perspectives de pérennisation d'EGEE, un atelier sur ce thème est organisé les 30 et 31 janvier 2006.

Budget et programmation pluriannuelle

Le budget consolidé nécessaire pour le fonctionnement de l'infrastructure EGEE au CNRS est actuellement de 3,24 M€ par an, se répartissant en trois postes : 1,5 M€ de salaires de personnel permanent, 1,5 M€ en provenance de l'Union Européenne servant à financer en totalité du personnel temporaire et 0,24 M€ de fonctionnement pour le personnel permanent et l'achat de ressources informatiques pour les différents nœuds. Le financement de l'Union Européenne est acquis jusqu'au 1^{er} Avril 2008, date de fin du programme EGEE-II qui vient d'être accepté par l'Union Européenne. Au delà, il est prévisible qu'un fort soutien de l'Union Européenne se manifesterait mais sa forme et son montant exacts dépendront beaucoup du schéma de pérennisation choisi. D'autre part, le soutien que le CNRS et ses partenaires voudront alors apporter au développement de l'infrastructure dépendra bien sûr des succès obtenus d'ici là dans l'appropriation de l'outil par les différentes communautés scientifiques. Le budget du programme proposé à partir de 2008 est basé sur le maintien d'une aide de l'Union Européenne à son niveau actuel et sur un soutien accru à l'infrastructure lié à un accroissement important du nombre d'utilisateurs. Les sommes indiquées ci-dessous correspondent au budget total du programme qui sera abondé, nous l'espérons, par le CNRS mais aussi par tous les autres partenaires institutionnels.

	2006	2007	2008	2009
Fonctionnement grille (hors personnel)	150	150	150	150
Soutien infrastructure	120	200	250	300
Aide aux applications	100	100	100	100
Animation recherche	30	30	50	50
Formation	50	50	100	100
Total	450	530	650	700

Sur les 450 k€ demandés en 2006, 240 k€ étaient déjà programmés au titre du simple soutien à EGEE (200 k€ de la DSG et 40 k€ de l'IN2P3). Le lancement du programme ICAD dès 2006 ne nécessiterait donc que 210 k€ supplémentaires.

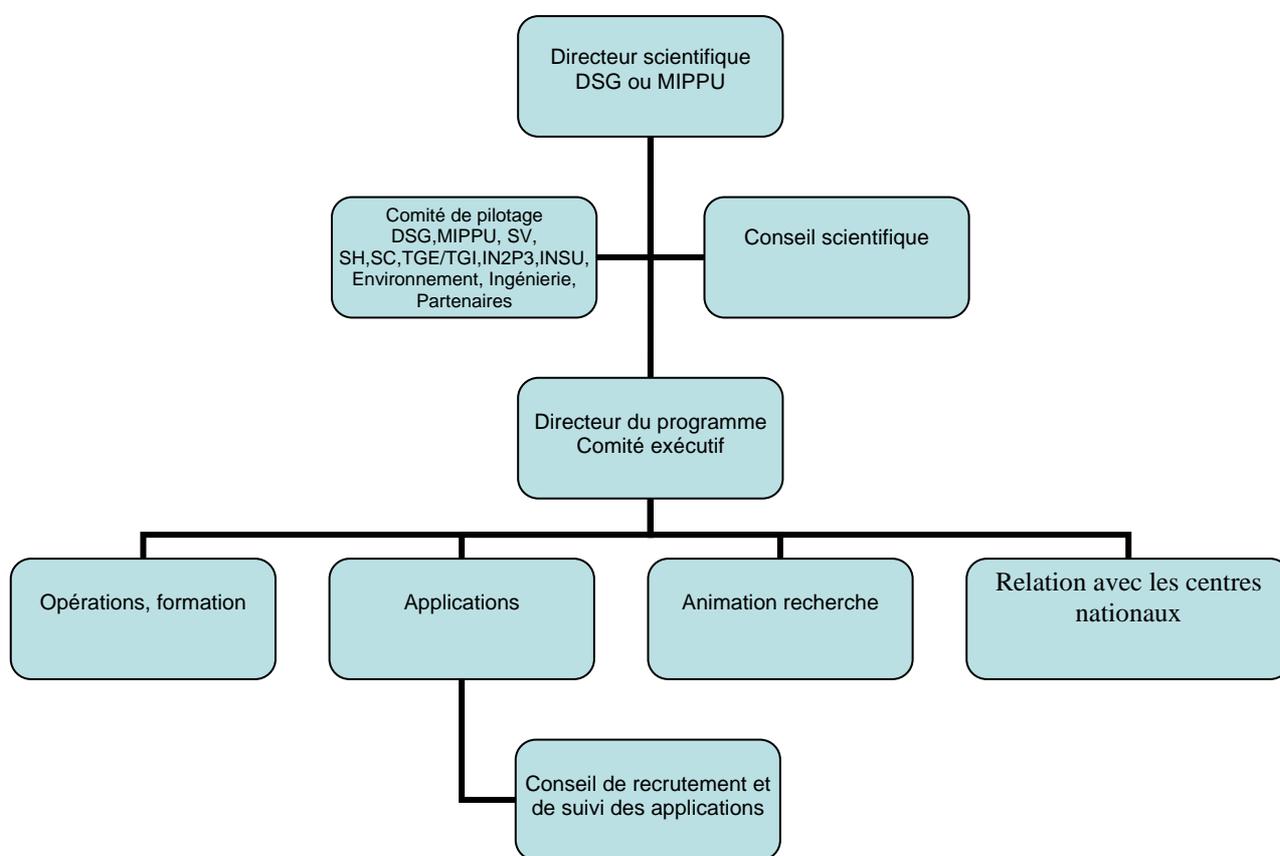
Direction

La structure de la direction du programme est organisée autour du responsable général du programme qui s'entoure d'un comité de pilotage comprenant le président du comité scientifique chargé du recrutement et du suivi des applications, le responsable opérationnel de la grille et celui de l'animation du secteur recherche.

Le représentant français dans EGEE pour le CNRS sera le responsable du programme et sera proposé aux partenaires comme représentant d'EGEE France.

Le comité de pilotage du programme comprend les représentants de la Direction Scientifique générale, des six départements et des deux instituts du CNRS, d'un représentant de la cellule TGE/TGI et d'un représentant de chaque organisme partenaire.

Le conseil scientifique d'ICAD est formé de personnalités extérieures au programme, françaises et étrangères, membres du CNRS ou non. Sa tâche essentielle est d'effectuer une fois par an l'évaluation des différents aspects du programme. Le comité exécutif chargé de la bonne marche du programme est formé du responsable du programme et des responsables des trois branches du programme. Le schéma global est indiqué ci-dessous.



Conclusion

Le dispositif indispensable de soutien et de pérennisation à l'infrastructure de production EGEE actuellement disponible au CNRS a été décrit en détail et est proposé sous forme de programme interdisciplinaire ; En effet, cette structure permet de remplir de façon optimale cet objectif tout en créant les conditions nécessaires à l'augmentation de la base des utilisateurs, à l'animation de la recherche en informatique associée et la coordination avec les autres ressources informatiques disponibles au CNRS. Il est très important, en effet, d'avancer de pair sur tous ces fronts pour que la France puisse disposer d'une véritable infrastructure pérenne de calcul distribué, insérée dans un réseau mondial, et capable de fournir à une très vaste communauté d'utilisateurs couvrant l'ensemble des disciplines concernées l'outil puissant, efficace et économique nécessaire pour les découvertes scientifiques de demain. Les besoins financiers pour lancer l'ensemble de ce programme ambitieux s'élèvent à 210 k€ en 2006 (en supplément des 240 k€ déjà programmés pour le soutien à l'infrastructure) pour atteindre 700 k€ en 2009, soit un total de 2,33 M€ sur 4 ans. Il est à noter que le CNRS ne contribuerait sans doute pas seul au financement de ce programme car il est très probable que certains organismes déjà partenaires d'EGEE (CEA, CS, CGG, Ecole Centrale, CNES) acceptent de le soutenir, et l'arrivée de nouveaux partenaires n'est pas à écarter (INRIA, Universités, Ministère)

Annexe 1. Informations complémentaires sur le projet EGEE

La structure interne d'EGEE

L'Union Européenne, grâce à son 6ème PCRD, est le principal bailleur de fonds d'EGEE. EGEE est un projet d'infrastructure I3 dépendant de la DG Recherche et financée à hauteur de 32 M€ pour la période 1/4/2004-31/03/2006. Les participants internationaux à EGEE sont décrits dans la note ci-dessus.

Le Project Management Board (PMB) d'EGEE est constitué d'un représentant par fédération, ainsi que d'un représentant du programme de réseau informatique GEANT et du programme de grille pour le LHC, WLCG. Le PMB est l'instance dirigeante d'EGEE qui nomme tous les responsables opérationnels, fixe les règles internes (décrites dans le Consortium Agreement signé par tous les partenaires), distribue le budget entre les différentes fédérations, etc. Le Consortium Agreement comprend notamment l'obligation pour chaque partenaire de contribuer au projet par des ressources internes d'un montant égal à ce qu'il reçoit de l'Union Européenne, quel que soit le type de sa gestion administrative vis-à-vis de l'UE.

La fédération française dans EGEE est formée des cinq entités suivantes : CNRS, CEA, Ecole Centrale de Paris, Communications et Systèmes, et Compagnie Générale de Géologie (CGG).

Ces cinq partenaires se réunissent plusieurs fois dans l'année et notamment avant chaque réunion importante du PMB pour harmoniser leur position. Les représentants des cinq partenaires désignent le représentant français au PMB et son suppléant. Compte tenu du poids très important du CNRS dans la fédération française (environ 75% de l'effort fourni), les représentants des différentes unités du CNRS participent le plus souvent aux réunions des représentants de la fédération française.

L'expérience a montré que cette structure tant au niveau français qu'au niveau d'EGEE permettait de résoudre de façon optimale tous les problèmes rencontrés jusque là.

La politique d'utilisation des ressources est à l'appréciation de chaque propriétaire des ressources d'un nœud. Il doit indiquer quelles organisations virtuelles (ie quelles communautés) il autorise et avec quelle priorité à venir utiliser ses ressources. La stratégie globale est bien sûr de promouvoir le maximum de souplesse sur l'utilisation des ressources de la grille. Les outils de comptabilité existent pour mesurer localement et globalement l'utilisation de la grille par chaque communauté.

Les relations EGEE-WLCG

Il est très important de bien comprendre les relations existant entre les projets EGEE et [WLCG](#). Ces deux projets n'ont pas la même finalité, la même durée, la même structure, la même couverture géographique mais partagent la même infrastructure et le même logiciel. On peut donc parler de siamois à un corps et deux têtes.

WLCG est le projet dont l'unique objectif est de fournir l'infrastructure de calcul aux expériences auprès du LHC. C'est donc un projet dédié à la seule physique des hautes énergies (en fait une partie de cette communauté), de très longue durée, coordonné par le CERN et organisé de façon similaire aux quatre expériences autour du LHC. Ses membres sont les laboratoires contribuant aux ressources matérielles pour le LHC, c'est-à-dire la plupart des laboratoires HEP du monde entier. Les grands centres américains sont donc partenaires de WLCG mais pas d'EGEE. Réciproquement, tous les laboratoires non HEP d'EGEE ne sont pas membres de WLCG.

Toutefois, l'infrastructure de grille est commune aux projets WLCG et EGEE et c'est la même personne qui est en charge de la faire fonctionner dans les deux projets. De même, l'intergiciel

développé par EGEE est celui déployé sur WLCG et la même personne en est responsable dans les deux projets.

En résumé, WLCG reçoit d'EGEE une fraction très conséquente du personnel nécessaire à l'opération de son infrastructure, la totalité du personnel nécessaire au développement de l'intergiciel ainsi qu'à l'utilisation des ressources non HEP que les propriétaires des nœuds acceptent de mettre à la disposition du LHC. EGEE reçoit de WLCG la fraction des ressources matérielles déployées pour le LHC que chaque propriétaire de nœud accepte de mettre à la disposition des autres communautés.

Ce scénario, certes un peu complexe, fonctionne pour l'instant de façon satisfaisante. Pour éviter des tensions dans le futur, il sera nécessaire, et c'est ce qui est prévu, que les communautés non LHC fortement utilisatrices de la grille y incorporent leurs ressources informatiques propres à un niveau voisin de leur utilisation moyenne.

Les mêmes relations étroites entre EGEE et WLCG existent aussi au niveau national entre EGEE-France et LCG-France. Une participation croisée des responsables dans les différents comités de pilotage de ces projets assure la coordination politique nécessaire.

L'évaluation d'EGEE

EGEE pour l'instant n'est évalué que par l'Union Européenne. Une revue détaillée a eu lieu en Février 2005 et a donné des résultats très positifs. Une seconde revue détaillée aura lieu à la fin du projet en Avril 2006 tandis qu'une revue intermédiaire plus légère a eu lieu en Décembre 2005. Il est à noter cependant qu'aucune structure française n'a évalué EGEE depuis sa création. J'ai transmis au directeur général du CNRS et aux directeurs des départements scientifiques concernés les commentaires issus de la revue EGEE ainsi qu'un [rapport annuel d'activité en 2004 et 2005](#).

Le projet EGEE-II

EGEE-II est le projet constituant la suite d'EGEE (du 1/4/06 au 31/3/2008) avec les nuances suivantes :

- le nombre de partenaires augmente de 70 à 91. Pour la France, le CNES et l'association HealthGrid rejoignent la fédération française.
- la mission essentielle d'EGEE-II est de réunir toutes les conditions de pérennisation de l'Infrastructure d'EGEE grâce à la conjonction de la création d'un effort structuré de grille nationale dans chaque pays participant et d'une structure paneuropéenne stable de coordination.
- le domaine applicatif d'EGEE-II est encore plus vaste que celui d'EGEE avec notamment l'ajout de la recherche en fusion nucléaire autour d'ITER.

Le budget demandé à l'UE pour EGEE-II est de 37 M€ et est en cours de négociations. Une réduction minimale est possible. La part française et celle du CNRS sont essentiellement inchangées (respectivement 4 et 3 M€ sur deux ans).

Les responsabilités du CNRS dans EGEE

Les équipes du CNRS impliquées dans EGEE proviennent de quatre départements (IN2P3, STIC, SDU, SDV), de 9 laboratoires (CC_IN2P3, CPPM Marseille, LPC Clermont-Ferrand, LAL, LAPP pour l'IN2P3, CREATIS, IBCP pour SDV, I3S, LRI, UREC pour STIC, IPSL pour SDU) pour un total de 85 FTE et 50 FTE. Les domaines de responsabilité sont très variés :

- tour de contrôle de partie française de la grille et une de celles de la grille complète
- responsabilité de la partie réseau et des relations avec GEANT
- responsabilité de l'assurance qualité
- responsabilité du [secteur des applications](#) avec la direction globale de ce secteur, et celle du secteur des applications biomédicales
- responsabilité du [Forum Industriel](#)
- responsabilité de l'[EGAAP](#), (EGEE Generic Applications Advisory Panel), le conseil scientifique chargé du « recrutement » de nouvelles applications.

Tous les objectifs quantitatifs fixés au début du projet concernant le nombre de nouvelles applications ont d'ores et déjà été atteints. Concernant l'Industry Forum, qui regroupe plus de cent entreprises de toutes tailles, il est à noter que les sociétés IBM et DATAMAT financent à 100% une responsable de l'animation de ce forum basée au LAL à Orsay.
Le CNRS reçoit d'EGEE 1.5 M€ par an, ce qui permet de financer vingt-cinq postes d'ingénieurs.