

ACTION DE RECHERCHE AMONT

MASSE DE DONNEES

**MODELISATION
SIMULATION
APPLICATIONS**

APPEL A PROJET 2005

Site web de l'action pour les dossiers de réponses et les adresses d'envois
<http://mdmsa.labri.fr>
mél : mdmsapres@listes.u-bordeaux1.fr

TABLE DES MATIERES

| | | |
|------|--|---|
| 1. | Contexte | 3 |
| 2. | Objectifs Scientifiques | 3 |
| 2.1. | Algorithmique efficace..... | 4 |
| 2.2. | Analyse et traitement de contenus non textuels | 4 |
| 2.3. | Bases de données..... | 4 |
| 2.4. | Fouille de données..... | 4 |
| 2.5. | Réalité virtuelle, réalité augmentée | 5 |
| 2.6. | Simulation efficace de phénomènes complexes..... | 5 |
| 2.7. | Visualisation de données hétérogènes..... | 5 |
| 3. | Types de projets | 5 |
| 4. | Rédaction et Sélection..... | 6 |
| 5. | Suivi des projets sélectionnés..... | 6 |

1. CONTEXTE

Dans un contexte scientifique et un cadre international très actif tant du point de vue académique qu'industriel, cette action de recherche amont sollicite des projets novateurs, se situant au meilleur niveau international et contribuant à faire avancer significativement la recherche dans les domaines mentionnés et à renforcer ainsi la place de la France dans ces thématiques de recherche sur les scènes européenne et internationale.

Cet appel encourage et considérera toute proposition originale de type pluridisciplinaire. Les communautés scientifiques en biologie, génomique, chimie, astronomie, météorologie, physique des particules, environnement, observations de la terre ou encore en sciences humaines et sociales sont encouragées à répondre à cet appel en liaison avec des partenaires de la communauté informatique.

Dans le prolongement de l'action concertée incitative Masse de Données, et conformément à l'esprit des Actions de Recherche Amont, cette action visera à développer une recherche fondamentale et s'inscrit donc en complémentarité des différents réseaux de recherche et d'innovation technologique (RRIT): le RIAM, le RNRT, le RNTL ou le RNTS dont l'objectif est de développer les interactions entre recherche académique et entreprises.

Cette année, ces réseaux lancent en particulier des appels concernant: Mobilité et Interopérabilité, Intelligence Ambiante, Sécurité, Systèmes Embarqués, Réalité Virtuelle - IHM, Réseau d'information et de connaissance, Nouveaux modes de Distribution, Numérisation de la production audiovisuelle et du cinéma.

Toute recherche amont relevant des thématiques de ces réseaux pourra être examinée dans le cadre de cette action dès lors qu'ils relèvent de la problématique des masses de données.

2. OBJECTIFS SCIENTIFIQUES

L'objectif de cette action de recherche est de fortement dynamiser la recherche fondamentale en Informatique sur l'ensemble des aspects liés aux masses de données et tout particulièrement la modélisation et la simulation liées à des applications. Ces applications utilisent de plus en plus de données multimédias. Toutes les facettes de ces recherches sont adressées (données multimédia, fouilles de données, réalité virtuelle,...) avec un accent particulier sur les applications à moyen ou long terme.

L'amélioration des dispositifs expérimentaux et la part grandissante de l'informatique et de la simulation dans la plupart des champs disciplinaires conduit à la production de données en quantité de plus en plus importante : biologie, génomique, chimie, physique des particules, astronomie, météorologie, environnement, observations de la terre ou encore en sciences humaines et sociales. Ces applications impliquent de plus en plus des données multimédias (image, son, vidéo) dont il faut traiter les aspects de visualisation qui sont essentiels et qui font appel à des techniques de traitement d'images, de réalité virtuelle ou augmentée mais aussi aux techniques d'interfaces intelligentes.

Tout ceci implique une algorithmique sophistiquée aussi bien pour les bases de données, qui sont le cœur du système, que pour le problème de passage à l'échelle toujours présent dès lors que l'on parle de masse de données. Par exemple, il est nécessaire d'exploiter les capacités du système visuel pour explorer de grandes quantités d'information (fouille de données, interrogation d'une base d'images) ou encore simuler des phénomènes physiques et naturels ; la réalité augmentée présente la particularité de faire co-habiter des objets virtuels et des environnements ou objets physiques réels avec souvent des contraintes d'interactions physique-virtuel en temps réel.

La mise en réseau de plusieurs milliers d'ordinateurs (grille informatique) est un des moyens pour atteindre les puissances nécessaires aux traitements des diverses applications

évoquées ci-dessus. Mais il faut surtout être capable de combiner des grands entrepôts de données distribués reliés entre eux par des réseaux à très haut débit. De telles grilles de données doivent être capables d'offrir plusieurs classes de services pour transporter des flux de différentes natures (voix, images, données) avec des contraintes de qualité de service spécifiques et pour traiter de grands volumes de données. Le changement d'échelle dans la taille et le volume des données mais aussi en matière de complexité des traitements reste un enjeu récurrent dans ce domaine de recherche. Par exemple, la construction des normes telles que MPEG7 et l'algorithmique qui lui est associée doit se concevoir dans des infrastructures logicielles utilisant pleinement la gestion et le traitement distribués des programmes et des données. Les paragraphes ci-dessous sont des exemples et non une liste exhaustive des recherches amont que cette action doit traiter.

2.1. Algorithmique efficace

L'algorithmique a un rôle crucial dans tous les traitements de masse de données et de nombreuses recherches doivent être menées afin d'améliorer des algorithmes ayant des convergences lentes ou pour adapter des algorithmes aux objets multi-échelles dans le contexte des masses de données. En particulier, la définition d'objets abstraits à partir de problèmes concrets rencontrés conduit à une algorithmique parallèle/distribuée efficace et l'étude approfondie de ces objets peut permettre d'obtenir des conséquences sur les applications envisagées notamment dans le contexte de la programmation parallèle et des grilles de calculs. L'étude des structures de données réparties et des coûts de communication pour le traitement des masses de données en est un exemple.

2.2. Analyse et traitement de contenus multimédia

Il existe aujourd'hui des capteurs nombreux et très divers (scanners 3D, satellites, dispositifs médicaux,...) qui fournissent des images ou des vidéos en très grandes quantités. Ces images ou ces vidéos doivent être stockées et organisées pour pouvoir être analysées, évaluées, retrouvées, transmises, indexées, améliorées,... On trouve donc ici tous les problèmes de bases de données multimédias, d'indexation, de compression, de traitement d'images, de recherche d'informations, d'extraction des connaissances. Tous ces aspects rentrent pleinement dans le cadre de cette action et adressent en particulier le problème du passage à l'échelle.

2.3. Bases de données

Les bases de données sont au cœur de la gestion des masses de données. En particulier, des progrès dans l'interrogation de bases de données semi-structurées doivent permettre des avancées notables pour des données qui par essence même sont réparties. La définition de langages de requêtes et leur optimisation sont des sujets centraux qui doivent être explorés dans le contexte réparti tout comme la réduction de la complexité des cubes de données. Plus largement, le partage de ressources qu'il s'agisse de données, programmes ou services, est au cœur de ce champ de recherche. Les infrastructures telles les architectures de médiation et les dispositifs pair-à-pair en sont des dispositifs incontournables. Par ailleurs, l'hétérogénéité dans le cadre de la grille de données adresse toutes les questions de compatibilités et de transparence de formats entre les documents. Les bibliothèques numériques, le Web ou les bases de données biologiques sont des exemples d'objets qui sont des enjeux d'importance dans le cadre des masses de données.

2.4. Fouille de données

La fouille de données a pour but de lutter contre le morcellement des expertises et la dilution des connaissances, et de restituer à l'expert humain la compréhension et la maîtrise des phénomènes latents dans les monceaux de données textuelles ou non textuelles partout où l'informatique est présente. La numérisation massive des fonds d'archive photos, vidéos et sons en sont autant d'exemples. Dans ce sens, elle participe pleinement à la recherche en

masse de données. Outre les bases de données qui en sont un outil essentiel, toutes les démarches algorithmiques faisant appel aux statistiques, à l'intelligence artificielle et à l'apprentissage doivent contribuer à terme à rendre celle-ci efficace dans des applications telles que la médecine, les prévisions météorologiques ou terrestres. Des questions doivent être adressées telles que la mesure de qualité, d'objectivité et de reproductibilité des fouilles effectuées. Notamment, pour des données réparties, l'aspect persistance et transparence de la localisation et flux de données sont des problématiques importantes.

2.5.Réalité virtuelle, réalité augmentée

L'évolution technologique en informatique du côté des processeurs graphiques ou des systèmes d'acquisition 3D, les nouveaux dispositifs de visualisation 3D associés à des dispositifs d'interaction, les recherches sur les interfaces sonores, haptiques, tactiles, laissent à penser que des évolutions considérables sont à attendre du côté de la réalité virtuelle et de la réalité augmentée. Parmi les problématiques envisagées, on peut citer :

- Les problèmes de visualisation couplés à la simulation des phénomènes concernés ;
- Les problèmes de modélisation et de simulation : nécessité de multi-modèles multi-échelles ;
- Les dispositifs d'interaction, et notamment leurs liens avec la perception et la cognition.

Tous les problèmes soulevés doivent prendre en compte le temps réel, car l'utilisateur participe à l'interaction. Ceci peut conduire à la répartition des calculs sur la grille par exemple mais aussi vers d'autres mécanismes.

2.6.Simulation efficace de phénomènes complexes

Les applications visées conduisent à simuler des objets ou des phénomènes hétérogènes, tout en intégrant en entrée des données et des connaissances a priori très diverses. La nécessité de couplage de ces simulations avec des dispositifs de visualisation et/ou d'immersion, conduit à la recherche d'algorithmes très efficaces, voire de temps-réel, permettant, par exemple à l'utilisateur de concentrer interactivement le calcul sur ses centres d'intérêt. Pour gérer cette complexité, la conception de multi-modèles couplés entre eux, qui devront souvent être mis en oeuvre à des échelles de temps et d'espace totalement différentes, est un problème crucial à aborder. De plus le développement de techniques de simulation multi-échelles, adaptatives, et/ou faisant appel à l'algorithmique distribuée s'avère indispensable.

2.7.Visualisation de données hétérogènes

Les systèmes actuels (satellites, capteurs par exemple) fournissent des données informatiques hétérogènes. Lorsque la taille des données est faible, le simple affichage d'une image (une courbe par exemple) peut aider un expert à visualiser une caractéristique. Il n'en est pas de même sur de grandes masses de données où l'on se trouve face à des données très bruitées ou à des milliers de courbes réelles. Il s'agit donc de construire des abstractions visuelles assez proches du secteur d'application pour permettre une véritable perception des phénomènes liés aux données. Ces abstractions peuvent permettre, par exemple, d'extraire des phénomènes (existants mais dont on ne sait pas a priori comment les mettre en évidence) par des dispositifs liés à la reconnaissance de motifs ou la fragmentation de données. L'enjeu est donc de réaliser des modèles intégrant cette diversité au sein de vues efficaces.

3. TYPES DE PROJETS

Il importe d'insister sur le caractère clairement amont des projets de recherche attendus qui viseront à faire réellement collaborer les équipes impliquées. Le caractère original et innovant du projet (concepts, technologies, expériences...) devra être explicité et la valeur ajoutée des coopérations entre les différentes équipes clairement identifiées. Les partenariats avec des centres de recherche européens sont acceptés et favorisés, dans le cadre de coopérations

bilatérales ou de programmes européens, dans la mesure où chaque partenaire étranger assure son propre financement dans le projet.

Les projets attendus, d'une durée maximale de 36 mois, pourront être de plusieurs types, par exemple (liste non limitative) :

- Projet de recherche : sur un domaine commun d'expertise clairement défini, basé sur la coopération active d'un nombre limité d'équipes.
- Projet de recherche multi-thématique : basé sur la coopération active d'équipes de recherche relevant de champs disciplinaires différents. Le projet proposera un programme de recherche commun, pluridisciplinaire, basé sur la complémentarité des équipes participantes.
- Projet de recherche avec infrastructure : grâce à son financement conséquent, cette Action peut être l'occasion de monter des projets de recherche nécessitant l'acquisition d'équipements hors de portée des financements classiques. De tels projets, s'appuyant sur un programme de recherche commun, devront être particulièrement argumentés. Le problème des moyens humains indispensables à l'utilisation et la maintenance des équipements devra être abordé, en explicitant les demandes faites dans le cadre de l'Action et, le cas échéant, les autres apports.

Une attention particulière sera portée aux projets visant à :

- l'implication affirmée d'équipes de champs disciplinaires cités ci-dessus,
- fournir des solutions globales aux problèmes posés en favorisant les collaborations et les synergies entre chercheurs d'équipes ou de structures différentes.
- la visualisation sous toutes ses formes (synthèse d'images réalistes, réalité virtuelle et augmentée, abstractions visuelles,...),
- développer et améliorer les mécanismes et algorithmes destinés au traitement des masses de données,
- la conception et le test de mécanismes permettant d'accélérer l'accès ou de partager efficacement des masses de données.

4. REDACTION ET SELECTION

Les propositions, rédigées en anglais, seront évaluées sur leur caractère novateur, sur la possibilité de faire collaborer des équipes de manière crédible en faisant progresser les connaissances de chacun, et sur l'intérêt des résultats potentiels.

L'évaluation des projets sera faite par un Conseil Scientifique indépendant (pour plus de détails voir le site <http://mdmsa.labri.fr>). Il s'appuiera sur l'avis d'expertises françaises et étrangères.

La labellisation des projets sera faite par l'Agence Nationale pour la Recherche.

5. SUIVI DES PROJETS SELECTIONNES

Chaque projet sélectionné sera suivi par deux membres du conseil scientifique nommément désignés et devra maintenir régulièrement une page web résumant l'ensemble de ses activités. Chaque projet fournira un rapport à mi-parcours permettant au Conseil Scientifique d'apprécier l'avancement des travaux, la coopération des équipes participantes et les moyens complémentaires que la labellisation lui a permis d'obtenir.

Il devra enfin rendre un rapport à la fin de l'exécution du projet présentant les résultats obtenus au regard des moyens mis en oeuvre et permettant d'apprécier le chemin accompli.