

**Action Concertée Incitative**  
***IMPbio***  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**Attention,**  
**toute proposition doit faire l'objet d'un résumé enregistré**  
**avant le 4 avril 2004**  
**directement sur le site web de l'ACI :**

<http://impbio.lirmm.fr>

**Chaque dossier doit être**

– **déposé avant le 18 avril 2004 sur le site web de l'ACI:**

<http://impbio.lirmm.fr>

– **envoyé par voie postale avec les signatures requises**  
**avant le 30 avril 2004**

(cachet de la poste faisant foi)

à

Ministère délégué à la Recherche et aux Nouvelles Technologies  
Direction de la Recherche  
Cellule ACI  
ACI IMPbio  
1, rue Descartes  
75231 Paris cedex 05

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**I - FICHE D'IDENTITE DU PROJET**

**Nom du Projet :** (maximum 20 caractères)

MEDIANTE

**Titre du Projet :** (maximum 3 lignes)

Développement de Méthodes et d'Outils en Ligne pour l'Analyse de Données d'Expériences sur les Pucés « oligos longs » distribuées par le Réseau National Genopole

**Type du Projet<sup>1</sup>:**

Projet de recherche	Logiciels, bases de données, serveurs
	<b>X</b>

**Durée du projet<sup>2</sup> : 36 mois**

**Description courte du Projet :** (une demi-page maximum)

Le projet MEDIANTE (<http://medcal.ipmc.cnrs.fr:8080/mediante>) sert d'appui bio-informatique au projet de production de puces pangénomiques chez l'homme et chez la souris, mis en place par le Réseau National Genopole. L'interface déjà disponible permet aux utilisateurs de construire un « panier » de sondes adaptées à leurs projets, tout en participant à l'amélioration du set global. L'objectif de la présente demande est de compléter le système MEDIANTE de façon à:

- 1) stocker l'ensemble des résultats d'expériences menées avec les puces pangénomiques de la ressource nationale (chez l'homme et chez la souris).
- 2) construire une base de connaissances à partir de descriptions de gènes stockées dans Mediante, à partir de résultats validés d'expériences transcriptome (cf point n°3), et en développant un outil d'analyse textuel permettant l'extraction d'informations à partir de références bibliographiques spécifiques fournies par les utilisateurs.
- 3) intégrer au système différents outils de fouille de données qui permettront l'analyse des expériences mémorisées, la recherche de profils type d'expériences et la confrontation d'une nouvelle expérience avec les profils déjà identifiés. Les informations stockées dans la base de connaissances guideront le processus de fouille, ce qui permettra notamment de focaliser l'analyse sur des gènes d'intérêt et d'assister l'utilisateur dans l'interprétation des résultats.
- 4) intégrer au système un médiateur ontologique permettant de gérer une ontologie locale dans un format unifié, servant de base terminologique à tous les services du système. Ce médiateur fournirait ainsi des outils complémentaires assurant l'évolution incrémentale de l'ontologie locale et l'adaptation du système vers des normes ontologiques de plus en plus expressives.

**Coordinateur du projet**

Nom	Prénom	Laboratoire (sigle éventuel et nom complet)
BARBRY	Pascal	Institut de Pharmacologie Moléculaire et Cellulaire, CNRS/UNSA UMR6097

**Organisme de rattachement financier pour le présent projet**

CNRS, Administration Déléguée n°20

**Equipes ou laboratoires partenaires (nom complet et éventuellement sigle)<sup>3</sup>**

Rose DIENG, INRIA de Sophia Antipolis, Projet ACACIA (Acquisition des connaissances pour l'assistance à la conception par interaction entre agents)

Martine COLLARD, I3S CNRS/UNSA UMR 6070, Projet EXeCO (EXtraction de connaissances à partir des données et ANalyse et CONception des systèmes d'information)

<sup>1</sup> Cocher la case correspondante au type du projet soumis.

<sup>2</sup> La durée d'un projet ne peut dépasser 36 mois. Des demandes de projets d'une durée plus courte devront être particulièrement argumentées.

<sup>3</sup> Insérer autant de lignes que nécessaire.

**Action Concertée Incitative**  
***IMPbio***  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**II - PRESENTATION DETAILLEE DU PROJET**

**A- IDENTIFICATION DU COORDINATEUR ET DES AUTRES  
PARTENAIRES DU PROJET :**

**A1- Coordinateur du Projet :**

*Un unique coordinateur doit être désigné par les partenaires.*

M. ou Mme. Prénom Nom <sup>4</sup>	Pascal BARBRY
Fonction <sup>4</sup>	Directeur de l'IPMC
Laboratoire (Nom complet et sigle le cas échéant) <sup>4</sup>	Institut de Pharmacologie Moléculaire et Cellulaire (CNRS/UNSA UMR 6097)
Adresse <sup>4</sup>	660, route des lucioles F06560 Sophia Antipolis
Téléphone <sup>4</sup>	33 (0)4 93 95 77 93
Fax	33 (0)4 93 95 77 94
Mél <sup>4</sup>	<a href="mailto:barbry@ipmc.cnrs.fr">barbry@ipmc.cnrs.fr</a> ; <a href="mailto:direction@ipmc.cnrs.fr">direction@ipmc.cnrs.fr</a>

---

<sup>4</sup> Champs obligatoires

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**A2- Equipes ou laboratoires partenaires du Projet <sup>5</sup>:**

**Identification de l'équipe ou du laboratoire**

Equipe ou Laboratoire	Institut de Pharmacologie Moléculaire et Cellulaire (CNRS/UNSA UMR 6097)
Adresse	660, route des lucioles F06560 Sophia Antipolis

**Organisme de rattachement financier de l'équipe pour le présent projet**

<b>Délégation Régionale n°20 du CNRS</b>
--

**Responsable du projet au sein de l'équipe ou du laboratoire**

M. ou Mme. Prénom Nom	Pascal BARBRY
Fonction	Directeur de l'IPMC
Téléphone	33 (0)4 93 95 77 93
Fax	33 (0)4 93 95 77 94
Mél	<a href="mailto:barbry@ipmc.cnrs.fr">barbry@ipmc.cnrs.fr</a> ; <a href="mailto:direction@ipmc.cnrs.fr">direction@ipmc.cnrs.fr</a>

**Membres de l'équipe participant au projet (y compris le responsable désigné ci-dessus)**

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
Barbry	Pascal	DR1 CNRS	25%
Le Brigand	Kevin	CDD INRIA	100%
1 IR2 CNRS à recruter au 01/09/2004		IR2 CNRS	100%

**Références :**

*Pour chaque (enseignant-)chercheur participant, liste de 3 à 5 publications, logiciels ou brevets les plus significatifs, en relation avec la thématique du projet.*

<ol style="list-style-type: none"> <li>1. <b>Moreilhon, C.</b>, Hologne, C., Bajolet, O., Cottrez, F., <b>Magnone, V.</b>, Merten, M., Groux, H., Puchelle, E., <b>Barbry, P.</b> (2004) Alive Staphylococcus aureus and Soluble Virulence Factors Induce Different Transcriptional Responses in Human Airway Cells. <i>Submitted</i>.</li> <li>2. Fromigué, O., Louis, K., <b>Dayem, M.</b>, Milanini, J., Tartare-Deckert, S., Pages, G., Ponzio, G., Hofman, P., <b>Barbry, P.</b>, Auberger, P., and Mari, B. (2003) Modulation of gene expression in normal human pulmonary fibroblasts following coculture with non small cell lung cancer cells: potential mechanism in the modification of tumor microenvironment. <i>Oncogene</i>, 22: 8487-8497.</li> <li>3. Karine Hugot, Marie-Pierre Rivière, <b>Chimène Moreilhon</b>, <b>Manal A.Dayem</b>, Joseph Cozzitorto, Gilles Arbiol, <b>Pascal Barbry</b>, Catherine Weiss, Eric Galiana (2003) Delineating developmental resistance pathways in plant by selection and characterization of leaf-expressed genes encoding extracellular proteins in flowering tobacco. <i>Plant Physiol.</i> 134: 858-870.</li> <li>4. <b>Dayem MA</b>, <b>Moreilhon C</b>, Turchi L, <b>Magnone V</b>, Christen R, Ponzio G, <b>Barbry P</b>. Early gene expression in wounded human keratinocytes revealed by DNA microarray analysis. <i>Comparative and Functional Genomics</i> 2003 4: 47-55.2 (<a href="http://www3.interscience.wiley.com/cgi-bin/fulltext/102531162/PDFSTART">http://www3.interscience.wiley.com/cgi-bin/fulltext/102531162/PDFSTART</a>)</li> </ol>
--

<sup>5</sup> Une fiche doit être remplie pour chaque laboratoire ou équipe partenaire

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**Equipe partenaire n°1 (nom complet et éventuellement sigle)<sup>6</sup>**

**Identification de l'équipe ou du laboratoire**

Equipe ou Laboratoire	Projet ACACIA (Acquisition des connaissances pour l'assistance à la conception par interaction entre agents)
Adresse	2004, Route des Lucioles BP 93 F06902 Sophia Antipolis Cedex France

**Organisme de rattachement financier de l'équipe pour le présent projet**

<b>Laboratoire I3S – Université de Nice-Sophia Antipolis – CNRS UMR 6070</b>
--

**Responsable du projet au sein de l'équipe ou du laboratoire**

M. ou Mme. Prénom Nom	Rose DIENG-KUNTZ
Fonction	Responsable scientifique du projet Acacia
Téléphone	04 92 38 78 10
Fax	04 92 38 77 83
Mél	Rose.Dieng@sophia.inria.fr

**Membres de l'équipe participant au projet (y compris le responsable désigné ci-dessus)**

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
Dieng-Kuntz	Rose	DR2	25%
Corby	Olivier	CR1	20%
Giboin	Alain	CR1	10%
Gandon	Fabien	CR2	10%
Khelif	Khaled	Doctorant	100%
X	Y	Ingénieur expert	100%

**Références :**

*Pour chaque (enseignant-)chercheur participant, liste de 3 à 5 publications, logiciels ou brevets les plus significatifs, en relation avec la thématique du projet.*

- 1. Corby O. & Faron-Zucker C.** (2001). Corese : A Corporate Semantic Web Engine. In World Wide Web Workshop on Real World RDF and Semantic Web Applications, Hawaii, 2001.
- 2. Dieng-Kuntz R., Corby O., Gandon F., Giboin A., Golebiowska J., Matta N., Ribiere M.,** Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire pour le "Knowledge Management", 2ème éd., Dunod, 2001.
- 3. Dieng-Kuntz R.,** Matta N. eds. Knowledge Management and Organizational Memories. Kluwer Academic Press, July 2002.
- 4. F. Gandon, R. Dieng-Kuntz, O. Corby, A. Giboin,** Semantic Web and Multi-Agents Approach to Corporate Memory Management, 17th IFIP World Computer Congress IIP Track, p. 103-115, August 25-30, 2002, Montréal, Canada
- 5. Golebiowska J. & Dieng-Kuntz R. & Corby O., & Mousseau D.** (2001), Building and Exploiting Ontologies for an Automobile Project Memory. In Proc.of ACM Conference on Knowledge Capture (K-CAP), Victoria, Canada, 2001.
- 6. Khelif K., Dieng-Kuntz R.,** Annotations sémantiques pour le domaine des biopuces. A paraître dans les Actes de la 14<sup>ème</sup> Conférence Francophone sur l'Ingénierie des Connaissances ( IC'2004), Lyon, Mai 2004.

<sup>6</sup> Insérer autant de lignes que nécessaire.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**Equipe partenaire n°2 (nom complet et éventuellement sigle)<sup>7</sup>**

**Identification de l'équipe ou du laboratoire**

Equipe ou Laboratoire	Projet EXeCO – Laboratoire I3S
Adresse	Les Algorithmes 2000 route des Lucioles – bât. Euclide BP.121 F06903 Sophia-Antipolis cedex

**Organisme de rattachement financier de l'équipe pour le présent projet**

<b>Laboratoire I3S – Université de Nice-Sophia Antipolis – CNRS UMR 6070</b>
--

**Responsable du projet au sein de l'équipe ou du laboratoire**

M. ou Mme. Prénom Nom	Martine COLLARD
Fonction	Maître de Conférences
Téléphone	04 92 94 27 58
Fax	04 92 94 26 98
Mél	Martine.Collard@unice.fr

**Membres de l'équipe participant au projet (y compris le responsable désigné ci-dessus)**

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
BRISSON	Laurent	Doctorant	40%
COLLARD	Martine	MC – UNSA	30%
LE THANH	Nhan	Professeur - UNSA	25%
PASQUIER	Nicolas	MC - UNSA	20%
MARTINEZ	Ricardo	Doctorant	100%
LE DUC	Chan	Doctorant	100%
NGUYEN	Thi Dieu Thu	Doctorant	40%
Post doc		A recruter	100%

**Références :**

*Pour chaque (enseignant-)chercheur participant, liste de 3 à 5 publications, logiciels ou brevets les plus significatifs, en relation avec la thématique du projet.*

<p>[1] L. Brisson, N. Pasquier, M. Collard, <i>Generic Rule Templates for Predicting Atherosclerosis Risk Factors</i>, ECML-PKDD'04, Discovery Challenge, Pise, Italie, 2004. Soumis.</p> <p>[2] C. Le Duc, N. Le Thanh, <i>Revision an Reasoning on Revision Rules in DL-based Knowledge Bases</i>, ECAI 2004, Valencia, Spain, August 2004. Soumis.</p> <p>[3] C. Le Duc, V. Viet Hoang, N. Le Thanh, <i>Les règles de révision et les logiques de description. L'application pour les services d'inférences non standard de ONDL</i>, Colloque CNRIUT04, Nice, France, mai 2004.</p> <p>[4] D. Francisci, M. Collard, <i>Multi-Criteria Evaluation of Interesting Dependencies according to a Data Mining Approach</i>, CEC'03, IEEE Congress on Evolutionary Computation, Canberra, Australie, 2003.</p> <p>[5] C. Le Duc, N. Le Thanh, <i>On the problems of representing least common subsumer and computing approximation in Description Logics</i>, DL'03 int. workshop, Rome, Italy, September, 2003.</p> <p>[6] D. Francisci, M. Collard, <i>Optimizing rule quality in a data mining context</i>, CIRAS'03, 2<sup>nd</sup> International Conference on Computational Intelligence, Robotics and Autonomous Systems, Singapour, 2003.</p> <p>[7] C. Le Duc, N. Le Thanh, <i>Combining Revision Production Rules and Description Logics</i>, KES 2003 Seventh International Conference on Knowledge-Based Intelligent Information &amp; Engineering Systems, Proc., Part I Springer, University of Oxford, United Kingdom, September 2003.</p> <p>[8] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, L. Lakhal, <i>Computing iceberg concept lattices with Titanic</i>, Data and Knowledge Engineering, 42(2):189-222, Elsevier Science, 2002.</p> <p>[9] C. Le Duc, N. Le Thanh, <i>Approximation from Description Logics with disjunctions to another without disjunctions</i>, KI-2002 : Workshop on Applications of Description Logics ADL'02, Aachen, Germany, September, 2002.</p> <p>[10] A. Barbar, M. Collard, <i>A distance-Based Approach for Database Re-engineering</i>, ACS/IEEE International Conference on Computer Systems and Applications, Beyrou, Liban, 2001.</p>
---

<sup>7</sup> Insérer autant de lignes que nécessaire.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**B - DESCRIPTION DU PROJET (10 pages maximum<sup>8</sup>)**

**B1 – Objectifs et contexte :**

*On précisera, en particulier, les verrous scientifiques et technologiques à dépasser, l'état de l'art ainsi que les projets concurrents ou similaires connus dans le contexte national et international, en particulier ceux auxquels les équipes du projet participent. Dans le cas des logiciels, banques de données ou serveurs, on détaillera l'impact national et international actuel.*

**OBJECTIFS**

Le projet MEDIANTE (<http://medcal.ipmc.cnrs.fr:8080/mediante>) sert d'appui bio-informatique au projet de production de puces pangénomiques chez l'homme et chez la souris, mis en place par le Réseau National Genopole. Une interface déjà accessible permet aux utilisateurs de construire un « panier » de sondes adaptées à leurs projets : leur travail de sélection sur une collection de sondes calculées participe à l'amélioration du set, à travers la comptabilisation des séquences les plus demandées et des oligos préférés de chaque utilisateur. L'interface actuelle permet également de distribuer les informations générales concernant les puces produites (schéma de dépôt, annotations des sondes). L'objectif général de la présente demande est de doter Mediante de modules d'aide à l'interprétation des expériences, s'enrichissant en continu des expériences effectuées sur les lames produites, de manière à fournir aux biologistes un cadre de référence pour l'analyse, la validation et l'interprétation de leurs résultats d'expériences.

L'élément central du système sera constitué d'une base de connaissances construite autour d'une mémoire d'expériences autour de ce type particulier de puces d'une part, et de références bibliographiques d'autre part. L'analyse sera guidée en exploitant la base de connaissances. L'interprétation sera facilitée par la confrontation des résultats avec des profils précédemment identifiés à partir de la même mémoire d'expériences.

La participation des biologistes au projet MEDIANTE permettra :

- l'enrichissement de la mémoire d'expériences (à travers la fourniture de références bibliographiques associées à un projet particulier, et de la fourniture profils d'expériences),
- la confrontation des résultats d'une expérience aux profils d'expériences précédemment identifiés et mémorisés,
- la validation des lames produites.

MEDIANTE constituera un entrepôt de données transcriptome , avec un module d'annotation et de recherche sur les expériences biopuces, un module de fouille de données et un module médiateur dédié aux ontologies. L'entrepôt de données permettra de stocker et d'organiser toutes les données représentant la mémoire d'expériences d'analyse du transcriptome (MEAT) sur les puces du réseau national. Le module MEAT-Annot&Search permettra de constituer une base de connaissances consistant entre autres en une base d'annotations sémantiques sur ces expériences (en particulier sur l'interprétation de leurs résultats) ou sur des articles de référence utiles pour l'aide à l'interprétation des résultats de ces expériences. Le module de fouille (MEAT-Miner) fournira des outils d'extraction de connaissances intégrant les informations mémorisées dans MEAT pour l'analyse des résultats d'expériences. Le module médiateur (MEAT-Onto) gèrera les ontologies locales nécessaires aux deux modules précédents.

**CONTEXTE**

L'émergence de la génomique fonctionnelle est récente, et découle de la réalisation du séquençage de plusieurs génomes entiers, dont le génome humain. Une activité importante

---

<sup>8</sup> 12 points, interligne simple.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

consiste maintenant à définir les fonctions associées à chacun des gènes identifiés et à préciser les réseaux de relations existant entre eux. Parmi les approches expérimentales permettant de réaliser cette tâche, l'une des plus prometteuses correspond aux biopuces d'ADN, qui permettent de quantifier simultanément l'expression de plusieurs milliers de gènes dans un échantillon. La technologie des puces à ADN est née au milieu des années 1990 avec le dépôt à haute densité de gènes sur des supports solides (membranes de nylon puis lames de verre). L'analyse du transcriptome sur lame de verre permet de mesurer les niveaux de fluorescence associés à l'expression de l'ensemble des ARNm d'une cellule ou d'un tissu donné dans une condition physiologique définie. La quantité d'informations apportée en une seule expérimentation est très importante : plusieurs dizaines de milliers de sondes pouvant être déposés sur une seule puce, il est possible de définir en une seule expérience des niveaux d'expression pour un génome complet. Le processus de récupération de l'information comprend de nombreuses étapes (expérience biologique sur l'échantillon d'intérêt, fabrication (éventuelle) des puces à ADN, hybridation, mesure du signal, quantification, uniformisation des données,...). Le stockage des données associées doit être conforme à un format d'échange des données d'expériences défini par un comité de normalisation international (MGED: <http://www.omg.org/docs/formal/03-02-03.pdf>). Plusieurs entrepôts d'expériences ont été mis en place (EMBL : ArrayExpress ;NCBI : GeneExpressionOmnibus) pour stocker l'ensemble des données disponibles sur toutes les plates-formes existantes. Alors que ces entrepôts joueront sans doute un rôle extrêmement important pour l'échange de données entre plates-formes distinctes, chaque centre producteur de données transcriptome doit néanmoins mettre en place ses propres outils pour lui permettre de vérifier la consistance des données produites, et pour analyser en continu les informations issues de ses propres projets. C'est dans ce contexte que se situe la présente demande.

La mise en place d'une ressource nationale en oligos longs, associée à la production de puces pangénomiques était réclamée par de nombreux sites français. Un projet collaboratif entre le Réseau National Genopole et le HGMP anglais de Tom Freeman (MRC, Hinxton, UK) fut mis en place en 2002, avec pour objectif de produire des puces oligos longs couvrant l'ensemble des transcrits existant chez l'homme et chez la souris. Le HGMP disposait au début de cette collaboration d'une expérience de production de puces « homme » et « souris » à large couverture, à partir de sets commerciaux (Sigma/Compugen). Il s'agissait dans ces conditions d'élargir l'expertise académique bi-nationale en créant une chaîne complète de détermination de sondes oligos longs, appuyées sur plusieurs sites de production de puces à ADN mis en place dans le cadre du programme Génomique du Ministère de la Recherche. L'achat en masse des collections d'oligos permettait une réduction importante des coûts de production des puces.

A cette dimension du projet s'ajoutaient : (1) la mise à disposition à tous les sites académiques des informations sur les collections d'oligonucléotides, pour permettre aux utilisateurs potentiels (notamment des spécialistes d'un champ thématique précis) de participer à l'amélioration ou à la conception des sets ; (2) la mise en place d'une procédure d'uniformisation des méthodes expérimentales. Des tests préliminaires réalisés sur huit plates-formes françaises (Montagne-Sainte-Genève, Evry, Strasbourg, Montpellier, Nice, Grenoble, Marseille, Toulouse) ont permis de vérifier la qualité des résultats obtenus avec des lames de verre hybridées en fluorescence, avec des sondes oligos issues de Mediant. Une corrélation significative avec des données Affymetrix a notamment indiqué la cohérence expérimentale entre des plates-formes de type lame de verre et une plate-forme Affymetrix. Ces expériences ont validé la démarche de sélection des oligonucléotides mise en place dans Mediant.

143.274 oligos de 50-52 nucléotides ont alors été calculés contre 28.074 transcrits humains distincts, et 118.307 oligonucléotides distincts ont été calculés contre 25.173 transcrits murins. Pour chacun des transcrits, une procédure de présélection a été mise en place afin de déterminer à partir de ces collections d'oligos longs le « meilleur » oligo disponible pour chaque transcrit.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

Dans quelques cas, 2 ou 3 oligos ont été finalement sélectionnés par un même transcrit (dans le cas notamment de quelques variants d'épissage). La procédure de sélection du « meilleur » oligo intègre plusieurs informations, dont la spécificité de chacun des oligos vis-à-vis de bases de transcrits ENSEMBL ou RefSeq, et le nombre d'EST reconnus spécifiquement par chaque oligo. L'interface mise en place dans le cadre du projet Mediante (<http://medcal.ipmc.cnrs.fr:8080/Mediate>) permet à des utilisateurs distants d'évaluer la qualité de toutes les sondes disponibles pour leur(s) gène(s) d'intérêt. L'application permet à l'utilisateur distant de constituer un « panier virtuel » de sondes qui correspond à la puce à ADN « idéale » dont il souhaiterait disposer. Les informations sont accessibles par mot clé et/ou par référence en utilisant des informations regroupées à partir des principales bases de données publiques (LocusLink - <http://www.ncbi.nlm.nih.gov/LocusLink> -, UniGene - <http://www.ncbi.nlm.nih.gov/UniGene> -, Ensembl - <http://www.ensembl.org> -, HUGO (<http://www.gene.ucl.ac.uk/nomenclature/> -, GENATLAS - <http://www.citi2.fr/GENATLAS/> -). Les demandes des utilisateurs sont exprimées *via* un ou plusieurs critère(s) de sélection (identificateur LocusLink, mot-clé,...). L'utilisateur peut également rechercher des sondes à partir de séquences nucléotidiques en lançant un blast contre tous les oligos disponibles dans la base. A l'issue de ces recherches, une liste des réponses acceptables est présentée à l'utilisateur qui valide alors toute ou partie de la liste, aboutissant à la « puce idéale de l'utilisateur X ». La comparaison entre la liste des demandes et la liste des sondes disponibles (sur la puce pangénomique, voire sur différentes puces commerciales) permet alors d'établir : le nombre de sondes répondant aux critères de recherche, et de préciser les séquences manquantes sur la dernière version disponible de la puce.

Des statistiques globales permettent de déterminer :

- les niveaux de consultation pour chaque gène,
- le taux de satisfaction vis à vis de la requête (nombre de sondes disponibles / nombre de gènes demandés),
- la séquence et les caractéristiques de chaque sonde.

La consultation de Mediante permet ainsi de vérifier si la puce « souhaitée » par l'utilisateur est correctement couverte par la puce pangénomique, voire, par une des puces disponibles localement auprès d'une des plates-formes du réseau. Dans tous les cas, des outils de transferts *ad hoc* permettent à l'utilisateur connecté de rapatrier toutes les informations utiles pour la réalisation autonome d'un projet transcriptome (séquence des oligos, annotations,...).

Les ~25000 oligos longs du set humain ont été produits par un industriel, et distribués aux plates-formes transcriptome d'Evry, de Nice/Sophia Antipolis, de Strasbourg, et d'Hinxton (UK). Ces 4 sites seront les quatre plates-formes chargées de la production des lames pangénomiques, et de leur diffusion aux laboratoires académiques français et anglais. La quantité de matériel distribuée aux trois sites français s'élevant à 12 nmol, sera suffisante pour produire 10 à 20 000 lames.

Les premiers résultats obtenus avec ces oligos indiquent un rapport signal/bruit élevé, des différentiels cohérents avec les modèles expérimentaux analysés, une bonne reproductibilité. Des expériences préliminaires indiquent par ailleurs la possibilité de ne travailler qu'avec 100ng d'ARN total (avec des protocoles d'amplifications linéaires) avec des résultats comparables à ceux obtenus avec 400ng ou 2,5 microgrammes d'ARN totaux.

Enfin, la production des oligos longs constituant la ressource nationale « souris » sera démarrée d'ici la mi-avril. Dans ce cadre, on peut raisonnablement penser que des lames pangénomiques humaines seront disponibles avant l'été 2004, et que des lames pangénomiques murines seront prêtes avant l'automne 2004. La cohérence du projet nécessite la mise en place d'un système d'information homogène entre les trois sites français de production, et des outils de stockage des résultats. C'est dans ce cadre que se situe la présente demande. Un rôle important a été joué depuis l'origine par la plate-forme de Sophia Antipolis dans la mise en place de ce projet. Cette

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

plate-forme de fabrication et d'analyse des biopuces a été créée en 1999 au sein de l'Institut de Pharmacologie Moléculaire et Cellulaire (CNRS UMR6097), en association avec l'INRA (INRA UMR1112) : elle est devenue en 2002 plate-forme du réseau national des Genopoles, et a été reconnue en 2003 comme plate-forme nationale lors de l'évaluation du RIO (réunion inter-organisme). Le projet de ressource nationale en transcriptome a aujourd'hui besoin de disposer d'un système d'informations adapté pour archiver les données expérimentales obtenues avec les puces produites. Le présent projet, issu d'une collaboration entre la plate-forme de Sophia Antipolis et des chercheurs en informatique et en statistique de l'INRIA et de l'IS3 de Sophia Antipolis, vise à développer les outils informatiques nécessaires.

**B2 – Description du projet :**

*Entre autres, le caractère innovant du projet (concepts, technologies, expériences ...) devra être explicité et la valeur ajoutée des coopérations entre les différentes équipes sera discutée. Dans le cas des logiciels, banques de données ou serveurs, on décrira précisément le niveau et le type de services que l'on souhaite atteindre.*

Les besoins, liés à l'activité du biologiste travaillant sur les biopuces, sont les suivants.

1. **Gestion de la Production des Biopuces et Archivage des Données Expérimentales.** La publication de données transcriptome va requérir désormais la publication dans des bases de données publiques de toutes les données associées. Cela conduit le biologiste à regrouper dans un format pré-défini un grand nombre d'informations sur les propriétés des sondes, le protocole expérimental de production des données biopuces, l'expérience biologique.
2. **Analyse des Données d'Expression.** Une fois l'expérience biologique effectuée, la première étape consiste à préparer les données en utilisant, en particulier, une méthode de régression ou une analyse en composantes principales pour réduire la dimension du jeu de données. Selon la problématique étudiée, l'étude de gènes différentiellement exprimés, la recherche de groupes de gènes, l'identification des gènes caractérisant au mieux une pathologie, des méthodes de clustering ou de classification supervisée seront appliquées. De nombreux algorithmes ont déjà été définis et sont implémentés sur des plates-formes existantes (BioConductor, GeneSpring, ...). Ces outils seront dans ce cas réutilisés.
3. **Aide à la Validation des Résultats Expérimentaux.** A partir de ces premiers résultats, une partie seulement des modifications observées (gènes induits ou réprimés) peut être expliquée par le modèle postulé initialement par le biologiste. Son souci est donc de relier les données expérimentales nouvelles générées par l'expérience aux informations contenues dans les sites de références tels que les sites NCBI (LocusLink, UniGene, OMIM, HomoloGene, GEO, PubMed,...), ou des sites plus spécifiques, produisant des informations sur l'expression spécifiques des gènes dans tel ou tel tissu (<http://www.ncbi.nlm.nih.gov/SAGE>; Osprey, ...). Au cours de cette activité, le biologiste fait une recherche d'information sur un site donné (RefSeq, GO, etc.), en utilisant le nom adéquat du gène qui l'intéresse, et récupère de façon plus ou moins systématique toute l'information concernant le gène étudié. Le but étant d'établir ce qui a été publié à ce jour sur un gène donné, sur les interactions que ce gène pourrait avoir avec un autre, sa contribution à une fonction moléculaire, etc... Ce type d'information une fois trouvé, est utilisé pour valider les résultats d'expériences, c'est-à-dire pour définir les profils d'expression qui sont cohérents avec les données privées (recherche d'information dans la base des expériences) ou de la littérature (recherche d'information dans des bases documentaires distantes).

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

- 3. Aide à l'interprétation des résultats d'expériences.** Au vu des résultats expérimentaux, le biologiste cherche à identifier des relations non encore connues entre gènes (même comportement, participation à un même phénomène ou un phénomène lié au phénomène principal, mêmes fonctions biologiques, intervention dans une pathologie etc.) et les relations entre expériences (validation ou non des expériences ultérieures, etc.). Une fois les expériences réalisées, un certain nombre de gènes sont identifiés pour leur activation ou leur inhibition. Ce processus permet de répondre aux questions suivantes :
- quelles nouvelles sondes serait-il intéressant d'ajouter pour compléter l'étude (protéines en « interaction » avec des gènes affectés par le traitement, variants d'épissage, sondes défectueuses,...) ?
  - quelles autres expériences n'impliquant pas la technologie des biopuces seraient-elles nécessaires pour renforcer les conclusions des expériences réalisées ?
  - quelles sont les conclusions de l'expérience en cours ?

Pour répondre à l'ensemble de ces besoins, MEDIANTE fonctionnera comme un entrepôt de données qui constituera une Mémoire d'Expériences pour l'Analyse du Transcriptome (MEAT). Ces données seront stockées via le web par les techniciens de la (des) plate(s)-forme(s), pour ce qui concerne la production des lames ; et par les utilisateurs des biopuces, pour ce qui concerne les hybridations. Les différentes étapes de l'approche que nous proposons sont les suivantes:

1. Vérification et validation par l'utilisateur biologiste des sondes disponibles sur la puce, et sélection assistée d'un sous-ensemble qui lui semble pertinent dans le cadre du projet.
2. Commande de lames pour lancer une expérience sur les biopuces.
3. A la même étape, soumission d'articles de revue (de type Annual Review, TIBS, ...) concernant les gènes supposés *a priori* intéressants pour la future expérience, ou les phénomènes biologiques étudiés.
4. Constitution d'un *corpus* de documents électroniques à partir de ces articles.
5. Création d'annotations sémantiques sur lesdits articles.
6. Après réalisation de l'expérience, stockage de sa description et des résultats associés dans MEDIANTE selon un format d'échange compatible avec Array Express (<http://www.ebi.ac.uk/arrayexpress/>). Possibilité pour les utilisateurs de déposer directement leurs expériences sur le serveur de l'EBI.
7. Analyse statistique des résultats : le biologiste utilisera des outils de fouille de données construits spécifiquement pour MEDIANTE (MEAT-Miner) ou bien issus de projets existants, comme BioConductor (<http://www.bioconductor.org>).
8. Interprétation des résultats, qui pourra impliquer des recherches bibliographiques supplémentaires guidées par les ontologies, afin de confirmer ou d'infirmer les hypothèses.
9. Ajout par l'utilisateur de nouvelles annotations sur l'expérience grâce à l'éditeur d'annotations.

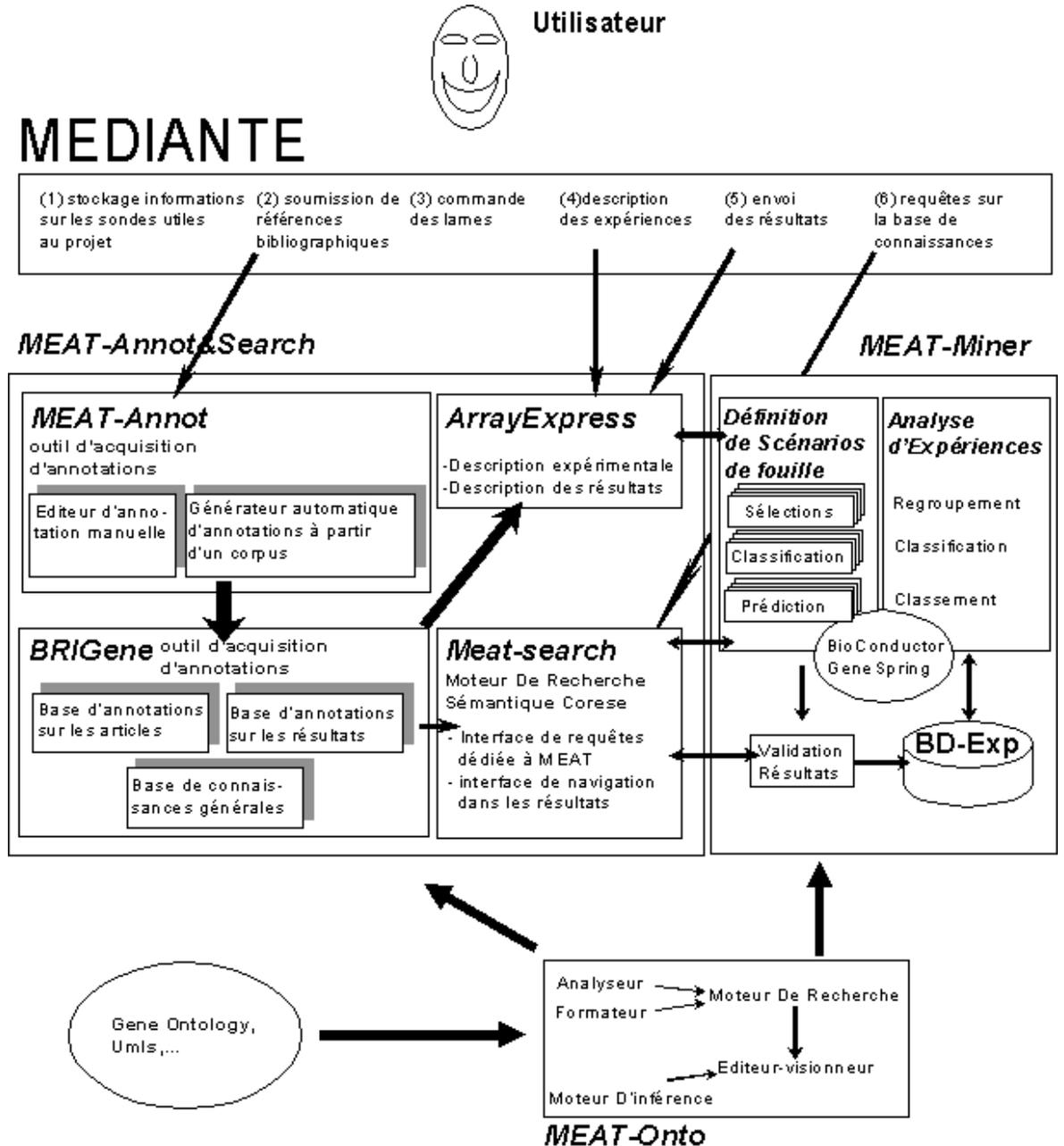
Ces différentes étapes seront réalisées à l'aide des quatre composantes suivantes :

1. l'interface de stockage, développé sous JSP, qui utilise un moteur de bases de données PostgreSQL.
2. le module d'annotation et de recherche sur la « mémoire d'expérience » (MEAT-Annot&Search), qui repose sur les technologies du Web sémantique.
3. le module de fouille de données (MEAT-Miner), qui implémente des techniques d'extraction de connaissances à partir de données (sélection, classification, prédiction,...) et propose des solutions adaptées à l'intégration d'informations fournies par MEAT-Annot,

**Action Concertée Incitative  
IMPbio  
Descriptif complet du projet  
(recherche multidisciplinaire, service bioinformatique)**

4. le module de médiateur ontologique (MEAT-Onto), qui repose sur des technologies de bases de données, de web sémantique et d'inférence logique.

Le schéma général du système est présenté sur la Figure n°1 :



La majeure partie de l'interface de stockage étant déjà en production (<http://medical.ipmc.cnrs.fr:8080/mediante>), nous détaillerons principalement chacun des trois derniers modules.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

- Le module MEAT-Annot&Search comprendra:
  - *L'outil d'acquisition d'annotations, composé d'un éditeur d'annotation manuelle et d'un outil de génération semi-automatique d'annotations à partir de l'analyse linguistique d'un corpus textuel.* Ces annotations consisteront entre autres en des instances particulières de concepts de l'ontologie (gènes, maladies, processus biologiques, composants cellulaires, etc.) ou en des relations sémantiques entre ces concepts (e.g. interactions entre gènes, relations entre gènes et maladies, participation directe ou indirecte de certains gènes à certains phénomènes biologiques, comportements de ces gènes vis à vis de traitements chimiques etc).
    - L'éditeur d'annotation manuelle sera guidé par une ou plusieurs ontologies biomédicales adéquates (Gene Ontology, UMLS) fournies par l'intermédiaire de MEAT-Onto. Il permettra d'associer à un élément de document particulier (texte ou image) une annotation traduite en RDF.
    - *Le générateur d'annotation à partir de corpus textuel* prendra en entrée le corpus textuel et intégrera divers outils de traitement automatique de la langue naturelle : outils existants comme l'extracteur de termes et de relations Syntex et l'outil GATE (General Architecture for Text Engineering), auxquels nous ajouterons des extensions dédiées à l'extraction de relations sémantiques impliquant des gènes. Guidé par les ontologies fournies par MEAT-Onto, et par des règles heuristiques basées sur les régularités observées dans le corpus textuel et représentatives des relations intéressantes dans le domaine, le générateur d'annotation permettra de reconnaître dans les textes les termes correspondant à des concepts de l'ontologie et à repérer dans les textes certaines relations sémantiques entre ceux-ci (e.g. interactions entre gènes ou interactions entre des gènes et d'autres entités). Le générateur d'annotation comprendra également une interface de validation permettant à l'expert de valider les annotations effectuées.
  - La base d'annotations (BRIGene) structurée en plusieurs bases dédiées aux différents projets, avec d'une part les annotations concernant les articles de référence de ces projets et d'autre part les annotations concernant les résultats de ces projets. Pourront aussi y figurer des connaissances générales en biologie, non associées à un document particulier mais susceptibles d'être utilisées pour faire des inférences sur les annotations lors de la recherche d'informations. Notons que les annotations pourront être contextuelles et dépendre du point de vue adopté par l'annotateur : de ce fait, le système devra pouvoir gérer certaines annotations contradictoires reposant sur des points de vue différents.
  - *Le moteur de recherche sémantique Corese* permettant, via une *interface de requête dédiée* à MEAT, de reposer sur l'ontologie et sur la base d'annotations éventuellement complétée grâce aux inférences permises par une base de règles, pour fournir des réponses pertinentes aux requêtes du biologiste. Nous développerons des extensions de Corese et en particulier des interfaces graphiques dédiées à MEAT : une interface de requête (en se basant sur les requêtes privilégiées par les biologistes recherchant des articles scientifiques pouvant les aider à interpréter les résultats de leurs expériences ou à valider les interprétations), une base de règles d'inférences dédiées à MEAT et une interface de présentation des réponses et de navigation dans celles-ci.

Remarque : Le module MEAT-Annot&Search exploitera l'ontologie choisie comme pertinente pour les expériences biopuces. Il pourra s'agir de l'ontologie Gene Ontology, (qui couvre les gènes, les processus biologiques et les composants cellulaires) ou d'UMLS si les annotations doivent par exemple porter sur les maladies. Les techniques linguistiques utilisées pour la génération semi-automatique des annotations sémantiques pourront également être utilisées en

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

cas de besoin pour l'enrichissement de l'ontologie utilisée (UMLS, Gene Ontology) voire pour la création d'une ontologie dédiée aux expériences biopuces.

## 2. Le module de fouille de données (MEAT-Miner).

▪ Le module de fouille de données MEAT-Miner mettra à disposition de l'utilisateur des outils de fouille de données permettant d'intégrer au processus de recherche de modèles, les connaissances issues de la base de connaissances BRIGene.

Son rôle sera

- de proposer à l'utilisateur différents scénarios pour l'analyse des données d'expression d'une expérience ; on entend ici par scénario, une suite de méthodes de fouilles de données se combinant de manière à obtenir un modèle à partir du jeu de données initial,
- d'exécuter les scénarios définis et d'accompagner les modèles ainsi extraits d'explications issues de la base de connaissances BRIGene,
- de construire une base de données stockant les données et les résultats d'expériences autorisant l'analyse comparative d'une nouvelle expérience,
- d'enrichir la base de connaissances BRIGene avec les résultats (modèles) d'expériences validés.

Le module MEAT-Miner comprendra

- *un interface de définition de scénarios* pour l'analyse des données ; l'utilisateur-biologiste pouvant ainsi définir plusieurs scénarios de fouille sur le même jeu de données d'expression. La mise en œuvre de différents scénarios partant par exemple de différents ensembles de gènes sélectionnés ou utilisant différents algorithmes de classification permettra d'étudier les modifications induites par ces paramètres sur les connaissances extraites et de définir la meilleure stratégie d'analyse.
- *une palette d'outils dédiés à l'analyse des données d'expression* et à l'utilisation de la base de connaissances créée par MEAT-Annot et MEAT-Onto pour :
  - *la sélection de gènes d'intérêt*. Contrairement aux jeux de données habituellement analysés en fouille de données, une des caractéristiques des données d'expression réside dans le nombre important de gènes en regard du nombre d'observations. La sélection d'un sous-ensemble de gènes permet de réduire le bruit induit dans les algorithmes d'extraction et de focaliser l'étude sur des gènes d'intérêt. Les techniques de réduction classiques se basent sur le nombre de valeurs manquantes ou sur des analyses de variances; les méthodes de classification permettent de regrouper des gènes ayant des profils d'expression similaires et ainsi d'identifier des gènes d'intérêt dans l'ensemble des observations.  
Dans MEAT-Miner, la sélection de gènes d'intérêt (MEAT-Selection) sera réalisée à partir plusieurs sources d'information : la liste des gènes d'intérêt fournie par le biologiste, l'analyse statistique des données d'expression, mais, également les annotations associées aux descriptions d'expériences créées par MEAT-Annot. En effet, lors de la définition d'une nouvelle expérience par le biologiste, le module MEAT-Search sera invoqué pour interroger les bases d'annotations et fournir les gènes d'intérêt en considérant la problématique décrit et, les références bibliographiques fournies par le biologiste.
  - *la recherche de groupes de gènes*, qui pourra être réalisée en faisant appel aux outils développés dans différentes plates-formes existantes (BioConductor, GeneSpring, ...) ou en utilisant un algorithme (MEAT-Cluster) dédié au système MEAT qui permettra de rechercher des groupes de gènes voisins selon des critères

## Action Concertée Incitative

### IMPbio

## Descriptif complet du projet

### (recherche multidisciplinaire, service bioinformatique)

définis relativement à leurs annotations. La notion de voisinage utilisée dans MEAT-Cluster, devra intégrer des critères multiples de similarité entre gènes : similarité des profils d'expression dans l'expérience et par exemple similarité dans la position chromosomique ou dans la fonction biologique.

- *la prédiction d'appartenance à une classe*, qui pourra également être réalisée en faisant appel à un module d'une plate-forme existante ou en utilisant un algorithme (MEAT-Predict) dédié au système MEAT qui permettra de découvrir un modèle de prédiction pour un traitement biologique, une fonction, une pathologie, etc... Dans le cas de l'algorithme MEAT-Predict, la recherche d'un modèle de prédiction, par exemple constitué de règles de prédiction, sera guidée par les données d'expression et par les informations issues de la base de connaissances.

Notons que :

- lors de l'exécution des algorithmes MEAT-Cluster et MEAT-Predict, le moteur de recherche MEAT-Search sera invoqué automatiquement pour la recherche d'annotations et autres informations issues de la base de connaissances. Aussi, les modèles générés seront naturellement accompagnés de ces informations et éviteront à l'utilisateur-biologiste, dans un premier temps, une recherche fastidieuse dans la bibliographie pour expliquer les résultats obtenus ;

- le recours à des algorithmes standard s'imposera dans le cas où les annotations et les références bibliographiques relatives à la problématique posée sont trop rares ou inexistantes ;

- l'exécution de scénarios alternatifs permettra de lancer de nouvelles analyses du jeu de données et d'observer les modifications induites sur les modèles par un changement de paramètre.

○ *l'éditeur de résultats d'analyse* qui permettra à l'utilisateur d'observer les connaissances extraites et de valider l'expérience et les résultats qui seront stockés dans la base BD\_EXP (cf. paragraphe suivant). Lorsque l'expérience sera validée, sa description et la description de ses résultats seront insérées dans la base de connaissances de MEAT-Annot.

○ *un outil d'analyse d'expériences* intégrant

- *la base d'expériences BD\_EXP* qui stockera les données d'expression sous forme de matrices et les résultats extraits de l'analyse lorsque l'expérience est validée par l'utilisateur.

Notons que la description des expériences gérées par MEAT-Annot ainsi que les informations stockées dans BD-EXP seront dans le format utilisé par la base de données ArrayExpress de l'EBI (<http://www.ebi.ac.uk/arrayexpress>).

- *un outil d'analyse* qui exploitera la base de données BD\_EXP pour caractériser une nouvelle expérience par rapport aux expériences précédemment validées. Les expériences seront regroupées par problématique et on maintiendra, pour chaque problématique, une classification basée sur les données d'expression. Cet outil d'analyse d'expériences permettra de déterminer la similarité de la nouvelle expérience avec une des classes identifiées..

### 3. Le module de médiateur ontologique (MEAT-Onto).

Le module MEAT-Onto gère l'interface de MEDIANTE aux bases d'ontologies externes. Il gère également les nouvelles connaissances détectées par MEAT-Miner et MEAT-Annot.

Le module MEAT-Onto comprendra:

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

- *L'outil d'acquisition d'ontologies, composé d'un analyseur de requête, d'un moteur de recherche, d'un transformateur automatique de format et d'un éditeur-visionneur d'ontologies.* Les bases d'ontologies locales sont alimentées par les ontologies externes récupérées depuis des réservoirs d'ontologies existantes (GO et UMLS en particulier). Les ontologies récupérées sont transformées automatiquement dans le format RDF (standard utilisé actuellement dans MEDIANTE). La transformation dans le format OWL (le standard qui succède à DAML+OIL, choisi par le Bio-Ontology Consortium comme format unifié) sera faite manuellement à l'aide d'un éditeur. Les nouvelles connaissances détectées par MEAT-Miner et MEAT-Annot seront également gérées dans les bases ontologiques sous les deux formats RDF et OWL.
  - *L'analyseur de requête* sera utilisé spécialement par le moteur de recherche. Il permettra la mise en forme d'une requête adéquate au réservoir d'ontologies destinataire (local ou externe).
  - *Le moteur de recherche* prendra en charge la coordination de la réalisation d'une demande venant de MEAT-Annot ou MEAT-Miner. Il prendra en entrée une demande de ces composants, effectuera d'abord une recherche dans les bases ontologiques locales, puis une recherche éventuelle dans une des bases ontologiques externes, en cas d'échec. Les ontologies locales sont indexées par une annotation sous format OWL, décrivant son contenu, sa provenance et son format d'origine.
  - *Le transformateur automatique* permettra de convertir une ontologie d'un format origine dans le format RDF. Il sera utilisé par le moteur de recherche ou par l'éditeur-visionneur d'ontologie.
  - *L'éditeur-visionneur d'ontologie fonctionnera manuellement.* Il permettra à l'utilisateur d'effectuer les actions suivantes :
    - Visionner une ontologie sous forme d'un graphe des concepts. L'utilisateur pourra naviguer dans ce graphe et afficher les caractéristiques attachées à chaque concept. Il pourra également réaliser une recherche des concepts selon ses caractéristiques sémantiques
    - Effectuer manuellement un chargement d'une ontologie locale à partir d'un réservoir d'ontologies externe. Ce travail sera réalisé à l'aide du moteur de recherche.
    - Modifier une ontologie locale : la modification des caractéristiques d'un concept, l'ajout d'un lien ou d'un rôle, suppression des concepts et des liens, etc. Pour protéger la base de connaissances, ce mode de travail pourra être verrouillé par un mot de passe personnalisé. Un journal de modification sera conservé pour pouvoir rétablir les états précédents de l'ontologie.
    - Editer un nouveau concept dans une ontologie locale. Le concept ajouté sera contrôlé par le moteur d'inférence (la consistance, la redondance, ...) avant la validation dans la base d'ontologie.
- *Le moteur d'inférence reposant sur la logique de description* disposant deux interfaces d'utilisation : une interface de programmation sous forme de composants logiciels (API servlet java et de JSP2) et une interface d'exploitation graphique pour les utilisateurs finaux. En effet, une représentation formelle des ontologies est la seule alternative permettant d'éviter l'inconsistance dans une ontologie de taille importante. Le Metathesaurus UMLS, dans la version 2002, contient 776 940 concepts et 11 138 000 liens sémantiques. A titre expérimental, Undo Hahn a réalisé une transformation dans la logique de description de deux sous ontologies d'UMLS : l'anatomie avec 38 059 concepts et la pathologie avec 50 087 concepts. Cette

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

transformation lui permet d'identifier 2 328 définitions de concept inconsistantes dans l'ontologie d'anatomie et 355 cycles dans l'ontologie de pathologie [*Comparative and Functional Genomics* 2003 ; 4 : 94-97]. Le moteur d'inférence logique ONDIL sera réalisé pour une triple utilité :

- Le fournisseur des services d'inférence pour les composants de MEAT-Onto en particulier pour la construction de l'ontologie unifiée sous le format OWL.
- Le fournisseur des services d'inférence complémentaires pour les biologistes pour exploiter les connaissances dans les ontologies, par exemple : la vérification de la relation d'appartenance entre les deux concepts, la recherche d'approximation à partir des connaissances partielles du concepts, la vérification de redondance et/ou d'inconsistance d'un nouveau concept, etc.
- Le fournisseur des services d'inférence pour les autres composants logiciels du système. Ces services pourront être utilisés pour faciliter la migration de MEDIANTE vers les nouveaux standards d'échange dans le domaine (projet GONG : *Gene Ontology Next Generation project* : <http://gong.man.ac.uk/>; projet TAMBIS : <http://imgproj.cs.man.ac.uk/tambis/>).

### **Caractère innovant**

Le monde médical est depuis longtemps sensibilisé au besoin de modéliser ses connaissances et d'explicitier ses terminologies. Il existe ainsi nombre de ressources terminologiques ou ontologiques dans le domaine médical : GALEN (Rector et al, 1994, 1993), SNOMED RT (Côté et al, 1993, Spackman et al, 1997), MENELAS (Zweigenbaum et al, 1995), UMLS (National Library of Medicine, 2003) sont des ontologies / thesaurus / meta-thesaurus qui modélisent une partie du domaine médical.

**MEAT-Annot&Search.** Les travaux prévus sur le module MEAT-Annot&Search peuvent se comparer avec (a) les travaux exploitant l'extraction d'information dans le domaine de la biologie (Staab, 2002), (b) ceux sur la génération d'annotations sémantiques pour le web sémantique (Handsuh et al, 2002).

Dans le domaine de la fouille d'articles en biologie, (Shatkay et al, 2002) propose des techniques statistiques et des algorithmes d'apprentissage pour découvrir des interactions entre gènes à partir de résumés d'articles en biologie dans la base PubMed. Notre approche reposera sur des outils linguistiques et diffère ainsi des approches d'apprentissage proposées par (Blasche et al, 2002), (Nédellec, 2002). Les techniques linguistiques que nous utiliserons différeront des techniques d'extraction d'information offertes dans (Proux et al, 2000a, 200b) pour extraire des interactions géniques à partir de textes. Nous adopterons la méthode d'acquisition des connaissances à partir de corpus textuel prônée par (Aussenac-Gilles et al, 2000).

Concernant la génération automatique d'annotations RDF, l'approche visée diffèrera de (Cao et al, 2003) qui repose sur la généralisation à partir d'un exemple d'annotation manuelle sur des documents web structurés. Elle diffèrera aussi de celle proposée dans (Golebiowska, 2002) où l'annotation générée sur un document consistait en des instances des concepts de l'ontologie qui avaient justement été générés à partir de ce document : la génération d'annotation était donc liée à l'enrichissement de l'ontologie. Or notre approche permettra de créer des annotations consistant non seulement en des instances de concepts, mais aussi en des instances de relations, et le tout en reposant sur une ontologie déjà existante.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

Enfin, le module MEAT-Annot&Search ne vise pas comme le module GenoAnnot de la plateforme Genostar (<http://www.genostar.org>) l'annotation de séquences génomiques.

Outre la génération semi-automatique des annotations sémantiques grâce à de multiples outils linguistiques à partir de textes, la manipulation d'annotations sémantiques contextuelles et multi-points de vue sera l'une des originalités du module MEAT-Annot.

L'un des aspects innovants du projet sera l'intégration, dans un même système, de techniques de fouilles de données avec des techniques de génération d'annotations sémantiques basées sur des ontologies.

**MEAT-Miner.** Du point de vue de la fouille de données, l'approche suivie dans MEDIANTE se place dans le contexte des travaux sur la recherche de modèles pertinents et "intéressants" pour l'utilisateur expert du domaine d'application. Le problème de l'évaluation de l'intérêt/utilité de la connaissance extraite est un problème central en fouille de données. En effet, les algorithmes mis en œuvre génèrent souvent des volumes d'informations (règles) très importants et ces informations doivent être filtrées pour en extraire à nouveau les plus pertinentes. Les premiers critères d'intérêt des modèles ont été définis de manière objective en termes de mesures statistiques représentant la *précision*, le *support* et la *confiance* d'une règle par exemple, ou le *taux d'erreur* d'un modèle de prédiction. MEAT-Miner se basera également sur ce type de mesures pour évaluer la qualité d'un modèle de prédiction. Mais par ailleurs, il est essentiel pour la pertinence de la démarche, d'étudier le caractère subjectif de l'intérêt des modèles en termes d'utilité pour les experts du domaine. Ce sujet soulève des difficultés car les facteurs d'utilité dépendent alors du domaine des données et visent à sélectionner des informations intéressantes pour un utilisateur donné ou innovantes en regard de connaissances déjà acquises sur le domaine. La plupart des approches proposées demandent explicitement à l'utilisateur quel type d'information est intéressant ou inintéressant et le système génère alors les modèles ainsi définis. (Klemetinen et al., 1994) ont proposé une approche basée sur des modèles génériques de règles. (Srikant, Vu et Agrawal, 1997) ont proposé que l'utilisateur définisse des contraintes. Plus récemment (Albert-Lorincz et Boulicaut, 2003) ont proposé l'extraction de séquences fréquentes vérifiant une expression régulière donnée alors que (Piatetsky-Shapiro et Matheus, 1994) se basaient sur un système expert spécifiquement construit pour extraire des informations concernant un système de santé.

Les données de biopuces offrent un champ d'application privilégié à ce sujet de recherche, car on dispose des sources d'information expertes avec les ontologies sur les gènes, les bases de connaissances médicales et les articles scientifiques en ligne. De nombreux travaux de fouille de données d'expression filtrent les informations extraites en se basant sur les ontologies du domaine et/ou propose des solutions hybrides avec la fouille de textes. L'approche suivie dans MEAT-Miner est cependant originale dans le cadre des travaux de fouille sur les données de biopuces, car d'une part, elle utilisera le moteur de recherche MEAT-Search qui fournit un condensé des informations nécessaires et d'autre part, elle vise à intégrer directement les informations issues de la base de connaissances dans le processus d'extraction.

**MEAT-Onto.** Le développement du module MEAT-Onto repose sur les résultats de recherche récents dans le domaine de représentation formelle des connaissances avec les logiques de description (cf par exemple <http://eden.i3S.unice.fr/ondil/>). L'intégration d'un tel moteur comme noyau de MEAT-Onto fait partie des premières tentatives d'utilisation des services d'inférence de Logiques de Description au sein d'une application réelle. Les moteurs d'inférence développés par les équipes ACACIA et Krystal (du projet EXECO) sont comparables à FACT (Fast Classification of Terminologies) développé à l'Université de Manchester en 2000 (<http://www.cs.man.ac.uk/~horrocks/FaCT/>), KAON développé à l'Université de Karlsruhe en 2002 (<http://kaon.semanticweb.org/>) ou RACER (Semantic Middleware for Industrial Projects) développé à l'université technologique de Harburg en 2003. Les aspects innovants dans le développement du module MEAT-Onto concernent:

<p><b>Action Concertée Incitative</b>  <b>IMPbio</b>  <b>Descriptif complet du projet</b>  <b>(recherche multidisciplinaire, service bioinformatique)</b></p>
---

- L'intégration *des inférences logiques* comme une méthode complémentaire dans les différents processus de recherche, de détection, de classification et d'échange d'expériences. En effet, d'une part les inférences logiques, de la manière triviale, permettent de :
  - vérifier la cohérence (consistance) et/ou la redondance d'une connaissance découverte par rapport à la base de connaissances définie (par exemple pour vérifier la consistance du résultat d'une fouille de données, etc.)
  - vérifier ou détecter l'appartenance d'un objet à une classe
  - proposer une classification d'un nouveau concept (nouvelle classe) à partir de ses caractéristiques partielles,
  - calculer l'approximation sémantique d'inter-ontologies par exemple entre un concept défini dans ULMS et un concept similaire défini dans GO
  - etc.
- L'utilisation des inférences logiques pour détecter des connaissances cachées, pour démontrer des hypothèses ou encore pour calculer la similarité entre deux expériences, etc.

### Références

- Adomavicius, G. and Tuzhilin, A. "Discovery of actionable patterns in databases: the action hierarchy approach." *KDD-97*, 1997, pp. 111-114.
- Agrawal, R. and Srikant, R. Fast algorithms for mining association rules. *VLDB-94*, 1994.
- Aussenac-Gilles N., Biébow B. & Szulman S., (2003). D'une méthode à un guide pratique de modélisation de connaissances à partir de textes. In TIA 2003, p. 41-53, Strasbourg
- Aussenac-Gilles N., B. Biébow, S. Szulman. Revisiting Ontology Design ; a methodology based on corpus analysis. In R. Dieng, O. Corby eds, Knowledge Engineering and Knowledge Management: methods, models and tools, Springer-Verlag, LNAI 1937, p. 172-188.
- Bayardo, R. and Agrawal, R. "Mining the most interesting rules." *KDD-99*, 1999.
- Berners-Lee T., Hendler J. & Lassila O (2001). The Semantic Web, *Scientific American*, 84(5) p. 34-43
- Bing Liu, Wynne Hsu, Shu Chen and Yiming Ma, "Analyzing the Subjective Interestingness of Association Rules," *IEEE Intelligent Systems*, 2000
- Blaschke C. & Valencia A., (2002). Molecular biology nomenclature thwarts information-extraction progress. *IEEE Intelligent Systems & their Applications*, p. 73-76, March/April.
- Bourigault D. & Fabre C. (2000), Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de grammaire*, Vol.25, pp.131-151
- Cao T-D., Gandon F., Dieng R., (2003), Intégration de sources extérieures dans un Web sémantique d'entreprise géré par un système multi-agents. In IC 2003, Laval
- Corby O. & Faron-Zucker C. (2002), Corese: A Corporate Semantic Web Engine. WWW11 Workshop on Real World RDF and Semantic Web Applications, Hawaii
- Cunningham H. & Maynard D. & Bontcheva K. & Tablan V. & Ursu C. & Dimitrov M. (2003), *Developing Language Processing Components with GATE (User Guide)*
- Cunningham H., Maynard D., Bontcheva K. & Tablan V. (2002). *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications*. ACL'02.
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. "From data mining to knowledge discovery: an overview," In: *Advances in knowledge discovery and data mining*, U. Fayyad, G. Piatetsky-Francisci D., M. Collard, *Multi-Criteria Evaluation of Interesting Dependencies according to a Data Mining Approach*, CEC'03, IEEE Congress on Evolutionary Computation, Canberra, Australie, 2003.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

Francisci D., M. Collard, *Optimizing rule quality in a data mining context*, CIRAS'03, 2<sup>nd</sup> International Conference on Computational Intelligence, Robotics and Autonomous Systems, Singapour, 2003.

Golebiowska J. , Dieng R., Corby O. & Mousseau D. (2002). Ontologies au service de la mémoire d'un projet-véhicule et de la recherche d'information, Doc. Numérique, p.173-192.

Golebiowska J., R. Dieng, O. Corby, D. Mousseau. Samovar : using ontologies and text-mining for building an automobile project memory, in R. Dieng-Kuntz et N. Matta eds, *Knowledge Management and Organizational Memories*, p. 89-102, Kluwer Academic Publishers, Boston, 2002

Handschuh S. , Koivunen M., Dieng R. & Staab S. (2003), eds, Proc. of KCAP'2003 Workshop on Knowledge Markup and Semantic Annotation, Sanibel, Florida, October 26.

Kamber, M., and Shinghal, R. 1996. "Evaluating the interestingness of characteristic rules." *KDD-96*, 1996, pp. 263-266.

Klemetinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A.I. 1994. "Finding interesting rules from large sets of discovered association rules." *Proceedings of the Third International Conference on Information and Knowledge Management*, 1994, pp. 401-407.

Le Moigno S., Charlet J., Bourigault D. & Jaulent M. (2002), Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale, In IC 2003, Rouen

National Library of Medicine (2003). UMLS Knowledge Source. 14th Edition, Jan. 2003 Doc. National Institute of Health - National Library of Medicine, Bethesda, Md, USA.

Nédellec C., (2002), Bibliographical Information Extraction in Genomics. *IEEE Intelligent Systems & their Applications*, p. 76-80, March/April.

Ng. R. T., Lakshmanan, L. V. S., Han, J., and Pang, A. "Exploratory mining and pruning optimizations of constrained association rules." *SIGMOD-98*, 1998.

Piatetsky-Shapiro, G., and Matheus, C. "The interestingness of deviations." *KDD-94*, 1994.

Proux D., Rechenmann F., Julliard L.: A Pragmatic Information Extraction Strategy for Gathering Data on Genetic Interactions. *ISMB 2000*: 279-285

Proux D., Rechenmann F., Julliard L.: Muninn: A Pragmatic Information Extraction System. *SPIRE 2000*: 236-241

Rector A. et al (1994). The Galen Model Schemata for Anatomy: Towards a Re-Usable Application- Independent Model of Medical Concepts. *Proc. MIE 94* , p. 229-233.

Rector A., Rogers J. & Pole P. (1996). The Galen High Level Ontology. *Proceedings of Medical Informatics in Europe MIE 96*, p. 174-178. Amsterdam: IOS Press.

Schulze-Kremer S. & Smith B. & Kumar A.(2002). Revising the UMLS Semantic Network

Shatkay H., Edwards S. & Boguski M., (2002), Information Retrieval Meets Gene Analysis. *IEEE Intelligent Systems & their Applications*, p. 45-53

Silberschatz, A., and Tuzhilin, A. "What makes patterns interesting in knowledge discovery systems." *IEEE Trans. on Know. and Data Eng.* 8(6), 1996, pp. 970-974.

Soo V., Lee C., Li C., Chen S. & Chen C. (2003). Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques, *JCDL'03*, Texas.

Spackman K. , Campbell K. & Côte R. (1997). SNOMED RT: A Reference Terminology for Health Care. In *Proceedings of the 1997 Amia Annual Symposium*. Nashville.

Srikant, R., Vu, Q. and Agrawal, R. "Mining association rules with item constraints." *KDD-97*, 1997, pp. 67-73.

Staab S. ed. (2002), Mining Information for Functional Genomics. *IEEE Intelligent Systems & their Applications*, p. 66-80, March-April.

Staab S., Maedche A. & Handschuh S (2001), An annotation framework for the semantic web, In *Proceedings of the First Workshop on Multimedia Annotation*, Japon

**Action Concertée Incitative**  
***IMPbio***  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

Zweigenbaum P. et al (1995). Menelas: Coding and Information Retrieval from Natural Language Patient Discharge Summaries. Advances in Health Telemat., p. 82-89, IOS Press.

**Intérêt de la coopération**

L'IPMC apportera :

Son expertise dans le domaine des puces ADN.

- Un cadre d'application pour l'adaptation, la mise en œuvre et l'expérimentation de méthodes et outils relatifs à la gestion des connaissances dans le contexte d'une mémoire d'expériences en biologie.
- Son expérience dans le domaine de la recherche d'informations en biologie. L'IPMC sera à même d'identifier les besoins, de proposer des fonctionnalités, et d'évaluer l'utilisation et l'intérêt de la mémoire d'expériences.
- Son expertise dans le domaine des bases de données et de l'algorithmique appliquées à la biologie.
- Son savoir-faire dans l'utilisation des technologies client-serveur pour la mise en place d'applications sur le WEB

ACACIA apportera :

- Son expertise dans le domaine de la mémoire d'entreprise, du travail collaboratif et du Web sémantique : matérialisation d'une mémoire dans un Web sémantique d'entreprise, construction d'ontologies et d'annotations à partir de textes, recherche d'information guidée par des ontologies.
- Son moteur de recherche sémantique : CORESE qui est actuellement utilisé dans de nombreux projets de collaboration d'ACACIA
- Son expérience dans le domaine de la recherche d'information dans des publications sur les gènes, via ses travaux antérieurs dans le projet ESCRIRE.

Le projet EXeCO apportera :

- Son savoir-faire dans le domaine des méthodes d'extraction de connaissances à partir de données et plus particulièrement sur la sélection d'informations pertinentes et l'intégration de connaissances expertes dans le processus d'extraction,
- Son expertise sur les algorithmes standard et ses recherches sur certaines meta-heuristiques,
- Son expérience sur l'analyse de jeux de données réelles.
- Son savoir-faire et son expertise dans le domaine de représentation formelle de connaissances par les logiques de description et sur les algorithmes d'inférence logique

**Action Concertée Incitative**  
***IMPbio***  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**B3 – Résultats attendus :**

*On détaillera l'échéancier des résultats et réalisations intermédiaires et finaux attendus. On précisera les risques scientifiques qui seront pris. On discutera de l'impact potentiel de ce projet sur les scènes européenne et internationale.*

*Impact national et international*

Ce projet s'intègre dans les objectifs européens de mise en place de standards dans le domaine de l'analyse transcriptome, et participera activement à la mise en place d'outils adaptés.

*Echéancier*

La partie d'archivage sera terminée dès les six premiers mois du projet. Elle permettra de proposer rapidement aux utilisateurs un outil d'archivage et de transfert des données. Parallèlement, les outils de MEAT-Annot, MEAT-Miner et MEAT-Onto seront développés, en sorte de disposer d'un premier prototype à J+18. Après six mois de tests, le cahier des charges de la version 2 du projet sera établie, en sorte qu'elle soit disponible à la fin de ce dernier (J+36 mois). Des rapports d'avancement annuels seront établis.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**C – MOYENS FINANCIERS ET HUMAINS DEMANDES PAR CHAQUE EQUIPE<sup>11</sup>**

*Comme indiqué dans les tableaux ci-dessous, on distinguera*

- *les financements via le Fonds National pour la Science qui peuvent inclure*
  - \* *du fonctionnement*
  - \* *de l'équipement*
  - \* *des mois de personnel temporaire (CDD) pour un montant ne pouvant dépasser 50% du financement total attribué. La durée du ou des contrat(s) prévus, qui ne peuvent être supérieures à 24 mois chacun, sera précisée.*
- *les demandes d'allocations de recherche*

*Les diverses possibilités concernant l'attribution de moyens pour recruter ou accueillir des personnels seront globalement très limitées pour l'ensemble des ACI. Leurs demandes devront donc être particulièrement justifiées. Si les bénéficiaires de ces demandes sont connus ou pressentis, les CV correspondants seront joints à la présente demande.*

*On présentera une justification scientifique des moyens demandés pour chacune des équipes impliquées dans le projet.*

---

<sup>11</sup> Une fiche C doit être remplie pour chaque laboratoire ou équipe partenaire

**Action Concertée Incitative  
IMPbio  
Descriptif complet du projet  
(recherche multidisciplinaire, service bioinformatique)**

**C1 - Demandes effectuées dans le cadre de l'ACI pour le présent projet :**

**Nom de l'équipe ou du laboratoire :**

Moyens demandés dans le cadre de la présente ACI (en K€ TTC) :

**Financements via le Fonds National de la Science :**

	2004	2005	2006	Total
Equipement	Equipe1 : 2K€	Equipe1 : 2K€	Equipe1 : 2K€	Equipe1 : 6K€
	Equipe2 : 2K€	Equipe2 : 2K€	Equipe2 : 2K€	Equipe2 : 6K€
	Equipe3 : 2K€	Equipe3 : 2K€	Equipe3 : 2K€	Equipe3 : 6K€
Fonctionnement (dont CDD décrits ci-dessous)	Equipe1 : 20K€	Equipe1 : 20K€	Equipe1 : 20 K€	Equipe1 : 60K€
	Equipe2 : 20K€	Equipe2 : 20K€	Equipe2 : 20K€	Equipe2 : 60K€
	Equipe3 : 20K€	Equipe3 : 20K€	Equipe3 : 20K€	Equipe3 : 60K€
<b>Total / année</b>	<b>66K€</b>	<b>66K€</b>	<b>66K€</b>	<b>198K€</b>

**Dépenses de personnels (CDD) <sup>12</sup>:**

**IPMC**

Nature de l'emploi (post-doc, ingénieur, assistant-ingénieur,...)	Ingénieur de Recherche
Durée de l'emploi (en mois) <sup>13</sup>	12 mois
Coût total de l'emploi	33K€

**INRIA**

Nature de l'emploi (post-doc, ingénieur, assistant-ingénieur,...)	Ingénieur Expert
Durée de l'emploi (en mois) <sup>14</sup>	8.5 mois
Coût total de l'emploi	33K€

**I3S**

Nature de l'emploi (post-doc, ingénieur, assistant-ingénieur,...)	Ingénieur
Durée de l'emploi (en mois) <sup>15</sup>	12 mois
Coût total de l'emploi	33K

**Allocations de recherche :**

<sup>12</sup> Un tableau doit être rempli pour chaque demande de CDD.

<sup>13</sup> Doit être inférieure à 24 mois

<sup>14</sup> Doit être inférieure à 24 mois

<sup>15</sup> Doit être inférieure à 24 mois

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

	2004-2005	2005-2006	Total
Nombre d'allocations de recherche débutant en :			

**Justifications scientifiques de l'ensemble des demandes :**

**C2 - Autres soutiens financiers :**

*On mentionnera les autres actions liées de près ou de loin au projet dans lesquelles l'équipe ou le laboratoire est engagé (projets européens, RNRT, RNTL, autres ACI, ...).*

L'équipe Acacia participe au réseau d'excellence Knowledge Web, L'équipe Acacia participe au réseau d'excellence Knowledge Web, à l'ACI Ligne de Vie pour la gestion de connaissances dans un réseau de santé, au projet RTNT KMP et dispose d'une demi-bourse régionale pour la thèse de Khaled Khelif.

**Action Concertée Incitative**  
**IMPbio**  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**D - RECAPITULATIF GLOBAL DES DEMANDES DU PROJET :**

**Financements via le Fonds National de la Science :**

	2004	2005	2006	Total
Equipement	6K€	6K€	6K€	18K€
Fonctionnement (dont CDD décrits ci-dessous)	60K€	60K€	60K€	180K€
<b>Total / année</b>	<b>66K€</b>	<b>66K€</b>	<b>66K€</b>	<b>198K€</b>

**Dépenses de personnels (CDD) :**

Nombre d'emplois	3
Durée cumulative des emplois (en mois)	32.5 mois
Coût total des emplois	99K€

**Allocations de recherche :**

	2004-2005	2005-2006	Total
Nombre d'allocations de recherche débutant en :	0	0	0

**Action Concertée Incitative**  
***IMPbio***  
**Descriptif complet du projet**  
**(recherche multidisciplinaire, service bioinformatique)**

**E - ENGAGEMENT DU COORDINATEUR DU PROJET :**

*La présente page ne sera remplie que dans la version sous forme papier.*

Je soussigné Pascal Barbry, coordinateur du projet MEDIANTE, m'engage dans l'hypothèse où le présent projet serait retenu à :

- fournir un rapport d'évaluation à mi-parcours permettant au Conseil Scientifique d'apprécier l'avancement des travaux et la coopération des équipes participantes.
- fournir un rapport à la fin de l'exécution du projet.
- maintenir régulièrement une page web résumant l'ensemble des activités du projet.

**Signature du coordinateur du projet :**



**Visa du Directeur du Laboratoire ou de l'Unité de Recherche auquel appartient le coordinateur du projet:**