



Formulaire – Appel à projets interdisciplinaires

Défi Imag'In

Identification

Civilité et Nom du porteur du projet	Professeur Thierry TATONI
Titre long (max 150 caractères)	Visualisation of Graph In transdisciplinary Global Ecology, Economy and Sociology data-Kernel
Acronyme	VIGI-GEEK

Résumé du projet :

" VIGI-GEEK " propose de construire un outil de représentation sous forme de graphes des données de différents champs disciplinaires (écologie, sociologie, économie) et d'élaborer des méthodes de création de scénarios par approches successives (coévolution de facteurs), basée sur des concepts actuellement décrits par les approches globales. L'objectif est de construire des graphes paramétrables avec des données hétérogènes (de la molécule à l'écosystème, en passant par les traits de vie, jusqu'aux paysages et aux interactions homme-milieu) concernant l'écologie méditerranéenne et d'analyser les données grâce à des algorithmes utilisés dans d'autres disciplines.

Contexte scientifique et adéquation du thème du défi

Force est de constater que la donnée utilisée par les scientifiques dans le domaine de l'écologie continue de se diversifier et n'est plus entièrement produite par les institutions scientifiques : ceux-ci s'appuient aussi sur des réseaux extérieurs d'acteurs et de compétences. De nouvelles disciplines comme la «protéomique», la «méta-bolomique», la «méta-génomique» viennent compléter les prismes d'observation déjà multiples de la biodiversité.

L'accessibilité de ces données est variable, et les processus de qualification qui évaluent leur utilisabilité et leur efficacité sont encore rares. Inversement, la production de données scientifiques est de plus en plus financée sous condition de mise à disposition (depuis plusieurs décennies pour les données de biologie moléculaire, mais encore de façon balbutiante pour les données écologiques et environnementales), sans que des outils appropriés à une analyse intégrative soient proposés.

Le constat est là, la révolution du *Big Data* en écologie tarde, alors qu'elle est considérée par la plupart des disciplines scientifiques et des industries produisant et utilisant de l'information comme la plus prometteuse des pistes de progrès et de découvertes. Les systèmes d'information reliant des objets grâce à des qualifications (des liens) sont omniprésents, le premier étant l'homme, le consommateur. L'entreprise « Google » a très largement contribué à la vulgarisation d'outil de recherche d'information à partir de pages internet et leurs liens. Très vite, les systèmes mercantiles ont compris la puissance analytique de la «fouille de données» (les assurances pour gérer les risques, la française des jeux pour trouver les tricheurs, les banques pour les investissements, les « traders » pour augmenter leur marge, les publicitaires et les réseaux sociaux pour augmenter leurs impacts...).

Aujourd'hui, une discipline nouvelle appelée « écologie statistique » [développé notamment dans le cadre du **Groupe de Recherche en Ecologie Statistique** (Gimenez et al. 2014)] propose de nouvelles méthodes d'analyse de ces données, et pourrait participer au développement d'approches intégratives et multimodales, prenant en compte sans priorisation plusieurs disciplines, notamment dans les domaines de l'économie et des sciences sociales. Mais, à cause d'un cloisonnement encore trop important entre disciplines, ce n'est pas encore le cas. Pourtant, l'urgence environnementale demande aujourd'hui de répondre, grâce à un système partagé et connecté, à des questions locales et globales à la fois, dépassant les questions scientifiques du type *cette dégradation est-elle liée à cette pression ?* mais répondant plutôt à des questions opérationnelles du type : *Comment améliorer/préserver l'état écologique d'un milieu de la manière la plus efficace en prenant en compte les interactions homme-milieu?*

Cette question revient à identifier les limites à ne pas dépasser pour un ensemble de conditions ayant les unes sur les autres des effets antagonistes ou potentialisateurs.

Objectifs

L'objectif de « VIGI-GEEK » est de produire un outil « multidisciplinaire » de construction et de visualisation de graphes dans le cadre d'un consortium multidisciplinaire appelé « IndexMed » (en cours de développement : www.indexmed.eu), puis de développer à moyen terme les usages de ces graphes pour l'aide à la décision en gestion environnementale dans le cadre d'un projet de recherche à soumettre aux appels à projets européens (BiodivERsA, FEDER, SeasEra, H2020).

Ces graphes seront construits à partir d'informations agrégées grâce à des points nodaux d'indexation massive et de qualification des données sur l'environnement littoral et marin méditerranéen, dans différentes disciplines [socio-écologie, économétrie, écologie (structure et fonctionnement), urbanisme, gestion...] et à l'échelle méditerranéenne. Le développement du prototype vise à rendre ces graphes paramétrables pour fouiller et visualiser ces données pluridisciplinaires en mettant sur le même plan des données de types socio-écologiques, économiques, écologiques, moléculaires et fonctionnelles (relations trophiques, traits fonctionnels...).

Ces points nodaux d'indexation seront «clonables» à volonté avec des règles d'enrichissement et de partage correspondant aux licences «creative common» du type «partage dans les mêmes conditions», autorisant les autres à reproduire, diffuser et modifier l'index, à condition qu'ils publient toute adaptation de l'index sous les mêmes conditions (open-source, open data). Ces règles devront favoriser l'émergence de standards pour améliorer l'interopérabilité des données et favorisera la participation de nouveaux laboratoires contributeurs en tenant compte de leurs possibilités techniques de contribution au fur et à mesure du développement du projet.

Les modèles et outils qui découleront de ce prototype ont pour objectif de normer, qualifier puis croiser des données de disciplines différentes et de proposer des méthodes d'élaboration de scénarios par approches successives (encadrement de valeurs), en se basant sur des concepts actuellement décrits dans le cadre des approches globales en écologie. Ils compléteront et augmenteront les domaines d'applications en cours de développement en écologie statistique.

Dans le cadre de ce défi, la première année a pour but la mise en place d'un prototype de visualisation de ces graphes et la conception d'une proposition H2020 s'appuyant sur les laboratoires compétents dans chaque champ disciplinaire.

Grâce à un premier niveau d'interopérabilité, le projet «VIGI-GEEK» expérimentera la formation et la visualisation en temps réel de «cubes de données» mettant en relation par exemple des aspects sociologiques, économiques et écologiques de la biodiversité méditerranéenne et s'appuiera sur les nouvelles disciplines de l'informatique et des mathématiques (algorithme des graphes, fouille de données) pour proposer des solutions aux problématiques transdisciplinaires qui concernent directement la biodiversité. Le partenariat avec les astronomes du LAM permettra d'adapter les outils fréquemment utilisés dans leur discipline pour construire ces «cubes de données» avec l'aide du LIF qui a déjà participé à la réalisation d'outils similaires. Le partenariat avec le laboratoire LIPADE, avec lequel les échanges seront très soutenus la première année, a pour objectif de proposer de nouvelles méthodes de paramétrage de ces graphes multidisciplinaires, et de permettre leur évolution en temps réel, grâce à des bases de données distantes et interconnectées.

Méthodes

Les approches bi-, tri- ou transdisciplinaires s'appuient sur des «modules de Bases De Données interopérables». Ce concept impose aux participants non seulement de développer leurs métadonnées, mais aussi de favoriser un enrichissement des données *via* des «ontologies» communes. Ces qualifications permettent de concevoir des graphes de données couplant les disciplines et de les fouiller en s'appuyant sur les savoirs faire existants (Aix-Marseille Université, Université de Montpellier, CESAB, ...)

L'architecture des systèmes d'informations déployés pour les projets en cours d'élaboration doit être **décentralisée** et consiste en l'indexation, la qualification, la cartographie et l'interfaçage des données de l'environnement littoral et marin méditerranéen pour la recherche en écologie et l'aide à la gestion du patrimoine naturel.

Une organisation «modulaire» (permettant l'administration d'un type d'objet ou de données de manière indépendante par l'acteur le plus compétent) doit être préférée aux systèmes centralisés (dans ce cadre interdisciplinaire s'appuyant sur des systèmes d'observation à large échelle, chaque participant ne peut pas regrouper les données de toutes les disciplines. Les données qui serviront de modèle concerneront un habitat marin et un habitat terrestre communs pour toutes les disciplines. Ce type de méthode pourra se décliner sur de nombreux modèles environnementaux (habitats terrestre et marins, communautés d'animaux, assemblages d'espèces).

Un défi technique et scientifique relevé par plusieurs communautés

Le prototype exploratoire proposé dans «VIGI-GEEK» sollicite, en plus de celui des écologues, socio-écologues et économistes, le concours de mathématiciens, de physiciens, d'informaticiens, d'archivistes, d'astronomes, tous membres du consortium IndexMed afin d'adapter au mieux les outils dont ils se servent et de s'appuyer sur leur expérience/expertise.

Les verrous scientifiques sont liés à la gestion de données rendues complexes par

- leur volume, et la dynamique de leur mise à jour, des mises à jour des référentiels et des normes qui permettent de les administrer
- leur hétérogénéité et leur complexité intrinsèques, notamment liées au croisement de données biodiversité et métriques contextuelles,
- l'hétérogénéité des utilisateurs, des réseaux d'acteurs producteurs et leurs motivations à maintenir et alimenter leurs systèmes d'information.

L'efficacité des services répondant aux contraintes contradictoires (modèles des bases de données, optimisation des flux grâce à des services d'administration et des outils de paramétrages, gestion de droits d'utilisation en fonction des statuts) sera gérée par une indexation globale appuyée par des méthodes de qualifications définies et validées par les acteurs sous forme de communautés thématiques. Un soin particulier sera apporté aux aspects suivants :

- objectifs de chaque qualification, rapport coûts / avantages,
- inter-calibration de protocoles, d'observateurs, typologie adaptée,
- qualités des tatouages de jeux de données / systèmes de traçabilité (qualification / marquage) et rapports coûts / avantages.

Deux principes seront mis en œuvre dans le cadre du projet : (i) les systèmes répartis d'informations (e.g. non centralisés) et (ii) les processus d'adaptation des modèles de données afin de couvrir les usages présents et d'anticiper les besoins des usages futurs.

Des ateliers impliquant les acteurs compétents de chaque discipline seront organisés pour aborder et résoudre différentes problématiques:

Concernant le modèle systémique de production de la donnée :

- Une évaluation des besoins antagonistes (i) de normalisation ou de suivi des normes existantes, (ii) de diversification de la donnée sur la biodiversité, de ses méthodes d'acquisition et de ses formats et de leurs conséquences respectives sur l'efficacité des types de recherche associées,
- Le rapport entre données très homogènes, notamment les données «calculées» ou mesurées *versus* les données «d'interprétation», plus sujettes à variation liée à l'observateur, les méthodes de valorisation des producteurs de données et les moyens de pérenniser des systèmes d'observations de données «interprétées» (et donc coûteuses en temps / homme),
- Les méthodes d'agrégation et de fouille de données pour des usages secondaires et les scénarios envisageables pour leur exploitation à différentes échelles en fonction du type de donnée,

Concernant les services et usages de la donnée :

- La problématique du cycle de vie de la donnée concernant la «biodiversité»,
- La valeur de la donnée et les indicateurs d'utilisation de cette donnée, comme un nouveau critère d'évaluation des chercheurs travaillant sur la biodiversité,
- Le contexte juridique de la donnée sur la biodiversité issue de la recherche, au sein des organismes de recherche mais aussi pour toute utilisation secondaire, externe au contexte «recherche»,
- Les outils nécessaires à une véritable traçabilité de la donnée, quelque-soient les transformations subies par les données brutes (agrégation, moyenne, division, requalification...)

Ces ateliers devront se poursuivre les années suivantes dans le cadre de l'instruction de la réponse à appel à projet H2020 avec des collaborateurs de chaque discipline de différents pays européens. Il est pour cela nécessaire de prévoir des déplacements fréquents en Europe pour l'équipe qui pilote le projet. Il est envisagé de réaliser un séminaire fin Mai couplé avec le séminaire organisé par PREDON (Préservation des données scientifiques en réponse à l'appel à projet MASTODONS du CNRS) afin de créer des synergies entre les problématiques de conservation des données étudiées dans le cadre de PREDON et celles d'usages et de représentation de la donnée proposées dans le cadre de VIGI-GEEK. Le projet VIGI-GEEK utilisera comme modèle :

- En Ecologie, les données issues de programmes de recherche français (thèses Romain David, Laure Thierry, Leïta Tschanz - IMBE) et européens dans le domaine de l'écologie marine (DCSMM, CIGESMED, DESVOTES...) et l'écologie terrestre.
- En économie, le modèle de donnée est encore à construire ; parmi les données économiques, certaines seront collectées auprès organismes spécialisés (INSEE, IFREMER, IRD), d'autres seront produites directement par l'équipe à l'aide d'enquêtes et de questionnaires, d'autres enfin pourront être obtenues auprès d'autres chercheurs en SHS ayant travaillé sur des sujets proches et qui seront sollicités pour leur expertise.
- Les données socio-écologiques, recueillies dans le cadre d'un post doctorat (demande de financement en cours) concernent les représentations sociales, la perception et la valeur sociale (des produits et des services liés à ces milieux), les interactions Homme-milieux. Un des modèles utilisé sera l'habitat coralligène mais sera généralisable à d'autres modèles qui décriront la relation Homme-Littoral.

- Les données sciences participatives : un programme de science participative développé en sein du projet CIGESMED permettra de mettre en place un suivi d'habitats coralligènes à l'échelle Méditerranéenne et, donc, d'obtenir une grande quantité des données biologique et écologiques, associés à des données cartographiques. Cet aspect est développé dans le cadre d'un post doctorat (Giulia GATTI - IMBE).

Le développement d'un réseau qui intègre des chercheurs en sciences sociales et humaines, et ce à l'échelle du bassin méditerranéen est indispensable. En effet, la question de la compatibilité des modèles et des données issues de différentes disciplines est une question fondamentale car la construction des modèles et la mesure des données pertinentes au sein de chaque discipline obéissent à des logiques différentes. Les ateliers permettront de construire des protocoles de mise en compatibilité entre les modèles et données écologiques, économiques et sociales ainsi que d'identifier les actions prioritaires d'accès aux données à insérer dans les graphes.

Pilote du projet : Thierry Tatoni – CV court 2014

49 ans, Docteur en Sciences (Biologie des populations et écologie)
 Professeur 1ère classe, Université d'Aix-Marseille
 UMR IMBE, Aix-Marseille Université, campus St-jérôme, case 421, 13397 Marseille cedex 20
 04 91 28 84 26 – thierry.tatoni@imbe.fr

Principales responsabilités :

Directeur de l'Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (UMR CNRS / IRD / Aix-Marseille Université / Université d'Avignon et des Pays du Vaucluse).
 Chargé de mission par l'Institut d'Ecologie et d'Environnement (INEE) du CNRS pour coordonner le DIPEE (Dispositif de partenariat en écologie et environnement) d'Aix-Marseille.

Autres implications dans l'administration de la recherche

Commission Recherche de l'OSU Institut Pythéas
 Bureau de l'Ecole Doctorale "Sciences de l'Environnement"
 Comité de Pilotage de l'Observatoire Homme/Milieu « Bassin minier de Gardanne » (INEE, CNRS) et de l'OHM "Littoral méditerranéen"
 Commission spécialisée du département « territoires » de l'IRSTEA (vice-président).
 Président de l'instance d'évaluation des chercheurs du département "Territoires" de l'IRSTEA.
 Comité de Pilotage de l'Observatoire des Pratiques Interdisciplinaires (OPI) du LabexMed

Principales relations avec le monde industriel et socio-économique

Membre du comité de suivi de la Stratégie Globale de la Biodiversité en PACA et du SRCE.
 Membre de la Commission Environnement de Marseille Provence Métropole (MPM) et du Conseil Consultatif d'Orientation de l'Agence Régionale Pour l'Environnement (ARPE PACA)
 Co-responsable du Diplôme d'Etudes Supérieures Universitaires (DESU) « Biodiversité, Villes et Territoires », CRVH Aix-en-Provence / Aix-Marseille Université

Principales activités de rayonnement

Conseil Scientifique du Programme Interdisciplinaire de Recherche Ville et Environnement (PIRVE)
 Conseil scientifique du programme « Biodiversité et gestion forestière » (GIP Ecofor)
 Co rédacteur des recommandations de la CPU pour le projet de loi Biodiversité
 Conseil de programmation et d'évaluation scientifique du MUCEM
 Membre de plusieurs CS de PNR (président du PNR Luberon) et de Parc Nationaux (vice-président du PN Calanques), du CSRPN PACA, CORP de la fédération des PNR, du CO de l'ATEN

Thèmes de recherche

Analyse de la vulnérabilité écologique en région méditerranéenne par des approches relevant de l'écologie des communautés et de l'écologie du paysage.
 Etude des fonctionnalités écologiques et approche globale des services rendus par la biodiversité.
 Participation aux comités de pilotage de l'observatoire régional de la biodiversité et de SILENE ; implication dans la validation des indicateurs de biodiversité pour la SNB.

3 Publications (noms soulignés = doctorants sous la direction de T. TATONI)

- Baumberger T., Affre L., Torre F., Vidal E., Dumas P.J. & Tatoni T., 2012. Plant community changes as ecological indicator of seabird colonies' impacts on Mediterranean Islands. *Ecological Indicators*, 15, 76-84.
- Lizée M.H., Manel S., Mauffrey J.F., Tatoni T. & Deschamps-Cottin M., 2012. Matrix structure and patch isolation influences override the species-area relationship for urban butterfly communities. *Landscape Ecology*, 27, 159-169.
- Tschanz L., T. Tatoni, J.J. Brun, 2013. Social-Ecological Systems: towards a global approach of biodiversity observation in a Regional natural park of alpine territory. In Conference Volume "5th Symposium for Research in Protected Areas", 10 to 12 June 2013, Mittersill, pages 765 – 768.

Autres scientifiques Séniors :

- Dominique AMI : Maître de Conférence Aix-Marseille Université, au GREQAM : Economie de l'environnement, économétrie appliquée, méthodes d'évaluation non marchandes,
- Joachim CLAUDET : chargé de recherche CNRS basé au CRIOBE, à Perpignan, travaille actuellement sur la compréhension et la modélisation des systèmes socio-écologiques. Il coordonne actuellement un ERA-Net BiodivERSa et une ANR ainsi que plusieurs thésards et post-doc sur cette thématique. Il travaille à l'interface entre les sciences naturelles et les sciences humaines, intéressé tout particulièrement sur comment les changements environnementaux induisent par rétro-action des changements d'activités humaines,
- Jean-Pierre FERAL : Océanographie, écologie côtière, biologie des populations, systématique (DREM CNRS, pilote national du descripteur 4 de la DCSMM (réseaux trophiques marins), co-directeur de l'OHM littoral méditerranéen, Directeur du projet européen CIGESMED, membre du programme européen DEVOTES,
- Olivier GIMENEZ : statisticien et démographe, conservation des grands mammifères (terrestres / marins), directeur du GDR écologie statistique (<https://sites.google.com/site/gdrecostat/>) et directeur du département biodiversité et conservation du Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175, Montpellier).

Références

- Ben Amor I., Benbernou S., Ouziri M., Medjahed B., Malik Z. (2015). Discovering Best Teams for Data Leak Aware Crowdsourcing in Social Networks, A paraître dans ACM Transaction of the Web 2015.
- Benbernou S., Huang X., Ouziri M. (2014). Management of big RDF graph in open data: uncertain big data. Journées Big Data Mining and Visualization, Lille, 23-24 Juin 2014.
- Berti-Equille L. (2012). La qualité et la gouvernance des données : au service de la performance des entreprises. Traité IC2, Lavoisier, Cachan.
- Chignard S. (2012). L'open data, comprendre l'ouverture des données publiques. Fyp Éditions, Limoges.
- Conruyt N., Sébastien D., Cosadia S., Vignes-Lebbe R., Touraivane T. (2009). Moving from biodiversity information systems to biodiversity information services. In: Information and Communication Technologies for Biodiversity Conservation and Agriculture, L. Maurer and K. Tochtermann (Eds.). Shaker Verlag: Aachen.
- Féral J.-P. (2001). Concepts and methods for studying marine biodiversity, from gene to ecosystem. Océanis. Institut Océanographique, Paris 24 (4): 1-420 [1998], Institut Océanographique, Paris
- Gachet S., Véla E., Tatoni T. (2005). BASECO: a floristic and ecological database of Mediterranean French flora. Biodiversity and Conservation 14:1023-1034.
- Gatti G., Bianchi C.N., Parravicini V., Rovere A., Peirano A., Montefalcone M., Massa F., Morri C. (2015). Ecological change, sliding baselines and the importance of historical data: lessons from combining observational and quantitative data on a temperate reef over 70 years. PLoS ONE 10(2): e0118581.
- Heip C., Hummel H., Van Avesaath P., Appeltans W., Arvanitidis C., Aspden R., Austen M., Boero F., Bouma T.J., Boxshall G., Buchholz F., Crowe T., Delaney A., Deprez T., Emblow C., Féral J.P., Gasol J.M., Gooday A., Harder J., Ianora A., Kraberg A., Mackenzie B., Ojaveer H., Paterson D., Rumohr H., Schiedek D., Sokolowski A., Somerfield P., Sousa Pinto I., Vincx M., Weslawski J., Nash R. (2009). Marine Biodiversity and Ecosystem Functioning. Printbase, Dublin.
- Klyne G., Carroll J.J. (2004). Ressources Description Framework (RDF): concepts and abstract syntax. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>
- Lacombe R., Bertin P.H., Vauglin F., Vieillefosse A. (2011). Pour une politique ambitieuse des données publiques : les données publiques au service de l'innovation et de la transparence. Rapport à la Délégation aux usages de l'Internet, Ecole de Ponts Paris Tech.
- Michener W.K., Jones M.B. (2012). Ecoinformatics : supporting ecology as a data-intensive science. Trends in Ecology & Evolution, 27: 85-93.
- Nimis P.L., Vignes Lebbe R. (2010). Tools for identifying biodiversity: progress and problems. EUT, Trieste.
- Gimenez O., Buckland S.T., Morgan B.J.T., Bez N., Bertrand S., Choquet R., Dray S., Etienne M.P., Fewster R., Gosselin F., Mérigot B., Monestiez P., Morales J., Mortier F., Munoz F., Ovaskainen O., Pavoine S., Pradel R., Schurr F.M., Thomas L., Thuiller W., Trenkel V., de Valpine P., Rexstad E. (2014). Statistical ecology comes of age. Biology Letters 10: 20140698.
- Smith D., Szekely B. (2005). LSID Best Practices: A guide to deploying life science identifiers. IBM Developer Works.

Laboratoires et disciplines impliqués et leurs contributions respectives au projet (année 2015):

Laboratoire - Structure	Tutelle - disciplines	Nom	personne. mois
CEFE	INEE – écologie statistique	Olivier GIMENEZ	1
CRIOBE	INEE – systèmes socio-écologiques	Joachim CLAUDET	1
GBIF	NGO – Données sur la biodiversité	Anne Sophie ARCHAMBEAU, Régine VIGNES-LEBBE	0.5
IdGC	IN2P3 – Informatique, Cloud-computing, Grilles de calcul	Geneviève ROMIER, Vincent BRETON	0.5
IMBE	INEE – écologie terrestre et marine, Phylo-géographie, Méta-génomique, Taxonomie, Sciences participatives,	Didier AURELLE, Denise BELLAN SANTINI, Gérard BELLAN, Anne CHENUIL, Romain DAVID, Alrick DIAS, Jean-Pierre FERAL, Giulia GATTI,	15,5

	aide à la gestion (possibles extensions à la protéomique et la métabolomique)	Dorian GUILLEMAIN, Thierry TATONI, Laure THIERRY DE VILLE D'AVRAY, Leïta TSCHANZ, Jean VACELET	
GREQAM	EHESS, AMU, CNRS, Ecole Centrale Marseille – Econométrie	Dominique AMI, Charles LAI TONG, Velayoudom MARIMOUTOU	3
LAM	INSU – cubes de donnée, astronomie	Christian SURACE	2
Laboratoire LIPADE	INSB - Université Paris Descartes – Exploitation des graphes sémantiques, bases de données, linked open data, crowdsourcing, Informatique	Salima BENBERNOU, Mourad OUZIRI, Soror SAHRI	4
OSU Pytheas	INSU – Structuration des systèmes d'information, hardware	Cyrille BLANPAIN, Frédéric ZUBERER, Julien CHARPIN, Julien LECUBIN, Adrien MALGOYRE	3
LIF	Aix-Marseille Université (AMU) – Bases de données, fouille de données	Noël NOVELLI	0.5
Qatar Computing Research Institute	IRD (ESPACE DEV U228) and LIF – Data quality exploration in extra-large databases, Visual analytics Incremental and distributed data mining	Laure BERTI-EQUILLES (adviser)	0.1
TOTAL ETP			31.1 p.mois

IndexMed est un consortium créé par l'axe Gestion de la biodiversité et des espaces naturels de l'IMBE (Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale). Son objectif principal est de développer la culture des bases de données et leur utilisation efficace dans le milieu de la recherche en écologie et biodiversité. Il s'est étendu à plusieurs UMR de disciplines différentes (environnement, astronomie). Il doit permettre de répondre à des appels à projet dans le domaine des bases de données en écologie méditerranéenne en favorisant l'interdisciplinarité et les collaborations avec d'autres entités du CNRS. Les projets qui y seront développés doivent s'appuyer sur les différentes démarches nationales et internationales et promouvoir un travail partenarial international. Ce consortium doit notamment servir de relais aux réseaux et démarches en place nationalement et internationalement et proposer une réponse aux obligations européennes (Aarhus, INSPIRE,...) auxquels les laboratoires de recherche travaillant dans le domaine de l'environnement sont et seront de plus en plus soumis.

L'objectif à court terme d'IndexMed est de mettre en place une plateforme d'indexation des données sur la biodiversité méditerranéenne et de paramètres environnementaux ayant un intérêt pour la recherche. Cette indexation utilisera les outils et méthodes préconisés nationalement (SINP, MNHN/GBIF, RBDD), internationalement (MedOBIS, OBIS, GBIF, LifeWatch, GEOBON...) et s'appuiera sur les catalogues développés à ce niveau (IDCNP du SINP, Réseaux d'acteurs de la FRB). Le projet VIGI-GEEK propose un outil opérationnel de visualisation de données environnementales s'appuyant sur les résultats de travaux de recherche concernant l'indexation des données environnementales (référentiels du SPN, GBIF, EU-BON).

Cette demande doit permettre prioritairement de financer les déplacements nécessaires au développement de la partie « représentation des données » du consortium IndexMed.

Le système d'information est réparti, néanmoins, le projet demande de disposer d'une capacité minimale d'hébergement de données pour les partenaires ne disposant pas de ressources en propre, ainsi que de serveurs hébergeant les services pour la représentation de ces données. La partie « matériel » de cette demande permettra en outre l'équipement informatique de 3 personnes (développement et GIS), l'achat de documentation.

Contributions animation (centré sur la représentation de données)	Prix Unitaire	Montant
8 participations-homme à des groupes de travail nationaux	300 euros	2 400 euros
20 Frais de déplacement « rencontres d'acteurs »	200 euros	4 000 euros
2 formations techniques	2 850 euros	5 700 euros
Séminaire d'échange au bout d'un an (80 part/2jours)		8 500 euros
8 missions*homme dans d'autres pays européens (moyenne de 3 jours)	800 euros	6 400 euros
TOTAL ANIMATION (partie prioritaire)		27 000 euros

Matériel	Prix Unitaire	Montant
2 postes développement/visualisation	2000 euros	4 000 euros
2 serveurs HQ	6 000 euros	12 000 euros
2 NAS	5 000 euros	10 000 euros
2 périphériques mobiles	650 euros	1 300 euros
Petit matériel		700 euros
Documentation		1 000 euros
TOTAL MATERIEL		29 000 euros

Total prévisionnel : 56 000 euros si l'ensemble est accepté.

Logiciel : Utilisation exclusive de programmes open source pour les bases, les traitements de données, les SIG.