

Romain David¹, Jean-Pierre Féral¹, Anne-Sophie Archambeau², Nicolas Bailly³, Cyrille Blanpain⁴, Vincent Breton⁵, Anna Cohen-nabeiro⁶, Aurélie Delavaud⁶, Alrick Dias¹, Sophie Gachet¹, Robin Goffaux⁶, Dino Ienco⁷, Romain Julliard⁸, Julien Lecubin⁴, Grégoire Lois⁸, Sophie Pamerlon², Geneviève Romier⁵, Christian Surace⁹, Thierry Taton¹ et la communauté IndexM(e)d

Contacts : Romain DAVID, Jean-Pierre FERVAL, Dino IENCO, Thierry TATONI romain.david@imbe.fr



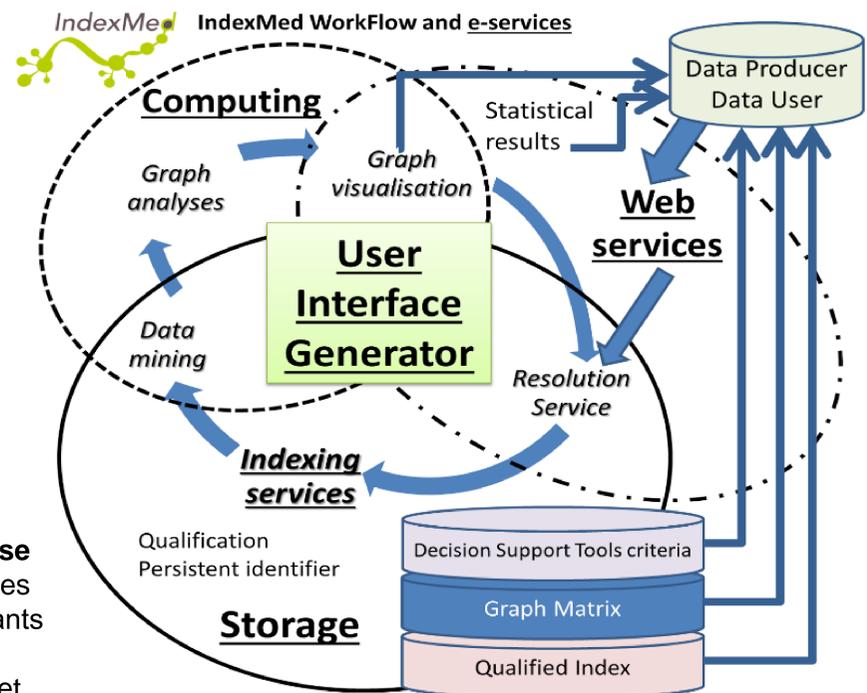
Le défi Comment fouiller des données hétérogènes non liées en écologie ?

Dans le domaine de l'écologie et de la biodiversité, l'augmentation des fréquences d'acquisition de données (missions de terrain, capteurs optiques, capteurs radar, systèmes de suivi de la qualité des eaux, recensement automatique ou semi-automatique des taxons, etc..) ont permis d'obtenir un volume de données considérable.

Récemment, les 3^{èmes} journées organisées en juin 2016 par le consortium IndexMed (indexmed2016.sciencesconf.org) ont mis en évidence le potentiel des approches basées sur les graphes, les lacunes en termes de compétences et d'expérience de la communauté des écologues pour adapter et utiliser ces méthodes.



Une première discussion a notamment permis de montrer que des **techniques à base de graphes** peuvent être adaptées pour modéliser plus efficacement les composantes d'interactions spatiales malgré l'hétérogénéité de ce type d'information. Les participants à ces journées (STIC et écologues) ont sollicité l'organisation de rencontres et le développement de collaborations entre la communauté des chercheurs en écologie et de la biodiversité et celle des chercheurs en sciences des données et STIC.



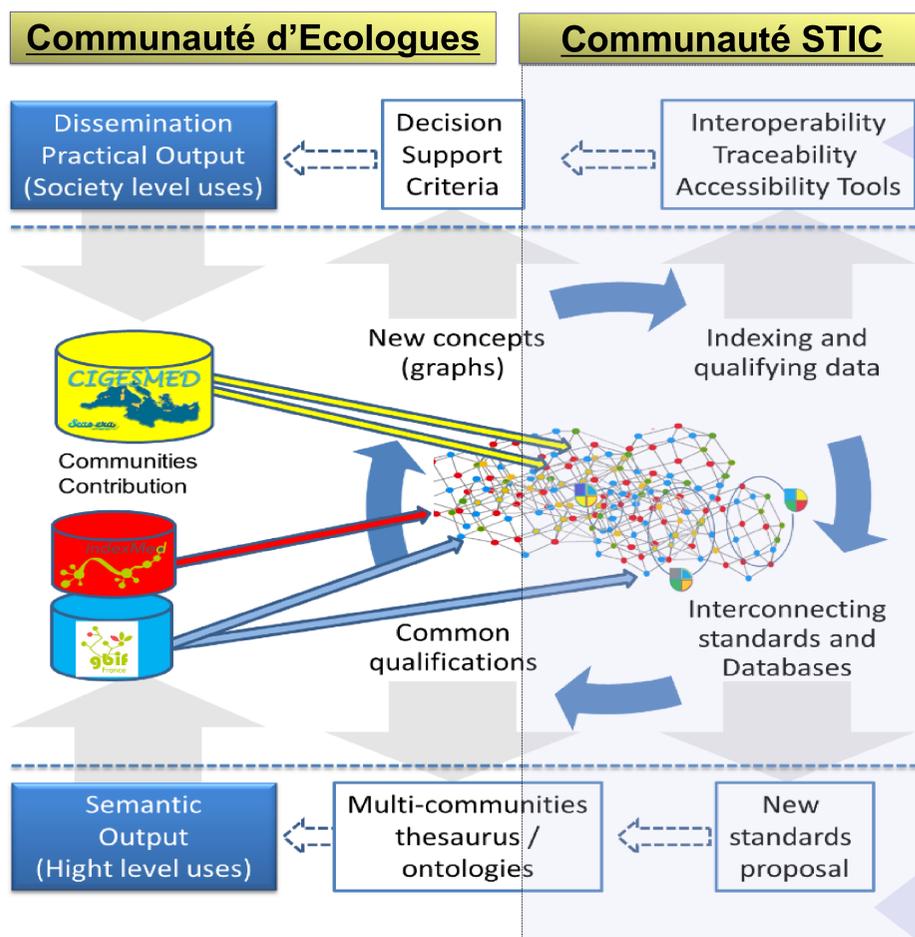
MANIFESTEZ votre intérêt pour ces rencontres!

romain.david@imbe.fr

Depuis le séminaire de juin, IndexMed devient **IndexMeed (Indexing for Mining Ecological and Environmental Data)**, un **nouveau consortium** ayant pour objectif d'indexer les données sur la biodiversité méditerranéenne issues de bases de données hétérogènes et distantes, afin de construire des graphes et de les analyser. Ces méthodes de fouille de données permettent de construire des **systèmes d'aide à la décision***

<http://www.indexmed.eu>

Une proposition d'action inter-communautaire :



OBJECTIFS

1/ La mise en place d'une dynamique d'échange entre des experts en écologie/biodiversité et des experts du domaine des STIC.

- Regrouper des représentants des deux champs disciplinaires permettra :
- de **formaliser des besoins** en terme d'analyses complexes de la part de la communauté écologie/biodiversité et
 - de **stimuler la recherche en STIC** afin de proposer des solutions plus adéquates pour l'analyse et la gestion des données écologiques dans le contexte du *Big Data* (prise en compte de la dimension temporelle et spatiale, données multi-échelles et hétérogènes).

En ce qui concerne les techniques et approches STIC étudiées et développées, la recherche sera dirigée vers des techniques de gestion et d'analyse de graphes qui puissent prendre en compte la complexité des données hétérogènes et notamment **passer à l'échelle sur des jeux de données volumineux** sans détériorer la qualité des résultats obtenus.

2/ La réalisation d'une première carte des compétences de laboratoires en informatique qui pourront apporter des outils méthodologiques ou des techniques algorithmiques adéquates pour l'analyse des données issue de l'écologie et de la biodiversité.

La définition de cette carte de compétences nous permettra, par la suite, de pouvoir détecter un ensemble de laboratoires (et des personnels associés) pour envisager des actions au plan international comme, par exemple, **une proposition de type M.R.S.E.I. (ANR)**. Cette MRESI permet de préparer une participation à des appels d'offres européens (BiodivERSA, SeasEra, H2020...) sur la thématique du Big Data, liés aux données de l'écologie et de la biodiversité.

Les verrous scientifiques à lever dans un projet de recherche interdisciplinaire issu de cette action devront être précisés par les experts du domaine des STIC, et concerneront notamment

- l'augmentation des fréquences et de la densité d'acquisition des observations (développement des méthodes de reconnaissance automatique et déploiement d'outils d'acquisition moins onéreux),
- la diversification des objets et des descripteurs d'objet intégrés dans les graphes,
- la normalisation des descripteurs de la donnée et les méthodes permettant d'intégrer les différents niveaux de qualité des données.

* DAVID et al 2016. David R., J.-P. Féral, A.-S. Archambeau, N. Bailly, C. Blanpain, V. Breton, A. De Jode, A. Delavaud, A. Dias, S. Gachet, D. Guillemain, J. Lecubin, G. Romier, C. Surace, L. Thierry de Ville d'Avray, C. Arvanitidis, A. Chenail, M.E. Çınar, D. Koutsoubas, S. Sartoretto, T. Taton, "IndexMed projects : new tools using the CIGESMED DataBase on Coralligenous for indexing, visualizing and data mining based on graphs". In : Sauvage S, Sánchez-Pérez J.-M., Rizzoli, A.E. (Eds.), *Proceedings of the 8th International Congress on Environmental Modelling and Software, Environmental modelling and software for supporting a sustainable future*, Vol. 3, pp.656-665, Toulouse, France, July 2016

(1) Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), CNRS, Aix Marseille Université, IRD, and Université d'Avignon, Station Marine d'Endoume, Chemin de la Batterie des Lions, 13007 Marseille, France. romain.david@imbe.fr, jean-pierre.feral@imbe.fr, alrick.dias@imbe.fr, sophie.gachet@imbe.fr, michelle.leydet@imbe.fr, thierry.tatoni@imbe.fr (2) GBIF-France, MNHN, CP 48, 43 rue Buffon, 75005 Paris, France. archambeau@gbif.fr, gbif@gbif.fr (3) HCMR/IMBBC Hellenic Centre for Marine Research, Institute of Marine Biology, Biotechnology & Aquaculture, LifeWatchGreece, Gouves, 71500 Heraklion, Crete, Greece. nbailly@hcmr.gr (4) Service informatique (SIP), OSU Pythéas, CNRS, Aix Marseille Université, 13007 Marseille, France. cyrille.blanpain@osupytheas.fr, julien.lecubin@osupytheas.fr (5) Institut des Grilles et du cloud (IDG) France Grilles % LPC Clermont-Ferrand 4 avenue Blaise Pascal 63178 Aubière Cedex bretton@idgrilles.fr, genevieve.romier@idgrilles.fr (6) FRB ECOSCOPE - Pôle pour l'observation et la diffusion des données de recherche sur la biodiversité, Fondation pour la Recherche sur la Biodiversité, 195 rue Saint-Jacques - 75005 Paris, France. aurelie.delavaud@fondationbiodiversite.fr, anna.cohen-nabeiro@fondationbiodiversite.fr, robin.goffaux@fondationbiodiversite.fr (7) UMR TETIS, Montpellier dino.ienco@teledetection.fr (8) CESCO - Centre d'Ecologie et des Sciences de la Conservation Muséum national d'Histoire naturelle romain.julliard@mnhn.fr, gregoire.lois@mnhn.fr (9) Laboratoire d'Astrophysique de Marseille (LAM), CNRS, Aix Marseille Université, rue Frédéric Joliot-Curie, 13013 Marseille, France. christian.surace@lam.fr