IndexMEED cases studies using ”Omics” data with graph theory
Romain David, Jean-Pierre Féral, Sophie Archambeau, Fanny Arnaud, David Auber, Nicolas Bailly, Loup Bernard, Laure Berti-Équille, Cyrille Blanpain, Vincent Breton, et al.

To cite this version:

HAL Id: hal-01761535
https://hal.archives-ouvertes.fr/hal-01761535
Submitted on 9 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution| 4.0 International License
IndexMEED cases studies using "Omics" data with graph theory


‡ Mediterranean Institute of Biodiversity and marine and terrestrial Ecology (IMBE), Aix Marseille Université/CNRS/IRD/Université d’Aix-Marseille, Station Marine d’Endoume, Marseille, France, Metropolitan
§ GBIF France, Paris, France, Metropolitan
¶ ENS Lyon, OHM Vallée du Rhône, Lyon, France, Metropolitan
# Hellenic Centre for Marine Research (HCMR), Gouves, Greece
ARCHIMEDE-UMR 7044, Université de Strasbourg CNRS, Strasbourg, France, Metropolitan
« IRD (ESPACE DEV U228) and LIF, Marseille, France, Metropolitan
SIP OSU Pythées, CNRS, Marseille, France, Metropolitan
IdGC – LPC, CNRS, France Grilles, Marseille, France, Metropolitan
MedMIB-UMR 7268, Université d’Aix-Marseille, Marseille, France, Metropolitan
ECOSCOPE, FRB, Paris, France, Metropolitan
MedMIB-UMR 7268, Université d’Aix-Marseille, Marseille, France, Metropolitan
ECOSCOPE, FRB, Marseille, France
ECOSCOPE, FRB, Marseille, France, Metropolitan
Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain
EDEn - Dept. of Architecture and Civil Eng., University of Bath, Bath, United Kingdom
P I3S (the laboratory of Computer Science of the University of Nice-Sophia Antipolis) and Inria, Nice, France, Metropolitan
A UMR TETIS, Montpellier, France, Metropolitan
Museum of Natural History (MNHN), Paris, France
CESCO - Centre d’Ecologie et des Sciences de la Conservation Muséum national d'Histoire naturelle, Paris, France
SIP OSU Pythées, Marseille, France, Metropolitan
European Grill Infrastructure, Amsterdam, Netherlands
Mediterranean Institute of Biodiversity and marine and terrestrial Ecology (IMBE), Aix Marseille Université/CNRS/IRD/Université d’Aix-Marseille, Marseille, France, Metropolitan
CESCO - Centre d’Ecologie et des Sciences de la Conservation Muséum national d'Histoire naturelle, Paris, France, Metropolitan
Université Bretagne Occidentale IUEM, Brest, France, Metropolitan
UMR 7268 ADES - Anthropologie Bioculturelle, Droit, Ethique et Santé Université d’Aix-Marseille / CNRS / EFS, Marseille, France, Metropolitan
LAM / CeSAM, Marseille, France, Metropolitan
UMR Espace DEV, Montpellier, Montpellier, France, Metropolitan

© David R et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Abstract

Data produced within marine and terrestrial biodiversity research projects that evaluate and monitor Good Environmental Status, have a high potential for use by stakeholders involved in environmental management. However, environmental data, especially in ecology, are not readily accessible to various users. The specific scientific goals and the logics of project organization and information gathering lead to a decentralized data distribution. In such a heterogeneous system across different organizations and data formats, it is difficult to efficiently harmonize the outputs. Few tools are available to assist. For instance standards and specific protocols can be applied to interconnect databases. Such semantic approaches greatly increase data interoperability.

This communication present the recent results and the consortium IndexMEED (Indexing for Mining Ecological and Environmental Data) activity that aims to build new approaches to investigate complex research questions, and support the emergence of new scientific hypotheses based on graph theory Auber et al. 2014). Current developments in data mining based on graphs, as well as the potential for relevant contributions to environmental research, particularly about strategic decision-making, and new ways of organizing data will be presented (David et al. 2015). In particular, the consortium makes decisions on how i) to analyze heterogeneous distributed data spread throughout different databases combining molecular and habitat characteristics data [3], ii) to create matches and incorporate some approximations, iii) to identify statistical relationships between observed data and the emergence of contextual patterns using a calculation library and distributed calculation center at the European level, iv) to encourage openness and sharing data while complying with the general principles of FAIR (Findable, Accessible, Interoperable, Re-usable and citable) in order to enhance data value and their utilization. IndexMEED participants are now exploring the ability of two scientific communities (ecology sensu lato and computer sciences) to work together, using several studies cases. The ECOSCOPE
project aims to meet the need to access structured and complementary omics-datasets to better understand biodiversity state and its dynamics. Indeed, the ECOSCOPE case study targets to visualize, through the graph approach, links between datasets and databases from genetics to ecosystems. Another case study, displaying anthropology fossils and omics on the same graph, will also be presented. DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) and CIGESMED (Coralligenous based Indicators to evaluate and monitor the “Good Environmental Status” of the MEDiterranean coastal water) European projects, conducted by IMBE, are focused on photo quadrats, cartography and omics data of the marine hard bottom in order to discover context patterns helpful to build decision support system building. Study case “65 Millions d’observateurs” French project is testing AskOmics to provide a graph-based querying interface using RDF (Resource Description Framework) and SPARQL technologies.

Scientific questions can be resolved by the new data mining approaches that offer new ways to investigate heterogeneous environmental data with graph mining (Muñoz et al. 2017). The uses of data from biodiversity research demonstrate the prototype functionalities (David et al. 2016) and introduce new perspectives to analyze environmental and societal responses including decision-making at large scale, both at the information system level and the observing system level than at the observed system level.

Keywords

Interdisciplinarity, Data qualification, Omics data, Graph, Thesaurus, Decision Support Tools

Presenting author

Romain David

Funding program

Labex DRIIHM (OHM Bassin Minier de Provence, OHM Vallée du Rhône, OHM Littoral méditerranéen), Fédération ECCOREV FR 3098, OSU Pythéas, and LabEx OT Med

Hosting institution

CESAB, ECOSCOPE, FRB, GBIF, IMBE, LAM,
References