



Research Project Title:

**PREGO - Text mining and data
integration to elucidate ecosystem
functioning: associating
organisms and environments with
biogeochemical and
anthropogenic impact processes**

Principal Investigator:
Evangelos Pafilis



African dust “veil” as captured in South Heraklion, Crete
March 2018

Popular Title:

PREGO (Process – environment – organism): building a knowledge network associating the different types of ecosystems with the microorganisms they contain, with the biological/environmental processes the latter are involved with

Scientific Field:

Environment and energy

Host Institution:

Greece



H.F.R.I.
Hellenic Foundation for
Research & Innovation

Process, Environment, Organism (PREGO), is a systems-biology approach to elucidate ecosystem function at the microbial dimension. Large-scale text-mining, data-mining, and network analysis are combined to this end.

To understand key functions of ecosystems, it is fundamental to study 'what biogeochemical processes' occur, in 'which environments' (where), and 'which organisms carry them out' (who).

Microbiology, molecular ecology and biodiversity address the above. Phylogenetic marker gene analyses, aim at deciphering the community composition of environmental samples. Sequence analysis pipelines assemble, cluster, and characterize environmental DNA, RNA, and protein sequences to infer community composition and to assign functions. Standards-compliant, expert-assigned, metadata annotations (like isolation source) provide valuable input too. Importantly, pieces of information missing from an experiment's data record metadata, or stored in fragmented computational analysis results, may be described in the accompanying literature. Thus, although valuable researcher input exists, it may just lie buried in free-text.

What-where-who associations, not observable previously, could become apparent once hidden evidence and fragmented data are all brought together. Thus, added value could be gained by combining the output of a range of existing computational analysis tools with expert-curated evidence, and automatically extracted facts of interest hidden in the vast body of biology literature. This is the motivation of PREGO, a one-stop-shop for researchers interested in searching and visually exploring such what-where-who associations.

The main aim of PREGO is to become a one-stop-shop for researchers interested in exploring ‘what process, which environment, which organism’ pieces of scientific evidence. PREGO’s audience ranges from the more specific microbial ecology, microbiology and biodiversity researcher audience, to broader pertinent third parties and stakeholders (e.g. environmental decision making authorities and biotechnology companies).

Unraveling which microorganisms occur in which environment and the processes they perform, can assist pure theoretical scientific studies to applied investigations. Organism and sampling environment prioritization for antibiotics discovery, pollution treatment, and pharmaceutical compound exploration are merely examples of pertinent scientific questions that can be supported.

Notably, PREGO mines associations from global literature and data repositories. Thus, it can assist researchers by providing them with novel perspectives in formulating hypotheses based on input from experiments conducted by scientists from all over the globe.

“



PREGO is based on previously successfully employed pieces of software and techniques; originally developed in a biomedical research context. In PREGO text mining, data integration, association statistics and graph theory, methods are combined to serve best microbiology, molecular ecology and biodiversity research. H.F.R.I. by funding PREGO allows more-than-a-decade years of experience to distil in a novelty-seeking project. H.F.R.I. via PREGO and similar projects assists: the uptake of the pertinent research methods by the Greek scientific community, the application of such methods in the study of key-Greek-interest types of ecosystems, the transfer of pertinent methodology and skills to the next generation of scientists.

Reoccurring calls for funding as well as the centrally orchestrated information and diffusion activities indicate that H.F.R.I. is not only listening carefully to the needs of the Greek scientific community, but also acts to implement means of long-term research support. This is a turning point

*The Principal Investigator,
Evangelos Pafilis*

Funding

Amount: **155,000 €**

Duration: **36 months**

Foundation: **H.F.R.I.**





H.F.R.I.
Hellenic Foundation for
Research & Innovation

CONTACT

127, Vasilissis Sofias Avenue
115 21 Athens, Greece
info@elidek.gr
www.elidek.gr



HELLENIC REPUBLIC
MINISTRY OF
DEVELOPMENT AND INVESTMENTS



GENERAL SECRETARIAT FOR
RESEARCH AND TECHNOLOGY