Available online http://amq.aiqua.it
ISSN (print): 2279-7327, ISSN (online): 2279-7335

Alpine and Mediterranean Quaternary, 32 (2), 2019, 167 - 184 https://doi.org/10.26382/AMQ.2019.11

BEyOND, A NEW TOOL FOR SAPROPEL S1 STUDIES IN THE MEDITERRANEAN SEA

Rubén Amezcua-Buendía [1], Claudia Diamantini [2], Domenico Potena [2], Alessandra Negri [1]

[1] Department of Life and Environmental Sciences. Polytechnic University of Marche, Ancona, Italy.
[2] Department of Information Engineering. Polytechnic University of Marche, Ancona, Italy.

Corresponding authors: R. Amezcua-Buendía <r.amezcua@pm.univpm.it> and D. Potena <d.potena@univpm.it>

ABSTRACT: Science advances and, with it, the storage of large amounts of data. The need to use this data efficiently, quickly and safely is possible thanks to the Big Data Analytics (BDA) that allows us to store and relate data in order to obtain new knowledge. In this paper we present and explain how we constructed the new database, "BEyOND", that provides a wide variety of organized and standardized paleoproxies relative to the past 20.000 years of Mediterranean Sea history. BEyOND makes available to all researchers the possibility to extract and analyze data. We focused on a specific interval of time corresponding to the deposition of the most recent sapropel (S1) and the potential uses offered by the tool.
BEyOND contains 126 sediment cores data from 79 scientific papers and a total of 1.678 different proxy related data that have been categorized in: geochemistry, isotopes, pollen, sediment grain size, coccolithophore, dinoflagellate and foraminifera.
Our work highlights the development of a new methodology to correlate data, including the cases where data regarding the precise age control for each core was missing.
It highlights as well the potential of using data analytics to extract hidden patterns and new knowledge also in the field of paleoceanography.

Keywords: Big data analytics, database, Mediterranean sea, sapropel, S1.

1. INTRODUCTION

The science of earth and environment has matured through two major phases and it is entering a third one. In the first phase, which ended two decades ago, science was oriented and focused on developing knowledge in geology, atmospheric chemistry, oceanography, ecosystems, and other fields of the Earth system. In the 1980s, the scientific community began to investigate these disciplines as interacting elements of a single system. In this second phase, we started to understand complex phenomena such as climate change, which links concepts from atmospheric and oceanic science, biology, and human behavior. In the emerging third phase, knowledge developed primarily for the purpose of scientific understanding, is being completed with the new knowledge created through data science and applications that allow us to acquire, manage, and make available data that helps us to take action and practical decisions (Dozier & Gail, 2009).

Nowadays, one of the most challenging tasks in Earth science is the possibility to use the vast amount of literature data by means of the new technology often referred to as the simple word BIG DATA. Big Data Analytics (BDA) allows to explore huge volumes of data, coming from heterogeneous data sources, in order to extract novel, useful and human understandable knowledge. To this end, BDA refers to methodologies coming from several research areas, like machine learning, data mining, statistics, artificial intelligence and database (Gandomi & Haider, 2015). The amount of data available nowadays is really huge and as by IBM, 90% of the data have been added in the last 2 years (IBM Marketing Cloud, 2017). Most scientific disciplines could be described as small data, or even data poor. Therefore, in these fields most of the experiments or studies are based on few data points. Many disciplines in geological sciences like palaeoceanography, paleoclimatology fall among them. However, although each single study is based on relatively few data, the whole set of studies potentially offers a considerable amount of data, provided that we are able to collect and correlate them. BDA technologies can provide us the toolsets to integrate these different data sources, organize them in a database, that supports data analysis.

In the last years many data have been stored and made accessible in several fields of research, including natural hazards (e.g., EM-DAT https://www.emdat.be/), chemistry (e.g., ChemSpider http://www.chemspider.com/), biology (e.g., GenBank https://www.ncbi.nlm.nih.gov/genbank/); as for the paleo studies domain, several databases are under construction and implementation and among them, the best known are PANGAEA (https://www.pangaea.de/), NOAA (https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets), IODP (http://www-odp.tamu.edu/database/) and NEOTOMA (https://www.neotomadb.org/).

The PANGAEA database contains data (geochemistry, isotopes, pollen, sedimentology, stratigraphy, microfossils) georeferenced in time and space regarding all the fields that comprise environmental sciences. The Janus database (IODP) stores and retrieves data collected on the drill ship JOIDES Resolution, while National Centers for Environmental Information (NCEI), that it's the world's largest provider of weather and climate data, provides the data to NOAA. Finally, NEOTOMA database is focused on the paleo data from the Pliocene to the Quaternary. This variety of databases potentially gives many benefits to researchers but implies the presence of some obstacles too. The availability of a wide variety and quantity of data provides researchers with new possibilities to compare and correlate their data, potentially without the need to perform new measurements. In addition, more accuracy is provided because of the possibility to choose, among many data, those that appear to be more meaningful for a given type of research. Nevertheless, there are several difficulties related to the heterogeneity of data sources (e.g., structured and related data in a database or unstructured data provided in the form of publication), the model of the data representation (e.g., XML, JSON, relational tables…), the forms of access (e.g., APIs, search forms, CVS downloads…), and the structure of the data (e.g., It is frequent that studies dealing with the same proxy, express the results in different units). This hinders the ability to correlate data and, therefore, the comparison of results. Looking towards the future it is hence critical that the scientific community becomes aware of the necessity to create standards that ensure a universal data format (McKay & Emile-Geay, 2016) that would increase the quality of the results and decrease the time of reworking them. According to Bell (2009) we are at a stage of development that is analogous to when the printing press was invented, and The Jim Gray's fourth paradigm, data intensive computing, is in an initial phase but cannot be neglected. Data analysis covers a whole range of activities throughout the workflow pipeline, including the use of databases (versus a collection of flat files that a database can access), analysis and modelling, and then data visualization. Jim Gray's recipe for designing a database for a given discipline is that it must be able to answer the 20 key questions that the scientist wants to ask. Gray called this new paradigm "*eScience*" and characterized it as "*IT meets scientists*". Whether you're a scientist or a technologist, this new data-intensive science is fascinating stuff (McFedries, 2011).

The most important problem when using a BDA approach in disciplines that use not only areal data, but sediment core data, is the need to uniform data and to refer it to events in time. This additional data type consists in the result of accumulation during time and therefore represents a record of processes occurring through time.

Project 418 is an ESSO-managed pilot project which addresses some of EarthCube's core activities envisioned for the EarthCube Cyberinfrastructure (Lingerfelt et al., 2018). These activities include Resource Registration, Data Discovery, and Data Access. Project 418 will serve as a pilot for the beginning point for these tasks, provide a foundation for future initiatives, as well as become a core component linking data facilities of EarthCube funded projects.

EarthCube supported LinkedEarth (Khider et al., 2018) project. It is a further development helping manifest a better future by creating an online platform that (1) enables the curation of a publicly-accessible database by paleoclimate experts themselves, and (2) fosters the development of community data standards, including an ontology. In turn, these developments enable cutting-edge data-analytic tools to be built and applied to a wider array of datasets than ever possible before, supporting more rigorous assessments of the magnitude and rates of pre-industrial climate change. More information at the LinkedEarth's site http://linked.earth show that there is a growing community sharing a key objective: the development of a community standard for paleoclimate data and metadata.

The LinkedEarth cyberinfrastructure also supported the development of the Past Global Changes past 2,000 years project (PAGES2k). In this contest, Emile-Geay et al. (2017) used a community-sourced database (PAGES2k temperature database) including a wide variety of proxies, like marine sediments, lake sediments, glacier ice or trees with the goal to reconstruct the global temperature of the Common Era.

Finally, Alberico et al. (2017), created the WDB-Paleo database that contains paleoclimatic proxies data of marine sediment core with the objective of studying the past climatic and environmental conditions of Mediterranean Sea.

The aim of the present paper is to describe the database BEyOND (Big palEo OceaN Data), that we have ideated and implemented using paleodata record of the last 20.000 years.

Method:

In order to design the database, we first defined our research questions as:

1. is it possible to correlate data at the minimum level of granularity (namely at sub-sample level), provided that sample range is heterogeneous in different studies?
2. is it possible to correlate proxies at different aggregation levels (e.g., cores from different areas, cores at different depths, proxies of the same core from different papers)?
3. is it possible to support an exploratory data analysis for Sapropel S1?

As a second step, we analyzed different sources, namely research papers and existing databases, to identify available data, their structure and format.

On the basis of collected requirements, BEyOND has been built following state-of-the-art design methodology, starting from a conceptual design.

Finally, the database has been populated by extracting, transforming and loading subsets of data from existing databases and, when needed, manually inserting those data published in research papers only. The databases will be described in Section 2.1.

To manage the different data that derive from different sources and standards, we used Data Analytics (DA). Ridge (2014) defines DA as any activity that involves applying analytical processes to data by using computer systems for the purpose of deriving insight from that data. DA allows us to explore huge amounts of data with the purpose of extracting new patterns or information that is comprehensible for further study. The objective of this work is to provide researchers, with BEyOND, a database which integrates heterogeneous paleodata from several sources and allows researchers to extract, correlate and aggregate data through the standard database query language SQL.

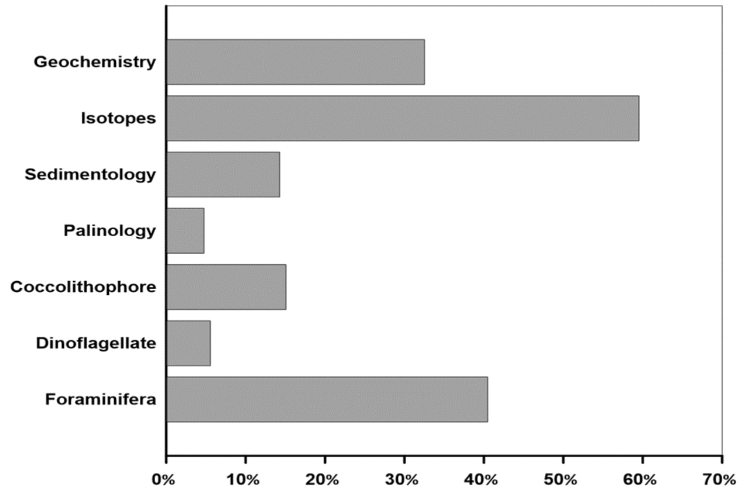

Fig. 1 - Percentage of cores in BEyOND, associated with the different proxies in the Mediterranean Sea.

2. DATABASE DESIGN

The database design is the data organization according to a database model, which determines its logical structure, that is, the manner in which data can be stored, organized and manipulated. To design our database, it was fundamental to know the data characteristics to be inserted and the goal to achieve with it. Yet, how databases are designed is another fundamental information. BEyOND mainly provides data of the Mediterranean Sea past 20.000 years history, but our study is focused on a specific interval of time corresponding to the deposition of the most recent sapropel (S1). Sapropels were defined by Kidd et al. (1978) as sharply-defined, dark-colored sedimentary layers with a C_{org} content >2 wt.% and a thickness >1 cm. The sapropel origin is still matter of discussion, centered on the (1) close relationships between sapropels and fluctuations of the Earth's orbital parameters (Hilgen, 1991; Lourens et al., 1996; Ziegler et al., 2010) and (2) the hydrographic (Rohling, 1991, 1994; Myers et al., 1998), biogeochemical and ecological changes (Sarmiento et al., 1988; Kemp et al., 1999; Sachs & Repeta, 1999; Negri & Giunta, 2001; Principato et al., 2003; Gallego-Torres et al., 2011) which led to sapropel deposition.

Over the last decade numerous studies (Incarbona et al., 2011; Bout-Roumazeilles et al., 2013; Triantaphyllou, 2014; Hennekam et al., 2014; Azrieli-Tal et al., 2014; Mojtahid et al., 2015; Tesi et al., 2017) provided more knowledge about sapropel deposition but the debate about its origin is still open. In fact, the scientific struggle between those invoking simple anoxia of the Mediterranean Sea (Olausson, 1961; Ryan, 1972; Cita et al., 1977; Thunell et al., 1977; Thunell, 1979; Rossignol-Strick, 1983, 1985) and those invoking increased export productivity causing anoxia (Calvert & Price, 1983; Pedersen & Calvert, 1990) is still unsolved, even if some papers point to a combination of effects (Myers et al., 1998; Martinez-Ruiz et al., 2000; Stratford et al., 2000; Rohling et al., 2015).

2.1 Data sources

The starting point for the design of BEyOND is the study of other already existing databases, and of the literature.

The databases used are:

- PANGAEA: it is one of the biggest digital data libraries for the earth system science and it has provided us with a large amount of data. All the data derive from marine sediment cores that are georeferenced in space (with latitude, longitude and depth below sea) and in time (geological age). PANGAEA provides a web interface for initial data search and different formats for data download.
- NOAA (National Oceanic and Atmospheric Administration): it contains current and historical data focused on the conditions of the oceans and atmosphere. The access of the data can be through traditional access methods as web-based or FTP (File Transfer Protocol) and they mainly provide the data in ".txt" or ".xls" format.
- Janus database: ODP (Ocean Drilling Program) and IODP (International Ocean Discovery Program) data are stored in the same database, Janus. This database is based on a relational data model and contains marine geoscience data collected onboard the drillship *JOIDES Resolution*. Data can be visualized through the Janus Web site and downloaded in ".txt" format.
- Global Pollen Database (GPD): it contains data of fossil and modern pollen records from natural archives where data are hosted by NOAA. Databases such as North American Pollen Databases (NAPD) or European Pollen Database (EPD) have been the precursors of this database.
- Neotoma Paleoecology Database: it is a multiproxy paleoecological database that covers the Pliocene-Quaternary, including modern microfossil samples, and that works like a centralized database with virtual constituent databases (NAPD, EPD, FAUNMAP) and which is implemented through Microsoft SQL Server. Data is accessible from Neotoma Explorer, an interactive web application for searching and downloading data in ".txt" format.

Of the above databases, we focused our analysis on the portion of data describing marine sediment cores.

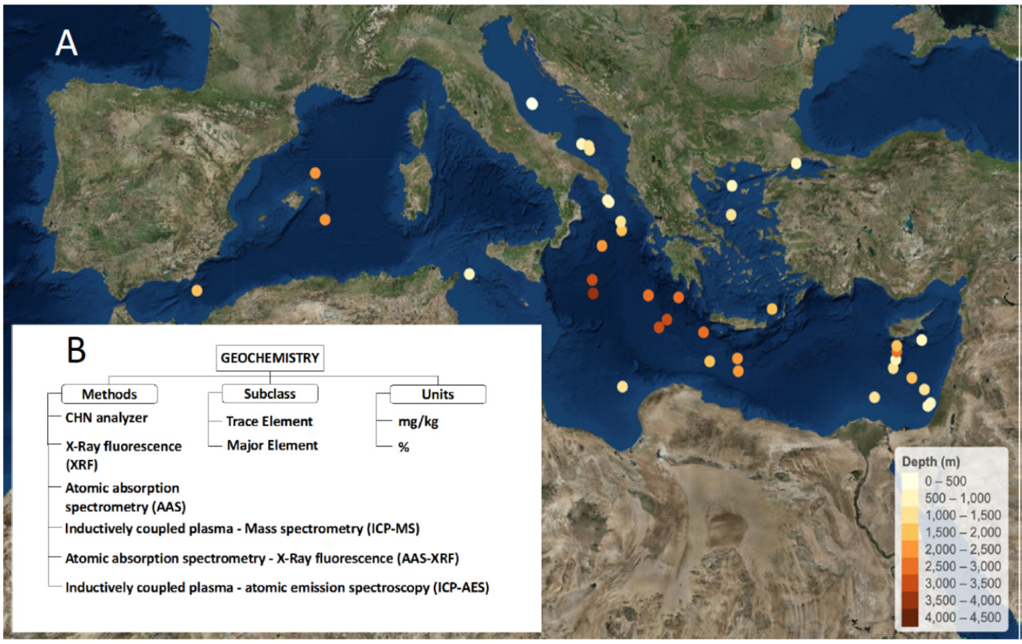

Fig. 2 - (A) Core associated with geochemistry data. (B) Diagram showing the geochemistry proxies classification used for the Database BEyOND.

As for the literature analysis, we started searching digital libraries for the following keywords: “sapropel”, “anoxia”, “Holocene”, “Mediterranean Sea” and “productivity”.

2.2 Categories of data

BEyOND has been organized around cores, which are geolocated and numbered in order to be identifiable. Cores are divided into samples (portion of the core between two discrete depth values) where several proxy measurements have been made in each sample.
The proxies inserted in the database are considered as a function of the category (e.g., geochemistry, sedimentology, isotopes…), but also of the method of analysis and measurement unit of the proxy itself.

In this regard, the first step has been the categorization of data in geochemistry, isotopes, sedimentology (sediment grain size and composition), palynology and micropaleontology (coccolithophore, dinoflagellate and foraminifera). Fig. 1 shows the distribution of proxies in

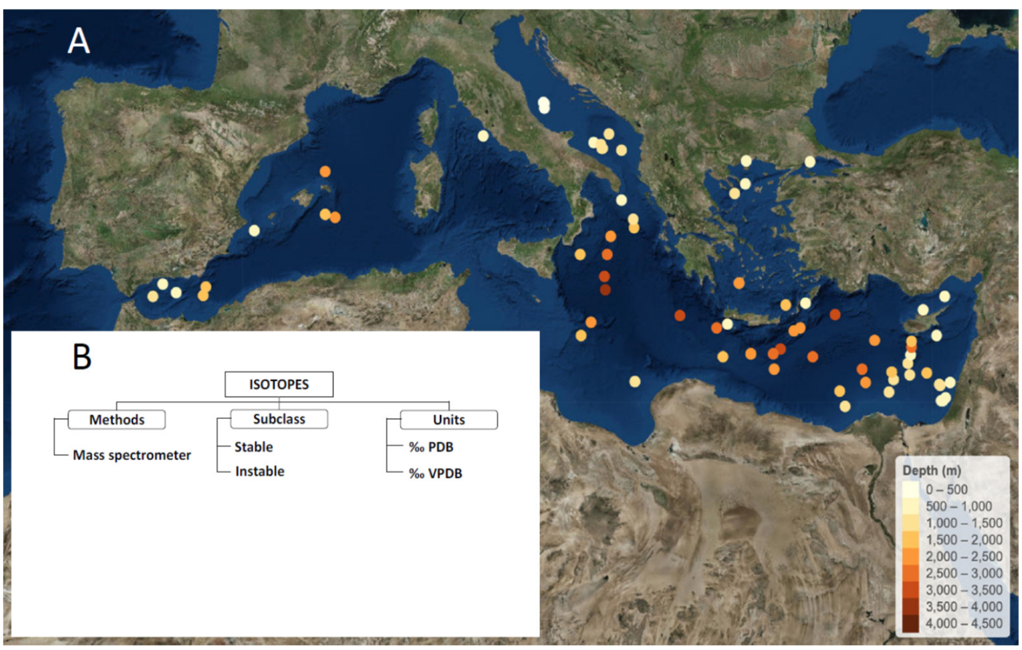

Fig. 3 - (A) Core associated with stable Isotopes data. (B) Diagram showing the stable isotope proxies classification used for the Database BEyOND.

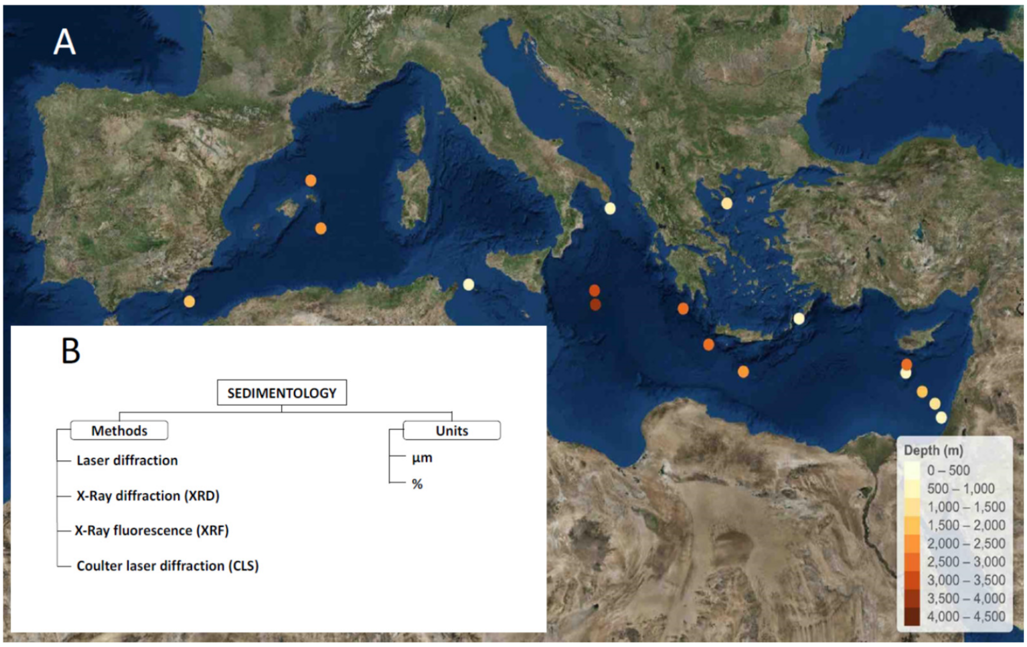

Fig. 4 - (A) Core associated with Sedimentology data. (B) Diagram showing the sedimentology classification used for the Database BEyOND.

the different cores that have been taken into account.

Some proxies, especially in geochemistry, show a issues related to the analytical method: in fact different methods can be used for the analysis (e.g., barium has been obtained by Inductively Coupled Plasma - Atomic Emission Spectroscopy (ICP-AES), Inductively Coupled Plasma - Mass Spectrometry (ICP-MS) or X-Ray Fluorescence (XRF). Another problem is the measurement unit of some proxies: in fact, even when the methodology is the same, results can be expressed in different units (e.g., Aluminum can be expressed in mg/kg or as a percentage of total weight).

Geochemistry:

In general, the geochemical proxies are very useful in the study of palaeoceanography and paleoclimate. In fact, the chemical composition of the marine sediment provides information about past climates, ocean circulation, sea-floor conditions and changes in the sediment over time (Calvert & Pedersen, 2007). Trace elements

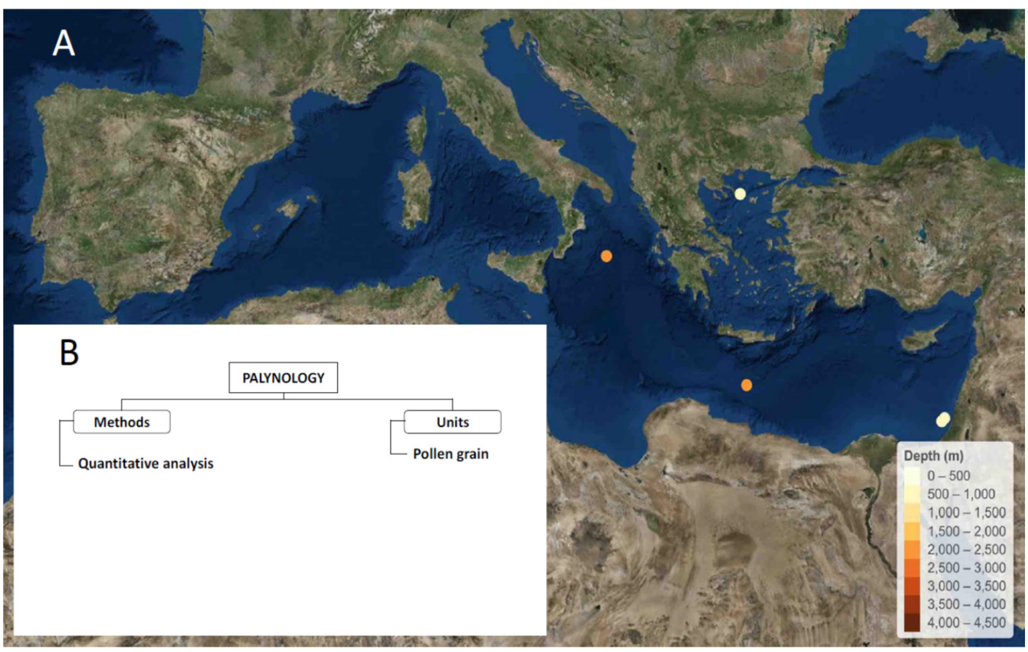

Fig. 5 - (A) Core associated with Palynology data. (B) Diagram showing the palynology proxies classification used for the Database BEyOND.

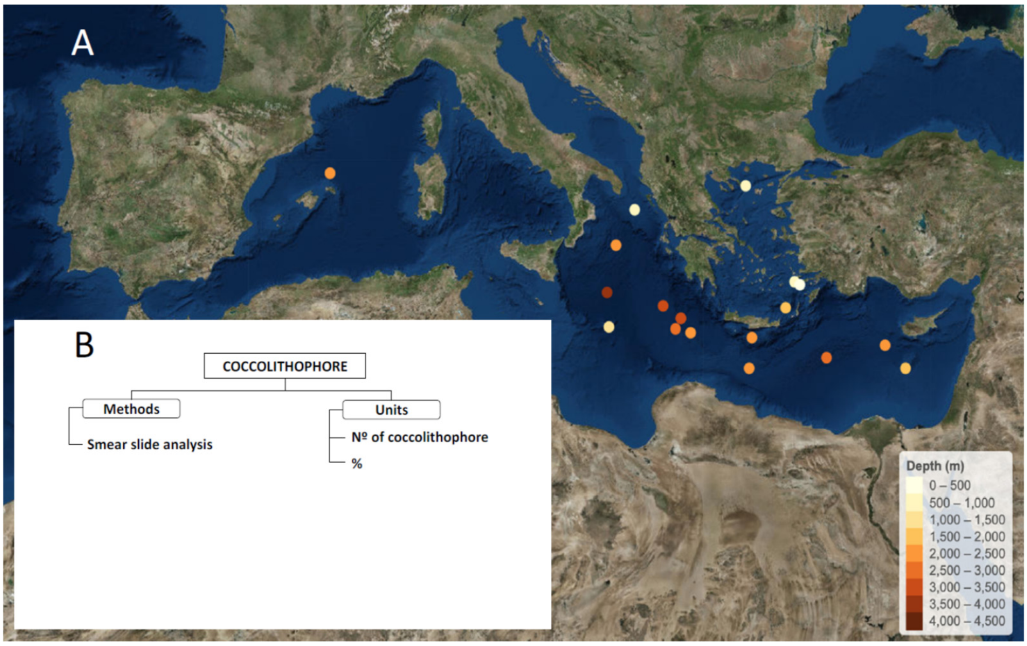

Fig. 6 - (A) Core associated with Coccolithophore data. (B) Diagram showing the coccolithophore proxies classification used for the Database BEyOND.

provide clues about the aeolian or river input to marine sediments (Jimenez-Espejo et al., 2007; Frigola et al., 2008) while other elements such as Barium, Aluminum and Phosphorus allow to reconstruct marine paleoproductivity (Möbius et al., 2010; Martinez-Ruiz et al., 2015). Redox-sensitive elements are used to recognize bottom water conditions (Martinez-Ruiz et al., 2000; Tribovillard et al., 2006; Tachikawa et al., 2015; Tesi et al., 2017). Fig. 2.A shows the position of cores providing geochemical data. A total of 41 cores are located in deep environment of the Mediterranean Sea. Fig. 2.B shows how we categorized the geochemical data according mainly to the measurement methods (ICP-AES, ICP-MS, AAS, XRF, AAS-XRF and CHN Analyzer) and the unit (mg/kg and percentage) in which the proxy data have been expressed.

Isotopes:

Stable oxygen and carbon isotope records of benthic and planktonic foraminifera are a powerful tool for

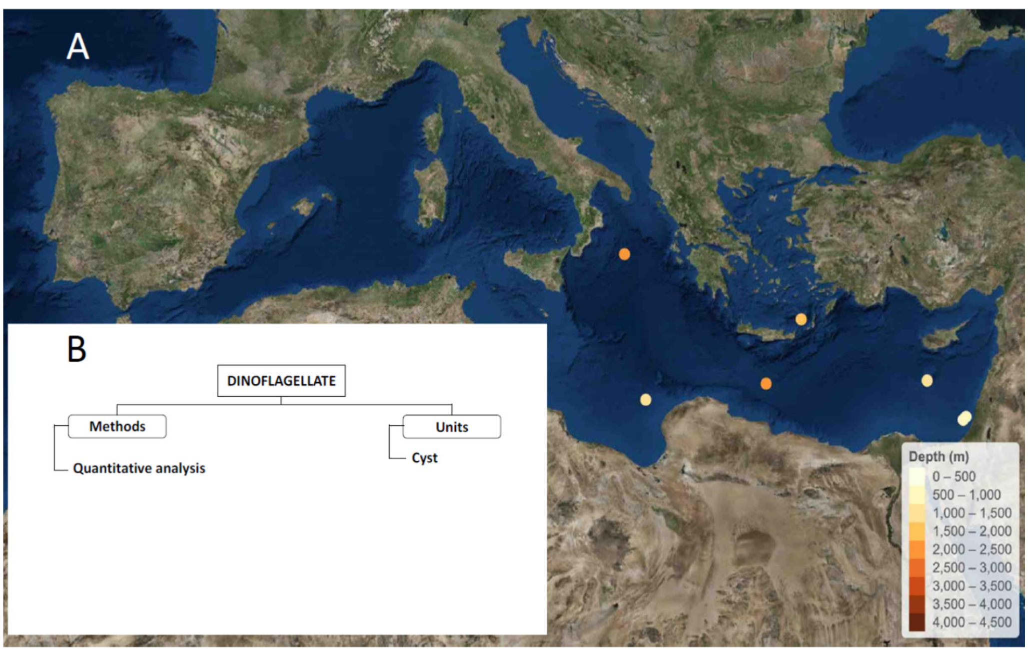

Fig. 7 - (A) Core associated with Dinoflagellate data. (B) Diagram showing the dinoflagellate proxies classification used for the Database BEyOND.

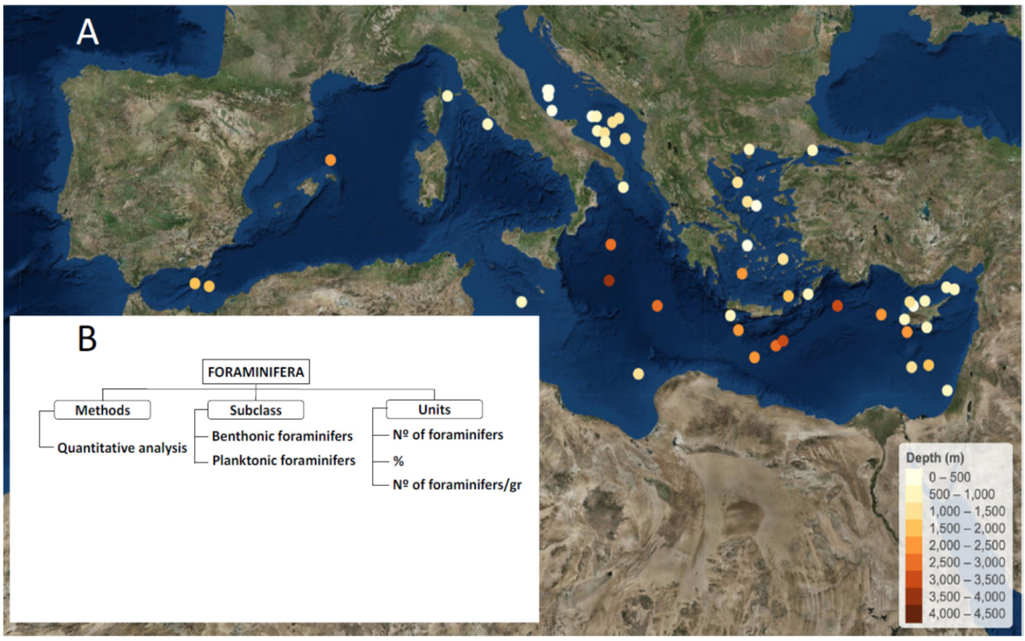

Fig. 8 - (A) Core associated with Foraminifera data. (B) Diagram showing the foraminifera proxies classification used for the Database BEyOND.

palaeoceanography and paleoclimate studies. The $\delta^{18}O$ signal shows fluctuations of global ice volume (Grazzini & Pierre, 1991; Jimenez-Espejo et al., 2008), changes in sea salinity (Fontugne et al., 1989) and the freshwater input (Dubois-Dauphin et al., 2017). The $\delta^{13}C$ signal provides information about deep water properties and its circulation, organic matter fluxes and carbon cycling of the oceans (Woodruff & Savin, 1985; Zahn et al., 1986; Sarnthein et al., 1994; Mackensen et al., 2001; Kuhnt et al., 2008). Fig. 3.A describes the position of the 75 cores providing isotopes data and Fig. 3.B the categorization of the data that have been measured by a single method (mass spectrometer) and expressed in two different units (‰ PDB and ‰VPDB).

Sediment grain size and composition:

Sedimentology helps to comprehend the systems that control paleoclimatic and palaeohydrological changes. The clay mineral study contributes to understanding processes related to oceanic circulation, wind directions,

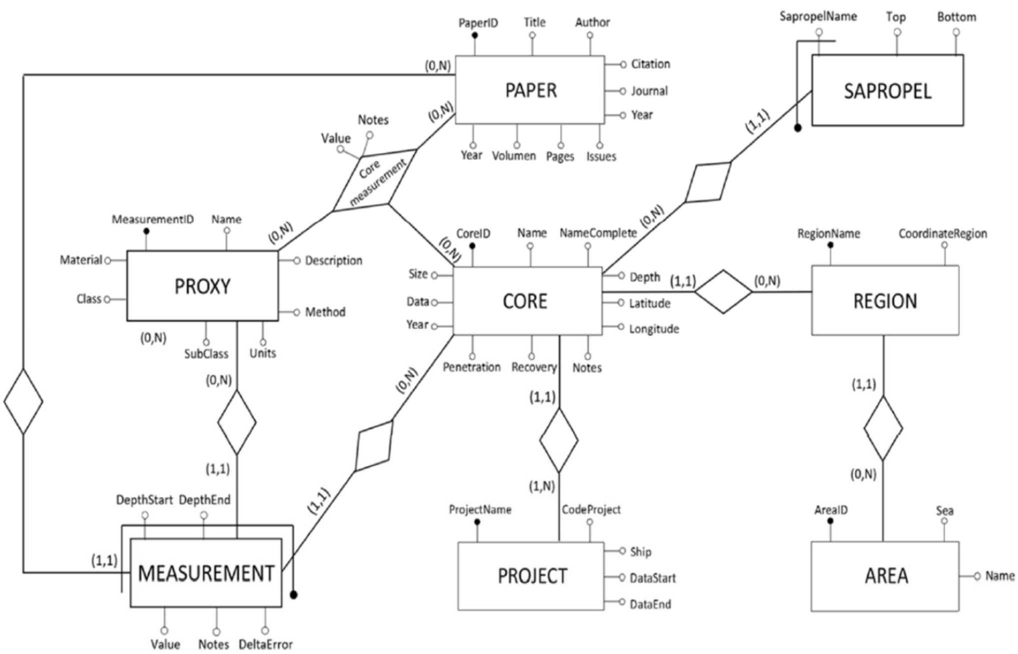

Fig. 9 - E-R Schema of BEyOND.

precipitations just like the dynamics river inputs (Ehrmann et al., 2007, 2016). In addition, the grain size distribution, which is conditioned by physical processes, helps to know the transport processes and source areas (Bout-Roumazeilles et al., 2013; Ehrmann et al., 2007). Fig. 4.A shows the positions of the 18 cores associated with the grain size and composition of the sediments, while in Fig. 4.B the methods utilized for these cores (Laser diffraction, CLS, XRF and XRD) and the measurement units (µm and percentage) are reported.

Palynology:

Pollen records in marine sediments provide accurate information on paleovegetation, palaeohydrology and paleoclimate on a local and global level. Pollen is studied in order to reconstruct changes in temperature and precipitation occurred in the past (Kotthoff et al., 2008; Langgut et al., 2011) but is also used to investigate processes such as productivity (Hennekam et al., 2015). As is evident from Fig. 5.A, however very few data (only 6 cores) are available for the investigated interval. Fig. 5.B shows the only methodology used for palynology, that correspond to quantitative analysis (counts of pollen grains) (Parker & Arnold, 1999), and the unit of measurement (pollen grain).

Coccolithophores:

Coccolithophores are planktonic microalgae living in almost all oceanic areas. According to the specialists, changes in coccolith abundances permits to assess primary productivity fluctuations and, based on these, water column modifications (Negri & Giunta, 2001; Principato et al., 2003). Furthermore, a relation between reworked specimens eroded from continental sediment brought to the ocean by the river runoff is hypothesized by Negri et al., 1999 which evidence increased river discharge during sapropel deposition. Finally, the presence or absence of selected species allows the recognition of oligotrophic or eutrophic environments (Crudeli et al., 2006). Fig. 6.A shows the location of the cores (19) reporting coccolithophore data. Fig. 6.B shows the methodology of analysis used in coccolithophore studies, that is quantitative on smear slide (Bown & Young, 1998), whose values are expressed in number of coccolithophore counted or percentage.

Dinoflagellate:

Dinoflagellates are a group of unicellular protists that can be identified using the light microscope, and are (usually) recognized by their golden-brown plastids, assimilative cell with indented waist, distinctive swimming patterns, and relatively large nucleus that contains visible chromosomes (Carty & Parrow, 2015). Resting Cyst association is closely related with parameters such as temperature, salinity and nutrient concentration (Meier et al., 2002, 2004). Also, they can give information about oxidation fronts, productivity and bottom oxygen conditions, and through this, on bottom water circulation (Zonneveld et al., 2001). Dinoflagellate data are scarce as shown Fig. 7.A (7 cores). Fig. 7.B displays the only method (quantitative analysis) and unit of measurement (Cyst) inserted in database.

Foraminifera:

Foraminifera are single-celled amoeboid protist comprising the order Foraminiferida (or Foraminifera of supergroup Rhizaria), characterized by reticulating pseudopods and typically a shell. They may be planktonic or benthic, are mainly marine and are found in all marine environments.

The study of foraminifera contributes to knowledge of several processes in palaeoceanography, paleoclimate, paleoproductivity and palaeohydrology. The abundance, assemblage and size of planktonic foraminifera are the result of increased or decreased productivity due to hydrological conditions (e.g., water column stratification, turbidity, salinity), climate (e.g., river input, temperature) (Principato et al., 2003; Triantaphyllou et al., 2010; Mojtahid et al., 2015; Tesi et al., 2017). Benthic foraminifera are indicators of the deep water conditions as oxygen content in the interstitial and bottom water as well as the food supply and availability (Schmiedl et al., 1998; Kuhnt et al., 2007; Abu-Zied et al., 2008; Gupta et al., 2008). The locations of the 51 cores that have studied the foraminifera data are presented in Fig. 8.A. In addition, Fig. 8.B shows the two subclasses (planktonic or benthonic). In both subclasses foraminifera are measured by "quantitative analysis" and expressed in three different measurement units (number of foraminifers, percentage or number of foraminifers/gr).

2.3 Design Methodology

The conceptual modelling is a very important phase in designing a quality database (Elmasri & Navathe, 2010). It provides an abstract representation of the application domain, focusing on the description of the major concepts involved and of their relationships, without being misled by database implementation details.

In particular, Fig. 9 shows the conceptual schema of the BEyOND database expressed in the Entity-Relationship (E-R) model. In the E-R graphical notation, rectangles represent entities, while diamonds represent relationships between entities. Entities are conceptual abstractions of a class of real world objects, which in turn are named instances (or occurrences) of the entity (e.g., the entity "core" has the real cores as instances). The description of all entities is reported in Appendix A. Instances of relationship are n-ples of entity instances that are someway related to each other (e.g., an instance of "core measurement" is a triple whose components are a particular measure or proxy, the particular core where the proxy has been taken, and the particular paper where the measurement has been published). Circles associated to entities or relationships represent attributes, e.g., relevant characteristics we want to record about them (e.g., a core is described by its name, latitude and longitude, the date of sampling, etc.). Finally, the black circle in each entity represents the identifier, e.g., an attribute whose value uniquely characterizes each entity occurrence (e.g., the CoreID). In general, an identifier can be defined through a set of attributes. In some cases, an identifier can not be built using only attributes of the entity, but has to include identifiers of other entities that are linked through relationships. This

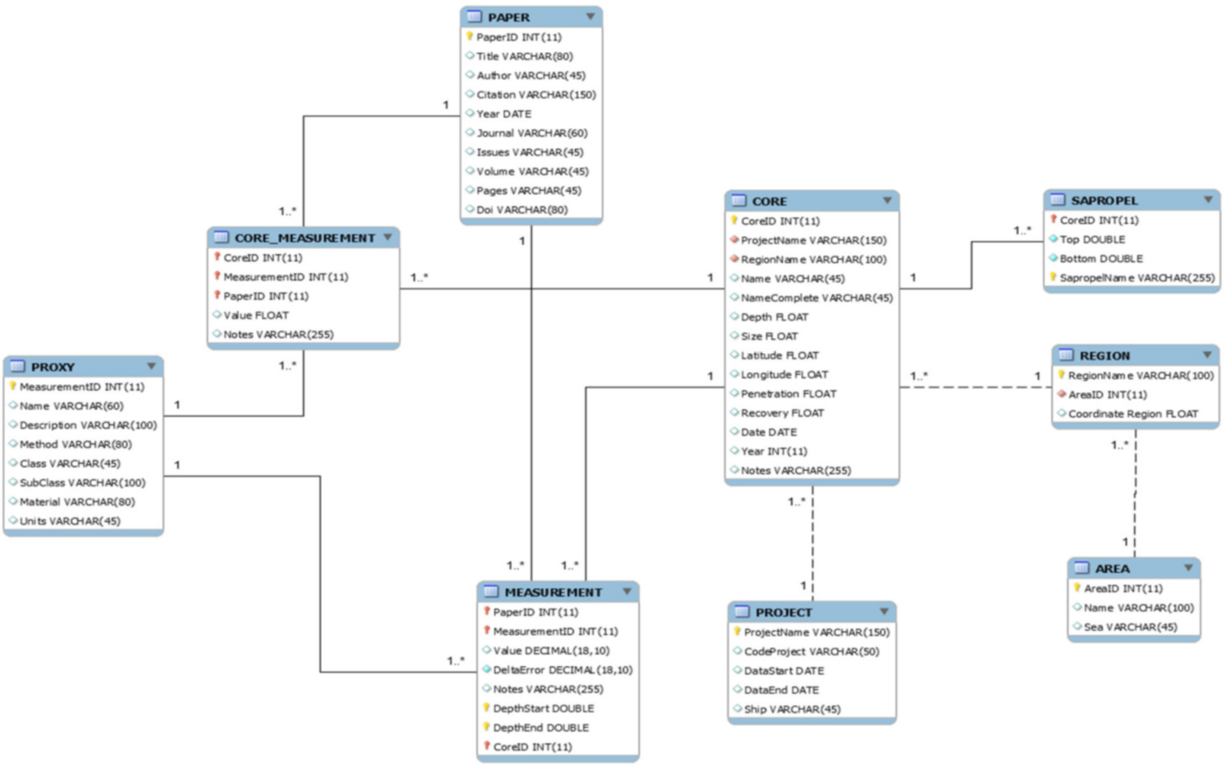

Fig. 10 - The relational schema of BEyOND. The key symbols are a primary key, while attributes with a red symbol are foreign keys.

latter is the case of the entity “measurement”, which represents the value of a given proxy measured between two depth values of a given core, as reported in a specific research paper. Hence, the identifier of “measurement” is formed by two internal attributes representing the start and the end depth of the sample (DepthStart and DepthEnd respectively) and the identifiers of “paper”, “proxy” and “core” entities respectively. In turn, an instance of “core” is characterized by both the investigation “project” in which the core has been collected, and the geographical “region” it belongs to. Each region is part of an “area”, namely a given sea and zone (e.g., the Ionian Sea in the Eastern Mediterranean area).

It is worth noting that the concept of “measurement” is central in the BEyOND schema, since it represents the finest measurement we can store in the database. Using relationships, we have defined between “measurement” and other entities, we are able to extract information which is not explicitly stored in the database. For instance, we are able to derive proxies values for a *virtual core*, namely the aggregation of all cores in a given region or area. Furthermore, proxies values can be aggregated on the basis of the core on which they were measured, the category of proxy, and so on. Moreover, we can also compare and correlate measurements related to the same sample but reported in different papers, and to the same sapropel but in different cores. Examples of application of analysis enabled by BEyOND are described in Section 3.

We like to note that the conceptual methodology adopted allows to obtain a compact, normalized, non-redundant schema that still maintains the same expressive capacity and information richness of existing databases. As an example, the problem of the variable "age": pre-existing databases can only correlate data where the age variable has been calculated, while in BEyOND we can insert data with and without the age variable because we use a different method to correlate them (see section 3).

2.4 Implementation

BEyOND has been implemented using the relational model. The adopted database management system is MySQL. The schema reported in Fig. 10 has been obtained from the E-R schema of Fig. 9 by following the rules for E-R to relational schema translation.

The database BEyOND is publicly available at http://beyond.dii.univpm.it/. It is focused on paleoclimatology and palaeoceanography studies where all data have been obtained from marine sediment cores. Whenever possible, we extracted and transformed data coming from other existing databases (see Section 2.1) and loaded them in BEyOND. Other data have been manually inserted starting from published papers, sometimes directly asking the authors to provide excel files.

Currently, we have inserted data from:

- 79 papers;
- 126 cores (see Fig. 11);
- 1.678 different proxy related data that correspond to all

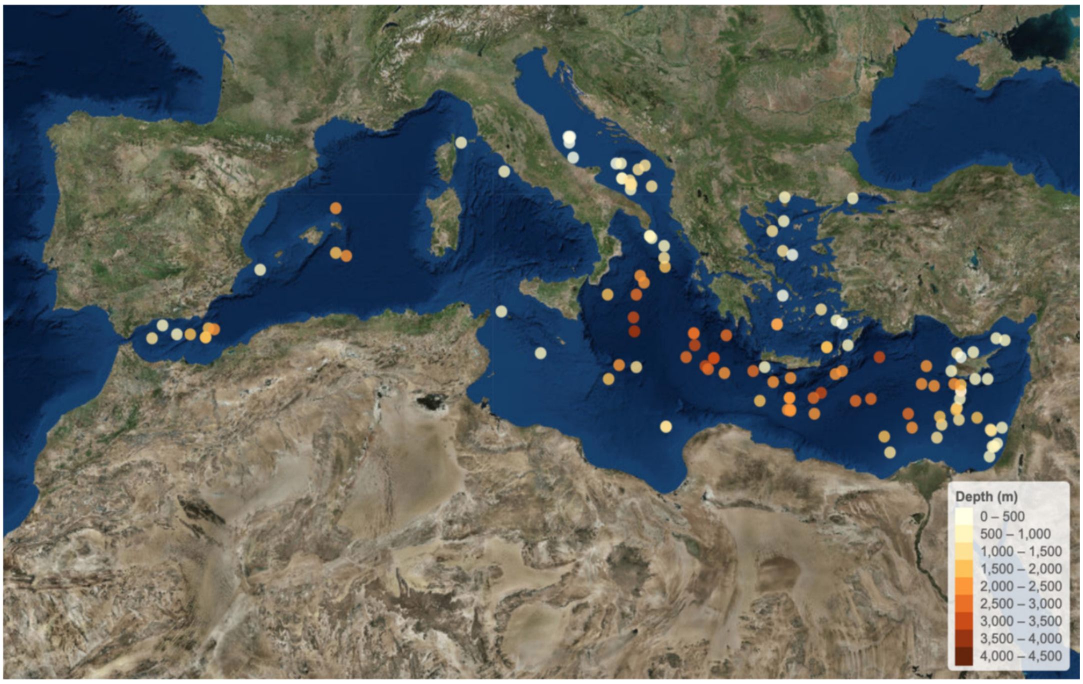

Fig. 11 - Map of the Mediterranean Sea with the locations and depths of cores used in BEyOND.

the different ways found in literature used to express the same proxy. See Tab. 1 that clarifies this concept: e.g., the Aluminum element is classified into 10 different ways because authors calculated and expressed the proxy with different methods and units);

- 283.135 proxy measurements.

The reasons for the creation of a new database are: (1) most of the data are available in repositories, like PANGAEA or NOAA, where data can be downloaded as tab-delimited texts but no functionality is provided to query the repository nor to combine data coming from different sources; (2) in literature, all databases are organized according to core-age where all the data are expressed in function to age. In BEyOND, age is represented as any other proxy, through the concept of "measurement" (see section 2.3), allowing us a greater flexibility. Indeed, we are able to store proxies' values of a (sample of a) core even if the age is not provided.

3. APPLICATIONS

To demonstrate the impact of BEyOND, in this section we show a set of possible applications to the reconstruction of a combined stack relative to a proxy in the S1 interval, supported by advanced analyses based on a rich set of integrated data.

In fact, to explain processes across a certain area during a certain interval of time we cannot focus on a single field of research (e.g., geochemistry) but we must relate different studies in order to complement and strengthen different discoveries. For this reason, it is of basic importance that every researcher can access and use the data, but also contribute populating the database with his/her data.

Furthermore, the creation of the database, allows us to use data analytics, to access and analyze data with the purpose of gaining and discovering useful patterns and trends in large data sets (Lussier & Hendon, 2016). We have managed to correlate cores located at different depths, presenting different thicknesses, that is to say, different sedimentation rates across the Mediterranean Sea. These correlations have been carried out by SQL queries. We have standardized the depth values of the cores on the basis of the depth values of the sapropel lithology using the following formulas:

$$Standardized\ Start = \frac{DepthStart - TopS1}{BottomS1 - TopS1}$$

$$Standardized\ End = \frac{DepthEnd - TopS1}{BottomS1 - TopS1}$$

Where DepthStart is the top part of the sample (usually of centimetric thickness) analyzed measured starting from the top of the core. Analogously, DepthEnd is the base of the sample analyzed. TopS1 is the top of the sapropel, while BottomS1 is its base.

Tab. 2 shows the new depths (Standardized Start and Standardized End) obtained with the use of these formulas. Then, the proxy values were averaged in the same standardized depth range.

Let C1 and C2 be two different cores, let $Bottom_1S_1$ and Top_1S_1 be the bottom and the top depths measured for S1 of the core Ci;

Let V1(x,y) and V2(x,y) be two functions returning the values of the same proxy measured between depths "x" and "y" on C1 and C2 respectively;

Let D1 and D2 be the sets of samples depths of C1 and C2 respectively, and let "m" and "p" be the cardinality of D1 and D2 respectively;

Name	Description	Method	Class	SubClass	Material	Units
Al	Aluminum	Energy dispersive polarization X-ray fluorescence spectrometer (EDP-XRF)	Geochemistry	Major Element		mg/kg
Al	Aluminum	Atomic absorption spectrometry (AAS)	Geochemistry	Major Element		%
Al	Aluminum	Atomic absorption spectrometry (AAS)	Geochemistry	Major Element		mg/kg
Al	Aluminum	Atomic absorption spectrometry - X-Ray fluorescence (AAS-XRF)	Geochemistry	Major Element		%
Al	Aluminum	Inductively coupled plasma - atomic emission spectroscopy (ICP-AES)	Geochemistry	Major Element		%
Al	Aluminum	Inductively coupled plasma - atomic emission spectroscopy (ICP-AES)	Geochemistry	Major Element		mg/kg
Al	Aluminum	Inductively coupled plasma - Mass spectrometry (ICP-MS)	Geochemistry	Major Element		mg/kg
Al	Aluminum	X-Ray fluorescence (XRF)	Geochemistry	Major Element		counts
Al	Aluminum	X-Ray fluorescence (XRF)	Geochemistry	Major Element		mg/kg
Al	Aluminum	X-Ray fluorescence (XRF-Mo tube)	Geochemistry	Major Element		%

Tab 1. Example of the different ways to express the aluminum proxy in BEyOND.

CoreID	DepthStart (m)	DepthEnd (m)	Proxy Value	Top S1 (m)	Bottom S1 (m)	Standardized Start (m)	Standardized End (m)
1	0,01	0,015	1	0,015	0,035	-0,25	0
1	0,02	0,025	2	0,015	0,035	0,25	0,5
1	0,04	0,045	0,5	0,015	0,035	1,25	1,5
2	0	0,005	0,5	0,005	0,02	-0,33	0
2	0,01	0,015	2,5	0,005	0,02	0,33	0,66
2	0,025	0,03	0,7	0,005	0,02	1,33	1,66
3	0,01	0,02	0,5	0,025	0,045	-0,75	-0,25
3	0,03	0,04	2	0,025	0,045	0,25	0,75
3	0,05	0,06	0,5	0,025	0,045	1,25	1,75

Tab 2. Example of averaging depth data: CoreID is the number identifying the core (1; 2; 3). DepthStart and DepthEnd are the beginning and the end of the measured sample starting from the core-top. Proxy Value is the value obtained in the measured sample for a particular proxy (e. g. TOC). Top S1 and Bottom S1 are the sapropel interval top and base depth respectively. Standardized Start and Standardized End are the new depths obtained with the formulas reported in text.

$$\forall d_i \in D1, dn_{1i} = \frac{d_i - Top_1 S_1}{Bottom_1 S_1 - Top_1 S_1}\ ; D1n = \{dn_{11}, dn_{12}, \dots, dn_{1m}\}$$

$$\forall d_i \in D2, dn_{2i} = \frac{d_i - Top_2 S_1}{Bottom_2 S_1 - Top_2 S_1}\ ; D2n = \{dn_{21}, dn_{22}, \dots, dn_{2p}\}$$

$$\forall d_k \in (D1_n \cup D2_n)$$

$$V'(d_k, d_{k+1}) = average\ \left(V1\ (dn_{1i}, dn_{1j}) + V2\ (dn_{2h}, dn_{2l})\right), \qquad s.t.\ (d_k \geq dn_{1i} \wedge d_{k+1} \leq dn_{1j}) \wedge (d_k \geq dn_{2h} \wedge d_{k+1} \leq dn_{2l})$$

The function V'(x,y) represents the average values of the proxy measured at normalized depths. As for C1 and C2, we can see V'(x,y) as the function returning values of the considered proxy in the interval [x,y] of a "virtual core" computed as average of C1 and C2 (see Fig. 12). Here, the value 1 identifies the beginning of sapropel and the value 0 corresponds to the top of it (as indicated by the different authors and therefore it is based on different methods such as color, micropaleontology, geochemistry, etc.). It is noteworthy that this way the ages of the base and the top of the sapropel are not important, because the query provides a stack, combining all the different sapropel thicknesses found in the database. This allows to compare the trend recorded in the different proxies regardless of age as illustrated in the following examples.

To demonstrate the effectiveness of different cores correlation for the sapropel interval, we carried out some analysis with BEyOND. Firstly, we evaluated the correspondence between age and Standardized depth. Fig. 13 considers 37586 age values, of which 22916 in the standardized interval [0,1], from 20 cores. It is evident that the sapropel (S1) interval show a linear trend where r=0.92. Hence, a direct correlation between age and standardized depth is observed.

Then we evaluated the correspondence between TOC, Age and Standardized depth. Fig. 14 shows the perfect trend overlay between TOC vs Age and TOC vs Standardized depth in the sapropel interval, thus supporting the effectiveness of our correlation method.

As for data visualization Fig 15. shows the TOC (Total Organic Carbon) trend in 24 cores across the eastern Mediterranean Sea. As it is obvious the difference between inside and outside the sapropel interval is evident. Moreover, it must be underlined that these values are the result of the averages calculated among all the cores of the Eastern Mediterranean Sea. For this reason, the plot shows high standard deviations indicating that the data points are spread out over a wide range of values. Despite all this, a TOC drop around "0,5

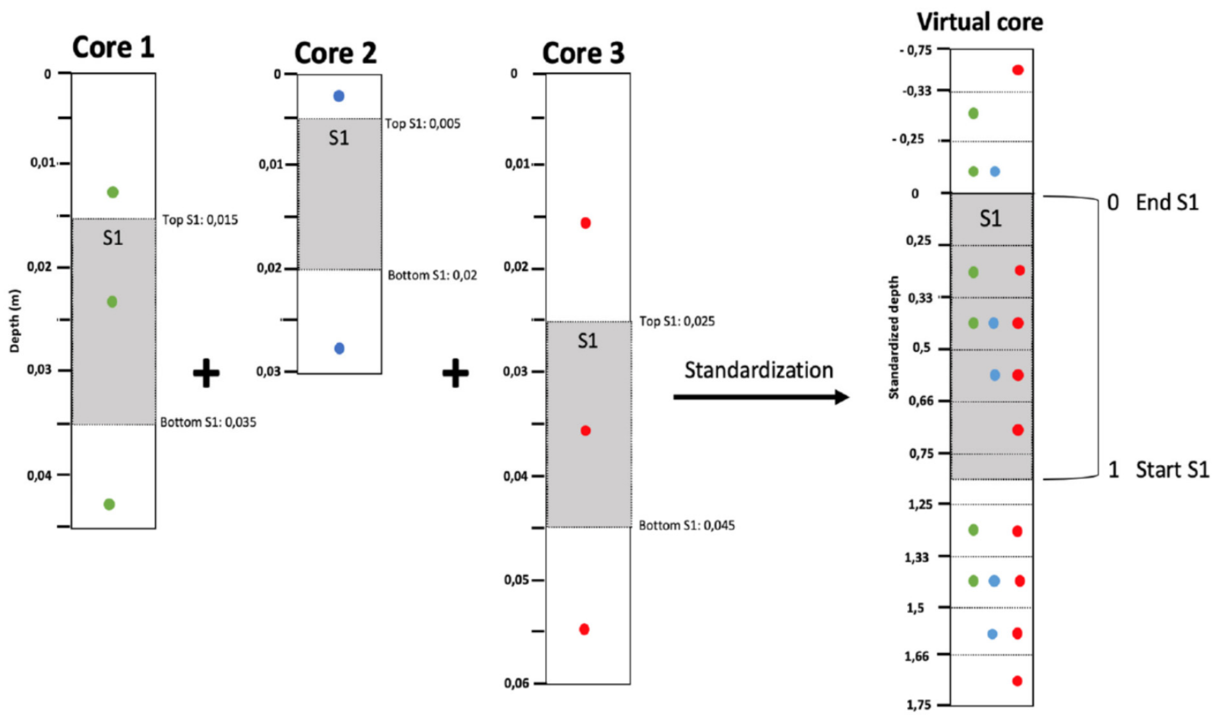

Fig. 12 - Standardization of 3 exemplified cores. The intervals S1 are at different depths and the sedimentation rate is different in every core. The green, blue and red dots are the points measured in the cores. Start S1 and End S1 correspond to the sapropel base and top respectively.

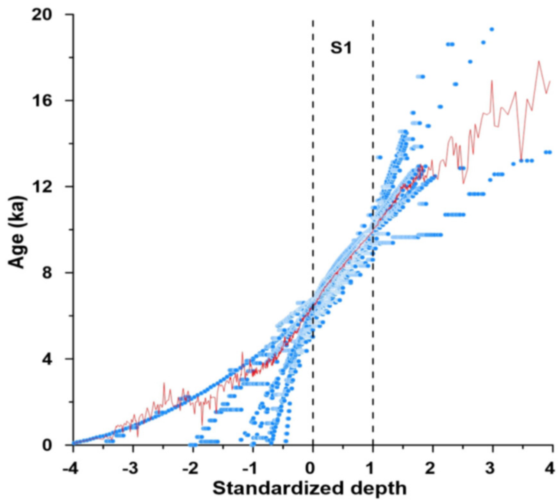

Fig. 13 - Age vs Standardized depth. (a) Blue dots correspond to Eastern Mediterranean Sea cores different ages obtained for the sapropel interval. (b) The red line shows the average age for these cores. Vertical dashed black lines represent sapropel interval (S1) where the beginning of sapropel is 1 and the top is 0 standardized depth.

standardized depth" corresponding to the sapropel interruption (Rohling et al., 1997; De Rijk et al., 1999; Myers & Rohling, 2000; Casford et al., 2003; Abu-Zied et al., 2008; Hennekam et al., 2014) is evident. Instead not all the studies cores show this interruption clearly.

Another example of output is the TOC trend, based on the core depth of the cores in the Eastern Mediterranean Sea (Fig. 16). Cores are shown in four depth interval: 1) 0-1000 m, 2) 1000-2000 m, 3) 2000-3000 m and 4) deeper than 3000 m. Fig. 16 evidences the clear TOC % characterizing each depth range. Obviously, this great variation is reflected in the high standard deviation

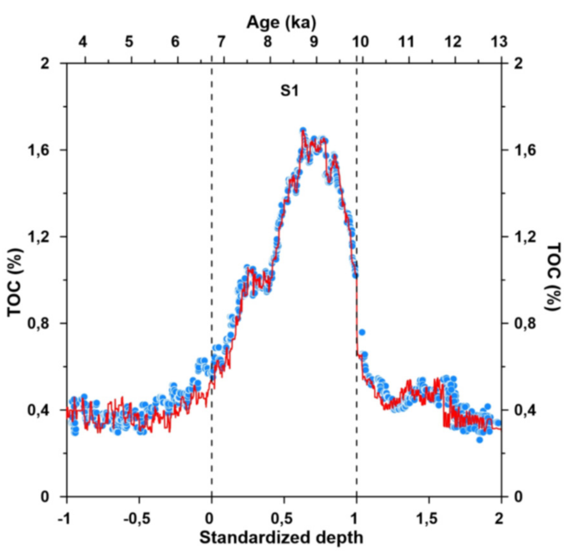

Fig. 14 - TOC trend in the Eastern Mediterranean Sea. (a) Blue dots show TOC vs Age. (b) Red line shows TOC vs Standardized depth. Vertical dashed black lines represent sapropel interval (S1) where the beginning of sapropel is 1 and the top is 0 standardized depth.

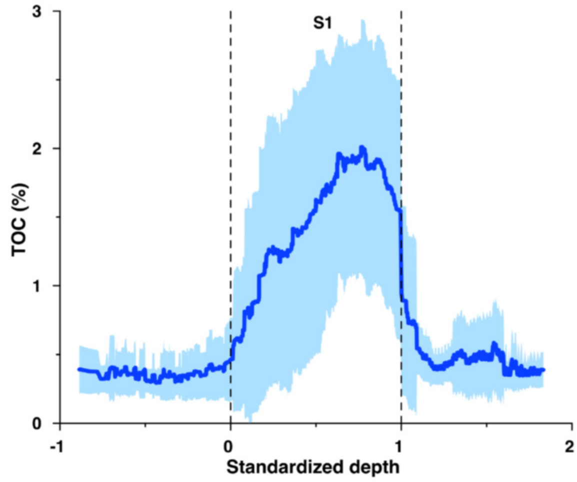

Fig. 15 - TOC trend across the Eastern Mediterranean Sea. The blue line shows the TOC value. The blue-shadow indicate the standard deviation of TOC. Vertical dashed black lines represent sapropel interval (S1) where the beginning of sapropel is 1 and the top is 0 standardized depth.

shown in Fig. 15.

The TOC plotted for different Mediterranean Sea areas is shown in Fig. 17. In this case, the three regions: Adriatic (Adriatic Sea), Levantine (Levantine Sea) and Ionian-Libyan (Ionian and Libyan Sea) clearly show different TOC behaviors. A clear bipartition is recorded in the Adriatic Sea, on the contrary, this is less clear in the Levantine and absent in the Ionian-Libyan Sea.

What we obtained is new when compared to previous works (Alberico et al., 2017; Emile-Geay et al., 2017). In fact, these databases only use records where the "age" is known, rejecting the records missing this variable. Instead, our database permit to manage core correlation in different region, at different depth and showing different sedimentation rates, using depth be-

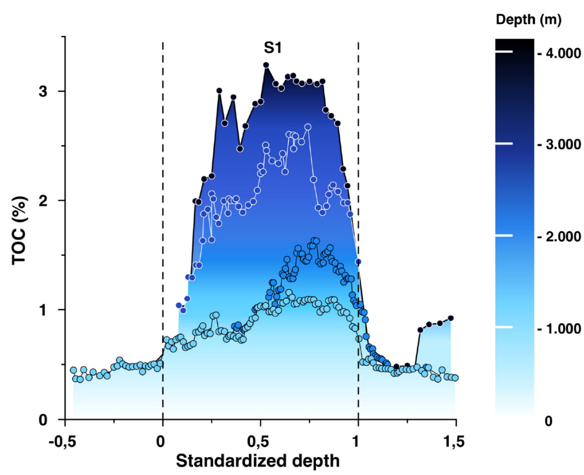

Fig. 16 - TOC trend across the Eastern Mediterranean Sea based on the core depth. The cores are grouped as: 0-1000 m (light blue), 1000-2000 m (blue), 2000-3000m (dark blue) and >3000 m (black). Vertical dashed black lines represent sapropel interval (S1) where the beginning of sapropel is 1 and the top is 0 standardized depth.

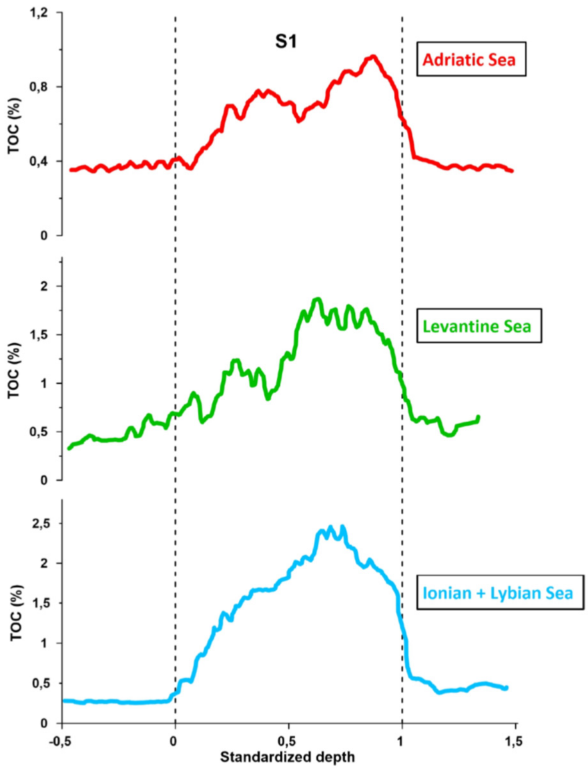

Fig. 17 - TOC trend across the Eastern Mediterranean sea based on the area from where the cores have been sampled. The cores are grouped in three regions: Adriatic sea (red), Levantine sea (green), Ionian and Libyan seas (blue). Vertical dashed black lines represent sapropel interval (S1) where the beginning of sapropel is 1 and the top is 0 standardized depth.

low sea floor (dbsf) and depths in the sapropel interval. Furthermore, our methodology does not suffer of unknown variable (Age) limiting the correlation of the cores. However, some issues still need to be addressed. In fact, as the aggregate results yield a mean value in the interval studied, then some events like the bottom ventilation or the burn-down interval evidenced by short lived spikes may result smoothed and the original lithological extension of the sapropel as indicated in Löwermark et al., (2006) may be less clear. On the other hand, the advantage to obtain a stack aggregating all the data across the Mediterranean allows a wider view and permit to appreciate large scale trends.

4. CONCLUSIVE REMARKS AND FUTURE WORK

Paleoceanographic studies are always based on selected sites where researchers perform analyses rarely muti-proxies based, more frequently based on selected a one (e.g., micropaleontology or geochemistry). BDA potentially allow to reverse the approach exploiting already published data to be considered and combined all together.

BEyOND, based on a collection of data from existing databases (PANGAEA, NOAA, Neotoma…) and on data compiled from the literature, the database:

- provides a wide variety of organized and standardized paleoproxies relative to the past 20.000 years. This organization allows to correlate data, to extract hidden patterns and new knowledge.
- enables the analysis of the results in different ways, that is, geographic area or basin depth, allowing to visualize the different proxies trends as a whole.
- provides a great flexibility in the introduction of new proxies as BEyOND is organized around the "measurement" entity.
- allows the possibility to bypass the age variable (not always considered when data are published) and the possibility to correlate cores even when the age is not known. This is the most innovative feature proposed in this paper.

Our work highlights the advantages and difficulties in the use of new technologies (database, data analytics...) to exploit (paleo) environmental data to reconstruct the dynamics behind the past climate and oceanography related changes. With our results, we show the high potential of using data analytics. Yet, the new technology tools such as machine learning, data mining, artificial intelligence will increase their importance for future "paleo" studies. However, before this will happen, we strengthen the need for data standardization as it is the main issue when working with a huge data volumes. We therefore encourage the scientific community to take this challenge in order to improve the exchange and usability of this methodology.

BEyOND is open accessible to all researchers to extract data. As a long term objective, it will be implemented in order to add new data in a way that favors the exchange of knowledge as other databases do (e.g., PANGAEA or NOAA). Furthermore, an important achievement will be friendly database where advanced analyses do not need informatic skills. In this regard, we have already developed user-friendly graphical interfaces and services to perform the above described operations (also reported in the Appendix B, in the form of SQL queries). New functionalities of exploratory analysis will be created in the next future. The output of this will be significantly increment in the data volume thanks to the whole contribution of the scientific community contribution. The rich set of integrated data then will be further exploited by the adopting of Data Mining techniques, which in the future will allow through automatic or semi-automatic methods the extraction of useful information from large amounts of data, and its scientific use.

ACKNOWLEDGEMENTS

We would like to acknowledge one anonymous reviewer, Caterina Morigi, Gianluca Marino and Editor Andrea Sposato for their constructive comments that were helpful to revise and improve our paper. Data presented in this study are available at beyond.dii.univpm.it/.

REFERENCES

Abu-Zied R.H., Rohling E.J., Jorissen F.J., Fontanier C., Casford J.S.L., Cooke, S. (2008) - Benthic foraminiferal response to changes in bottom-water oxygenation and organic carbon flux in the eastern Mediterranean during LGM to Recent times. Ma-

rine Micropaleontology, 67(1-2), 46-68.
Doi:10.1016/j.marmicro.2007.08.006

Alberico I., Giliberti I., Insinga D.D., Petrosino P., Vallefuoco M., Lirer F., Ferraro L. (2017) - Marine sediment cores database for the Mediterranean Basin: A tool for past climatic and environmental studies. Open Geosciences, 9(1), 221-239.
Doi:10.1515/geo-2017-0019

Azrieli-Tal I., Matthews A., Bar-Matthews M., Almogi-Labin A., Vance D., Archer C., Teutsch N. (2014) - Evidence from molybdenum and iron isotopes and molybdenum-uranium covariation for sulphidic bottom waters during Eastern Mediterranean sapropel S1 formation. Earth and Planetary Science Letters, 393, 231-242.
Doi:10.1016/j.epsl.2014.02.054

Bell G. (2009) - Foreword. The Fourth Paradigm: Data-Intensive Scientific Discovery, xi-xv.

Bout-Roumazeilles V., Combourieu-Nebout N., Desprat S., Siani G., Turon J.L., Essallami L. (2013) - Tracking atmospheric and riverine terrigenous supplies variability during the last glacial and the Holocene in central Mediterranean. Climate of the Past, *9*(3), 1065-1087.
Doi:10.5194/cp-9-1065-2013.

Bown P.R. Young J.R. (1998) - Techniques. In: Bown, P.R., Ed., Calcareous Nannofossil Biostratigraphy (British Micropalaeontological Society Publications Series), Chapman and Kluwer Academic, London, 16-28.

Calvert S.E., Price N.B. (1983) - Geochemistry of Namibian shelf sediments. In Coastal Upwelling Its Sediment Record. Springer, 1, 337-375.
Doi:10.1007/978-1-4615-6651-9_17

Calvert S.E., Pedersen T.F. (2007) - Chapter Fourteen Elemental Proxies for Palaeoclimatic and Palaeoceanographic Variability in Marine Sediments: Interpretation and Application. Developments in Marine Geology, 1(07), 567-644.
Doi:10.1016/S1572-5480(07)01019-6

Carty S., Parrow M.W. (2015). Dinoflagellates. Freshwater Algae of North America, pp. 773-807.
Doi:10.1016/B978-0-12-385876-4.00017-7

Casford J.S. L., Rohling E.J., Abu-Zied R.H., Fontanier C., Jorissen F.J., Leng M.J., Thomson J. (2003) - A dynamic concept for eastern Mediterranean circulation and oxygenation during sapropel formation. In Palaeogeography, Palaeoclimatology, Palaeoecology, 8, 103-119.
Doi:10.1016/S0031-0182(02)00601-6

Cita M.B., Vergnaud-Grazzini C., Robert C., Chamley H., Ciaranfi N., d'Onofrio S. (1977) - Paleoclimatic record of a long deep sea core from the eastern Mediterranean. Quaternary Research, 8(2), 205-235.
Doi:10.1016/0033-5894(77)90046-1

Crudeli D., Young J.R., Erba E., Geisen M., Ziveri P., De Lange G.J., Slomp C.P. (2006) - Fossil record of holococcoliths and selected hetero-holococcolith associations from the Mediterranean (Holocene-late Pleistocene): Evaluation of carbonate diagenesis and palaeoecological-palaeocenographic implications. Palaeogeography, Palaeoclimatology, Palaeoecology, 237(2-4), 191-212.
Doi:10.1016/j.palaeo.2005.11.022

De Rijk S., Hayes A., Rohling E.J. (1999) - Eastern Mediterranean sapropel S1 interruption: An expression of the onset of climatic deterioration around 7 ka BP. Marine Geology, 153(1-4), 337-343.
Doi:10.1016/S0025-3227(98)00075-9

Dozier J., Gail W.B. (2009) - The emerging science of enviroment applications. The Fourth Paradigm: Data-Intensive Scientific Discovery, 13.

Dubois-Dauphin Q., Montagna P., Siani G., Douville E., Wienberg C., Hebbeln D., Colin C. (2017) - Hydrological variations of the intermediate water masses of the western Mediterranean Sea during the past 20 ka inferred from neodymium isotopic composition in foraminifera and cold-water corals. Climate of the Past, 13(1), 17-37.
Doi:10.5194/cp-13-17-2017

Ehrmann W., Schmiedl G., Hamann Y., Kuhnt T., Hemleben C., Siebel W. (2007) - Clay minerals in late glacial and Holocene sediments of the northern and southern Aegean Sea. Palaeogeography, Palaeoclimatology, Palaeoecology, 249(1-2), 36-57.
Doi:10.1016/j.palaeo.2007.01.004

Ehrmann W., Schmiedl G., Seidel M., Krüger S., Schulz H. (2016) - A distal 140kyr sediment record of Nile discharge and East African monsoon variability. Climate of the Past, 12(3), 713-727.
Doi:10.5194/cp-12-713-2016

Elmasri R., Navathe S.B. (2010) - Fundamentals of Database Systems (6th ed.). Addison-Wesley Publishing Company, USA, pp. 199.

Emile-Geay J., McKay N.P., Kaufman D.S., Von Gunten L., Wang J., Anchukaitis K.J., Zinke J. (2017) - A global multiproxy database for temperature reconstructions of the Common Era. Scientific Data, 4, 170088.
Doi:10.1038/sdata.2017.88

Fontugne M.R., Paterne M., Calvert S.E., Murat A., Guichard F., Arnold M. (1989) - Adriatic deep water formation during the Holocene: Implication for the reoxygenation of the deep eastern Mediterranean Sea. Paleoceanography, 4(2), 199-206.
Doi:10.1029/PA004i002p00199

Frigola J., Moreno A., Cacho I., Canals M., Sierro F.J., Flores J.A., Grimalt J.O. (2008) - Evidence of abrupt changes in Western Mediterranean Deep Water circulation during the last 50 kyr: A high-resolution marine record from the Balearic Sea. Quaternary International, 181(1), 88-104.
Doi:10.1016/j.quaint.2007.06.016

Gallego-Torres D., Martinez-Ruiz F., Meyers P.A., Paytan A., Jimenez-Espejo F. J., Ortega-Huertas M. (2011) - Productivity patterns and N-fixation associated with Pliocene-Holocene sapropels: Paleoceanographic and paleoecological significance. Biogeosciences, 8(2), 415-431.
Doi:10.5194/bg-8-415-2011

Gandomi A., Haider M. (2015) - Beyond the hype: Big

data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.
Doi:10.1016/j.ijinfomgt.2014.10.007

Grazzini C.V., Pierre C. (1991) - High Fertility in the Alboran Sea Since the last Glacial Maximum. Paleoceanography, 6(4), 519-536.
Doi:10.1029/91PA00501

Gupta A.K., Das M., Clemens S.C., Mukherjee B. (2008) - Benthic foraminiferal faunal and isotopic changes as recorded in Holocene sediments of the northwest Indian Ocean. Paleoceanography, 23(2).
Doi:10.1029/2007PA001546

Hennekam R., Jilbert T., Schnetger B., De Lange G. J. (2014) - Solar forcing of Nile discharge and sapropel S1 formation in the early to middle Holocene eastern Mediterranean. Paleoceanography, 29(5), 343-356.
Doi:10.1002/2013PA002553

Hennekam R., Donders T.H., Zwiep K., De Lange G.J. (2015) - Integral view of Holocene precipitation and vegetation changes in the Nile catchment area as inferred from its delta sediments. Quaternary Science Reviews, 130, 189-199.
Doi:10.1016/j.quascirev.2015.05.031

Hilgen F.J. (1991) - Astronomical calibration of Gauss to Matuyama sapropels in the Mediterranean and implication for the Geomagnetic Polarity Time Scale. Earth and Planetary Science Letters, 104(2 -4), 226-244.
Doi:10.1016/0012-821X(91)90206-W

IBM Marketing Cloud (2017) - 10 Key Marketing Trends for 2017 and Ideas for exceeding customer expectations.
https://www-01.ibm.com/common/ssi/cgibin/ssialias?htmlfid=WRL12345USEN

Incarbona A., Ziveri P., Sabatino N., Manta D.S., Sprovieri M. (2011) - Conflicting coccolithophore and geochemical evidence for productivity levels in the Eastern Mediterranean sapropel S1. Marine Micropaleontology, 81(3-4), 131-143.
Doi:10.1016/j.marmicro.2011.09.003

Jimenez-Espejo F.J., Martinez-Ruiz F., Sakamoto T., Iijima K., Gallego-Torres D., Harada N. (2007) - Paleoenvironmental changes in the western Mediterranean since the last glacial maximum: High resolution multiproxy record from the Algero-Balearic basin. Palaeogeography, Palaeoclimatology, Palaeoecology, 246(2-4), 292-306.
Doi:10.1016/j.palaeo.2006.10.005

Jimenez-Espejo F.J., Martinez-Ruiz F., Rogerson M., González-Donoso J.M., Romero O.E., Linares D., Claros J.A.P. (2008) - Detrital input, productivity fluctuations, and water mass circulation in the westernmost Mediterranean Sea since the Last Glacial Maximum. Geochemistry, Geophysics, Geosystems, 9(11).
Doi:10.1029/2008GC002096

Kemp A.E.S., Pearce R.B., Koizumi I., Pike J., Rance S.J. (1999) - The role of mat-forming diatoms in the formation of Mediterranean sapropels. Nature, 398(6722), 57.
Doi:10.1038/18001

Khider D., Zhu F., Hu J., Emile-Geay J. (2018) - LinkedEarth/Pyleoclim util: Pyleoclim release v0.4.0.
Doi:10.5281/zenodo.1205662.

Kidd R.B., Cita M.B., Ryan W.B.F. (1978) - Stratigraphy of Eastern Mediterranean Sapropel Sequences Recovered during DSDP Leg 42A and Their Paleoenvironmental Significance. Initial Reports of the Deep Sea Drilling Project, 42A, 421-433.
Doi:10.2973/dsdp.proc.42-1.113-1.1978

Kotthoff U., Pross J., Müller U.C., Peyron O., Schmiedl G., Schulz H., Bordon A. (2008) - Climate dynamics in the borderlands of the Aegean Sea during formation of sapropel S1 deduced from a marine pollen record. Quaternary Science Reviews, 27(7-8), 832-845.
Doi:10.1016/j.quascirev.2007.12.001

Kuhnt T., Schmiedl G., Ehrmann W., Hamann Y., Hemleben C. (2007) - Deep-sea ecosystem variability of the Aegean Sea during the past 22 kyr as revealed by Benthic Foraminifera. Marine Micropaleontology, 64(3-4), 141-162.
Doi:10.1016/j.marmicro.2007.04.003

Kuhnt T., Schmiedl G., Ehrmann W., Hamann Y., Andersen N. (2008) - Stable isotopic composition of Holocene benthic foraminifers from the Eastern Mediterranean Sea: Past changes in productivity and deep water oxygenation. Palaeogeography, Palaeoclimatology, Palaeoecology, 268(1-2), 106-115.
Doi:10.1016/j.palaeo.2008.07.010

Langgut D., Almogi-Labin A., Bar-Matthews M., Weinstein-Evron M. (2011) - Vegetation and climate changes in the South Eastern Mediterranean during the Last Glacial-Interglacial cycle (86 ka): New marine pollen record. Quaternary Science Reviews, 30(27-28), 3960-3972.
Doi:10.1016/j.quascirev.2011.10.016

Lingerfelt,E., Fils,D., Shepherd,A. (2018) - Project 418: A Funded Project of the EarthCube Science Support Office. In AGU Fall Meeting Abstracts. IN31B-22.

Lourens L.J., Antonarakou A., Hilgen F.J., Van Hoof A. A.M., Vergnaud-Grazzini C., Zachariasse W.J. (1996) - Evaluation of the Plio-Pleistocene astronomical timescale. Paleoceanography, 11(4), 391-413.
Doi:10.1029/96PA01125

Löwemark L., Lin Y., Chen H.F., Yang T.N., Beier C., Werner F., Lee C.Y., Song S.R., Kao S.J. (2006) - Sapropel burn-down and ichnological response to late Quaternary sapropel formation in two ∼ 400 ky records from the eastern Mediterranean Sea, Palaeogeography, Palaeoclimatology, Palaeoecology, 239(3-4), 406-425.

Lussier R.N., Hendon J.R. (2016) - Fundamentals of Human Resource Management: Functions, Applications, Skill Development, (1st ed.). SAGE Publications, Los Angeles.

Mackensen A., Rudolph M., Kuhn G. (2001) - Late pleistocene deep-water circulation in the subantarctic

eastern Atlantic. Global and Planetary Change, 30(3-4), 197-229.
Doi:10.1016/S0921-8181(01)00102-3
Martinez-Ruiz F., Kastner M., Paytan A., Ortega-Huertas M., Bernasconi S.M. (2000) - Geochemical evidence for enhanced productivity during S1 sapropel deposition in the eastern Mediterranean. Paleoceanography, 15(2), 200-209.
Doi:10.1029/1999PA000419
Martinez-Ruiz F., Kastner M., Gallego-Torres D., Rodrigo-Gámiz M., Nieto-Moreno V., Ortega-Huertas M. (2015) - Paleoclimate and paleoceanography over the past 20,000yr in the Mediterranean Sea Basins as indicated by sediment elemental proxies. Quaternary Science Reviews, 107, 25-46.
Doi:10.1016/j.quascirev.2014.09.018.
McFedries P. (2011) - The Coming Data Deluge, IEEE Spectrum, 48(2), 19-19.
Doi:10.1109/MSPEC.2011.5693066
McKay N.P., Emile-Geay J. (2016) - Technical note: The Linked Paleo Data framework - A common tongue for paleoclimatology. Climate of the Past, 12(4), 1093-1100.
Doi:10.5194/cp-12-1093-2016
Meier K.J.S., Janofske D., Willems H. (2002) - New calcareous dinoflagellates (Calciodinelloideae) from the Mediterranean Sea. Journal of Phycology, 38(3), 602-615.
Doi:10.1046/j.1529-8817.2002.t01-1-01191.x
Meier K. J.S., Zonneveld K.A.F., Kasten S., Willems H. (2004) - Different nutrient sources forcing increased productivity during eastern Mediterranean S1 sapropel formation as reflected by calcareous dinoflagellate cysts. Paleoceanography, 19 (1), 1-12.
Doi:10.1029/2003PA000895
Möbius J., Lahajnar N., Emeis K.C. (2010) - Diagenetic control of nitrogen isotope ratios in Holocene sapropels and recent sediments from the Eastern Mediterranean Sea. Biogeosciences, 7(11), 3901-3914.
Doi:10.5194/bg-7-3901-2010.
Mojtahid M., Manceau R., Schiebel R., Hennekam R., De Lange G.J. (2015) - Thirteen thousand years of southeastern Mediterranean climate variability inferred from an integrative planktic foraminiferal-based approach. Paleoceanography, 30(4), 402-422.
Doi:10.1002/2014PA002705.
Myers P.G., Haines K., Rohling E.J. (1998) - Modeling the paleocirculation of the Mediterranean: The last glacial maximum and the Holocene with emphasis on the formation of sapropel S1. Paleoceanography, 13(6), 586-606.
Doi:10.1029/98PA02736.
Myers P.G., Rohling E.J. (2000) - Modeling a 200-Yr interruption of the Holocene Sapropel S1. Quaternary Research, 53(1), 98-104.
Doi:10.1006/qres.1999.2100
Negri A., Giunta S. (2001) - Calcareous nannofossil paleoecology in the sapropel S1 of the Eastern Ionian sea: Paleoceanographic implications. Palaeogeography, Palaeoclimatology, Palaeoecology, 169(1-2), 101-112.
Doi:10.1016/S0031-0182(01)00219-X
Olausson E. (1961) - Studies of deep-sea cores. Rep. Swed. Deep Sea Exped. 1947-1948, 8, 353-391.
Parker W.C., Arnold A.J. (1999) - Quantitative methods of data analysis in foraminiferal ecology.Modern Foraminifera, 5, 71-89.
Doi:10.1007/0-306-48104-9_5
Pedersen T.F., Calvert S.E. (1990) - Anoxia vs. productivity: what controls the formation of organic- carbon-rich sediments and sedimentary rocks? American Association of Petroleum Geologists Bulletin, 74(4), 454-466.
Doi:10.1306/0C9B232B-1710-11D7-864500010-2-C1865D
Principato M.S., Giunta S., Corselli C., Negri A. (2003) - Late Pleistocene-Holocene planktonic assemblages in three box-cores from the Mediterranean Ridge area (west-southwest of Crete): Palaeoecological and palaeoceanographic reconstruction of sapropel S1 interval. Palaeogeography, Palaeoclimatology, Palaeoecology, 190, 61-77.
Doi:10.1016/S0031-0182(02)00599-0.
Ridge E. (2014) - Guerrilla Analytics: A Practical Approach to Working with Data. Guerrilla Analytics: A Practical Approach to Working with Data. Morgan Kaufmann.
Doi:10.1016/C2013-0-14458-3
Rohling E.J. (1991) - Shoaling of the Eastern Mediterranean Pycnocline due to reduction of excess evaporation: Implications for sapropel formation. Paleoceanography, 6(6), 747-753.
Doi:10.1029/91PA02455
Rohling E.J. (1994) - Review and new aspects concerning the formation of eastern Mediterranean sapropels. Marine Geology, 122(1-2), 1-28.
Doi:10.1016/0025-3227(94)90202-X
Rohling E.J., Jorissen F.J., De Stigter H.C. (1997) - 200 Year interruption of Holocene sapropel formation in the Adriatic Sea. Journal of Micropalaeontology, 16(2), 97-108.
Doi:10.1144/jm.16.2.97
Rohling E.J., Marino G., Grant K.M. (2015) - Mediterranean climate and oceanography, and the periodic development of anoxic events (sapropels). Earth-Science Reviews, 143, 62-97.
Doi:10.1016/j.earscirev.2015.01.008
Rossignol-Strick M. (1983) - African monsoons, an immediate climate response to orbital insolation. Nature, 304(5921), 46.
Doi:10.1038/304046a0.
Rossignol-Strick M. (1985) - Mediterranean Quaternary sapropels, an immediate response of the African monsoon to variation of insolation. Palaeogeography, Palaeoclimatology, Palaeoecology, 49(3-4), 237-263.
Doi:10.1016/0031-0182(85)90056-2
Ryan W.B.F. (1972) - Stratigraphy of late Quaternary sediments in the eastern Mediterranean. In: Stanley, D.J., Ed., The Mediterranean Sea: a

natural sedimentation laboratory: 149-169, Stroudsburg, Pa.: Dowden, Hutchinson & Ross.

Sachs J.P., Repeta D.J. (1999) - Oligotrophy and nitrogen fixation during eastern Mediterranean sapropel events. Science, 286(5449), 2485-2488. Doi:10.1126/science.286.5449.2485

Sarmiento J.L., Herbert T., Toggweiler J.R. (1988) - Mediterranean nutrient balance and episodes of anoxia. Global Biogeochemical Cycles, 2(4), 427-444. Doi:10.1029/GB002i004p00427

Sarnthein M., Winn K., Jung S.J.A., Duplessy J.C, Labeyrie L., Erlenkeuser H., Ganssen G. (1994) - Changes in East Atlantic Deepwater Circulation over the last 30,000 years: Eight time slice reconstructions. Paleoceanography, 9(2), 209-267. Doi:10.1029/93PA03301

Schmiedl G., Hemleben C., Keller J., Segl M. (1998) - Impact of climatic changes on the benthic foraminiferal fauna in the Ionian Sea during the last 330,000 years. Paleoceanography, 13(5), 447-458. Doi:10.1029/98PA01864

Stratford K., Williams R.G., Myers P.G. (2000) - Impact of the circulation on sapropel formation in the eastern Mediterranean. Global Biogeochemical Cycles, 14(2), 683-695. Doi:10.1029/1999GB001157

Tachikawa K., Vidal L., Cornuault M., Garcia M., Pothin A., Sonzogni C., Revel M. (2015) - Eastern Mediterranean Sea circulation inferred from the conditions of S1 sapropel deposition. Climate of the Past, 11(6), 855-867. Doi:10.5194/cp-11-855-2015

Tesi T., Asioli A., Minisini D., Maselli V., Dalla Valle G., Gamberi F., Trincardi F. (2017) - Large-scale response of the Eastern Mediterranean thermohaline circulation to African monsoon intensification during sapropel S1 formation. Quaternary Science Reviews, 159, 139-154. Doi:10.1016/j.quascirev.2017.01.020

Thunell R.C., Williams D.F., Kennett J.P. (1977) - Late Quaternary paleoclimatology, stratigraphy and sapropel history in eastern Mediterranean deep-sea sediments. Marine Micropaleontology, 2, 371-388. Doi:10.1016/0377-8398(77)90018-4

Thunell R.C. (1979) - Pliocene-Pleistocene paleotemperature and paleosalinity history of the Mediterranean Sea: Results from DSDP Sites 125 and 132. Marine Micropaleontology, 4, 173-187. Doi:10.1016/0377-8398(79)90013-6.

Triantaphyllou M. V. (2014) - Coccolithophore assemblages during the Holocene Climatic Optimum in the NE Mediterranean (Aegean and northern Levantine Seas, Greece): Paleoceanographic and paleoclimatic implications. Quaternary International, 345, 56-67. Doi:10.1016/j.quaint.2014.01.033

Triantaphyllou M.V., Antonarakou A., Dimiza M., Anagnostou C. (2010) - Calcareous nannofossil and planktonic foraminiferal distributional patterns during deposition of sapropels S6, S5 and S1 in the Libyan Sea (Eastern Mediterranean). Geo-Marine Letters, 30(1), 1-13. Doi:10.1007/s00367-009-0145-7

Tribovillard N., Algeo T.J., Lyons T., Riboulleau A. (2006) - Trace metals as paleoredox and paleoproductivity proxies: An update. Chemical Geology, 232(1-2), 12-32. Doi:10.1016/j.chemgeo.2006.02.012

Woodruff F., Savin S.M. (1985) - $\delta^{13}C$ values of Miocene Pacific benthic foraminifera: Correlations with sea level and biological productivity. Geology, 13 (2): 119-122. Doi:10.1130/0091-7613(1985)13<119:CVOMPB> 2.0.CO;2.

Zahn R., Winn K., Sarnthein M. (1986) - Benthic foraminiferal $\delta^{13}C$ and accumulation rates of organic carbon: *Uvigerina Peregrina* group and *Cibicidoides Wuellerstorfi*. Paleoceanography, 1(1), 27-42. Doi:10.1029/PA001i001p00027.

Ziegler M., Tuenter E., Lourens L.J. (2010) - The precession phase of the boreal summer monsoon as viewed from the eastern Mediterranean (ODP Site 968). Quaternary Science Reviews, 29(11-12), 1481-1490. Doi:10.1016/j.quascirev.2010.03.011

Zonneveld K.A.F., Versteegh G.J.M., De Lange G.J. (2001) - Palaeoproductivity and post-depositional aerobic organic matter decay reflected by dinoflagellate cyst assemblages of the Eastern Mediterranean S1 sapropel. Marine Geology, 172(3-4), 181-195. Doi:10.1016/S0025-3227(00)00134-1

Ms. received: March 15, 2019
Final text received: July 19, 2019