

The foraminifera.eu database: concept and status

Michael Hesemann

ABSTRACT

The foraminifera.eu project with its 120 avocational and professional scientific contributors has built a popular online image database containing information about fossil and recent foraminifera. Foraminiferal data available from over 200 years of scientific studies and publications were first analyzed in order to create a robust database using structurable and available data-attributes. A freely accessible database with online interfaces was established to improve the accessibility of foraminiferal data. So far, the database contains 9,800+ records of specimens with images and accompanying metadata. Thirty data attributes were chosen, and discrete values were defined and applied to each dataset covering the areas of identification, geographical and stratigraphical positioning, morphology, synonyms, references, and collection related data. The attributes were chosen based on their usefulness to both avocational and professional scientists and on their general availability. A discussion and review process has been established to enlarge the database by adding more records and to implement new data attributes and interfaces. In 2015, the database was accessed on average by 220 users daily (excluding bots) and 0,8 gigabytes of data were downloaded each day. The numbers indicate the utility and relevance of the foraminifera.eu database, and this is also acknowledged by senior scientists and major institutions through their contributions.

Michael Hesemann. Foraminifera.eu Project, Waterloostrasse 24, 22769 Hamburg, Germany.
hesemann@foraminifera.eu

INTRODUCTION

Over 200 years of scientific studies and thousands of publications have resulted in a huge amount of information on foraminifera and up to 80,000 described species (Lipps and Finger, 2010). However, the information is scattered and difficult to access. In the fundamental area of classification, major works are out of print and only accessible at academic institutions (e.g., Ellis and

Messina, 1940-2014; Kennett and Srinivasan, 1983; Loeblich and Tappan, 1988). The basic idea of the foraminifera.eu database (FEUDAT) is to improve the accessibility of foraminiferal data. While the available data are complex, unstructured, sometimes contradictory, and incomplete, a database needs to be structured and comprehensive in order to provide the intended improved access. As a consequence, data attributes need to

Keywords: foraminifera; database; paleoinformatics; identification key

Michael Hesemann. Foraminifera.eu Project, Waterloostrasse 24, 22769 Hamburg, Germany. hesemann@foraminifera.eu

Hesemann, Michael. 2015. The foraminifera.eu database: concept and status. *Palaeontologia Electronica* 18.3.48A: 1-14
palaeo-electronica.org/content/2015/1333-the-foraminifera-eu-database

be identified which can be used to structure foraminiferal data. Such structurable data attributes are found in the areas of classification, bibliography, collection management, distribution lists, and monographs - just to name the most important.

FEUDAT is created and run by the foraminifera.eu project (FEUPRO). It was founded in December 2007 by the avocational foraminiferologists M. Hesemann, D. Ketelsen, and K.-O. Bock (all members of the AG Mikropaläontologie im Naturwissenschaftlichen Verein Hamburg). It is supported by G. Rosenfeld, head of the Mikrobiologische Vereinigung Hamburg, who runs a scanning electron microscope (SEM) LEITZ AMR 1600 T. B. Ottway, Lagos, Portugal, joined the team in 2012. FEUPRO is based on the enthusiasm of naturalists and is strictly non-commercial. Neither the contributors nor the team get any financial compensation.

The mission of FEUPRO is to foster interest in foraminifera. Through FEUPRO, avocational and professional scientists get a platform on the internet to show their results and obtain virtual and real help. In the beginning, there was no intention to build a database but only to show images of foraminifera in our collections to other avocational scientists to encourage them to get in contact. In order to receive interesting samples, FEUPRO offered, and still offers, to process samples and to produce optical and SEM images of material sent to us for free. The offer and website worked very well and within a couple of months many samples had come in and dozens of SEM-images needed to be made. In April 2008, F. Frontalini, University of Urbino, now running the International Foraminiferal School, was the first scientific contributor and sent about 80 images of foraminifera from the Adriatic (Frontalini and Coccioni, 2007).

Through participation in the 2nd Course of the International Foraminiferal School in Urbino 2009, FEUPRO established contacts with senior scientists. M.A. Kaminski has supported FEUPRO since then continuously with his advice, expertise, and with images through his role as editor of the Grzybowski foundation publications. FEUPRO was presented in a poster session at the FORAMS 2010 meeting in Bonn, where a plenary session was held on foraminifera in the internet (Hesemann, 2010). Since then, FEUPRO has been supported by B.W. Hayward and T. Cedhagen, the driving forces for the World Foraminifera Database (WFD) (Hayward et al., 2015). T. Cedhagen sent the plates of the Brady's Challenger Report with the revised taxonomy by Jones (Brady, 1884; Jones, 1994). As a result, hundreds of images were made

and contributed. They needed to be shown in a proper way, and the idea to build a contributor oriented image database was born. Over the course of time, FEUPRO has had about 120 contributors ranging from avocational to senior professional scientist and institutions. The popularity of the foraminifera.eu website is reflected in 2015 by 220 daily visitors and 800 megabytes of downloaded data. To be of lasting value, it is essential that online databases have continuity beyond the demise of the originators (De Clerck et al., 2013). The FEUPRO team will put measures in place to ensure that this is the case.

FEUPRO is Not a Citizen Science Project

Citizen science projects are run by scientists and give non-scientists the opportunity to make meaningful contributions to science. The main purpose is to enlist the public to obtain a large quantity of data and enlarge scientific knowledge at low cost. Additionally, most projects strive to educate their contributors on the observed objects and on scientific processes. The topic of most projects is biodiversity monitoring (Bonney et al., 2009).

The main purpose of FEUPRO is to foster the interest in foraminifera and help all interested parties in their study of foraminifera. Only additionally FEUPRO strives to support the scientific process by helping junior scientists, providing a useful image database, and allowing senior scientists to communicate their work. The goal is to build a bridge between science and community (Hesemann et al., 2014). In contrast to citizen science projects, FEUPRO puts first priority on educational and motivational aspects and strives to build a community helping each other in the study of foraminifera. In contrast to citizen science, it is not limited to certain aspects of the observed object. With foraminifera the object itself is not easily observable and suitable for data collection by the public. Furthermore, avocational scientists run FEUPRO and FEUDAT with no pressure to publish, shifting the emphasis from scientific output towards usability, community building, and communication. Thus it must be concluded that FEUPRO and FEUDAT lie outside the normal definition of citizen science projects.

THE DATABASE CONCEPT OF FEUDAT

Database

FEUDAT is built to offer access to structured information on foraminifera. "Databases are a repository for information. They excel at managing

and manipulating structured information" (Davis and Phillips, 2006). A wide range of proprietary and standard database solutions is used in life sciences. The market for database solutions is very competitive and innovative, which makes it difficult to choose a sustainable setting. FEUPRO does not favor a specific solution. The data are held in a very simple way in order to ensure their easy usage in any kind of database setting. In FEUDAT the data are stored as text files in ASCII format. Each data record is represented by a single text line, with the single data elements separated by a separator, for example a comma. The data elements and their values are well defined as discussed later. The text file holding all the data is split into a main and six special text files in order to reduce redundancy. Data records in the different text files are referred to each other with referential data elements. Each data record contains, for example, a literature reference. While the main text file only holds a referential data element, the literature text file contains specific data elements on the publication such as author, title, year, etc. plus a referential data element. By setting the same unique value into the corresponding referential data elements both text files are referred to each other. The text files may be used in any type of database (relational or flat) and edited with any kind of office software. At present there is a main text file and six specific text files on references, morphological features, contributors, higher taxonomy above the genus level, synonyms, and localities.

Data Management

Evaluating and selecting adequate technical tools that meet the requirements of a project needs a full examination of many conflicting factors and is described as a demanding task (Jadhav and Sonar, 2009). The FEUPRO team found five requirements to be essential for the FEUDAT data management. They were prioritized as follows:

1. Easy accessibility of the data by users
2. Functionality for users
3. Functionality / ease of programming
4. Sustainability and readiness for future developments
5. Low cost

The testing and evaluation of different solutions is an ongoing process for the FEUPRO team reflecting dynamic development of computing and new demands by the growing community of users. FEUDAT has seen a development from a static to a

dynamic website and from not accommodating queries to facilitating quite sophisticated ones. Recently responsive webdesign has been implemented, which allows users with mobile devices and tablets to obtain limited but easy access.

At present (July 2015) the FEUPRO team uses the World Wide Web, HTML5, PHP 5.4, and MySQL 5.1.73 for data storage and output. Hypertext markup language (HTML) is the standard programming language to encode webpages. Its current version is HTML5. PHP (PHP: Hypertext Preprocessor) is a scripting language, which is commonly used to create HTML code and build dynamic webpages. MySQL is a popular and free Open Source database management system (Zal-ecka et al., 2015). "MySQL is a fast and powerful yet easy-to-use database system that offers just about anything a website might need in order to find and serve up data to browsers. When PHP allies with MySQL to store and retrieve this data, you have the fundamental parts required for the development of social networking sites and the beginnings of Web 2.0." (Nixon, 2012). Text editors such as Notepad++®, HTML5 editors such as Brackets® and spreadsheet software such as Microsoft Excel® or LibreOffice Calc® can be used to create new data records for upload. The main reasons for these choices will now be discussed.

The World Wide Web has become the mass medium of the twenty-first century (Dolven and Skjerpen, 2011). In contrast to print media or sharing of files it entails negligible costs. By using HTML encoded webpages, easy access is given to users worldwide and with different hardware platforms (servers, personal computers, tablets, and mobile devices) and operating systems (MS Windows®, Linux, (MAC-)OS X, Android®, and IOS®). Sharing of files or usage of the Usenet is seen as too complicated by most users. In 2008 and 2009, the upload of FEUPRO data to other online databases such as Micro*Scope (Marine Biological Laboratory, 2015), EOL (Encyclopedia of Life, 2015), MARBEF (Flanders Marine Institute VLIZ, 2015), GBIF (The Global Biodiversity Information Facility, 2015), WoRMS/WFD (Hayward et al., 2015) and BioLib.cz (Zicha et al., 2015) was tested. These were considered to be too restrictive in their functionality for users and programmers. Proprietary solutions were evaluated and found not to meet most of the requirements. For example PaleoTax, not designed for the World Wide Web, is based on Windows and maintained by only one person (Löser, 2004). The features of Taxomorph are seen as suitable, but it was only published as a

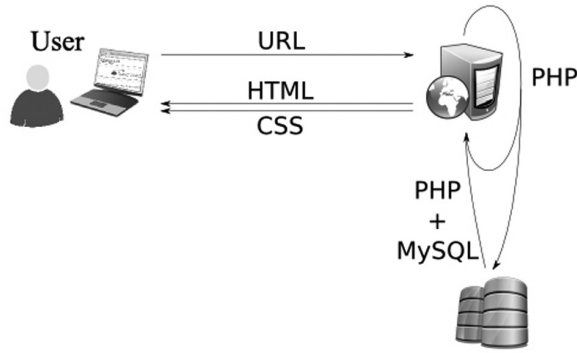


FIGURE 1. HTML-PHP-MySQL interface.

monograph on South Pacific foraminifera (Albani et al., 2001).

The choice for PHP and MySQL was made as they are popular in database management for the Internet. Most Internet providers offer inexpensive PHP/MySQL solutions. It is Open Source with many programmers working on its functionality and giving support in forums and webpages such as php.net (Arntzen et al., 2015). The many active programmers assure sustainability and readiness for the future. It is free and gives HTML output. PHP and MySQL are easy to learn, for example with the video tutorial by K. Skoglund (Skoglund, 2008).

FEUDAT runs with PHP and MySQL as a server side dynamic webpage, as illustrated in Figure 1. The server is run by FEUPRO, and FEUDAT is accessible only via Internet. The user is not obliged to run other software such as Java or JavaScript but just needs an HTML browser and Internet access. The user enters the data request into a webpage form containing PHP code. The PHP code is hidden to the user. The user request is sent to the server, which runs a PHP interpreter. The PHP code contains MySQL query statements, which are passed by the interpreter to the MySQL database. The MySQL database server sends back the requested data to the PHP interpreter, which transfers the result into a predesigned HTML output page. The HTML output page is shown to the user, with the PHP code hidden.

The HTML-PHP-MySQL tool is not only used for user query pages but also to create the standardized pages such as on the single specimen (Figure 2), species (Figure 3), genus (Figure 4), and contributor (Figure 5). Changes in the data are automatically reflected in all output pages as they retrieve the data from the database.

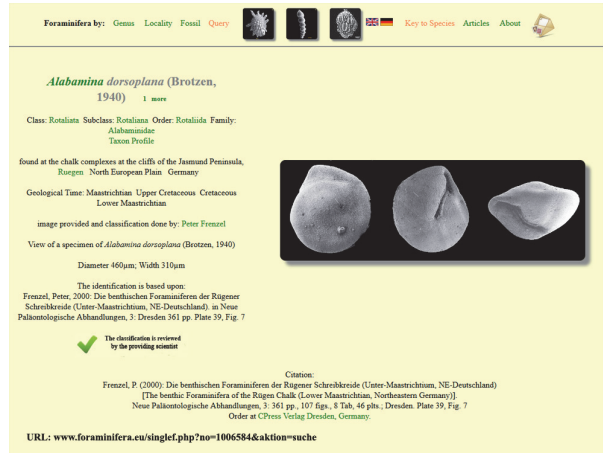


FIGURE 2. Single specimen page.

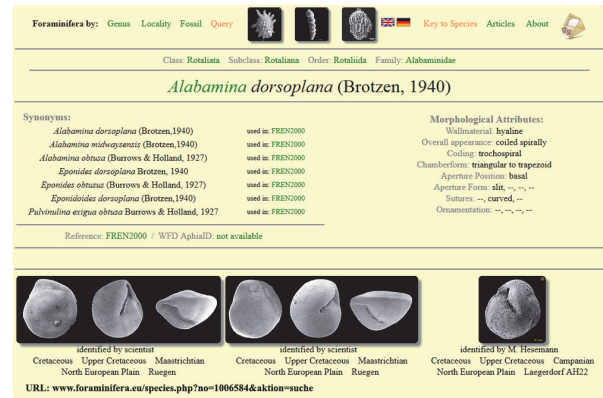


FIGURE 3. Species page detail.



FIGURE 4. Genus page detail.



FIGURE 5. Contributor page detail.

Data entry and editing is done by uploading the text files into spreadsheet programs. The data records are exported to a text file, which is uploaded to the MySQL database.

FEUDAT may also be installed and run locally on a single computer or within a network for example with XAMPP. XAMPP is a completely free, easy to install Apache distribution containing MySQL, PHP, and Perl (Apache Friends, 2015).

DISCUSSION OF THE FEUDAT DESIGN

FEUDAT started with the intention of showing the specimens in the collections of the FEUPRO team members. The single specimen was arbitrarily chosen as the basic data element. It turned out that this anchor is not only helpful as a starting point but also useful when building a more sophisticated database on foraminifera. The single specimen found in a specific sample with good documentation on its geographical and stratigraphical positioning is an incontrovertible element of information. While its classification, stratigraphical assignment, original life environment, and even its morphological description may change over time and differ amongst authors, the specimen itself persists.

Illustration

A single specimen, even if housed in a public collection, may not be easily accessed by interested people. Although its illustration may be easily distributed. A good illustration is seen as a key element for FEUDAT. Three dimensional (=3D), high quality Illustrations allowing 3D printing are seen

as the best available technical solution. However due to the high production costs, 3D images of foraminifera are rare and just a few provided by the American Museum of Natural History in New York will be integrated. All other illustrations used are two-dimensional. Two to three views are needed to get an idea of the overall morphology. Details of the surface, aperture, and ornamentation may need to be illustrated separately. To image, the internal structures will need special techniques, and the specimen itself may need to be sectioned and partially or even completely destroyed. FEUPRO cannot always produce perfect illustrations of specimens for FEUDAT. For example the scanning electron microscope available to us requires fixing and coating of the specimens, which hinders the imaging from different views. Being aware of the limitations, workarounds are established. If available, several specimens of the same kind in a sample will be imaged. Illustrations will be added even if a species is already well documented. Optical imaging and drawings are encouraged by sharing material with relevant experts and by providing free of charge training on optical imaging and stacking procedures. Written descriptions are commonly used to point out key morphological features, differences between closely related species, and as a workaround for unavailable illustrations. They are used in FEUDAT only to a limited extent so far and are discussed in the section on morphological attributes.

To give an overall and instant impression, different illustrations of a single specimen are arranged into one image or in some cases links lead in a loop to more images of the same specimen. When multiple images of one species are presented, the images are commonly ranked by type specimen, those identified by a professional scientist and then those identified by an avocational scientist. If a user chooses to get only one or two images, a good image of a type specimen and/or the most illustrative image(s) are shown.

Requirements for the Data

As mentioned earlier, the content of FEUDAT consists of illustrations of foraminifera and the data associated with the illustrated specimens. A data record is incorporated if good illustrations and the required data are available. The requirements for the data need to be balanced between the demands of the users, the efforts to obtain the data, and technical limitations of the database system used. These aspects will be discussed starting with the demands of the users.

TABLE 1. Useful data attributes as seen by different user groups and availability. Interest: *** high; ** medium; * little; - no; ? needs to be investigated.

Attribute	Beginner	Avocational Scientist	Student BSc, MSc, PhD	Senior Scientist	Collection Manager	Contributor	Availability
IDENTIFICATION	***	***	***	***	***	***	***
Genus	***	***	**	**	***	***	***
Species Author, Year	***	***	***	***	***	***	***
Higher classification	**	**	**	*	*	**	**
Synonyms Genus and Species	*	*	***	**	**	**	*
Source of Identification	*	***	***	**	*	***	*
Importance of specimen e.g., type-specimen	*	**	***	***	***	***	***
KEY TO SPECIES	**	**	***	*	**	***	**
Wall material	*	**	***	*	**	***	**
Chamber arrangement	**	**	***	*	**	***	**
Text description	***	***	**	**	**	***	*
More detailed features	**	**	***	*	**	***	*
VISUALIZATION	***	***	**	**	**	***	*
Illustration quality + amount	***	***	**	**	**	***	**
TIME RANGE	**	**	**	*	*	**	**
period	***	***	***	**	***	***	**
age, stage formation	***	***	***	**	***	***	** *
formation	-	**	***	***	**	***	*
LOCALIZATION	**	***	***	***	***	***	*
Locality of the Sample	***	***	***	***	***	***	***
Geographical distribution	**	***	***	***	**	**	-/?
Faunal distribution	*	***	***	***	*	**	-/?
TECHNICAL							
Collection related data	-	*	*	**	***	***	**
Illustration type	*	*	*	*	*	*	***

Useful Data Attributes

A data attribute is seen as useful if it structures the data in a way which potential users may ask for. Potential users are all people who are interested in foraminifera, from the novice, avocational scientist and student to collection managers, senior scientists, publishers, and artists. They may live in a highly developed country with easy access to all literature and information or in a less developed country with little access to publications. Users may be contributors or just consumers. They

may want to get background information or find citable sources. The attributes need to be chosen taking all these different requirements into consideration. Data on user requirements are compiled in Table 1. The data are based on discussions amongst the FEUPRO team, personal conversations and emails with contributors representing different groups, a questionnaire, and the analysis of website accesses. They are presented as soft data only.

TABLE 2. List of attributes. Attributes used for the technical database management are not listed. ¹values are not set for all data-records; ²only used for planktonic species; ³only used for benthic species.

Taxonomy	Distribution and Geological Time	Morphology	Morphology continued	Technical and literature
Class	Period	Wall material	Overall appearance ^{1 3}	Collection / Contributor
Subclass	Epoch	Coiling	Wall material ^{1 3}	Inventory number ¹
Order	Age ¹	non-/spinose ^{1 2}	Chamber Arrangement ^{1 3}	Type of illustration
Family	Formation ¹	Wall Micro-structure ^{1 2}	Chamber Form ^{1 3}	Amount of images to be shown
Genus	First Occurrence Date ^{1 2}	Wall Texture ^{1 2}	Position of primary aperture ^{1 3}	Author(s) / Editor(s)
Species	Last Occurrence Date ^{1 2}	Chamber Arrangement ^{1 2}	Aperture Form ^{1 3}	Title
named by	Ocean Continent	Height of Spire ^{1 2}	Type of Sutures ^{1 3}	Year
Genus ¹ (Synonym)	Area	Umbilical Depth ^{1 2}	Ornamentation ^{1 3}	Journal (Volume, Issue)
Species ¹ (Synonym)	Country	Chamber Form ^{1 2}	Projections ^{1 3}	Publisher ¹
named by ¹ (Synonym)	Locality	Keel ^{1 2}		Publication deals with epoch
Reliability of Identification	Remarks on Locality	Tooth Teeth ^{1 2}		
Type specimen	benthic	Lip ^{1 2}		
Reference (see Literature)	planktonic			

Consistency of the Database

Only a consistent database will produce a reasonable output on a query. One aspect of consistency is the design and structure of the attributes. The possible values for an attribute should be unique and precisely differentiated from each other. FEUDAT uses sets of valid values for each attribute. The sets are presented to users as drop-down lists to choose from. Another aspect of the consistency is the completeness of data in each data record. If a data record has no value for an attribute, it will not be found when searching this attribute. The output would be inconsistent and may be even misleading for a user. For this reason FEUDAT incorporates in the first place attributes for which data for all specimens are available or may be available with little effort. They are shown in the user interface in black color. Attributes that are seen as useful, but not fully available are treated as additional information only and are

shown in grey. A list of attributes currently incorporated is given in Table 2.

SINGLE ATTRIBUTES

Classification and Identification

The nomenclature and data on foraminiferal classification are seen as most useful by users and are available for every data record. They form the basis of foraminiferal research and communication. FEUDAT incorporates the following classification attributes: class, subclass, order, family, genus, species, author and year, synonyms, and reliability (of identification). Addressing values to these attributes for a single specimen.– the single data record.– depends on an agreed nomenclature and specialists who are capable of a proper identification. Both are only available to a limited extent.

There is no consensus on the higher classification of foraminifera (Hayward et al., 2015). As for all organisms there is no single authority for the

nomenclature of the estimated 80,000 taxa ever described. For a database it means that structured information is not fully available. Different names are used by different and even by the same author for the same taxon. For example the Cretaceous taxon *Gavelinella pertusa* (Marsson, 1878) has at least 16 synonyms (Frenzel, 2000). FEUPRO supports the initiative of Bruce Hayward and others to establish an authority on nomenclature with the World Foraminifera Database (WFD) (Hayward et al., 2015). In FEUDAT the nomenclature of WFD is used for the attributes family, genus, species (including synonyms), author, and year. As of July 2015, data on about 10,500 taxa are available in WFD. For taxa not yet covered by WFD the nomenclature in publications of recognised experts is used. Synonyms are incorporated and presented on the single species pages (Figure 3). As of 2015 synonyms are not searchable.

Until now, for the higher classification (class, subclass, and order), FEUDAT has applied the concept of Valeria I. Mikhalevich: "In contrast to the majority of classification schemes adopted, the basic morphological characteristics of the test are considered as having preferential significance compared with the character of the shell wall. The latter having important but subordinate value" (Mikhalevich, 2004). She concluded from her investigations on foraminiferal evolution that the overall appearance and chamber arrangement is the predominant feature in foraminiferal evolution whereas the wall composition is subordinate. For avocational scientists and beginners, overall appearance is easier to use and recognizable with a specimen at hand than wall material. With the same intention, Debenay in his guide on New Caledonian foraminifera presented the classification as "the morphology as it appears at first sight" despite of it "might be puzzling to specialist" (Debenay, 2012). In 2016, it is planned to switch to the WFD concept on higher classification. It complies with the new supraordinal classification derived from genetic sequence analyses (Pawlowski et al., 2013).

Morphological Attributes

The classification of foraminifera has been largely constructed on the basis of shell morphology, the nature of the wall, overall appearance, chamber arrangement, and specific morphological features, as pointed out in Loeblich and Tappan, 1988. Since the 1990s DNA sequencing has been introduced to the study of foraminifera. The analyses of molecular data of extant species show that

morphological features do not simply correlate with the DNA and vice versa. Morphologically distinct specimens are found to be closely related in their DNA as in a few extant unilocular lineages (Pawlowski et al., 2003). Specimens distinct in their DNA can show a similar morphology as exemplified in extant *Orbulina universa* d'Orbigny, 1839 (de Vargas et al., 1999). It is concluded that the morphological classification has a limited value in an evolutionary and biological sense (Pawlowski et al., 2013). Lineages built on changes in morphology over time may not reflect evolutionary processes. DNA data are not available for extinct species and thus are lacking for more than 90% of all species. Epigenetic studies of other organisms show that cellular and physiological processes are driven only to a limited extent by DNA sequences (Simmons, 2008). It is unknown to which degree test building and morphology in foraminifera are affected. Foraminiferal tests and their morphological attributes do form a stable, timeless, and comprehensive source of data, regardless of alterations in their biological interpretation. This comprehensive set of data needs to be clearly separated from its biological interpretation.

In FEUDAT the morphological attributes are primarily used to build two keys to species, one for benthic and one for planktonic foraminifera. The keys are made for non-experts and described later. The intention is to find a balance between ease of use and getting a result. Accordingly, only a limited number of the many available attributes are used. The attributes are chosen from the taxonomic parts of several publications (Kennett and Srinivasan, 1983; Olsson et al., 1998; Frenzel, 2000; Kaminski and Gradstein, 2005; Pearson et al., 2006; Petrizzo et al., 2011; Debenay, 2012). Following requests from beginners, an attribute 'overall appearance' is created and used. The morphological attributes are also shown on the single species pages (Figure 3). It is intended to add specific text descriptions on the single species pages after 2020.

The attributes used for benthic species are overall appearance, wall material, chamber arrangement, chamber form, position of the primary aperture, aperture form, sutures, and ornamentation features (hispid/pustulose, striae/costae/reticulate, spinose, spinose projections, fistulose). The attributes used for planktonic species are spinose/non-spinose, wall microstructure, wall texture, chamber arrangement, height of spire, umbilical depth, chamber form on the umbilical

side, keel, lip, and tooth. Thumbnail illustrations are given in the keys to species (Figures 6 and 7).

To keep it simple for the user, a single data record – a specimen – may only carry one value for each attribute, and the number of values is limited. In a few cases, due to these limitations, values may overlap and are not clearly differentiated from each other. A choice has to be made between usability and exactness. The established attributes and values are applicable to all data records and meet the requirement of a high availability. They are seen as most useful by non-experienced users and to a lesser extent by experts. As of July 2015 FEUDAT contains values on morphological attributes for 632 benthic and 142 planktonic species.

Geological Time

For geological time FEUDAT applies the attributes period, epoch, and stage. They allow narrowing down the results of a selection on the time scale. The values comply with those of the International Chronostratigraphic Chart v 2015/01 of the International Commission on Stratigraphy (Cohen et al., 2015). The data for each record are derived from the positioning of the sample after excluding reworking.

The attributes last and first occurrence dates have been applied so far to 142 planktonic species, mainly of Neogene age. It is planned to enlarge the coverage. The data are extracted from (Kennett and Srinivasan, 1983; Olsson et al., 1998; Pearson et al., 2006; Gradstein et al., 2012; Chronos, 2015). The data are accessible through the key to planktonic species as described later.

Localization and Distribution

The distribution of foraminifera may be split into geographical and ecological distribution. FEU-PRO has so far not managed to find meaningful attributes on both kinds of distribution. A major obstacle is the lack of data. Thousands of publications deal with the geographical distribution of recent foraminifera. They range from worldwide (e.g., Murray, 2006), regional (e.g., Culver and Buzas, 1982, 1985), to local investigations (e.g., Polovodova et al., 2009). Nonetheless, an estimated 80% of the needed data values were seen as missing. Based on a split of the world's oceans into 500 subareas and an estimated 4,000 modern species the amount of values needed would be 2,000,000. Some authors estimate the number of modern species as up to 10,000 (Lipps and Finger, 2010; Debenay, 2012). The WFD is making the effort to report the distribution range of each mod-



FIGURE 6. Key to benthic species entry page.

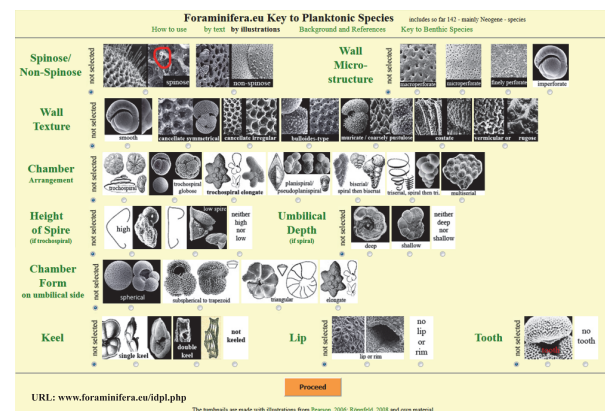


FIGURE 7. Key to planktonic species entry page.

ern species. The problem is that the task is enormous, and the listed distributions are far from comprehensive. For example, even though there are 60 distributional records for *Ammonia beccarii* (Linnaeus, 1758) in WFD, most of the occurrences we know from our samples are not mentioned. The data availability on the geographical distribution of fossil foraminifera is much poorer. Also, the question has to be asked whether to add data on their distribution in ancient oceans or in modern fossil sediments. Another obstacle is that foraminiferal distribution can change over very short time periods. A prominent example is that of invasive species such as *Amphistegina* spp. in the Mediterranean Sea (Langer et al., 2012; Caruso and Cosentino, 2014).

So far, FEUDAT uses only distributional data values based on the samples in which the imaged specimens were found. The information content



FIGURE 8. Localities entry page.

may be seen as poor as it is far from giving an overview of the real distribution of a species. The data are fully available and consistent. Besides locality, the attributes area and continent/ocean are established in order to offer more choices to the user. An interface based on Google Maps is used to visualize the position of the localities on a map of the world allowing zooming in and out (see Figure 8).

It is intended to improve the information content on the geographical and ecological distribution of recent foraminifera by splitting both the geographical and ecological distributions into only about 50 areas. The ecological areas will be derived from six water depth related zones multiplied by six latitude related zones plus sediment related factors. The problem with this approach is, of course, that the occurrence of a species in a single locality will be presented as though it occurs throughout the whole geographic or ecological area. Furthermore it is not clear, whether or how the dominance of a species can be integrated into FEUDAT.

Technical Attributes

Technical attributes comprise collection related data, type of illustration, and attributes needed to sort and limit the data output. The collection related data are added in according to the requirements of the contributors. For example, the American Museum of Natural History, New York, and the Smithsonian Institution National Museum of Natural History, Washington, asked that the number from their microfossil catalogue be shown on the single specimen pages. Entry pages are created for each incorporated monograph, which allows the user to sort the output amongst other choices to "as in the publication."

The attribute type of illustration is primarily added for contributors, who may produce SEM or optical images, drawings, images of thin sections, and anaglyph 3D images. They will get an idea about the quality required and learn about the limitations of each illustration method. A small percentage of users are more interested in the technical issues of illustrating tiny objects rather than in foraminifera per se. Some offer their expertise in return for having this attribute to choose from. The attribute amount of images allows the

TABLE 3. List of material (illustrations and data) integrated as of July 2015. SM= Scientific Monograph; TSC= Type Specimen Collections; SA=Scientific Articles; PS=from processed samples.

Type	Contributor/Publication	Subject	Data records
SM	Bossard, D. 19thcenturyscience.org.: (Brady, 1884)	recent	1,479
SM	Frenzel, P.: (Frenzel, 2000)	Benthics, Cretaceous	557
SM, SA	Geological Survey of Austria: (Papp and Schmidt, 1985), (Gebhardt, 2010)	Miocene, type specimens	519
SA	Grzybowski Foundation: (Setoyama et al., 2011), (Szarek et al., 2000), (Neagu, 2000)	Agglutinates	~390
TSC	American Museum of Natural History, New York	type specimens	270
TSC	Geological Survey of Austria, Vienna: (Gebhardt, 2010)	type specimens	195
TSC	Smithsonian Institution National Museum of Natural History, Washington	type specimens	111
SA	IOPD, ODP and DSDP scientific results publications	all	820
SA	various professional junior to senior scientific contributors	all	~1,700
PS	various avocational and professional scientists, own samples	all	~3,770
		TOTAL:	9,800+

user to narrow down the output to the most relevant illustrations. Other implemented technical attributes that serve for internal database purposes or minor topics will be not described here.

Quality Control

A main challenge for the quality control of FEUDAT data is to get professional foraminiferologists involved. Researchers are evaluated by the number and citations of their publications. Participating in online database initiatives, especially if run by avocational scientists, results in losing valuable "publishing time" (Dolven and Skjerpen, 2011). FEUPRO has not managed to get professional scientists involved in checking data on a regular basis although we have managed to obtain permission to use images and data from publications of professional scientists and institutions including those on type specimens. These data records are peer reviewed and account as of July 2015 for more than 60% of all records. A list of major sources is given in Table 3. It is intended to increase the share of data records from peer-reviewed publications to 75%. They are checked once for errors in data extraction from the source. Less than 40% of the data records are derived from activities of avocational scientists, namely our own sampling, extraction of foraminifera and imaging. The quality control of these data is conducted by members of the FEUPRO team— all avocational scientists—in accordance with the four eyes principle and relevant literature at hand. The attributes family, genus, species, and name giver are checked on entry against the WFD (Hayward et al., 2015). Once a year in December, the validity of

genus and species names is checked against the WFD using their automated check.

Reliability

The reliability of data largely depends on the expertise of those who generate them. The attribute reliability is used to categorize each data record as: type specimen, identified by professional scientist or identified by M. Hesemann. This attribute gives the user the opportunity to choose and for example exclude data records with the value identified by M. Hesemann. On the species pages the results are sorted by type specimen, identified by professional scientists and identified by M. Hesemann (Figure 3).

NEW KEYS TO SPECIES

Two new keys to species each for benthic and planktonic species have been developed and implemented. They combine the ease of "the morphology as it appears at first sight" approach (Debenay, 2012) with the capabilities of an interactive computer interface. The attributes with thumbnail images of each value are arranged in rows (Figure 6 and Figure 7). A user is obliged to compare the different values for an attribute and choose one for the specimen to be identified. A choice for three attributes will give a reasonable result and allows choosing only those values observed with confidence. Within seconds the user will see a results page arranged as table with images of taxa meeting the chosen criteria. This two-step concept of a key to species pointing to a table of possible results was used by Cushman in

1928 (Cushman, 1928). From here the user may click on each image to view the species page with more details. The user may also narrow down the results from a text-based interface on top of the results page. The text interface adds new choices such as in the key to planktonic species a time interval for first and last occurrence dates.

OUTLOOK

With its highly motivated team, 120 contributors and more than 5,000 more illustrations from scientific publications to be integrated, FEUPRO will continue to enlarge the coverage of FEUDAT in space and time. It is planned to add 1,500 data records each year. A priority will be the integration of illustrations of type specimens and from the publications of renowned senior scientists. Images of lesser significance, for example multiple images of the same species and those identified by junior or avocational scientists are seen as important, too, and will be added to the database. The mission of FEUPRO is to foster the interest in foraminifera, help those persons without adequate imaging equipment, and to show the results on a contributor's page as a motivation. It is seen as essential by the FEUPRO team not just to fill a database but also to work in the field, on samples, specimens, and to digest relevant publications. Efforts will continue to structure and integrate data on the geographical, faunal, and stratigraphical distribution of foraminifera.

It is encouraging and noteworthy that none of the people involved receive any financial compensation for their work and are driven purely by their enthusiasm as naturalists. They will be the driving force for new features and the allocation of resources.

SUMMARY

Based on arbitrarily received material from avocational and professional scientists (samples, illustrations, and data) FEUPRO has built the foraminiferal database FEUDAT. As of July 2015, it contains 9,800 data records on foraminiferal specimens, each consisting of an illustration and data values for 57 attributes. The attributes are chosen based on user demands and availability. They comprise taxonomical, geographical, stratigraphical, morphological, technical, and collection related information. More than 60% of the data records come from peer-reviewed scientific publications. Two new keys for the identification of benthic and

planktonic foraminifera have been developed and cover more than 780 taxa.

The database is freely accessible on the internet through user-oriented interfaces. The main interface allows the user to choose from 24 attributes and get an instant result as a set of thumbnail images with accompanying data in the form of a plate. By clicking on a single thumbnail the user is directed to the webpage for the single specimen where a bigger image and all available data are displayed. Special interfaces are created for queries based on localization, contributors, collections, and to serve as keys to species.

A motivated team, 120 avocational and scientific contributors and more than 5,000 images from scientific sources waiting to be integrated assure the enlargement of the database in space and time.

ACKNOWLEDGMENTS

The publication is inspired by B. Ottway, whom I am most grateful for reviewing different stages of the manuscript and suggesting numerous improvements. I am grateful to the other members of the foraminifera.eu project team K.-O. Bock, D. Ketelsen, S. Raveling, and G. Rosenfeld for the mutual work over many years on hundreds of samples. Special thanks go to D. Ketelsen for critically reviewing dozens of identifications based on the digestion of a multitude of publications. B. Ottway and D. Ketelsen continue to give detailed feedback on the features and interfaces of FEUDAT. I have to thank the 120 avocational and scientific contributors to the foraminifera.eu project for their overwhelming support, provision of samples, image rights, and feedback. I am very grateful to W. Roennfeld, who permits the usage of his excellent drawings in the keys to species (Roennfeld, W., 2008). The author thanks the two anonymous reviewers whose comments helped to improve the original manuscript.

REFERENCES

- Albani, A., Hayward, B.W., and Grenfell, H.R. 2001. Taxomorph: Foraminifera from the South-west Pacific. An inter-active catalogue, CD-ROM. University of New South Wales and Australian Biological Resources Study.
- Apache Friends 2015. XAMPP. Accessed at <https://www.apachefriends.org/index.html> on 2015-05-25.
- Arntzen, T.C., Bakken, S., Caraveo, S., and Gutmans, A. 2015. PHP.net. Accessed at php.net on 2015-05-25.
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., and Shirk, J. 2009. Citizen Science: A Developing Tool for Expanding Science

- Knowledge and Scientific Literacy. *BioScience*, 59:977-984.
- Brady, H.B. 1884. Report on the foraminifera dredged by HMS Challenger, during the years 1873-1876. *Report on the Scientific Results of the Voyage of HMS Challenger*, *Zoology* 9:1-814.
- Caruso, A. and Cosentino, C. 2014. The first colonization of the Genus *Amphistegina* and other exotic benthic foraminifera of the Pelagian Islands and south-eastern Sicily (central Mediterranean Sea). *Marine Micropaleontology*, 111:38-52.
- Chronos 2015. Chronos MesozoicDB. Accessed at portal.chronos.org/gridsphere/gridsphere.jsessionid=D5D17C0C530157DD9CDC71CAE3CFB5B1?cid=res_taxondb on 2015-05-22.
- Cohen, K.M., Finney, S., Gibbard, P.L., and Fan, J.-X. 2015. The ICS International Chronostratigraphic Chart 2015/01. IUGS.
- Culver, S.J. and Buzas, M.A. 1982. Recent benthic foraminiferal provinces between Newfoundland and Yucatan. *Geological Society of America Bulletin*, 93:269-277.
- Culver, S.J. and Buzas, M.A. 1985. Distribution of Recent Benthic Foraminifera off the North American Pacific Coast from Oregon to Alaska. *Smithsonian Contributions to the Marine Sciences*, 26:1-234.
- Cushman, J.A. 1928. Foraminifera, their classification and economic use. *Special Publ. No. 1, Cushman Lab. Foram. Res.* 1-401.
- Davis, M. and Phillips, J. 2006. *Learning PHP and MySQL*. O'Reilly, Sebastopol, USA.
- De Clerck, O., Guiry, M.D., Leliaert, F., Samyn, Y., and Verbruggen, H. 2013. Algal taxonomy: a road to nowhere? *Journal of Phycology*, 49:215-225.
- de Vargas, C., Norris, R., Zaninetti, L., Gibb, S.W., and Pawlowski, J. 1999. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl. Acad. Sci. USA*, 96:2864-2868.
- Debenay, J.P. 2012. *A Guide to 1,000 Foraminifera from Southwestern Pacific: New Caledonia*. Institut de recherche pour le développement, Marseille.
- d'Orbigny, A. 1839. Foraminifères, 224 pp.. In De la Sagra, R.M. (ed.), *Histoire physique, politique et naturelle de L'île de Cuba*. Bertrand, A., Paris.
- Dolven, J.K. and Skjerven, H. 2011. Paleoinformatics: Past, Present and Future Perspectives, p. 45-52. In Elewa, A.M.T. (ed.), *Computational Paleontology*. Springer, Heidelberg.
- Ellis, B.F. and Messina, A. 1940-2014. *Catalogue of foraminifera*. Micropaleontology Press, American Museum of Natural History, New York.
- Encyclopedia of Life 2015. EOL. www.eol.org on 2015-05-25.
- Flanders Marine Institute VLIZ 2015. Foraminifera Photogallery MARBEF. www.marbef.org/modules.php?name=Photogallery&album=418 on 2015-05-25.
- Frenzel, P. 2000. *Die benthischen Foraminiferen der Ruegener Schreibkreide (Unter-Maastrichtium, NE-Deutschland)*, 3. Neue Palaeontologische Abhandlungen, Dresden.
- Frontalini, F. and Coccioni, R. 2007. Benthic foraminifera for heavy metal pollution monitoring: A case study from the central Adriatic Sea coast of Italy. *Estuarine, Coastal and Shelf Science*, 76:404-417.
- Gebhardt, H. 2010. Foraminiferal holotypes, lectotypes, and syntypes stored in the paleontological collection of the Geological Survey of Austria. *Jahrbuch der Geologischen Bundesanstalt*, 150:245-262.
- Gradstein, F.M., Ogg, J.G., Schmitz, M.D., and Ogg, G.M. 2012. *The Geologic Time Scale 2012, 2-volume set*. Elsevier, Oxford, U.K.
- Hayward, B.W., Cedhagen, T., Kaminski, M., and Gross, O. 2015. World Foraminifera Database. www.marine-species.org/foraminifera on 25 May 2015.
- Hesemann, M. 2010. Concept for a foraminiferal database. In Bonn, O.C.i. (ed.), *Forams 2010, International Symposium on Foraminifera*. Rheinische Friedrich-Wilhelms-Universitaet, Bonn.
- Hesemann, M., Ottway, B., and Legoinha, P. 2014. The Foraminifera.eu Project - A paleoinformatics bridge between science and community. *Comunicações Geológicas*, 101, Especial I:447-450.
- Jadhav, A.S. and Sonar, R.M. 2009. Evaluating and selecting software packages: A review. *Information and Software Technology*, 51:555-563.
- Jones, R.W. 1994. *The Challenger Foraminifera*. Oxford University Press, Oxford, U.K.
- Kaminski, M.A. and Gradstein, F.M. 2005. Atlas of Paleogene cosmopolitan deep-water agglutinated foraminifera. *Grzybowski Foundation Special Publication*, 1-547.
- Kennett, J.P. and Srinivasan, M.S. 1983. *Neogene planktonic foraminifera. A phylogenetic atlas*. Hutchinson Ross Publishing Company, Stroudsburg.
- Langer, M.R., Weinmann, A.E., Lotters, S., and Rodder, D. 2012. "Strangers" in Paradise: Modeling the biogeographic range expansion of the foraminifera *Amphistegina* in the Mediterranean Sea. *Journal of Foraminiferal Research*, 42:234-244.
- Linnaeus, C. 1758. *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, 1, 10th edition*. L. Salvii, Holmiae, Stockholm.
- Lipps, J.H. and Finger, K.L. 2010. How many foraminifera are there ?, pp. 132. In Langer, M., Weinmann, A., Söntgerath, B., Göddertz, P., and Heumann, G. (eds.), *Forams 2010, International Symposium on Foraminifera*. Organizing Committee in Bonn, Bonn.
- Loeblich, A.R. and Tappan, H. 1988. *Foraminiferal genera and their classification*. Van Nostrand Reinhold, New York.
- Löser, H. 2004. PaleoTax - a database program for palaeontological data. *Computer and Geosciences*, 30:513-521.

- Marine Biological Laboratory 2015. Micro*Scope. pinkava.asu.edu/starcentral/microscope/ on 2015-05-25.
- Marsson, T. 1878. Die Foraminiferen der weissen Schreibkreide der Insel Rügen. Mitteilungen des naturwissenschaftlichen Vereins Neu-Vorpommern und Rügen in Greifswald. 10:115-196, Berlin.
- Mikhalevich, V.I. 2004. On the heterogeneity of the former Textulariina (Foraminifera), p. 317-349. In Bubik, M., and Kaminski, M. (eds.), *Proceedings of the Sixth International Workshop on Agglutinated Foraminifera*. Grzybowski Foundation Special Publication, Kraków.
- Murray, J.W. 2006. *Ecology and applications of benthic foraminifera*. Cambridge University Press, Cambridge.
- Neagu, T. 2000. Lower Cretaceous calcareous agglutinated Foraminifera from Southern Dobrogea, Romania. Part II. Early Cretaceous Cuneolinidae, p. 363-386. In Hart, M.B., Kaminski, M.A., and Smart, C.W. (eds.), *Proceedings of the Fifth International Workshop on Agglutinated Foraminifera*. Grzybowski Foundation Special Publication, Kraków.
- Nixon, R. 2012. *Learning PHP, MySQL, JavaScript, and CSS: A Step-by-Step Guide to Creating Dynamic Websites*. O'Reilly Media, Sebastopol, USA.
- Olsson, R.K., Hemleben, C., Berggren, W., and Huber, B.T. 1998. Atlas of Paleocene planktonic foraminifera. *Smithsonian Contributions to Paleobiology* 1-252.
- Papp, A. and Schmidt, M.E. 1985. Die fossilen Foraminiferen des Tertiären Beckens von Wien. Revision der Monographie von Alcide d'Orbigny (1846). *Abhandlungen der Geologischen Bundesanstalt*, 37:1-311.
- Pawlowski, J., Holzmann, M., Berny, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., and Bowser, S.S. 2003. The evolution of early Foraminifera. *Proceedings of the National Academy of Sciences*, 100:11494-11498.
- Pawlowski, J., Holzmann, M., and Tyszka, J. 2013. New supraordinal classification of Foraminifera: Molecules meet morphology. *Marine Micropaleontology*, 100:1-10.
- Pearson, P.N., Olsson, R.K., Huber, B.T., Hemleben, C., and Berggren, W.A. 2006. Atlas of Eocene Planktonic Foraminifera. *Cushman Foundation Special Publication*, 41.
- Petruzzo, M.R., Premoli Silva, I., and Verga, D. 2011. Practical Manual of Cretaceous Planktonic Foraminifera, New Revision. *International School on Planktonic Foraminifera, 3rd Course 2004*.
- Polovodova, I., Nikulina, A., Schoenfeld, J., and Dullo, W.-C. 2009. Recent benthic foraminifera in the Flensburg Fjord (Western Baltic Sea). *Journal of Micropalaeontology*, 28:131-142.
- Roennfeld, W. 2008. *Foraminiferen, ein Katalog typischer Formen, 3rd edition*. Self published, Tuebingen.
- Setoyama, E., Kaminski, M.A., and Tyszka, J. 2011. Late Cretaceous Agglutinated Foraminifera and Implications for the Biostratigraphy and Palaeobiogeography of the southwestern Barents Sea, p. 251-309. In Kaminski, M.A., and Filipescu, S. (eds.), *Proceedings of the Eighth International Workshop on Agglutinated Foraminifera*. Grzybowski Foundation Special Publication 16, Kraków.
- Simmons, D. 2008. Epigenetic influence and disease. *Nature Education*, 1(1):6.
- Skoglund, K. 2008. PHP with MySQL Essential Training. Lynda.com.
- Szarek, R.K., Klosowska, B., Prokoph, A., Kuhnt, W., and Wagner, T. 2000. Upper Albian agglutinated foraminifera of Northeast Germany, p. 445-463. In Hart, M.B., Kaminski, M.A., and Smart, C.W. (eds.), *Proceedings of the Fifth International Workshop on Agglutinated Foraminifera*. Grzybowski Foundation Special Publication, Kraków.
- The Global Biodiversity Information Facility 2015. GBIF. www.gbif.org/ on 2015-05-25.
- Zalecka, K., Wrzolek, T., and Granier, B. 2015. Simple and practical techniques to manage small databases, illustrated by a case study: bibliographic data from the "Fossil Cnidaria & Porifera" newsletter (1972-2010). *Carnets de Géologie / Notebooks on Geology*, 15(2):13-19.
- Zicha, O., Hrb, J., Manas, M., and Novak, J. 2015. BioLib.cz. www.biolib.cz/en/main/ on 2015-05-25.