PalyWeb:
A “lazzarus” palynomorph database project on the web

by Sophie WARNY, Philippe STEEMANS, Pierre BREUER, and Elodie PETUS

Key words: palynology; database; internet

Introduction

Since the sixties, palaeopalynology has been experiencing a spectacular growth in publication. Already by the end of 1966, KREMP (Lentin et al., 1996) had calculated that about 330 new articles on palynology were published each year in more than 200 journals, and about 4200 had already been published. In a paper published during the 90’s, Jansonius & McGregor (1996) made an estimation of the amount of palynological papers published each year. Although they noticed that the data set is incomplete, they estimated that about 400 new articles are published a year. One of the unfortunate consequences of the abundance of published information is a taxonomic chaos for some palynological groups. In addition, numerous palynomorphs have been published invalidly, some with redundant names. Several taxa have been described in publications with a limited distribution through different countries or in languages not accessible to most palynologists.

Adding to this problem is the decrease of palynological laboratories around the world, and the loss of some important palynological collections. When a curator or professor of palynology retires or passed away, and is not being replaced, one has to worry about the preservation of the palynological database established at that specific institute, university or museum.

Palynologists around the world are increasingly convinced that establishing a broad access database with open access for research and education is urgently needed to preserve our field of research, and to educate and serve future generations of palynologists. Several of us already have our own curation programs, but these data are not accessible unless you work with this specific group. We need to create open access online datasets to not only manage the mountain of data published in the past and in years to come, but also to increase the accessibility of existing collections. Such dataset already exist, for instance PALYNODATA, DinoSys, Dinoflaj, etc. and these are well publicized. But others are for the most part not as well known by the community, such as the Palaeozoic-centric database Phytopal for the acritarchs and ChitinoVosp for the Chitinozoans.

One possibility that we would like to bring to the attention of the palynological community is to adopt a technology that any one of us could access for free, wherever we are located in the world; i.e., the Wiki technology.

The WikiMedia technology

WikiMedia is a free software originally written for the well-known Wikipedia encyclopaedia freely accessible on internet (http://en.wikipedia.org/wiki). This technology is increasingly part of the education in countries such as the U.S., as it supports projects of the non-profit Wikimedia Foundation (http://en.wikipedia.org/wiki/Wikimedia_Foundation) as well as many other wikis. The term wiki is a shorter form of « wiki wiki » which is derived from the native language of Hawaii, meaning something quick or fast. On the internet, a wiki is a type of website that allows users to easily add, remove, or edit all content, very quickly and easily.

With this in mind, PalyWeb was created. PalyWeb is a palynomorph database project based on WikiMedia technology. It is a web based free-content palynological database in open access. To our knowledge, it is the only online database using this mode of data acquisition and management in palynology. It is also the only project aiming to collect datasets on all fossilised palynomorph taxa. Its conception is similar to the Wikipedia encyclopaedia and, therefore, can be considered as a wiki website. The ease of interaction and operation makes a wiki website an effective tool for collaborative scientific writing and education. PalyWeb is designed to be a database filled collaboratively by volunteers, allowing most articles to be changed by anyone with access to a computer, web browser and Internet connection. However, users need to be logged in to avoid external vandalism, and inconsistencies. PalyWeb is built on the expectation that collaboration among users will improve articles over time. Although the site will be "open access", the fact that "peer-review" and entry editing are carried on by users, means that the palynological database need to be restricted (for contributions and edits) to palynologists around the world, and not be fully opened to the general public.

To give estimation on the volume of information that the operating system can manage, the Wikipedia website provides some statistics: on May 2006, Wikipedia had 1,122,525 articles. That number excludes redirects, discussion pages, image description pages, user profile pages, templates, help pages, portals, articles without links to other articles, and pages about Wikipedia. At that time, users had made 53,926,940 edits, an average of 12.98 per page, since July 2002. Today (see graph below from Wikipedia source), there are more than 4 millions articles on the site.

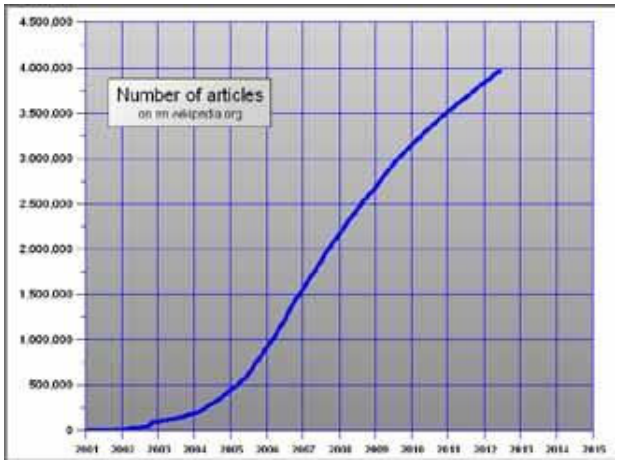

Fig. 1. Wikipedia content evolution (source Wikipedia)

The PalyWeb project

The PalyWeb project is housed on the web server of the University of Liège, in Belgium. A daily backup is carried out in order to safeguard all information in the event of failure of the system. Its internet address is: http://www.palyweb.ulg.ac.be/wiki.

The PalyWeb project began in 2005. Because of the field of research of the lead author (P. Steemans), the pilot project focussed on the Palaeozoic. But obviously, there are no technical limits. The only limits are the one that we will create ourselves as scientists; limits are those of the management of the database by volunteers and contributions by fellow scientists.

Since its conception, although many palynologists are interested in this database, few had had the time to fill the files with data. The PalyWeb is a “lazzarus” project because interests about it are manifested regularly before to be forgotten some months after. One way to solve this would be to have one staff that would be solely focusing on the database. Obviously, the key problem is purely financial. If all palynologists, as petroleum societies, institutes, universities etc. could unite to pool funds, it should be possible to hire one staff member to manage the database. This should be a subject of discussion for IFPS, IPC, AASP, etc.

PalyWeb is formatted to host the description of a maximum of available palynomorph taxa published in conventional scientific reviews in accordance with the Botanical and Zoological Nomenclature Codes. It was officially presented to the scientific community at the Commission International de la Microflore du Paléozoïque (http://www.cimp.ulg.ac.be) at the General meeting of Prague (2006). PalyWeb is not an online publication, new taxa cannot be published through this website as it is not accepted by the BNC. There is no limitation in number of pages or in length. Text, figures and pictures may be up- and downloaded but sizes of the pictures are limited. It is also forbidden to up-load pictures protected by a copyright without referencing to the original. The navigation through pages may be done by clicking on the internal links or in typing keywords in search engine (e.g. in typing a name of species, etc.).
At the moment, the database is organized in three sections: the main section contains the descriptions of the taxa previously published in scientific reviews, the second contains unpublished taxa left in open nomenclature for which information from the PalyWeb community is requested, and the third is devoted to enigmatic

palynomorphs. Those last two sections have to be considered as a discussion forum instead of a section of the database. Of course, this structure can be updated in the future based on feedback from the community.

The flexibility is one of the advantages of the PalyWeb site. Each page (article) is articulated around four main components:

a) The first one is the “page” itself which is named with a title located on top of the screen.

b) The second component is obtained by selecting the heading “to modify” (edit page), where the page can be changed immediately, without restrictions.

c) A third component is the heading “discussion”, where PalyWeb users can exchange their ideas on the contents of the pages.

d) The last component is the "page history”. All former versions of the pages are stored with the user names and the authors of modifications. Each previous version can be restored (a very useful tool as this prevents mistakes and allows readers to follow how the page content has evolved following modifications made by various contributors).

The general architecture of the website is illustrated on Figure 2, using the example of the page organization related to the cryptospore genus *Tetrahedraletes*. The blue words are internal links to existing pages. The words in red are links that have been created but the pages do not exist yet. All useful words for the database objective can be transformed into links related to different PalyWeb pages (external links are also possible). Most of the links are to taxonomic pages, but it is possible to link to other kinds of pages such as personal pages, etc. Each page can be associated to categories. This is a very important tool, not only to navigate the website, but also to organize pages in relation to different criteria. It is particularly interesting to group taxa by morphological characteristics, by stratigraphic levels or by geographic distribution and so on. Categories are made possible by grouping, alphabetically or by criteria.

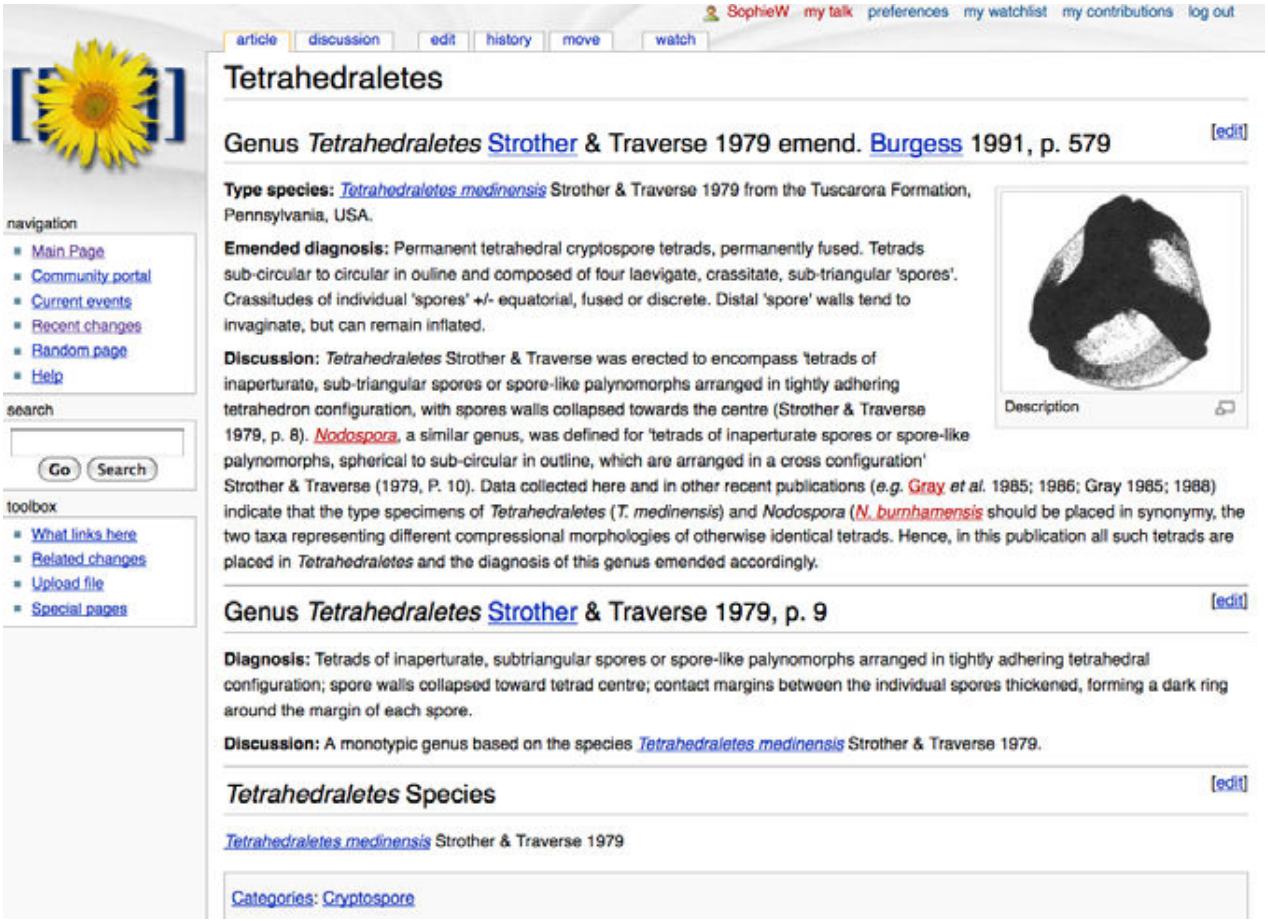

Fig. 2: Example of a PalyWeb page containing the description of the *Tetrahedraletes* genus. Words marked in blue in the descriptive text are linked to other pages such as bibliographic references or species description.

One page can belong to different categories. New categories are created or modified as are any other pages. This said, the high flexibility of such a database can also generate problems. To avoid inconsistency, it is important to follow some syntax rules when naming internal links. For example “cryptospore” and “cryptospores” are two different pages. Therefore, it is necessary to maintain some strictness in the way one contributes to the database. But again, if mistakes are made, all users may correct them.

The advantages of PalyWeb

The advantages of such a database are numerous. The whole palynological community has free access to the website and everybody is allowed to improve or to increase its contents. The free distribution, constant updates, diverse and detailed coverage, by numerous professional palynologists would guaranty the high quality of data on the website. The database is accessible from any kind of computer with an internet connexion (Windows, Mac, Linux). A common web browser is sufficient to work with the database (Internet Explorer, Firefox, ...). Therefore, it is not necessary to buy new software and to keep it up to date. The database is accessible wherever you are. To work with a colleague in a foreign country, you would not need to carry a copy of the database and there would be no problem of compatibilities. Information entered in the database is immediately available to all the community. To the contrary of other usual databases, it is not necessary to wait once the database is considered complete to make it accessible. The database would be a "work in progress" as long as the community would work on it. In addition, as new palynological works become available, PalyWeb would continually and immediately be upgraded. Again, there is no limitation in size and contents. The content of the database is highly flexible and its structure opens the door to discussions. The improvement of the database is not a function of the database creator or of a small group of persons; it will be the result of the worldwide collaboration of the whole palynological community. It does not require the organisation of meetings which are usually limited to people who have subsidies to travel.

The future of PalyWeb

To survive and be usefull, Palyweb needs to grow and be used. One simple possilibity is to use this site to share our own database. Undergraduate students could be hired to copy/paste data from individual database and transfer them to the PalyWeb.

Another possibility would be to start building PalyWeb via our education programs. Most of us in academia are teaching classes on palynomorphs. One possibility to increase the content and traffic on PalyWeb would be to assign each student in our classes a number of palynomorphs to decribe online using the WikiMedia technology. This is already being done at various universities such as LSU as part of the graduate students' "Distinguished Communicators" training, for instance, students are encouraged to write Wikipedia pages on the topic of their research. If we were to implement that with PalyWeb and our microfossil classes worldwide, the best would be to select palynomorphs well known by the faculty in charge of a specific classroom. The student could build a wiki page on PalyWeb following the model of the existing PalyWeb pages. If the semester is 20 weeks long, the student would have as a weekly "homework" the task of creating one page per week on one microfossil. They would do so by compiling data from the literature and in house database. With 10 students, at the end of the semester, 200 palynomorphs could be addressed. The role of the faculty would be to verify the information before making it public. The experts worldwide would then have the opportunity to comment on the validity of the information and improve it as needed. If this is repeated each year, in each institution teaching palynology, the database would build very quickly with a wide range of palynomorph types of all ages and provenance.

References

Jansonius, J., McGregor, D.C., 1996. Chapter 1. Introduction, in: Jansonius, J., McGregor, D.C. (Eds.), Palynology: Principles and applications. Salt Lake City: American Association of Stratigraphic Palynologists Foundation, 1-10.

Lentin, J.K., Davis, O.K., Muncey, T.S., Piel, K.M., 1996. Chapter 24. Personnal computers in palynology, in: Jansonius, J., McGregor, D.C. (Eds.), Palynology: principles and applications. Salt Lake City: American Association of Stratigraphic Palynologists Foundation, 961-984.

Steemans, P., Breuer, P., 2007. PalyWeb: A palynomorph database project on the web, in: Steemans, P., Javaux, E.J. (Eds.), Recent Advances in Palynology. Brest: Carnets de Géologie / Notebooks on Geology, 48-52.

Note: This article is an updated version of a paper published in 2007 by Steemans and Breuer.