

A NEW ATTEMPT TO ESTABLISH THE INTERNATIONAL RADIOCARBON SOILS DATABASE (IRSDB)

Peter Becker-Heidmann

University of Hamburg, Department of Geosciences, Hamburg, Germany. Corresponding author.
Email: p.becker-heidmann@uni-hamburg.de.

Pascal Heidmann

Reinbek, Germany.

ABSTRACT. Twenty years after the first International Radiocarbon Database Workshop, and 13 yr after the setup of a preliminary structure for a worldwide database on ^{14}C dates of soils, sound reasons and excuses for not establishing a real and globally accessible database have diminished. Climate change itself is widely accepted as reality now, and the strong demand of the modeling community for reliable data of the carbon pool—especially in soils—has been maintained. With the steadily increasing capacity of ^{14}C dating facilities, these data can be and are produced. Nevertheless, they still cannot be accessed easily and equally enough. Now, decreased costs of hardware and recent developments of the internet enable the IRSDB to be implemented, in a joint effort. As a seed, a test server has been set up, with open-source software, housing the database in alpha-stage, a web interface, and a community portal. Thus, the development of the design as well as the data input of the database is done in close collaboration of the users of the database, the laboratories, soil scientists, archaeologists, modelers, other scientists, and interested laypersons. In order to guarantee the longtime independence of the availability and usability of the database from vendors or changing standards, only widely used open-source software and open standards are used. Therefore, the development of plug-ins for data input from laboratory databases or output to different required formats as well as interfaces to GIS and other software is possible. A version control system takes care of the integrity of the data.

INTRODUCTION

Since radiocarbon dating became a common tool in archaeology, charcoal and wood and more and more soils and soil material were dated (Scharpenseel and Becker-Heidmann 1992). Worldwide, more than ~15,000 data have been published in different journals. Many ^{14}C dates of soils were measured to answer questions from other disciplines than archaeology, like geochemistry, soil genesis, or climate change research, and therefore were distributed in diverse publication media. This makes it very difficult to find and access these data whenever they could be useful to solve a problem within a discipline different from the original one. For example, ^{14}C dates of a soil profile collected to explain soil genesis at a certain location might also help to calculate turnover rates of soil organic matter for a regional climate change model, or give valuable archaeological information on regional land-use in ancient times.

During the 1980s, relational database software was developed and increasingly used also by ^{14}C laboratories (e.g. Englesman et al. 1986; Kalin and Long 1989; Michczyński and Pazdur 1989) and archaeological institutions (e.g. Moffett and Webb 1983; Lavell et al. 1992). Kra (1988) initiated the International Radiocarbon Data Base (IRDB), which was planned to store all ^{14}C data electronically, closely associated with the *Radiocarbon* journal. But only the introduction of the World Wide Web as hypertext-linked information delivered by a network of computers (Berners-Lee and Cailliau 1990) and, after the Scientific and Technology Act of the US American Congress in 1992, the connection of the separate scientific and commercial networks on national and international levels to the internet lead to the possibility of worldwide access to the proposed databases. At the same time, the demand for soil ^{14}C data from the climate modeling community increased rapidly, as it became clear that soil might be the missing carbon sink in the global carbon cycle models (Tans et al. 1990). Thus, during a workshop in Tucson in 1992 sponsored by NASA, the basic structure of the IRSDB was defined (Becker-Heidmann 1996) and a steering committee elected.

At that time, setting up a database meant to buy and setup a workstation, usually with a UNIX[®] operation system; to choose, buy, and install a database software; to develop or modify an appropriate database management application; to search the literature and other sources; and to initially and then continuously fill the database with the data sets. Altogether, this is an expensive, time-consuming, and long-lasting enterprise, with endangered success because of the large number of dependencies, and with no scientific output in itself. Due to these difficulties, and in spite of maintained or even increasing interest, the IRSDB did not come to reality until 2008. Contributions were published therefore in paper-based form so far (Scharpenseel et al. 1996; Becker-Heidmann et al. 1996).

PREMISES

Meanwhile, the conditions have changed significantly. The cost of computer hardware has been decreasing continuously and at the same time, its power has increased. A cheap standard off-the-shelf PC is sufficient to run a web server, database server, file server, mail server, and other software at the same time. It can be driven by a variety of widely used operating systems, has a multiple of the storage capacity needed for a project like the IRSDB, and a fast connection to any network. The development of open-source software (OSS; Raymond 1998) for nearly any kind of application reduces the effort of necessary programming to a minimum. OSS is characterized by free availability of the source code, the right to modify it, and the right to use it in any way. Advantages of OSS comprise no cost of purchase, independence from a distinguished software developer or company, reliability in terms of continuous maintenance and development by a highly motivated community, and the possibility to modify the software for individual needs. The greatest improvement was probably the development of the World Wide Web into Web 2.0 (O'Reilly 2005), i.e. the change from unidirectional presentation of contents to a bidirectional, interactive communication between the users. The client-server model, in which static text and data are put on a server computer by a group of providers and delivered to and displayed on client computers of the group of consumers, was superseded by a model where the contents are dynamically changed and where the roles of server and client, of provider and consumer, are blurred. With this technique, often called AJAX (Asynchronous JavaScript and XML), the server administers the data in a database, and the users communicate with the database and with each other by scripts running on and interacting between server and clients. The contents can be produced in collaboration of all interested users. Recently, the OSS initiative is supplying web developers with ready-to-use web application routines and with application programming interfaces (API) to communicate with existing external applications, whereby even beginners in web programming become able to quickly setup simple applications. A very active community of like-minded users helps each other via discussion boards, reducing the need to engage professional programmers in many cases. These developments altogether reduce the necessary manpower for individual projects dramatically.

IMPLEMENTATION

When starting with the setup and programming of the IRSDB, we seized many suggestions from already existing databases on the web, especially from those listed on the web page of the *Radiocarbon* journal (<http://www.radiocarbon.org/Info/index.html#databases>). The database program of the Council for British Archaeology uses a compact data input form (<http://www.britarch.ac.uk/info/c14.html>). A clear and simple output format is a characteristic attribute of the Oxford ¹⁴C date list (http://users.ox.ac.uk/~orau/dl_am2.html). The Canadian Archaeological Association presents an ideal detailed search form for the ¹⁴C database of North American archaeological sites (<http://card.canadianarchaeology.ca/>), and stands out in terms of user friendliness with a detailed introduction and additional informative help pages. For a location search form, which we assume to be

essential for the IRSDB, the search form of the NGDC of NOAA for volcano locations is a good model (<http://www.ngdc.noaa.gov/nndc/struts/form?t=102557&s=5&d=5>).

In a “no or at least low budget” approach, a test server to host the IRSDB was set up with a discarded, but for our purpose sufficiently equipped, PC (AMD Athlon[®], 1GHz, 384MB processor, 40GB hard drive) and connected to the internet. We acquired the domain irsdb.de, which points to our server. The next steps were the installation of basic web server and database software, and a software development framework. Basic data input and output routines were programmed, and a knowledge base using a Wiki software installed. These are described subsequently. Programming of the user administration module and the integration of map representation of locations via an available application programming interface (Google[®] Maps API, <http://code.google.com/>) are planned before we open the IRSDB to the public. At the moment, the web page of the IRSDB offers just the Wiki.

The server PC is operated by the commercial Windows[®] XP Professional, just because we are not used to the OSS operating system Linux[®], which would be the alternative free of charge. We equipped the PC with the OSS-Package XAMPP (acronym for Apache + MySQL + PHP + Perl, <http://www.apachefriends.org/en/xampp.html>), which is intended for developing purposes and is available and preconfigured for different operating systems. XAMPP comprises the widely used Apache web server and MySQL[®] database server software, with a file transfer server and a mail server software. Its main advantage over the single components and over similar packages is that its installation is very simple. The web site of the software is a clearly arranged page with download links, instructions for installation, use, and de-installation, as well as security instructions and further informations. An important feature is the easy server control, a distinguished application for start, stop, and administration of each of the components of the package. No complicated configuration is necessary, except setting passwords to secure the servers, which is clearly addressed in the welcome screen after successful installation.

The first important step in the establishment of the IRSDB is the setup of the database. To reduce the workload and possibility of introducing errors during the design phase, we used the OSS MySQL Workbench community edition (<http://dev.mysql.com/downloads/workbench/5.1.html>), which enables us to visually design, generate, and manage the database (Figure 1). Additions and changes of data fields and relations can be conducted either in the graphical model or the table description forms. Eventually, the database model can be exported as an SQL command file, which can then be executed to set up the database on the database server. The standard administration tool for the MySQL database via a web interface is the PHP script phpMyAdmin, which comes with the XAMPP package. Besides its main purpose, the direct manipulation of data and structure, it can set up a database and also provides a graphical representation similar to MySQL Workbench. We use this program to import the SQL command file being exported by MySQL Workbench, and thus create the IRSDB on our server.

The widely used standard technique for controlling the data flow between database, web server, and web browser software is using PHP scripts. Usually, the necessary programming is simplified by using existing development frameworks. We chose the object-oriented OS framework Zend[®] (<http://framework.zend.com/>). It consists of a set of prefabricated and tested program modules that are suitable for the needs of the IRSDB. Among those are routines for input/output forms and for connecting to and integration of geographical services, esp. the Google Maps API. With this framework, it was possible to develop in a very short time a simple form for entry of new data sets and a simple form for search and display of data sets. We can use further included modules for necessary secure user administration routines.

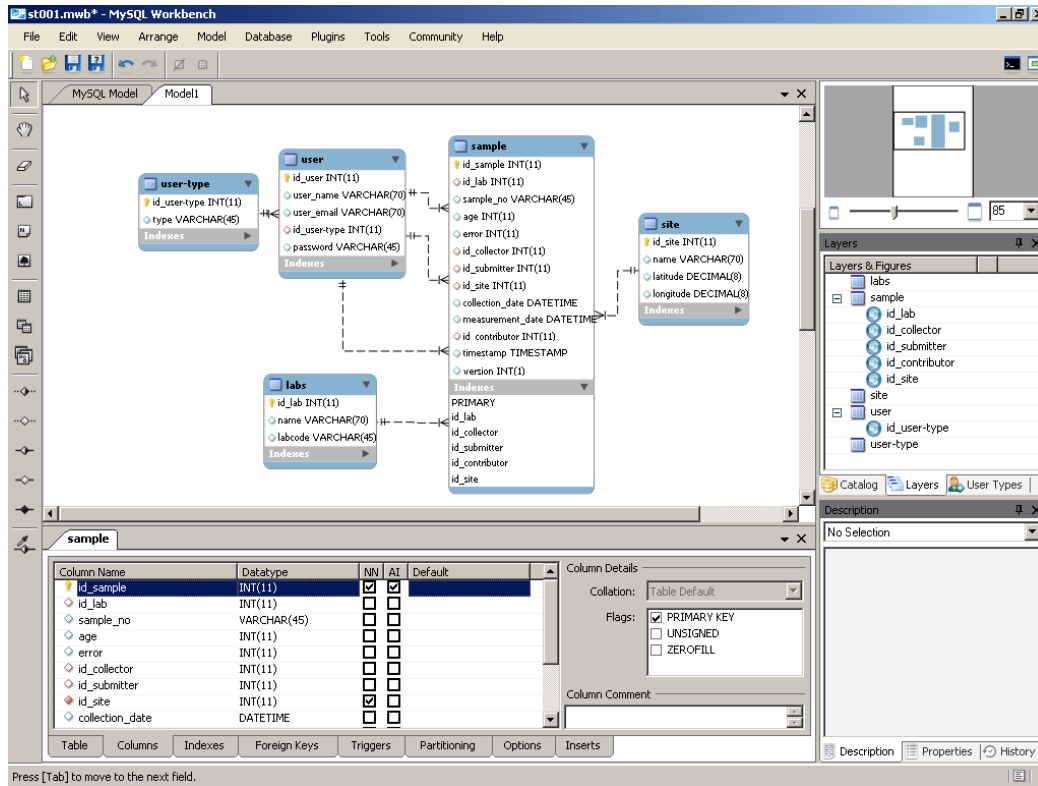


Figure 1 First steps of database model building with MySQL Workbench

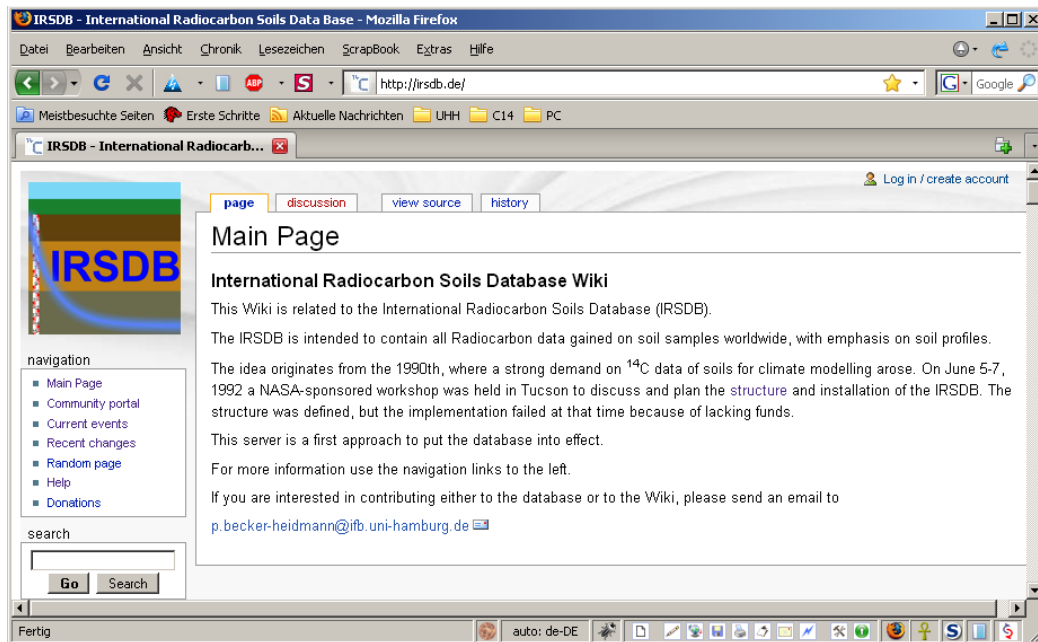


Figure 2 The main web page of the IRSDB is a Wiki on the database

For the basic idea of this new attempt to establish the IRSDB as a collaboration of the database users, we installed a knowledge-base using the OSS MediaWiki (<http://www.mediawiki.org/wiki/MediaWiki>) on our server PC, which is used as the entry page to the database, too (Figure 2) and informs about the structure of the IRSDB and its development (Figure 3).

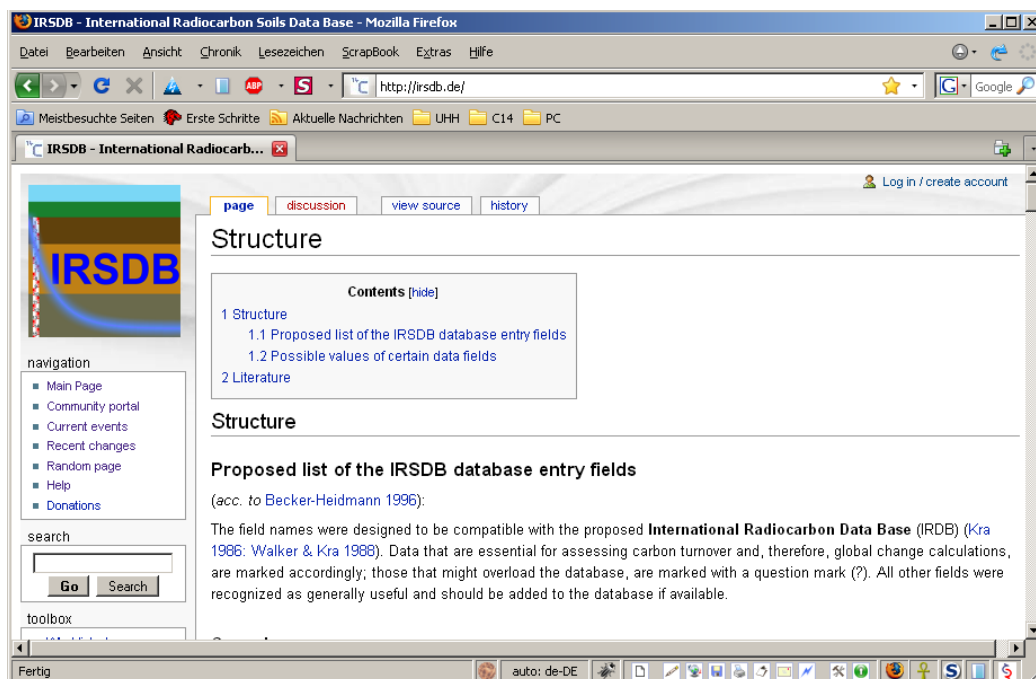


Figure 3 The structure of the IRSDB is described in the Wiki page

CONCLUSIONS

We show that with recently available techniques, a project such as the IRSDB could be launched with minimum effort in a short time and brought up to a state where other users can easily join in further development or data delivery. Still, some essential tasks are to be done. The thus-far programmed user administration routines have to be released in order to initiate the intended collaboration. Also, the actually rudimentary data input, search, and presentation forms have to be improved. Systems to support the interaction of the collaborating users have to be introduced (e.g. a bulletin board). A possible introduction of a version control system for the developing programs has to be discussed. Last but not least, we want to work together with the colleagues who operate the existing ^{14}C databases in different laboratories to develop APIs for data transfer of soil ^{14}C dates to the IRSDB.

ACKNOWLEDGMENTS

We are thankful to the innumerable highly motivated people who worked on the OSS described and enabled us to begin the IRSDB. Several colleagues encouraged us to dare this attempt; other colleagues offered their support and/or data delivery. The University of Hamburg implicitly closed the ^{14}C dating laboratory for reconstruction work on the institute's building, thus chaining the first author to his office and computer. Daniela had to endure the increased digital passion of her husband and son.

REFERENCES

- Becker-Heidmann P. 1996. Requirements for an international radiocarbon soils database. *Radiocarbon* 38(2): 177–80.
- Becker-Heidmann P, Scharpenseel HW, Wiechmann H. 1996. Hamburg radiocarbon thin layer soils database. *Radiocarbon* 38(2):295–345.
- Berners-Lee T, Cailliau R. 1990. WorldWideWeb: Proposal for a hypertext project [WWW document]. URL: <http://www.w3.org/Proposal.html>. Accessed 27 July 2009.
- Englesman FMR, Taayke E, Mook WG. 1986. Groningen ¹⁴C database. *Radiocarbon* 28(2A):788–96.
- Kalin RM, Long A. 1989. Radiocarbon data base: Q & A—an artificial intelligence database file management program. *Radiocarbon* 31(1):1–6.
- Kra R. 1988. Updating the past: the establishment of the International Radiocarbon Data Base. *American Antiquity* 53(1):118–25.
- Lavell C, Otlet RL, Walker AJ. 1992. The CBA/RCD computer database of radiocarbon dated sites. *Antiquity* 66(253):969–74.
- Michezyński A, Pazdur MF. 1989. Local microcomputer database for radiocarbon dates. *Zeszyty Naukowe Politechniki Śląskiej, Seria Matematyka-Fizyka*, z.61. *Geochronometria* 6:27–35.
- Moffett JC, Webb RE. 1983. Database management systems, radiocarbon and archaeology. *Radiocarbon* 25(2):667–8.
- O'Reilly T. 2005. What Is Web 2.0? Design patterns and business models for the next generation of software [WWW document]. URL: <http://oreilly.com/web2/archive/what-is-web-20.html>. Accessed 27 July 2009.
- Raymond ES. 1998. Goodbye, “free software”, hello, “open source” [WWW document]. URL: <http://www.catb.org/~esr/open-source.html>. Accessed 27 July 2009.
- Scharpenseel HW, Becker-Heidmann P. 1992. Twenty-five years radiocarbon dating of soils: paradigm of erring and learning. *Radiocarbon* 34(3):541–9.
- Scharpenseel HW, Pietig F, Schiffman H, Becker-Heidmann P. 1996. Radiocarbon dating of soils: database contribution by Bonn and Hamburg. *Radiocarbon* 38(2):277–93.
- Tans PP, Fung IY, Takahashi T. 1990. Observational constraints on the global atmospheric CO₂ budget. *Science* 247(4949):1431–8.