

Adam MICHCZYŃSKI

Mieczysław F. PAZDUR

Laboratorium C-14

Politechnika Śląska, Gliwice

GLIWICKA BAZA DANYCH DATOWAN RADIOWĘGLOWYCH - STAN AKTUALNY

Streszczenie. Artykuł przedstawia aktualny stan prac nad systemem komputerowej bazy danych. Wprowadzone modyfikacje polegają na zmianie struktury zapisu informacji i doprowadzeniu do powstania struktury w maksymalny sposób zbliżonej do tzw. trzeciej postaci normalnej bazy danych. Uzyskano w ten sposób znaczne zmniejszenie obszaru pamięci potrzebnego do przechowywania zgromadzonych danych.

GLIWICE RADIOCARBON DATA BANK - PRESENT STATUS

Summary. The paper gives a description of the actual status of the data bank. Recently introduced modifications include change of the structure of data bank. This modified structure is very close to the theoretical concept of the so-called normal structure of the data bank. A new form of the questionnaire of samples submitted for radiocarbon dating is presented.

КОМПЬЮТЕРНАЯ СИСТЕМА ИНФОРМАЦИИ РАДИОУГЛЕРОДНОЙ ЛАБОРАТОРИИ В ГЛИВИЦЕ

Резюме. Авторами представлено состояние работ по созданию компьютерной системы сохранения и обработки информации об результатах радиоуглеродных датировок проводимых в Радиоуглеродной лаборатории в Гливице. Создана система почти совпадает с теоретической нормой так называемой нормальной структуры. В соответствии с этой структурой была подготовлена тоже новая форма бланка для образцов датированных радиоуглеродным методом.

WSTĘP

Konieczność stworzenia zbioru wyników wszystkich datowań radiowęglowych wykonanych w laboratorium gliwickim oraz zapewnienia szybkiego i łatwego dostępu do informacji zawartych w tym zbiorze stała się przyczyną zapoczątkowania w 1987 roku prac nad lokalną gliwicką bazą datowań radiowęglowych, określaną w skrócie jako GdRDB (Pazdur, Porwoł, 1987). Na III Konferencji "Metody Chronologii Bezwzględnej" w 1989 roku prezentowana była pierwsza, próbna realizacja tej bazy, stworzona z wykorzystaniem języka programowania dBase III, zawierająca około 250 dat (Michczyński, Pazdur, 1989). Doświadczenia nabyte od tamtego czasu przy rozbudowie bazy GdRDB oraz tworzeniu niezależnej od GdRDB bazy danych datowań radiowęglowych kultur archeologicznych z rejonu Ańdów Środkowych (Krzanowski i in., 1993) uświadomiły konieczność istotnej przebudowy struktury bazy, a także takiego pokierowania systemem gromadzenia danych o aktualnie wykonywanych datowaniach, aby uzyskać jak najbardziej kompletne wypełnienie tejże struktury.

PIERWOTNA STRUKTURA GdRDB

Aby przedstawić problemy, które doprowadziły do powstania koncepcji nowej struktury GdRDB, należy pokrótce przypomnieć pierwotną strukturę bazy, przedstawioną w 1989 roku (Michczyński, Pazdur, 1989). Zawartość bazy zapamiętana jest w jednym pliku - RDB.DBF, zawierającym wszystkie dane dotyczące dat radiowęglowych. Informacje dotyczące każdej z dat tworzą jeden rekord omawianej bazy. Rekord ten składa się z 47 pól, w których zapisane są różnego rodzaju dane. Jego szczegółowa struktura, tj. nazwy wszystkich pól, ich typy (tekstowy, numeryczny lub zawierający datę) oraz długości (liczbę znaków, jaką mogą zawierać poszczególne pola) przedstawiona została w tabeli 1.

Pierwsze dwa pola każdego rekordu (LABCODE, LAENO) zawierają informację o kodzie laboratoryjnym i numerze, jaki nadano danej dacie radiowęglowej (np. Gd-1963). Następne trzy pola (SERIES, SITE, SAMPLE) identyfikują datowaną

Tabela 1

Struktura rekordu w bazie danych GdRDB

Field name	Type	Width	Field name	Type	Width
1 LABCODE	Character	4	25 COUNTRY	Character	25
2 LABNO	Numeric	5	26 ADMUNIT	Character	20
3 SERIES	Character	26	27 TOWN	Character	20
4 SITE	Character	21	28 REGION	Character	60
5 SAMPLE	Character	20	29 LAT	Character	10
6 DTYPE	Character	2	30 LONG	Character	10
7 AGEBP	Numeric	5	31 COLLDATE	Character	25
8 ERMAX	Numeric	4	32 COLLYBY	Character	25
9 ERMIN	Numeric	4	33 SUBMBY	Character	25
10 DC13	Numeric	6	34 SUBMINST	Character	80
11 MATERIAL	Character	15	35 RECDATE	Date	8
12 MATEXT	Character	25	36 MEASDATE	Date	8
13 IDAS	Character	80	37 REPDATE	Date	8
14 IDBY	Character	50	38 JOB	Character	12
15 FRACTION	Character	25	39 PROJECT	Character	25
16 TREATMENT	Character	50	40 SPONSOR	Character	50
17 CONTAM	Character	50	41 INDEXNO	Numeric	4
18 CONTEXT	Character	128	42 RPT	Character	6
19 SITETYPE	Character	25	43 DATELIST	Character	6
20 SITEDESC	Character	190	44 COMMENT	Character	125
21 ENVIRON	Character	190	45 LABCOMMENT	Character	80
22 PERIOD	Character	25	46 COUNTER	Numeric	2
23 PHASE	Character	25	47 REFERENCES	Character	80
24 DISCIPLINE	Character	25	* * Total * *1715		

próbkę. Zawierają one odpowiednio: nazwę serii, z której pochodzi próbka, nazwę stanowiska oraz symbol konkretnej próbki. Cztery kolejne pola zawierają informację o otrzymanej dacie, w tym jej typ (DTYPE), wartość (AGEBP) oraz wartość błędę "+" (ERMAX) i "-" (ERMIN). Do pola DC13 wpisywana jest wartość współczynnika $\delta^{13}\text{C}$, a jeśli pomiar tej wielkości nie był wykonany, wprowadzana jest wartość -25‰ równa wartości dla drewna współczesnego. Kolejne dwa pola (MATERIAL, MATEX) zawierają informacje o materiale próbki. Jeśli materiał ten został szczegółowo zidentyfikowany (np. określony został gatunek datowanego kawałka drzewa lub ziaren) to identyfikacja ta zapisywana

jest w polu IDAS, a kawałka drzewa lub ziaren) to identyfikacja ta zapisywana jest w polu IDAS, a dane personalne osoby przeprowadzającej identyfikację w polu IDBY. W przypadku gdy datowany był nie cały materiał próbki, lecz tylko pewna jego frakcja, informacja o użytej do pomiaru wieku frakcji umieszczona jest w polu FRACTION. Następne w kolejności pole TREATMENT zawiera dane o stosowanej preparatyce, a pole CONTAM określa rodzaj zanieczyszczeń datowanego materiału.

Dalsze cztery pola zawierają informacje o miejscu, z którego pobrana została próbka. Pole SITETYPE ma charakter pola kluczowego zawierającego informacje o pochodzeniu próbki. Pola CONTEXT, SITEDESC i ENVIRONMENT podają opisową informację dotyczącą odpowiednio: geologicznego lub archeologicznego kontekstu, z jakiego pobrana została próbka, rozumianego jako najbliższe otoczenie próbki, geologicznego i geomorfologicznego usytuowania miejsca pobrania próbki oraz geograficznego położenia tego miejsca.

Po kolejnych trzech polach informujących o okresie i fazie (PERIOD, PHASE), z jakiego prawdopodobnie pochodzi próbka oraz dyscyplinie naukowej, na potrzeby której datuje się próbkę (DISCIPLINE), następuje grupa sześciu pól zawierających dane o położeniu miejsca pobrania próbki. Są to pola: COUNTRY, w którym zawarta jest nazwa kraju, ADMUNIT - jednostka administracyjna, TOWN - najbliższa miejscowość, REGION - region geograficzny oraz LAT i LONG - szerokość i długość geograficzna. Cztery dalsze pola informują o dacie pobrania próbki (COLLDATE), osobie pobierającej (COLLBY) oraz osobie i instytucji, która dostarczyła próbkę do laboratorium (SUBMBY i SUBMINST). Po nich następuje grupa pól o charakterze pól kluczowych z informacjami istotnymi głównie dla laboratorium radiowęglowego. Są to: RECDATE - data otrzymania próbki przez laboratorium, MEASDATE - data zakończenia pomiarów, REPDATE - data raportu, JOB - symbol pracy naukowo-badawczej, w ramach której przeprowadzono datowanie, PROJECT - nazwa projektu, w ramach którego przeprowadzono datowanie, SPONSOR - nazwa instytucji finansującej badania, INDEXNO - numer,

pod jakim próbka figuruje w rejestrze próbek laboratorium, RPT - numer raportu o wyniku datowania, DATELIST - numer listy datowań, w której wynik został opublikowany, COUNTER - numer laboratoryjny licznika, na którym dokonano pomiaru. Polami kończącymi rekord są pola COMMENT - zawierające komentarz do otrzymanej daty, LABCOMMENT - zawierające komentarz laboratorium oraz REFERENCE - wskazujące publikacje dotyczące danej daty.

W trakcie uzupełniania tak zaprojektowanej struktury bazy coraz to większą ilością nowych danych ujawniły się istotne nowe problemy. Po pierwsze jak można łatwo zauważyć w tabeli 1, długość pojedynczego rekordu wynosi 1715 bajtów. Zatem rozmiar pliku, zawierającego 850 rekordów wynosi około 1.5 MB, a plik zawierający wyniki wszystkich datowań wykonanych w gliwickim laboratorium (kilka tysięcy) miałby rozmiar kilkunastu megabajtów. Oczywiście, operowanie na tak wielkich plikach, w szczególności wyszukiwanie w nich odpowiednich danych, byłoby utrudnione i zajmowałoby sporo czasu. Po drugie, spora część informacji zapisanych w niektórych polach jednych rekordów powtarza się także w analogicznych polach innych rekordów (np. opis geograficznego położenia miejsca pobrania próbki będzie identyczny dla wszystkich próbek z jednego stanowiska). Powoduje to wielokrotne zajmowanie miejsca w bazie danych przez tę samą informację.

NOWA STRUKTURA SYSTEMU GdRDB

Przedstawione powyżej problemy są typowymi problemami pojawiającymi się przy projektowaniu relacyjnych baz danych. Aby je usunąć, należy tak przekonstruować strukturę bazy danych, aby z tzw. pierwszej postaci normalnej bazy stworzyć trzecią postać normalną (Cellary, Królikowski, 1988). Należy podkreślić, że przebudowa ta nie pociąga za sobą zmian zakresu przechowywanej informacji, a prowadzi jedynie do zmiany formy, w jakiej informacja ta jest przechowywana. Baza danych w trzeciej postaci normalnej nie stanowi już zwykle jednego pliku, lecz jest systemem kilku plików zawierających pola pierwotnej postaci bazy wiążące się ze sobą w pewne grupy (np. wszystkie pola zawiera-

jące informację o stanowisku, z którego została pobrana próbka do datowania). Pliki te powiązane są z kolei między sobą za pomocą pól - kluczy (lub kilku pól tworzących klucz). Pola te, zazwyczaj o stosunkowo małej długości, występują w dwóch plikach systemu - w jednym z nich pozwalają one na jednoznaczne zidentyfikowanie rekordu (np. nazwa lub identyfikator stanowiska archeologicznego lub geologicznego), w drugim pełnią rolę odsyłacza wskazującego, gdzie w pierwszym pliku znaleźć dodatkowe informacje (np. pole zawierające nazwę lub identyfikator stanowiska, z którego pobrano próbkę do datowania, zawarte w pliku stanowiącym zbiór informacji o datach wskazuje, gdzie w pliku z informacjami o stanowiskach znaleźć kompletną informację o miejscu pobrania próbki). Zatem zadanie usunięcia przedstawionych wcześniej problemów oznaczało w istocie sprowadzenie bazy GdRDB opartej na pierwotnej strukturze do trzeciej postaci normalnej. Efekt tego postępowania, czyli nową strukturę systemu GdRDB przedstawia rysunek 1.

Bazę GdRDB stanowi teraz system 7 plików:

- DATES.DBF - plik zawierający podstawowe informacje o datach radiowęglowych, stanowiący macierzysty plik całego systemu,
- SAMPLES.DBF - plik zawierający informacje o próbkach,
- SITE.DBF - plik zawierający informacje o miejscach (stanowiskach), z których były pobierane próbki,
- REGION.DBF - plik zawierający informacje o regionach geograficznych, w których znajdowały się miejsca pobrania próbek,
- INDX.DBF - plik będący odpowiednikiem laboratoryjnego rejestru próbek, notujący m.in. kto i kiedy dostarczył próbkę do datowania, kto pokrywał koszty datowania itp.,
- SUBMINST.DBF - plik zawierający informacje o instytucjach dostarczających próbki do datowania,
- REFERENC.DBF - plik zawierający dane bibliograficzne publikacji dotyczących dat zawartych w bazie.

Pliki te wiążą się ze sobą za pomocą pól - kluczy takich, jak:

INDEX oraz INDEX1 - pola określające numer próbki w laboratoryjnym rejestrze próbek,

SITE - nazwa stanowiska, z którego pobrano próbkę,

powtarzaniu się miejscu pobrania próbki itd. Taka struktura zapobiega powtarzaniu się informacji (np. informacje o danym stanowisku zapisane są tylko raz w jednym z rekordów pliku SITE.DBF, do którego może się jednak odwoływać wiele rekordów pliku DATES.DBF), a tym samym prowadzi do oszczędności miejsca zajmowanego na dysku przez system GdRDB. W przypadku bazy zawierającej informacje o 350 datach zmiana struktury doprowadziła do redukcji obszaru pamięci na dysku zajmowanego przez bazę o około 37%. Można przypuszczać, że dla baz zawierających większe ilości dat redukcja ta będzie jeszcze większa.

W tym miejscu należy jednak powiedzieć, że przedstawiona struktura nie jest ściśle systemem baz w trzeciej postaci normalnej, gdyż warunek określający trzecią postać normalną nie jest spełniony dla bazy DATES.DBF. Jest to świadome odstępstwo podyktowane tym, aby w jednym pliku znajdował się komplet podstawowych informacji o danej dacie. Powoduje to wprawdzie nieznaczne zwiększenie obszaru pamięci zajmowanego przez system w porównaniu z rozwiązaniem ściśle zgodnym z regułami, ale daje to taką korzyść, iż w celu uzyskania podstawowych informacji nie trzeba tworzyć połączeń pomiędzy plikami systemu. Pozwala to z kolei na uzyskanie znacznego przyspieszenia operacji wyszukiwania elementarnych danych.

Przedstawione tutaj rozwiązanie nie jest pierwszą propozycją wieloplikowego, relacyjnego systemu bazy danych dla datowań radiowęglowych z gliwickiego laboratorium. Pierwsza koncepcja takiego systemu była przedstawiona już w 1987 roku (Pazdur, Porwoł 1987), jednakże ze względu na konieczność uzgodnienia struktury z projektowaną międzynarodową bazą danych radiowęglowych - IRDB (Pazdur, 1988) trzeba było z niej zrezygnować. Obecne rozwiązanie łączy w sobie cechy rozwiązań z 1987 (wieloplikowość) oraz 1989 roku (zgodność ze strukturą IRDB). Trzeba podkreślić, iż podobne systemy baz danych istnieją bądź też powstają w laboratoriach radiowęglowych całego świata (największe z nich są w Harwell, Groningen i Oxfordzie), a schemat struktury IRDB, mimo iż sama ta baza jeszcze nie funkcjonuje, staje się podstawą wymiany danych pomiędzy laboratoriami radiowęglowymi (Walker i in., 1990).

NOWA POSTAĆ ARKUSZA INFORMACYJNEGO PRÓBKII PRZEZNACZONEJ DO DATOWANIA

W trakcie uzupełniania bazy informacjami o coraz to większej ilości dat wyszło na jaw, że wiele pól przewidzianych w strukturze pozostaje pustych. Wynikało to głównie z faktu, iż dane o próbkach przesyłanych do datowania do Laboratorium C-14 w Gliwicach częstokroć nie są kompletne. Drugą przyczyną była niezupełna zgodność informacji, które podawane były w dotychczasowo wykorzystywanym arkuszu informacyjnym próbki przeznaczonej do datowania z informacjami jakie należało wpisać do bazy danych. Aby zapobiec tym niezgodnościom oraz udoskonalić system gromadzenia danych o próbkach zaprojektowany został nowy wzór arkusza informacyjnego próbki. Jest on przedstawiony na rysunku 2. W odróżnieniu od dotychczas stosowanego, arkusz ten sporządzony jest w języku angielskim. Podyktowane to zostało względami planowanej międzynarodowej wymiany informacji. Warto podkreślić, że podobnie postępują laboratoria w Europie Zachodniej, np. arkusz informacyjny próbki stosowany przez laboratorium w Groningen (Holandia) również sporządzony jest w języku angielskim. Nowy arkusz skonstruowany jest tak, aby uzyskać maksymalną ilość danych potrzebnych do wypełnienia informacjami pól bazy GdRDB. Rubryki arkusza odpowiadają większości pól GdRDB, a tam, gdzie w bazie występują pola, przy użyciu których ma być prowadzone wyszukiwanie informacji, arkusz sugeruje wybór jednego z określeń wymienionych w liście. Porównując nową strukturę GdRDB i wzór arkusza informacyjnego próbki można zauważyć, że poszczególne jego punkty lub grupy punktów odpowiadają plikom tworzącym system GdRDB.

DALSZE PERSPEKTYWY SYSTEMU GdRDB

W chwili obecnej zakończono testowanie funkcjonowania bazy GdRDB oparte na zmodyfikowanej strukturze. Dalszym niezbędnym krokiem jest opracowanie, napisanie i uruchomienie oprogramowania służącego do obsługi nowej wersji bazy, analogicznego do oprogramowania obsługującego starą wersję GdRDB (Michczyński, Pazdur, 1989). Oprogramowanie to będzie umożliwiać zmianę

Radiocarbon Laboratory
Silesian Technical University
Krzywoustego 2
PL-44-100 Gliwice, Poland
tel./fax: (+48 32) 372254, tel.: (+48 32) 372662
E-mail: pazdur@igleto2.gliwice.edu.pl

RADIOCARBON SAMPLE INFORMATION

THIS SPACE IS FILLED IN BY RADIOCARBON LABORATORY

Sample No

Date of submission:

Date of report:

Report No

Gd-

Age BP:

1. SUBMITTER:

Name

Institution

Address

Telephone Fax

Telex E-mail

2. COLLECTOR:

Name

Date of collection

3. SAMPLE DESCRIPTION:

Sample name

Material of sample:

- | | | |
|---|--|--|
| <input type="checkbox"/> charcoal | <input type="checkbox"/> gyttja | <input type="checkbox"/> bone(s) |
| <input type="checkbox"/> wood | <input type="checkbox"/> grain(s) | <input type="checkbox"/> humus |
| <input type="checkbox"/> peat | <input type="checkbox"/> terrestrial shell | <input type="checkbox"/> plant remains |
| <input type="checkbox"/> organic sediment | <input type="checkbox"/> fresh-water shell | <input type="checkbox"/> detritus |
| <input type="checkbox"/> mud | <input type="checkbox"/> marine shell | <input type="checkbox"/> tufa |
| <input type="checkbox"/> soil | <input type="checkbox"/> speleothem | <input type="checkbox"/> other |

Additional information about material of sample:
(botanical or lithological identification, species etc.)

.....

.....

..... Carbon content (approx. %)

Contamination:

- | | |
|------------------------------------|--|
| <input type="checkbox"/> none | <input type="checkbox"/> organic detritus |
| <input type="checkbox"/> probable | <input type="checkbox"/> humic |
| <input type="checkbox"/> rootlets | <input type="checkbox"/> organic preservatives |
| <input type="checkbox"/> carbonate | <input type="checkbox"/> other |

4. SAMPLE LOCATION: Country

Province/district

Town

Region

Geographical coordinates vs. Greenwich

Lat N/S Long E/W

Altitude m a s l.

Short geographical description of the environment of site (the nearest lake, mountain etc.)

.....

5. **CONTEXT OF SAMPLE:** Depth below surface:.....m

Short description of the geomorphological and stratigraphical context of sample, details about position of sample within

site:.....

Type of site: other excavation
 exposure trench core
 museum finding

6. **EXPECTED AGE /PERIOD/ PHASE:**

7. **FIELD OF STUDY:** archeology geology
 geography geomorphology geochemistry
 oceanography history botany
 limnology calibration

8. **SIGNIFICANCE OF PROBLEM TO THE SOLVING OF WHICH THE DATING OF SAMPLE SHOULD CONTRIBUTE:**

9. **OTHER EXISTING DATES OF THIS SITE, LAYER, OBJECT** (lab. code and age):

10. **REFERENCES TO LITERATURE** (please provide reprints if available):

11. Please provide the sketch with scale and position of sample within site, pollen diagram (use additional sheet if necessary).

Rys. 2. Wzór arkusza informacyjnego próbki
 Fig. 2. The sample information sheet

uporządkowania rekordów w bazie, wyszukiwanie interesujących dat lub grup dat, wyświetlanie informacji o dacie w postaci spójnego raportu (w wersji pełnej oraz zgodnej ze strukturą IRDB) oraz dopisywanie nowych rekordów. Ponadto konieczna jest kontynuacja prac polegających na wprowadzaniu do bazy danych o wykonanych datowaniach. Planuje się również rozpoczęcie systematycznej rejestracji w bazie aktualnie wykonywanych datowań, z czym wiązać się będzie m.in. wprowadzenie do powszechnego użytku opracowanego formularza opisu próbek.

Opracowanie niniejsze wykonano w ramach pracy BK-545/RMF-1/92.

LITERATURA

- Cellary W., Królikowski Z., 1988, Wprowadzenie do projektowania baz danych - dBase III. WNT, Warszawa.
- Krzanowski A., Michczyński A., Pazdur M. F., Ziólkowski M. S., 1993, Baza danych datowań radiowęglowych kultur archeologicznych rejonu Andów Środkowych; Zeszyty Naukowe Politechniki Śląskiej, Seria Matematyka -Fizyka Geochronometria nr 10, s. 139-151.
- Michczyński A., Pazdur M. F., 1989, Lokalna mikrokomputerowa baza danych datowań radiowęglowych; Zeszyty Naukowe Politechniki Śląskiej, Seria Matematyka-Fizyka, z. 61, Geochronometria nr 6, s. 27-35.
- Pazdur M. F., 1989, Międzynarodowa komputerowa baza danych datowań radiowęglowych; Zeszyty Naukowe Politechniki Śląskiej, Seria Matematyka -Fizyka, z. 61, Geochronometria nr 6, s. 21-26.
- Pazdur M. F., Porwoń M., 1987, Microcomputer database system for radiocarbon dates: a pilot project; Zeszyty Naukowe Politechniki Śląskiej, Seria Matematyka-Fizyka, z. 56, Geochronometria nr 4, s. 19-25.
- Struzińska-Walczak A., Walczak K., 1991, dBase IV - Programowanie SQL. Komputerowa Oficyna Wydawnicza "Help", Warszawa.
- Walker A. J., Otlet R. L., Housley R. A., Van der Plicht J., 1990, Operation of the Harwell UK ^{14}C data base and its expansion through data exchange with other laboratories; Radiocarbon, t. 32, s. 31-36.

Wpłynęło do Redakcji: 15 maja 1992

Recenzent: Prof. dr hab. Bolesław Nowaczyk

Abstract

The research intended to develop a computerised system for storage and retrieval of information concerning results of radiocarbon datings performed in the Gliwice Radiocarbon Laboratory has been started several years ago. The present paper gives a description of the actual status of the data bank. Recently introduced modifications include change of the structure of data bank. The new structure of the databank includes seven .DBF files (DATES, SAMPLES, SITE, REGION, INDX, SUBMINST, REFERENC) linked with aid of several keywords (INDEX, INDEX1, SITE, SITEID, SINO, REF). This modified structure is very close to the theoretical concept of the so-called third normal form of the data bank. A new questionnaire of samples submitted for radiocarbon dating is presented.