Operation of the Harwell UK 14C Data Base and its Expansion through Data Exchange with Other Laboratories

Walker, A.J.; Otlet, R.L.; Housley, R.A.; Plicht, Johannes van der

Published in:
Radiocarbon

Publication date:
1990

[RADIOCARBON, VOL 32, NO. 1, 1990, P 31-36]

OPERATION OF THE HARWELL UK ^{14}C DATA BASE AND ITS EXPANSION THROUGH DATA EXCHANGE WITH OTHER LABORATORIES[1]

A J WALKER[2], R L OTLET[3], R A HOUSLEY[4]
and
JOHANNES VAN DER PLICHT[3]

ABSTRACT. The use of computer data bases for storage and retrieval of ^{14}C data is a logical application for the rapidly expanding numbers of ^{14}C determinations. Harwell has established a data base for all samples originating from sites in the United Kingdom and Eire. The core of the data is the Council for British Archaeology's published Index of Radiocarbon Dates which we are expanding to include all Harwell UK dates released for publication by the submitters plus dates from other laboratories both within and outside the UK. As a demonstration of the feasibility of direct database-to-database communication, cooperation has been sought from Groningen and Oxford to transfer computer files containing ^{14}C results for UK sites. Neither of these laboratories use the same system as Harwell for their in-house data base and this exercise highlights the importance of a transfer language for both the national and international schemes as it is no longer practical to carry out such procedures through keyboard typing.

INTRODUCTION

Over the last five or more years, there has been continuing discussion regarding the use of database systems for storage of ^{14}C results and a considerable amount of effort has been devoted to their design and structure (Gulliksen 1983; Moffett & Webb 1983a, b; Gupta & Polach 1985; Engelsman, Taayke & Mook 1986; Engelsman, Mook & Taayke 1987). The data bases serve three different purposes (Fig 1).

First is the specific laboratory data base. This has probably been in existence the longest time and is designed primarily to serve the local needs of the measuring laboratory. All administration of the samples submitted, *ie*, their scheduling, movement through the measurement process, construction of measurement reports and preparation of laboratory date lists for publication in *RADIOCARBON*, is masterminded from this data base. A number of laboratories have already established such data bases, eg, ANU, Groningen, Harwell and Oxford, and many others are in the process of so doing.

Second is the worldwide data base. This seeks to enter all ^{14}C results, regardless of the discipline for which they are measured, into one data base for the benefit of any researcher who needs to use ^{14}C. This is particularly exemplified by the International Radiocarbon Data Base (IRDB) being established by Renee Kra (in press; 1989), who intends it to receive data from all the ^{14}C measuring laboratories in the world. There may also be national data bases, as proposed by Wilcock (1986) for the UK and by PACT (Mook, 1986) for Europe.

The third is a type that sets geographical boundaries on the origin of the samples, not on the laboratory producing the results. The Harwell data base is set up in this way and is devoted to archaeological ^{14}C results from sites in the UK and Eire (Walker *et al*, in press). At the Stavanger Museum, Norway, a similar data base is being set up to store geological samples originating from Norway, Sweden and Denmark (Selsing & Simonsen 1986). The laboratory at Lyon, France is establishing a data base system for archaeological ^{14}C results and will be making these available to French archaeologists through their national telephone network information retrieval system (Mintel).

With so many data bases now being established, it is becoming increasingly important that there should be a means of communication among them and it is the purpose of this paper to report an example demonstration of such exchange using the HLF transfer format.

PRESENT STATUS OF DATA BASES

The philosophy behind the establishment of data bases, for whatever discipline, is to make more information more easily available to more people than is currently possible. The

[1] This paper was presented at the 13th International Radiocarbon Conference, June 20-25, 1988 in Dubrovnik, Yugoslavia.

[2] Isotope Measurements Laboratory, Harwell, UK
[3] Radiocarbon Accelerator Laboratory, Oxford, UK
[4] Centrum voor Isotopen Onderzoek, Rijksuniversiteit, Groningen, The Netherlands

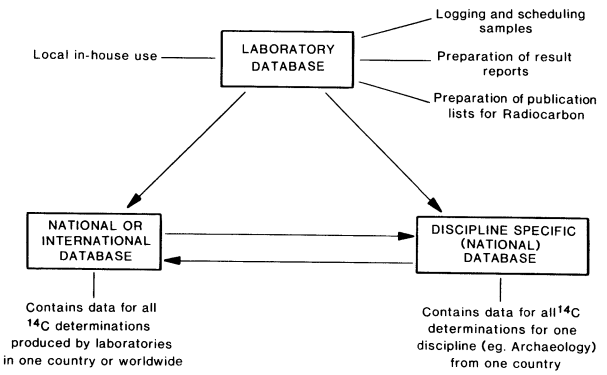

Fig 1. Three types of radiocarbon data bases

number of ^{14}C laboratories now using databases for their data storage is growing. As the dated samples continue to increase and PCs and inexpensive data management packages become more widely available, the number of different bases can be expected to increase dramatically in the next few years. If these data are ever to be made available for future research, a structured search route must be established, entailing 1) an overall, or central, data storage system to which a researcher can refer for results and further reference, and 2) a means of communication among data bases, which will allow data to be collected in comprehensive, logical units through direct transmission (*ie*, using a suitable computer-readable medium) in common transfer format.

The IRDB is intended to answer the first need and considerable progress has been made in its establishment over the last three years. At a workshop held at the 2nd Symposium on Archaeology and ^{14}C in Groningen in 1987, it was decided that the IRDB would not attempt to hold *all* relevant information for each determination but would act as an index in which a minimum amount of data would be held for each result (Walker & Kra 1988). Within each file entry would be a reference field which would act as a signpost, telling the user where the full details on the result could be found (Kra 1989). This signpost field may refer the searcher to another data base, a publication, eg, date list, an archaeological site report or scientific paper, or simply to the measuring laboratory. The advantage of this index-type system is that it takes less space and, therefore, costs less and also makes it more feasible to consider loading a significant number of the many ^{14}C results published over the last 40 years.

It is now widely recognized that, with the ever-increasing numbers of dates produced each year, an efficient system for data exchange and communication has become extremely important.

THE INTERNATIONAL EXCHANGE FORMAT

Practical Demonstration of Data Exchange. We have made a feasibility study of transfers of data from other sources. So far, data have been successfully transferred from the laboratory data bases of Oxford and Groningen, which all use different database management systems. At the time of transfer, Oxford used dBase III on an Amstrad PCW microcomputer (the data have now been transferred to a different system on a VAX); Groningen used their own custom-made database system on an Apple computer and Harwell used STATUS on an IBM PC compatible.

Oxford and Groningen supplied Harwell with a floppy disk containing the most basic of output files, in ASCII format. Oxford's data came on the Amstrad 3″ formatted disks and Groningen's in IBM-compatible format. We then wrote the necessary translation programs to convert the data into our version (HLF) of the proposed exchange format. We developed a general conversion program which will be readily adaptable for other format translations.

Translation of Oxford Data. The Oxford data consists of a single line for each entry, each line containing 12 fields. There was no coding of the fields but they were arranged in a set order and consistency of field numbering is maintained by including every stated field even if it is blank. The strict ordering made the conversion to the HLF format fairly simple. Look-up tables were required 1) to relate a given field number with a keyword heading, 2) to pick out the heading using the number and 3) to collect only dates from the UK and Eire.

The translation of 300 entries from the submitted file of ca 800 entries took only 3.5 minutes on our AT-type PC. The result was a reorganized file (Fig 2), which contains the same data coded according to the HLF fields. This was then transferred to the local STATUS format (Fig 2).

Translation of Groningen Data. The data supplied by Groningen consisted only of entries for the UK and Eire, so no sorting on country was required, and was similar to the Oxford data in that each entry consisted of a constant number of fields given in a set order. The Groningen data consisted of many more fields and required more involved assembly of items to bring them together for the HLF fields (Fig 3). Part of the Groningen information is laboratory treatment for which no provision has been made in the exchange format, except, perhaps in Laboratory Comment.

We feel that translation from almost any format to the exchange HLF format should not pose insurmountable problems. Ideally, translation should be done by the sending laboratory but, in order to complete our UK data base, we are prepared to receive data in

```
589,'Gough's New Cave','BM-2183','bone','horse (E.ferus)','uPal','England','Jacobi',12340,'  150',-20,'AM3'
```

Transfer Format (HLF)

```
*LABCODE                    OxA-589 OXFORD
*SERIES                     Gough's New Cave
*SUBSAMREF                  BM-2183
*SAMMATERIAL                bone
*IDAS                       horse (E.ferus)
*PERIOD                     uPal
*COUNTRY                    England
*SUBM                       by Jacobi
*AGEC                       12340 +/- 150
*DC13                       -20
*MACHINE                    AMS
*REFERENCE 0:               Gowlett et al, 1986a
*LABCOMMENT                 Assumed delta C-13
```

Local Harwell Format (STATUS)

```
$$T
OxA-589
$$T
$$N LABDAT <
#MACHINE=AMS,#DC13=-20,#AGE=12340,#AGETYPE=+/-,#ERRORP=150,#ERRORM=150, <
$$N SAMDAT <
#LABCODE=OxA,#LABNAME=OXFORD,#LABNUM=589, <
Gough's New Cave
$$N SAMMTL <
#SAMMATERIAL=bone, <
horse (E.ferus)
$$N SITELOC <
#COUNTRY=England, <
$$N PERIOD <
uPal
$$Ξ PUBLICATIONS <
Gowlett, J A J, Hall, E T, Hedges, R E M and Perry, C, 1986, Radiocarbon dates
from the Oxford AMS system: Archaeometry datelist 3: Archaeometry 28, 1, p 116-125
$$N LABCOMMENT <
Assumed delta C-13
$$A
```

Fig 2. Oxford data transfer

```
7400|    3801|Ossom's Cave    |R.M.Jacobi     |D.Bramwell | --/--/56
    | 7/11/75|Manifold Valley     |Staffordshire       |Great Britain|
    |        |         |         |bone |         |Palaeolithic|
    |        excavation|settlement|          |archaeology|acid Longin
    |collagen                |-19.13|           |           | 27.08|#0.22
    | 26.76|#0.22 |   10000|-  11000           |   10590|#   70
    |              |         |Cresswell           |818       |Jesus College
    |              |Cambridge CB5 8BL        |United Kingdom         |
```

Transfer Format (HLF)

*LABCODE	GrN-7400 GRONINGEN
*SITENAME	Ossom's Cave
*SUBM	--/--/56 by R.M.Jacobi
*COLL	7/11/75 by D.Bramwell
*TOWN	Manifold Valley
*REGION	Staffordshire
*COUNTRY	Great Britain
*SAMMATERIAL	bone, collagen
*PERIOD	Palaeolithic
*DISCIPLINE	archaeology
*LABCOMMENT	acid Longin
*DC13	-19.13
*PERMOD	27.08 +/- 0.22
*AGEC	10590 +/- 70
*SITETYPE	Cresswell
*SUBMADDR 0:	Jesus College, Cambridge, CB5 8BL
*SUBMADDR 1:	United Kingdom
*MACHINE	GC

Fig 3. Groningen data transfer

any computer-readable form. We are even working on the possibility of translating word-processed text in *RADIOCARBON* publication style format. The British Museum has been kind enough to supply us with a floppy disk containing their last *RADIOCARBON* date list, which, because of its well-defined format, may be a viable transfer process.

RELATIONSHIP OF THE EXCHANGE FORMAT TO THE INTERNATIONAL RADIOCARBON DATA BASE

Our HLF exchange format is fully compatible with that of the IRDB (Fig 4). The main difference is in the field concerned with the ^{14}C determination, to which six fields are allocated in the HLF, to cover all possible forms, whereas the IRDB has one. Other differences are in the geographical coordinates, where the HLF has one field but the IRDB two, and in the comments section where HLF has two fields, one for the submitter and one for the laboratory, but the IRDB has one. It is the suggestion of this laboratory that data being provided for the IRDB should be provided in the full HLF format and that the extraction into the form required by the IRDB package should be made at the destination point, *ie*, Arizona. Data superfluous to the IRDB need not be removed from the HLF files supplied for IRDB entry.

THE WAY AHEAD AND CONCLUSIONS

The successful transfer of the Groningen and Oxford data to the Harwell data base was useful because 1) the data base was easily enlarged without any manual entry, which is not only time-consuming and labor-intensive but can also be a source of introduced errors and 2) it demonstrated that the exchange format is extremely versatile and might with modifications lend itself to a universally acceptable exchange system. Data from the HLF can be transferred directly to the form required for the IRDB, thus making the task of submitters easier.

In conclusion, the Harwell data base has been set up as a collection of all archaeological dates from samples taken at UK and Irish sites, and, if it is to fulfill this objective, it must be expanded from its original core, namely the dates published in the CBA Index (Lavell 1971).

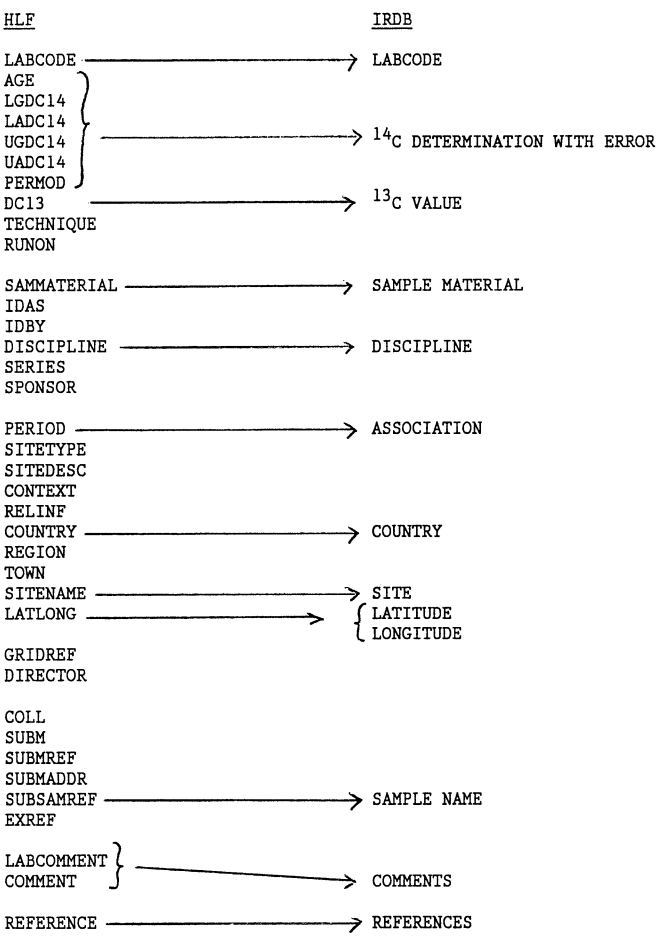

Fig 4. Exchange format (HLF) and IRDB format

Expansion through data exchange has been proved possible by the Oxford and Groningen examples and further data are actively sought from those laboratories which hold in their system UK and Irish dates. The key position held in this expansion by the exchange format reinforces the need to encourage others to accept it as the core of a universally accepted format which can then be used in any exchange situation. Unless data bases have the means to transfer data and thus to assemble single data sets for the researcher they will be of only limited long-term value.

References

Engelsman, FMR, Mook, WG and Taayke, E, in press, Operation of the Groningen ^{14}C data base. *In* Internatl symposium, Archaeology and ^{14}C, 2nd, Proc. *PACT*.

Engelsman, FMR, Taayke, E and Mook, WG 1986 Groningen ^{14}C data base. *In* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 12th, Proc. *Radiocarbon* 28 (2A): 788–798.

Gulliksen, S 1983 Radiocarbon data base, a pilot project. *In* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 11th, Proc. *Radiocarbon* 25 (2): 661–666.

Gupta, SK and Polach, HA 1985 *Radiocarbon dating practices at ANU, Handbook*. Radiocarbon Lab, Australian Natl Univ, Canberra, Australia.

Kra, RS, in press, Establishing a radiocarbon data base. *In* Internatl symposium, Archaeology and ^{14}C, 2nd, Proc. *PACT*.

—— 1988 Updating the past: The establishment of the International Radiocarbon Data Base. *Am Antiquity* 53 (1): 118–125.

—— 1989 The International Radiocarbon Data Base: A progress report. *In* Long, A and Kra, RS, eds, Internatl ^{14}C conf, 13th, Proc. *Radiocarbon* 31 (3): 1062–1069.

Lavell, C, compiler 1971 Archaeological site index to radiocarbon dates for Great Britain and Ireland. *British Archaeol Abs Supp*, Council Br Archaeol, London.

Moffett, JC and Webb, RE 1983a Database management of radiocarbon dates. *In* Aspinall, A and Warren, SE, eds, *Symposium on archaeometry*, 22nd, Proc. Univ Bradford, UK: 67–72.

—— 1983b Data base management systems, radiocarbon and archaeology. *In*, Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 11th, Proc. *Radiocarbon* 25 (2): 667–668.

Mook, WG 1986 Proposal of European scientific and technical network in the field of radiocarbon. Anniversary mtg, *PACT study group*, 10th, Louvain-la-Neuve, Belgium.

Selsing, L and Simonsen, A 1986 Using dBase II for retrieval of radiocarbon data. *Radiocarbon* 28 (3): 1055–1059.

Walker, AJ, Charlesworth, SA, Otlet, RL and Lavell, C, in press, Setting up the CBA Index of UK radiocarbon dates as a micro- computer data base. *In* Internatl symposium, Archaeology and ^{14}C, 2nd, Proc. *PACT*.

Walker, AJ and Kra, R 1988 Report on the International Radiocarbon Data Base (IRDB) workshop, archaeology and ^{14}C conference, Groningen, The Netherlands. *Radiocarbon* 30 (2): 255–258.

Wilcock, JD, Otlet, RL, Walker, AJ, Charlesworth, SA and Drodge, J 1986 Establishment of a working data base for the international exchange of ^{14}C data using universal transfer formats. *In* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 12th, Proc. *Radiocarbon* 28 (2A): 781–787.