[RADIOCARBON, VOL 28, NO. 2A, 1986, P 781–787]

ESTABLISHMENT OF A WORKING DATA BASE FOR THE INTERNATIONAL EXCHANGE OF ^{14}C DATA USING UNIVERSAL TRANSFER FORMATS

J D WILCOCK*, R L OTLET**, A J WALKER**,
S A CHARLESWORTH†, and J DRODGE‡

ABSTRACT. A high-level record structure for the international communication of ^{14}C data bases is proposed. The record structure is based on the high-level communication format first proposed by the authors at the 23rd International Symposium on Archaeometry, Naples, 1983 and does not require the abandonment of existing systems. A description is given of an implementation of the high-level system at Harwell, with examples of retrieval in an international format (the *Radiocarbon* date list format) and a keyword-organized local format.

INTRODUCTION

The need for a single, high-level, internationally accepted structure for ^{14}C result data bases has been frequently discussed. Recently, for example, papers have been presented and discussed at Brookhaven (Otlet & Walker, 1981), Groningen (Otlet & Walker, 1983), Birmingham (Moffett & Webb, 1982), Bradford (Moffett & Webb, 1983a), Seattle (Gulliksen, 1983; Moffett & Webb, 1983b), and Naples (Wilcock, Otlet & Walker, 1983). In the UK, and it is expected that elsewhere too, there has been much local discussion regarding the compatibility, completeness, and structure of individual adopted systems.

It would seem to us now that the probability of expecting a newly-proposed system to be universally adopted in place of established local systems is rather small. The problem still exists, however, that with a diversity of systems, communication between them is improbable, and retrieval by an individual from all of them is a complicated, if not impossible task.

The solution proposed, therefore, does not require abandonment or modification of the existing systems, but seeks to pull all the data from the separate systems into a common high-level communication structure via a series of individually tailored translation programs. In principle, no institution need be concerned with the high-level structure providing they can communicate with it, in both directions, using the language and structure of their own local system.

It is believed that this may be possible because there is an underlying similarity of the data being stored by established systems, which is largely dictated by the *Radiocarbon* publication format. A record struccture for the high-level communication format is here proposed, and a description given of an implementation which translates records prepared for storage in the structure, with examples of subsequent retrieval in an international format (the *Radiocarbon* date list) and a keyword-organized local format. The proposed solution using a high-level format does not require modification of

* Research Centre for Computer Archaeology, North Staffordshire Polytechnic, Stafford, UK
** Isotope Measurements Laboratory, AERE, Harwell, UK
† Isotope Measurements Laboratory, Harwell and School of Archaeological Sciences, University of Bradford, UK
‡ Computer Sciences Division, AERE, Harwell, UK

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

any existing local recording system; all that is required is the writing of two translation programs per installation.

THE PROPOSED HIGH-LEVEL INTERNATIONAL RECORD STRUCTURE FOR ^{14}C DATA BASES

The proposed record structure was evolved from the Harwell ^{14}C data base partly for local data processing needs, *viz,* the automatic preparation of Harwell ^{14}C dating certificates and *Radiocarbon* date lists. These operations have been discussed by Otlet and Walker (1981, 1983). The same procedure has been applied to the database structures of other ^{14}C laboratories readily at hand. It is now believed that the data categories listed below include the majority of categories used by ^{14}C laboratories internationally:

1. LABORATORY RESULTS

LABCODE	Laboratory reference number followed by source of data (*Radiocarbon,* other journal, laboratory direct), eg, HAR-1657 HARWELL
MACHINE	Measurement system, specific reference
RUNON	Measurement date
RECEIVEDON	Date received by measuring laboratory
AGE	Conventional ^{14}C age bp with errors, eg, 2500 ± 50, or $2500 \begin{smallmatrix}+51\\-49\end{smallmatrix}$, or MODERN, or >40,000
DC13	$^{13}C/^{12}C$ measurement in per mil
LGDC14	$\delta^{14}C$ – as in Stuiver and Polach (1977)
LADC14	$d^{14}C$ – all with 1σ error term;
UGDC14	$\Delta^{14}C$ – use same convention as for
UADC14	$D^{14}C$ – AGE
PERMOD	pM as in Stuiver and Polach (1977)

2. SAMPLE DETAILS

SAMMATERIAL	Sample material, eg, bone
IDAS	Identified as
IDBY	Identified by
DISCIPLINE	Discipline, eg, archaeology
SERIES	Series title, as defined in *Radiocarbon*
SPONSOR	Sponsor/Institution
PUBSTAT	Publication status, eg, PUBLISHED, UNPUBLISHED BUT READY, UNPUBLISHED AND NOT READY
SECGRAD	Security grading, eg, PUBLIC, PRIVATE, REFER TO SUBMITTER

3. SITE RELATIONSHIPS

PERIOD	Archaeologic/geologic period or culture
SITETYPE	Site type, eg, CAVE, FORT
SITEDESC	Concise site description
CONTEXT	Context, *ie*, stratigraphic location of sample
ENVIRONMENT	Factual environmental data, eg, geologic description of site (soils); palynology and other associated dating studies
COUNTRY	Country
COUNTY	Region, eg, County (UK)/State (USA)
TOWN	Town
SITENAME	Site name
LATLONG	Latitude/Longitude
GRIDREF	Local grid reference
DIRECTOR	Director of project

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

4. COLLECTION AND SUBMISSION		
	COLL	⟨Collection date⟩ by ⟨Collector⟩
	SUBM	⟨Submission date⟩ by ⟨Submitter⟩
	SUBMREF	Laboratory's reference no. for submitter
	SUBMADDR	Submitter's address
	SUBSAMREF	Submitter's concise sample reference
	EXREF	Extra referencing data
5. COMMENTS		
	LABCOMMENT	Laboratory comments
	COMMENT	Submitter's comments
6. PUBLICATIONS		
	AUTHOR	Author surname, initials, year and order within year, eg, Coles, J M, 1977b (NB: full reference details must be supplied as a separate list, for incorporation in a publications file). AUTHOR fields may be repeated as often as necessary to cover many publications of one sample.

The main headings in the data structure are intended to cover the principal stages in the documentation and progressing of the sample, *ie,*

1) Laboratory processing
2) Sample identity
3) Geographic origin
4) Collection and submission
5) Comments after processing
6) Publication

In a final internationally-accepted record structure, these concepts will be allocated uniquely defined mnemonics; the mnemonics used above are initial suggestions only. Some concepts will be irrelevant for a particular sample and may be left blank: since a keyword structure is used, blank concepts will not take up storage space. The implementation of these concepts at any local laboratory is left entirely to the laboratory concerned. What is important is that the local systems should be able to communicate with all the others via the high-level structure. There have been adverse comments in the past about the omission of important data categories in some ^{14}C records (eg, see Moffett & Webb, 1983). Clearly, the information provided should be as complete as is reasonably possible. There should always be good reason for leaving any of the proposed data categories blank. However, it is asserted that it is better to provide communication with a possibly incomplete local data base than to invest considerable effort and expense in completing the records for the backlog. Note that this high-level structure is independent of any data base management system in use. The authors have used Harwell STATUS for convenience, but any other DBMS could be used locally.

EXAMPLES OF THE USE OF THE PROPOSED DATA STRUCTURE

Local translation programs have been implemented between the proposed high-level format and the Harwell ^{14}C dating certificate, Harwell STATUS, and *Radiocarbon* formats, and other local printout formats. Whatever the source of the data, each sample is expressed as a packet of data concepts in the high-level record structure. The packets are then available

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

for any further data processing needs, eg:

1) communication to any other local system, probably by magnetic tape or disk, but perhaps using data transmission

2) information retrieval

3) automatic preparation of local and international publication formats.

INFORMATION RETRIEVAL

Information retrieval can be carried out on the packets using any logical combination of alphanumeric substring, numerical and numerical range checks using the operators AND, OR, and NOT. The most common requests will probably be logical combinations of the following fields:

Laboratory name, eg, HAR-
Range of Labcodes
Range of Processing dates
Laboratory reception date
Sample material
Age (before, after, range)
Discipline
Series
Sponsor
Laboratory's reference no. for submitter
Director
Collector
Submitter
Submitter's sample reference number
Identification
Identifier
Period
Site type (eg, cave)
Lat/Long or Local grid reference
Site name
Town
Region (COUNTY)
Country
Environmental data
Context
Author and date
Publication status

Although keyword retrievals of the above types are the most efficient, the Harwell STATUS package also allows the retrieval of records with any specified alphabetical string occurring anywhere in the data.

PUBLICATION FORMATS

These are a matter for individual design if the needs are local only. For national and international publication there will be some stipulated format,

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

Search criteria: Hambledon Hill series—⟨THIS IS THE DBMS SEARCH COMMAND TO PRODUCE THIS OUTPUT⟩

Hambledon Hill series ⟨SERIES⟩
Samples from Hambledon Hill, ⟨SITENAME⟩
Neolithic causewayed enclosure, ⟨SITETYPE⟩
Dorset, England ⟨TOWN⟩,⟨COUNTY⟩, ⟨COUNTRY⟩
(50° 54′ 34″ N, 02° 12′ 38″ W, Natl Grid Ref (ST852123). ⟨LATLONG⟩,⟨GRIDREF⟩
All samples coll and subm by R Mercer (1980). ⟨COLL⟩,⟨SUBM⟩

HAR-1882. HH752134 **4560 ± 90**
⟨LABCODE⟩, ⟨SUBSAMREF⟩,⟨AGE⟩
⟨DC13⟩ $\delta^{13}C = -24.9‰$
Charcoal, ⟨SAMMATERIAL⟩
id by CA Keepax ⟨IDBY⟩
as ashy gray wash, 50% Hawthorn type–not twiggy, ⟨IDAS⟩
AML760800, ⟨EXREF⟩
from ditch segment II, Layer 11. ⟨CONTEXT⟩,⟨ENVIRONMENT⟩
Coll Oct 1975 and subm Feb 1976. ⟨COLL⟩,⟨SUBM⟩
Comment (RM): interpretation of layer is problematic—possibly dumped organic deposit similar to that in site F, or richer, more organic silt deposit that incorporated cultural material in fm. ⟨LABCOMMENT⟩,⟨COMMENT⟩

HAR-1885. HH741245 **4480 ± 130**
⟨LABCODE⟩, ⟨SUBSAMREF⟩,⟨AGE⟩
⟨DC13⟩–assumed $\delta^{13}C = -25‰$
Charcoal, ⟨SAMMATERIAL⟩
id by CA Keepax ⟨IDBY⟩
as 50% Hawthorn type and *Prunus* sp from fairly large timbers plus one twig fragment of Blackthorn, ⟨IDAS⟩
AML760792, ⟨EXREF⟩
from ditch area II. ⟨CONTEXT⟩,⟨ENVIRONMENT⟩
Subm Sept 1976. ⟨COLL⟩,⟨SUBM⟩
Comments: $\delta^{13}C$ assumed; (RM): ditch was filled with gray organic soil, charcoal flecks and some burned bone, 1 pot and burned flint. Feature I is roughly oval and cuts into middle fill, overlain by secondary silt deposits. ⟨LABCOMMENT⟩ ⟨COMMENT⟩

Fig. 1. The high-level categories are marked using ⟨ ⟩ brackets to illustrate how easily the *Radiocarbon* format may be generated from the high-level format

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

Search criteria: Hambledon Hill series—		⟨THIS IS THE DBMS SEARCH COMMAND TO PRODUCE THIS OUTPUT⟩
Discipline:	Archaeology	⟨DISCIPLINE⟩
Sitetype:	causewayed enclosure	⟨SITETYPE⟩
Period:	Neolithic	⟨PERIOD⟩
Hambledon Hill series		⟨SERIES⟩
Sponsor:	AML	⟨SPONSOR⟩
Hambledon Hill, Dorset, England		⟨SITENAME⟩,⟨TOWN⟩, ⟨COUNTY⟩,⟨COUNTRY⟩
50° 54′ 34″ N, 02° 12′ 38″ W, ST852123		⟨LATLONG⟩,⟨GRIDREF⟩
All samples coll and subm by R Mercer (1980)		⟨COLL⟩,⟨SUBM⟩
HAR-1882 4560 ± 90		⟨LABCODE⟩,⟨AGE⟩
Coll Oct 1975 and subm Feb 1976, HH752134		⟨COLL⟩,⟨SUBM⟩, ⟨SUBSAMREF⟩
AML760800		⟨EXREF⟩
Charcoal,		⟨SAMMATERIAL⟩
id by CA Keepax		⟨IDBY⟩
as ashy gray wash, 50% Hawthorn type– not twiggy		⟨IDAS⟩
from Ditch segment II, Layer 11.		⟨CONTEXT⟩,⟨ENVIRONMENT⟩
Comment (RM): interpretation of layer is problematic–possibly dumped organic deposit similar to that in site F, or richer, more organic silt deposit that incorporated cultural material in its formation.		⟨LABCOMMENT⟩,⟨COMMENT⟩
HAR-1885 4480 ± 130		⟨LABCODE⟩,⟨AGE⟩
Subm Sept 1976, HH741245		⟨COLL⟩,⟨SUBM⟩, ⟨SUBSAMREF⟩
AML760792		⟨EXREF⟩
Charcoal,		⟨SAMMATERIAL⟩
id by CA Keepax		⟨IDBY⟩
as 50% Hawthorn type and *Prunus* sp from fairly large timbers plus one twig fragment of Blackthorn,		⟨IDAS⟩
from ditch area II.		⟨CONTEXT⟩,⟨ENVIRONMENT⟩
Comment (RM): ditch was filled with gray organic soil, charcoal flecks and some burned bone, 1 pot and burned flint. Feature I is roughly oval and cuts into middle fill, overlain by secondary silt deposits.		⟨LABCOMMENT⟩,⟨COMMENT⟩

Fig 2. The high-level categories are marked using ⟨ ⟩ brackets to illustrate how easily the formats may be generated from the high-level format

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press

eg, the *Radiocarbon* format. All these formats can be produced automatically by different print routines operating on the same high-level record structure. Included here are examples of retrieval listings printed in *Radiocarbon* (Fig 1), and local format (Fig 2).

High-quality printers, such as the daisy-wheel typewriters, can produce camera-ready copy of these publication formats. Alternatively, the computer output could be processed by a photo-typesetter or a microfilm/fiche plotter, or recorded on a magnetic tape or floppy disk for transport to other laboratories.

CONCLUSIONS

A proposed high-level record structure for the international communication of ^{14}C data bases entailing a minimum of interference with existing local systems has been demonstrated. It is hoped that this high-level structure, or one very like it, will be accepted by the international community of ^{14}C laboratories for communication purposes. Indeed, the high-level record structure concept proposed is so flexible that it can be adapted to accommodate all ^{14}C users' requirements with very little extra work. Proposals to establish inter-laboratory communications using this system are actively solicited.

References

Gulliksen, S, 1983, Radiocarbon database: a pilot project, *in* Stuiver, M and Kra, R S, eds, Internatl ^{14}C conf, 11th, Proc: Radiocarbon, v 25, no. 2, p 661–666.

Moffett, J C and Webb, R E, 1982, Database management systems and ^{14}C dating, *in* Laflin, S, ed, Computer applications in archaeology conf, Proc: Univ Birmingham, UK, p 76–78.

———— 1983a, Database management of ^{14}C dates, *in* Aspinall, A and Warren, S E, eds, Symposium on archaeometry, 22nd, Proc: Univ Bradford, UK, p 67–72.

———— 1983b, Database management systems, radiocarbon and archaeology, *in* Stuiver, M and Kra, R S, eds, Internatl ^{14}C conf, 11th, Proc: Radiocarbon, v 25, no. 2, p 667–668.

Otlet, R L and Walker, A J, 1981, A computer system for ^{14}C data management and the writing of laboratory reports for Radiocarbon in their specific publication format: Paper presented at Symposium on archaeometry, 21st, Brookhaven Natl Lab, Upton, New York, May 18–22.

———— 1983, The computer writing of Radiocarbon reports and further developments in the storage and retrieval of archaeological data, *in* Mook, W G and Waterbolk, H T, eds, ^{14}C and archaeology, Internatl symposium, 1st, Proc: PACT, v 8, p 91–106.

Stuiver, M and Polach, H A, 1977, Discussion: reporting of ^{14}C data: Radiocarbon, v 19, p 355–363.

Wilcock, J D, Otlet, R L and Walker, A J, 1983, Proposals for a high-level international record structure for radiocarbon databases, *in* Linington, R, ed, Symposium on archaeometry, 23rd, abs: Fondazione Lerici, p 122–123.

https://doi.org/10.1017/S003382220000802X Published online by Cambridge University Press