3.e

MEMO 4-C 2/76 ~~75~~

From : Luigi Lesca, Anton Schofield and Gérard Danet

To : Distribution

Subject: CINDA Coverage Control System

19th October, 1976

I. INTRODUCTION

1. The need for a suitable coverage control system of programs for CINDA has been stressed repeatedly in past memos by all Centres. The matter has also been discussed on the occasion of 4-C meetings and, more specifically and deeply at last year's CINDA Readers' Seminar held at Saclay, 17-18th November, 1975.

The main objectives aimed at in setting up the coverage scheme were :

(a) Minimum coding effort
Only 'negative' information is coded, that is, information which states that a given publication is not in CINDA. Information such as 'Issue x of journal xxx has been scanned and papers therein have been entered in CINDA' ('positive information') is redundant with CINDA, and past experience has shown that chances are that papers within issue x have been missed anyway.

(b) Automatic maintenance of the coverage file.
This is now possible, as references are now coded in CINDA in a more standardized fashion, and a reference key can be calculated for a high percentage of reference codes in CINDA (*). An order relationship can be established between all references within the coverage file, and maintenance can be carried out as described further on.

(c) Automatic cross-checking against CINDA
Positive coverage information can be generated from CINDA and matched with the negative information of the coverage file, thus ensuring coherence between both files.

(*) The systematic clean-up of remaining non-standard reference-codes will have to be carried out anyway at the CCDN for independent reasons.

2. The program is now operational and it has undergone a first series of tests.

Before describing the operations it performs some definitions are recalled.

II. PUBLICATION UNIT DEFINITION

3. A publication unit (p.u.) consists of :

- for journals, reports or progress reports : an issue
- for conferences : a volume
- for theses : a volume
- for books : a volume

III. ZZ - ENTRY DEFINITION

4. Unlike the normal CINDA entries, the ZZ-entries may refer to a single paper in a p.u. - "punctual entry" -, to an entire p.u. or to a range of p.u's.

5. YES : A YES-entry indicates that the given reference appears in CINDA. This is not strictly speaking coverage information, as 'YES' entries are extracted from CINDA, and not coded separately by the Reader. However, these entries are incorporated into the ZZ file to ensure consistency between this file and CINDA. The YES-entries are produced by machine at the CCDN from the CINDA file and as such they comply with the rules for coding the reference field in the CINDA entries.

ZERO : A ZERO-entry means that the given p.u. or the given range of publication units has been scanned but no CINDA relevant information has been found.

SAME : A SAME-entry refers to a p.u. that carries a double or multiple reference and can be quoted under any of them. Multiple - n - references require the preparation of (n-1) ZZ - SAME entries.

GAP : A GAP-entry points out that a p.u. - or a range of them - has not been scanned.

6. NOTE :
- a YES-entry is a "punctual" entry and in the case of journals it always carries a page number;
- a SAME-entry, generally required only for reports or progress reports, refers always to one and only one p.u. and to the entire p.u.;
- a ZERO-entry refers always to one p.u. or a range of them.

- a GAP-entry usually refers to one p.u. or a range of them. See paragraph 10 for an exception to this rule.

IV. ZZ-ENTRY CODING

7. The ZZ-entries have a format very similar to that of the normal CINDA entries with "ZZ" in the element field, YES, ZERO, SAME or GAP in the quantity field and the reference in the usual place. In order to reduce the number of coverage entries, information on a range of references can be coded on one ZZ-entry. (This entry will be internally split-up into two coverage file records which will be tagged as upper and lower boundaries, and dealt with by the maintenance algorithm).

The format adopted for "range" ZZ-entries is displayed in the examples shown in Appendix 1.

8. Similar coding, with two references, will be used for ZZ-SAME entries.

9. It is important to realize that a perfect functioning of the coverage system is assured only if the strictest coherence is used when coding the reference field in the normal CINDA entries - from which the YES-entries are produced - and the corresponding ZZ-GAP or ZERO-entries.

For instance if the reference field for the CINDA relevant article at page 411 of Nucl. Phys./A., Vol. 224, issue 2 is coded :

REFERENCE TYPE	REFERENCE SOURCE	REFERENCE VOLUME (ISSUE)	REFERENCE PAGE	DATE MONTH	DATE YEAR
J	NP/A	224	411	S	74

i.e., without the issue number given, a GAP entry for the next issue must be coded as follows :

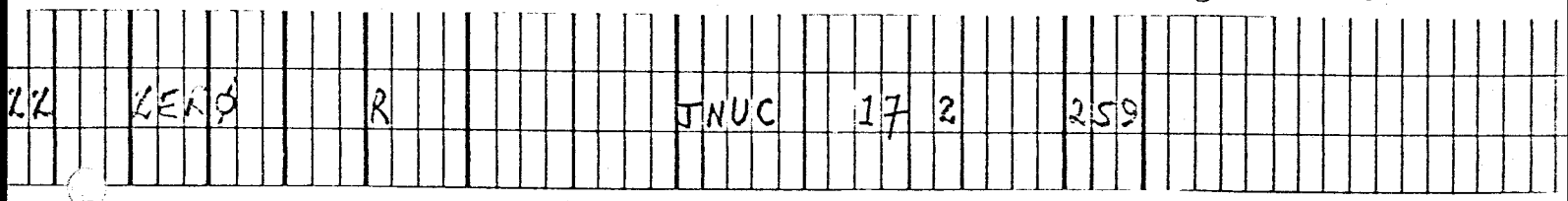
ELEMENT S	ELEMENT A	QUANTITY Q	LABORATORY	BLOCK NUMBER	READER OPERAT. HIERARCHY WORKTYPE	ENERGY MIN.	ENERGY MAX.	REFERENCE TYPE	REFERENCE SOURCE	REFERENCE VOLUME (ISSUE)	REFERENCE PAGE	DATE MONTH	DATE YEAR	Punched
Z		GAP			R			J	NP/A	224	429	S	74	NP/A 224 595 597

where the issue has been identified by its boundary pages and not by its issue number (3). Let us suppose, in fact, that the coder adopts the admittedly more simple compilation:

ELEMENT S	ELEMENT A	QUANTITY Q	LABORATORY	BLOCK NUMBER	READER OPERAT. HIERARCHY WORKTYPE	ENERGY MIN.	ENERGY MAX.	REFERENCE TYPE	REFERENCE SOURCE	REFERENCE VOLUME (ISSUE)	REFERENCE PAGE	DATE MONTH	DATE YEAR	Punched
Z		GAP			R			J	NP/A	224	3	S	74	

which is also formally correct but not coherent with the CINDA entry reference field compilation for this periodical. When, on the occasion of one of the next up-dates, a ZZ-YES entry will be produced for an article at page, say, 550 of the same volume 224, the system will be unable to realize that this page is in issue No. 3 and the two entries - the old GAP and the new YES - will co-exist in the ZZ-file and appear one next to the other (since they carry the same publication date and since the date is included in the sorting key).

Needless to say that for those publications for which the issue number is compulsory in CINDA entries (e.g. Nucleonics) the issue No. must be coded also in the ZZ-entries. In this case a ZZ-ZERO (or GAP) entry may be coded without the boundary page numbers given as in



10. Exceptionally a GAP-entry, which by the definition given above refers to at least an entire p.u. may also be 'punctual', that is, may refer to a specific paper of a p.u. which has been covered but where the article in question has been overlooked. This GAP-entry will be called a "secondary" GAP-entry and may be produced, for instance, from an automatic cross-checking between two bibliographic files.

When such an entry is met either of two cases are possible :

(a) the entry falls in a p.u. for which YES-entries exist : it will be accepted and stored in the ZZ file;

(b) the entry falls in a p.u. for which a ZERO-entry has been prepared : the program will notify the user that an inconsistency has been found and will stop. The entry itself will be rejected and the up-dating operation cannot continue until the incompatibility has not been settled.

11. For the sake of completeness (and detail) we mention a problem which may arise and which has not, until now, been settled. It appears for the periodicals for which the issue number is not given in the reference field. Let us suppose that the ZZ-file contains a YES-entry which falls between two ranges of p.u.'s with a ZERO or GAP code.

If there is more than one issue between these two ranges and one of them has been entirely overlooked, the detection of the omitted issue may appear to be not a simple matter.

To make the remark more concrete, let us take as an example Phys. Rev. C, Vol. 11 which consists of 6 issues and 2121 pages.

Let us suppose that the first two issues do not contain papers relevant to CINDA. The corresponding ZZ-ZERO-entry will be

ELEMENT S (1–2)	A (3–5)	QUANTITY Q (6–8)	LABORATORY (9–11)	BLOCK NUMBER (12–14)	READER OPERAT. HIERARCHY WORKTYPE (15–18)	ENERGY MIN. ● ± (19–22)	ENERGY MAX. ● ± (23–26)	TYPE (27)	REFERENCE SOURCE (28–31)	VOLUME (ISSUE) (32–36)	PAGE (37–41)	DATE MONTH (42)	DATE YEAR (43–44)	45–61
Z		ZER	Ø					J	PR/C	11	1	1	75	PR/C 11 64127

Punched

Now, supposing that the third issue has a CINDA entry on page 869 and that the next ZZ-entry in the coverage file looks like this,

Z		ZER	Ø					J	PR/C	11	1493	5	75	PR/C 11 212167

it may appear impossible to realize that nothing shows up in the ZZ-file concerning issue 4 with pages going from 1071 to 1492. Often, the answer to this remark is given by the fact that, for these publications the periodicity is quite regular and that the date may allow us to pin-point the omission.

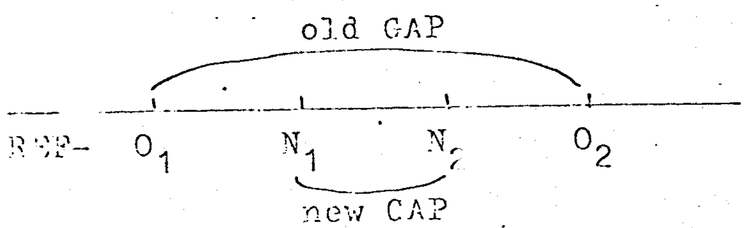
Admittedly, the issue number in these cases would constitute a more explicit information, but :

- it is clearly too much to ask to add the issue number to past CINDA entires;
- it would be annoying for the indexers to code, in addition to normal CINDA entries (which will become YES-entries in the ZZ-file) other entries (YES) to say that the entries have been coded. (Refer to the introduction, paragraph 1).

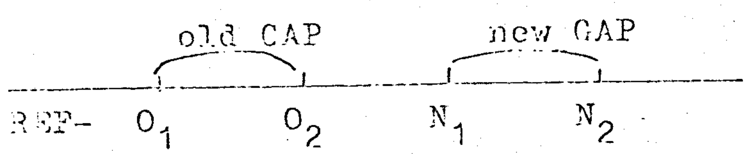
12. In addition to the coding rules outlined in the previous paragraph, the following general criteria must be fulfilled. In what follows O (for "old") refers to ZZ-entries already in the ZZ-file, N (for "new") to the entries to be introduced in the current up-date. Note that a range of p.u.'s may consist of one p.u. only.

(a) a GAP-range defined in a batch of new ZZ-entries must lie totally inside or outside an old GAP range.

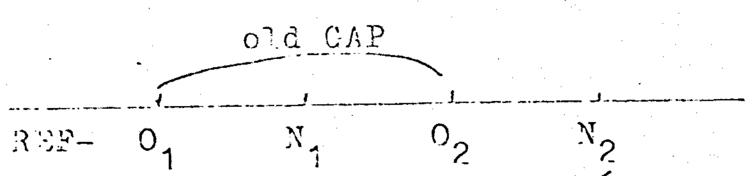
Examples :



right



right



wrong !

(b) two new GAP-ranges must not overlap each other;

(c) a new ZERO-range must not overlap an old YES. It may instead overlap an existing ZERO, for instance, to broaden it.

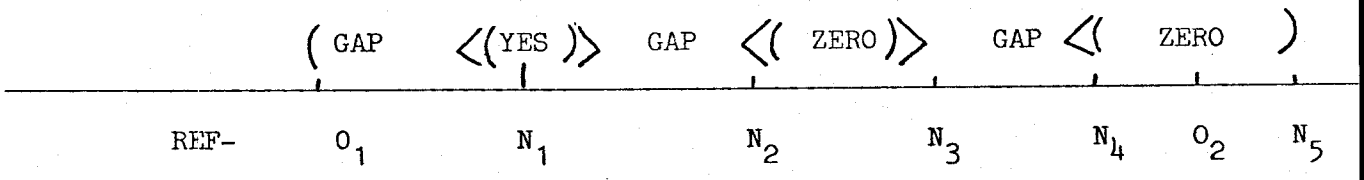
V. OPERATIONS CARRIED OUT BY THE PROGRAM

13. Within the rules layed down under IV all operations or combinations of operations are allowed.

An example will better explain the main features of the program.

Let us suppose that the range REF-0_1 to REF-0_2 is defined as a GAP and that at the next update at REF-N_1, a ZZ-YES entry is produced, a part, between, say, REF-N_2 and REF-N_3 is covered by ZERO-entries, another part, between REF-N_4^2 and REF-N_5^3 (REF-N_5 > REF-0_2) is covered again by ZERO entries, whereas the rest is left unscanned.

The situation which will appear once the up-date has been performed may be visually represented by :



where :

- the brackets refer to ranges with boundaries included;
- < means that the range upper limit is excluded;
- > means that the range lower limit is excluded.
- | means punctual entry,

and can be read as follows :

REF-0_1 - REF-N_1 (REF-0_1 included, REF-N_1 excluded)

is still a GAP.

REF-N_1 appears in a CINDA entry.

REF-N_1 - REF-N_2 (REF-N_1 and REF-N_2 excluded)

is still a GAP,

and so on.

14. At the next update each of the three GAP-sub-ranges thus created can in turn be split in any number of sub-sub-ranges and so on.

15. The reason why sometimes the range boundaries are excluded is, obviously, that there is no simple rule to determine, once REF-N is known, what REF-(N-1) or REF-(N+1) are.

VI. PROGRAM OUTPUT

16. A sample of the program printout with entries sorted according to the Reader code and the reference is given in Appendix 2.

VII. FUTURE DEVELOPMENTS

17. The introduction of YES entries for a p.u. (or a range of p.u.'s) may be worth while considering at a later stage, but only if they are to bring extra information, i.e., the completeness of the coverage of the p.u.'s. It may be questioned whether this is realistic (see end of paragraph 1.a). Some form of completeness is implied of course in ZERO entries : 'everything' in a p.u. is irrelevant to CINDA. Logically, YES and ZERO entries for p.u.'s are on the same footing as far as coverage information is considered. As a first step, only ZERO entries for p.u.'s have been included, as they generate reference keys which do not appear in CINDA, and, together with GAP entries, are indispensable to maintenance operations.

VIII. CONCLUSION

18. We welcome criticisms, questions or comments on what is presented in this memo.

We would also like to know whether the other Centres intend to contribute to the CCDN ZZ-file.

19. The ZZ-entries so far received from our indexers are still scarce. We are trying to extract the YES-entries from the CINDA file at the present time.

The operation is less simple than it may appear at first sight because of the inconsistencies in the reference field accepted by the previous versions of the CINDA checking program. All entries which do not meet the present standard will be rejected and systematically corrected at the CCDN and will appear in the book - 1978 edition.

Once this has been done, the YES-entries stemming from each area will be sent to the Centres.

NDA ENTRY FORM

Page _____

Punched _____ Checked _____

ELEMENT S	ELEMENT A	QUANTITY Q	LABORATORY	BLOCK NUMBER	READER OPERAT. HIERARCHY WORKTYPE	ENERGY MIN.	ENERGY MAX.	REFERENCE TYPE	REFERENCE SOURCE	REFERENCE VOLUME (ISSUE)	REFERENCE PAGE	MONTH DATE	YEAR DATE	COMMENTS
Z		ZERØ			5			R	SGAE-PH-	2407		175		SGAE-PH-2503 D75
										2514		Ø75		2524 N75
										2560		176		2566 276
		↓							↓	2577		276		2645 876
		SAME						R	INDC(SEC)-51			Ø75		NEANDC(ØR)144LØ75
									↓					SGAE-PH-2513 Ø75
								R	SGAE-PH-2525			N75		IAEA-169 3 74
								R	SGAE-PH-2567			276		IAEA-169 3 74
		↓						R	NEANDC(ØR)1404			774		SGAE-PH-2339 974
		GAP			↓			R	SGAE-PH-2334			974		
		ZERØ			2			JNP/A	259		1893	76		NP/A 259 3643 76
									260		1894	76		260 3484 76
									261		1894	76		261 3644 76
									261		3655	76		261 5395 76
									263		3656	76		263 5396 76
									264		3657	76		264 5397 76
									265		3657	76		265 5377 76
									266		2698	76		266 5378 76
↓		↓			↓				↓	267	3658	76		↓ 267 5398 76

APPENDIX 1

Part of

Memo 4C-2/76

Text pages have been stolen
by HDL.