

MEMO 4-C 2/83

To : See distribution
From : Peter Johnston PS
Subject : The choice of a cut-off date for the
Archival volume of CINDA to be prepared
in March, 1978.

8th February, 1977

This memo. describes a statistical survey of the CINDA file in January, 1977. The choice of a suitable cut-off date for the separation of CINDA into an archival volume and subsequent cumulative supplements depends on the balance between the size of the supplements and the degree of overlapp between entries in the supplements and 'blocked' entries in the archival volume.

The block structure in CINDA means that whatever the cut-off date, some blocks will exist, and will be created, with entries dated before and after the chosen date. The statistical survey presented here is based on the assumption that only blocks with no entries dated after the cut-off date will appear in the Archival volume, and that all blocks with entries dated since the cut-off date will appear in the cumulative supplements.

The alternatives of splitting blocks, or including the existing 'overlapping' blocks in both the archival volume and the supplements, appear less appropriate; the first owing to the loss of information, and the second because of duplication (approximately 14,000 entries).

Distribution

Dr S. Pearlstein
Dr V. Manokhin
Dr J.J. Schmidt —

cc/ *three*
Kimmel
Kessler
Kosow
Okamoto
Schmidt
Schwartz
Smith
Yaghubian
Kammert

Table 1 and Figure 1 show a simple breakdown of CINDA, ignoring the block structure, showing the number of CINDA entries for each publication date.

TABLE 1

Total entries January, 1977 138241

Publication Date	Entries	Publication Date	Entries
* 1976	4755	1959	2176
* 1975	7401	1958	4116
1974	8725	1957	2607
1973	10325	1956	1933
1972	9321	1955	1620
1971	9787	1954	570
1970	11067	1953	728
1969	8077	1952	515
1968	8004	1951	577
1967	7429	1950	628
1966	6653	1949	455
1965	7246	1948	151
1964	5205	1947	353
1963	3471	1946	118
1962	2958	before 1946	1337
1961	3541		
1960	3684		

* incomplete.

Table 2 presents a summary of the structure of CINDA blocks for which the most recent entry is dated between January, 1972 and January, 1975 (i.e. dates 72, 73, 74). The figures presented show the percentages of entries in these blocks dated before the most recent entry. For example, YEAR - 2 refers to entries dated 1971 in blocks for which the most recent date is 1973.

TABLE 2

Date	% of Total number of entries	Cumulative % of entries before given date
Most recent year	60.7	
Year - 1	10.7	39.3
Year - 2	9.0	28.6
Year - 3	5.0	19.6
Year - 4	3.5	14.6
Year - 5	2.8	11.1
Year - 6	2.1	8.3
Year - 7	1.7	6.2
Year - 8	1.6	4.5
Year - 9	1.2	2.9
Year - 10	0.9	1.7
Before Year - 10	0.8	0.8

These figures on the date distribution of CINDA blocks can be used to estimate the extent to which new entries in cumulative supplements would be blocked with entries in an archival volume.

The following assumptions have been made :

- (1) the number of new entries to CINDA remains constant over the period 1978-1983.
- (2) The date structure of CINDA blocks remains the same as for the period 1972-1974.

TABLE 3

<u>YEAR</u>	<u>PERCENTAGE DUPLICATION IN 1983</u>
Year of cut-off for the Archival volume (All blocks containing entries earlier than the given date)	Percentage of duplicated supplement entries dated earlier than the cut-off, but included because of blocking with 1978-1983 entries. (*)
1970	2.3%
1971	3.1%
1972	4.6%
1973	6.4%
1974	8.8%
1975	12.0%
1976	16.3%

(*) This assumes that all entries between the cut-off date and 1978 are already blocked with all appropriate older entries before the separation is made into archival volume and supplements, so that all such blocked older entries do not appear in the archival volume.

This extent of duplication between the archival volume and the supplements must be balanced against the increased size of the supplements for earlier cut-off dates.

TABLE 4

<u>Cut-off date</u>	<u>Archival volume</u>		<u>1st Supplement</u> ^(*)	<u>5th Supplement</u> ^(*)
	<u>No. Entries</u>	<u>Pages</u> (60/page)	<u>CINDA 1978</u>	<u>CINDA 1983</u>
1972	87,000	1,450	66,000	116,000
1973	96,000	1,600	56,000	106,000
1974	109,000	1,800	44,000	94,000
1975	120,000	2,000	33,000	83,000
1976	132,000	2,200	21,000	71,000

(*) Estimated on the basis of the annual growth between 1970 and 1975.

On the basis of the figures in Table 4, a choice of date between 1974, 1975 and 1976 would seem preferable. 1975 provides a compromise between excessive duplication and the size of the supplements.

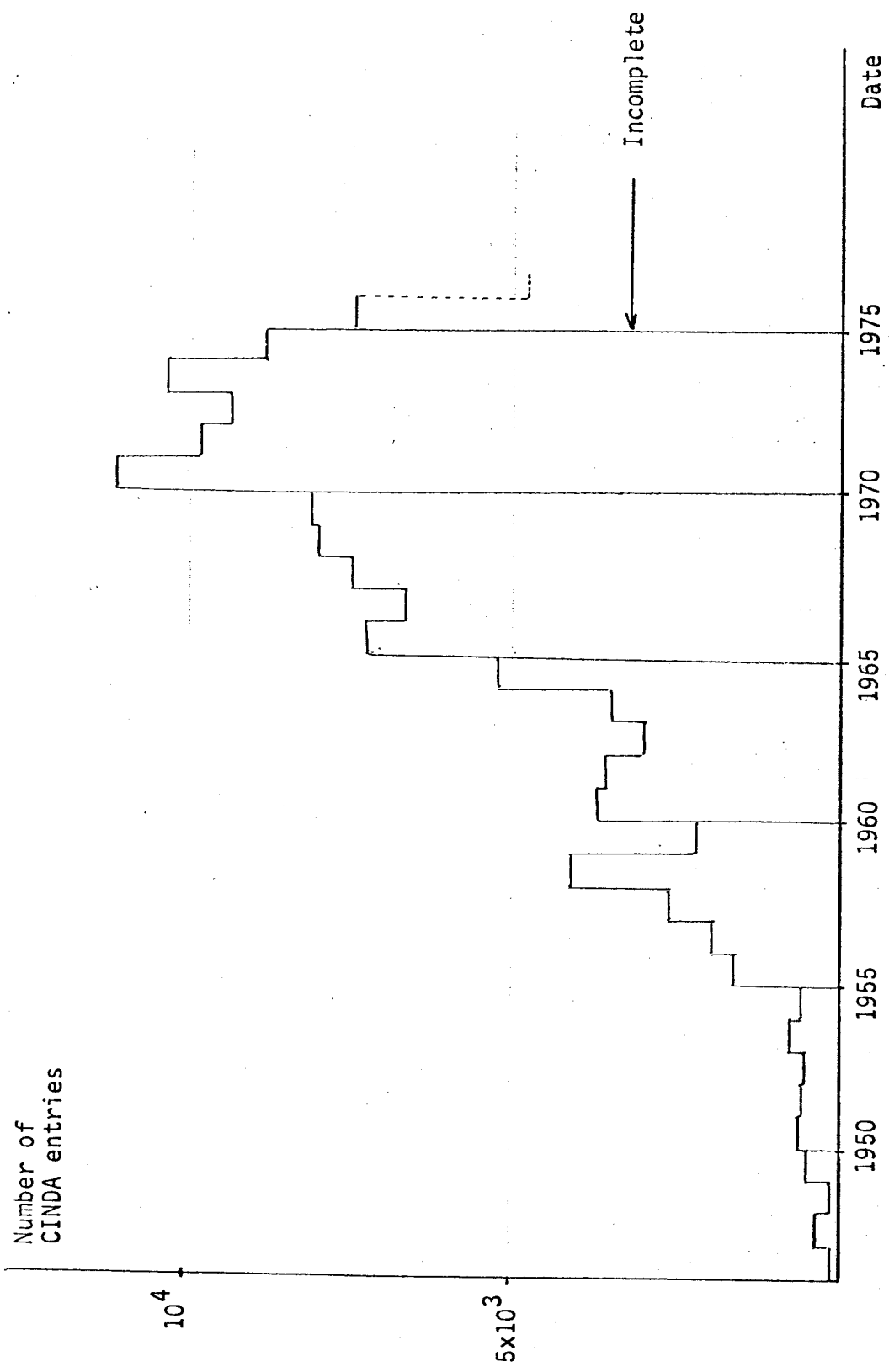


Figure 1. THE NUMBER OF CINDA ENTRIES FOR EACH YEAR SINCE 1946. (CINDA - JANUARY, 1977)