

To : Professor H. Goldstein
Dr. V. Manokhin
Dr. S. Pearlstein
Dr. J.J. Schmidt

From : Fritz Fröhner and Nigel Tubbs

Date : 8th October, 1974

Subject : Comparison CINDA versus INIS and NSA

1. History

In order to find out where we stand with CINDA relative to the rapidly developing INIS system, we decided earlier this year to compare test retrievals from INIS with CINDA. After mentioning this project at the 4C-Meeting in May, Nigel Tubbs had an action put on him to communicate the results to the other three centres. The following memo. is written in response to this action (No.12 in Appendix N of INDC(NDS)-58/G).

2. Retrievals requested

Because ZAED (Zentralstelle für Atomkernenergie-Dokumentation) in Karlsruhe offered adequate retrieval possibilities from the INIS file, we asked them to perform 4 retrospective searches :

- (1) all references to neutron physics;
- (2) all references to work on $^{54}\text{Fe} + n$;
- (3) all references to work on $^{27}\text{Al} + n$;
- (4) all references to work on $^{27}\text{Al}(n,p)$ and $^{27}\text{Al}(n,\alpha)$;

with publication dates in July, 1972 or later.

ZAED fulfilled these requests with retrievals made (free of charge) on 19th July, 1974 from their data base which consists of INIS, NSA and IKK (Informationen zur Kernenergie und Kerntechnik, which is essentially the West German input to INIS. For the purpose of the present comparison IKK was considered as part of INIS).

3. Preparatory work

In order to be able to compare, we had to put ZAED and CINDA references on an equal footing. This involved :

- identification of multiple CINDA entries referring to a single document (ZAED entries corresponded to single documents, not ZAQ Ref. combinations);

- elimination of the effect of different reference date conventions : CINDA uses the conference date, ZAED the publication date of conference proceedings, so that, e.g., 71 Kiev occurs before the search period according to CINDA but not according to ZAED;
- correction of occasional malfunctioning of the ZAED retrieval programs : some references were listed which were published definitely outside the search period, or duplicate entries in INIS and NSA were not always recognised as such so that the same document was listed twice.

The amount of this preparatory work was such that we decided not to include the voluminous listing generated by request (1) but to restrict the comparison to the more manageable requests (2), (3) and (4). The CINDA file was used as it existed in July, 1974 - not very different from CINDA74.

4. Number of references found

After elimination of invalid and duplicate entries, the results listed in Table 1 were obtained.

Table 1 - Number of different references found in the three files.

File(s)	Quantity	Retrieval		
		$^{54}\text{Fe}+n$	$^{27}\text{Al}+n$	$^{27}\text{Al}(n,p), ^{27}\text{Al}(n,\alpha)$
all 3 files, valid	n_T	49	85	21
INIS + NSA, valid		17	35	7
invalid		5	22	7
<u>Single occurrences, (valid)</u>				
INIS	n_1	9	16	2
NSA	n_2	15	34	7
CINDA	n_3	42	71	19
<u>Double occurrences, (valid)</u>				
INIS-NSA	n_{12}	7	15	2
NSA-CINDA	n_{23}	9	20	6
CINDA-INIS	n_{31}	7	11	2
<u>Triple occurrences, (valid)</u>				
INIS-NSA-CINDA	n_{123}	6	10	2

The number of invalid ZAED entries is seen to be quite high. The reasons were mentioned above : indexing deficiencies and occasionally, malfunctioning duplication checks in the retrieval programs. The number of references found in CINDA is about twice that from NSA, which in turn is approximately twice

that in INIS. Most of the INIS entries are also found in NSA. This high correlation between INIS and NSA is, of course, to be expected since part of the input for the two files comes from the same compilers. On the other hand, the overlap between CINDA and NSA/INIS is surprisingly small.

4. Coverage estimates

The unknown number n of actually existing references can be estimated from the single and double occurrences in two uncorrelated files : the probability p_{12} to find a given reference in both files is the product of the probabilities p_1 and p_2 to find it in file 1 and file 2,

$$P_{12} = P_1 P_2.$$

Equating these probabilities with the fractions of references occurring in the files, $p_1 = n_1/n$, $p_2 = n_2/n$, $p_{12} = n_{12}/n$, one gets the following estimate for the true number of references, n :

$$n = n_1 n_2 / n_{12}.$$

The data on n_1, n_2 and n_{12} in Table 1 were used to calculate n for the pairs CINDA-INIS and CINDA-NSA, for which significant correlations appear to be absent. Because of the limited sample sizes, the uncertainties of the estimates (cf. Table 2) are of the order of 10 - 20%. The n values from both file pairs can therefore be considered as consistent. Their average, \bar{n} , was used to calculate the fractions n_i/\bar{n} ($i = 1,2,3$) which, of course, are estimates of the coverage achieved in the three files. These estimates are also shown in Table 2. The last line shows the estimated coverage which could be achieved by combining all three files.

Table 2 - Estimate of the true number of references and of the file coverages.

File(s)	Estimated Quantity	Retrieval		
		$^{54}\text{Fe}+n$	$^{27}\text{Al}+n$	$^{27}\text{Al}(n,p), ^{27}\text{Al}(n,\alpha)$
<u>Estimated true number of references</u>				
CINDA-INIS	n	54	103	19
CINDA-NSA	n	70	121	22
Average	\bar{n}	<u>62</u>	<u>112</u>	<u>21</u>
<u>Coverage estimates</u>				
INIS	n_1/\bar{n}	0.15	0.14	0.10
NSA	n_2/\bar{n}	0.24	0.30	0.33
CINDA	n_3/\bar{n}	0.68	0.63	0.90
INIS-NSA-CINDA	n_T/\bar{n}	0.79	0.76	0.95

5. Conclusions

Although the estimated coverages have appreciable uncertainties (of the order of 10 - 20%) it appears that in the cases $^{54}\text{Fe}+n$ and $^{27}\text{Al}+n$ CINDA covers about two-thirds of the literature, NSA about one-third, and INIS at the present time about one-sixth. The apparently better CINDA coverage in the case of $^{27}\text{Al}(n,p)$, $^{27}\text{Al}(n,\alpha)$ may be due to the special effort devoted to threshold reactions at the CCDN during production of the "atlas" EANDC 95"U". In fact, this particular retrieval was requested with the aim of checking whether there was a noticeably improved coverage. The percentage of references missing in CINDA (of the order of 30%) is disturbingly high. Closer inspection of those which appear in NSA or INIS shows three things.

First, the missing references are by no means of marginal interest only. It is true that they are often dissertations or progress reports or abstracts of talks presented at local meetings, belonging to some already existing block of CINDA entries, but there are also quite important reports (in the samples investigated there were, e.g., ORNL, ANL, AE, CAE, and INDC reports on work by Labauve, Harvey, Ford III, Etemad, Cabe, Almen) or regular journal articles (e.g., NP/A, NIM, YF referring to work by Bhatia, Mubarakmand, Savelev).

Second, the references missing in CINDA tend to be quite recent, and the fact that they already occur in INIS or NSA seems to indicate that the average time delay is smaller in these files than in CINDA (on the other hand, one may then expect that eventually a part of the missing entries will find their way into CINDA).

Third, the gaps in CINDA concern all four service areas.

Generalizing boldly one can perhaps summarize by stating that :

- INIS is developing rapidly and has an impressively short time delay, but its coverage was only about half as good as that of NSA for publication dates from July, 1972 to July, 1974. Admittedly, INIS became fully operational only by January, 1973, and its usefulness will certainly increase steadily.
- NSA still contains almost all references found in INIS, the correlation between the two files is quite high.
- The time delay between publication of a document and occurrence of the reference seems to be somewhat longer in CINDA than in INIS and NSA. CINDA offers better coverage (about 60 - 70% for fairly recent references, perhaps somewhat better on the long term) and far superior retrievability.

6. Practical Consequences

Regular checks against NSA + INIS should reduce time delays and improve coverage of CINDA. A reasonably rapid coverage of 75 - 85% appears to be achievable in this way.

We are at present looking into the practical problems involved. We shall also try to fill gaps in CINDA discovered during the comparison as far as they are not filled in the meantime (i.e. in the CINDA74 Supplement) and as far as the CCDN service area is involved. References to be treated by the other centres will be listed and the lists distributed after completion of CINDA operations for the supplement.

Needless to say, INIS and NSA might benefit even more from similar comparisons with CINDA.

BY FROM IAEA REGISTRY = COPY FROM IAEA REGISTRY = COPY FROM IAEA REGISTRY = COPY
BY FROM IAEA REGISTRY = COPY FROM IAEA REGISTRY = COPY FROM IAEA REGISTRY = COPY

4-C Memo

DA-1324-0

NOTE FROM CENTRAL REGISTRY

ATTACHED ENCLOSURE WAS RECEIVED WITHOUT COVERING NOTE.

FROM: *DELD*

ADDRESSED TO : *Schmidt*

DESPATCHED ON : *74109/30*

SUBJECT: *4-Center Memo*

- ~~Schmidt~~
- Calamand
- Attree
- ~~Dunford~~
- Lemmel
- Lenley