MEMO 4-C 2/77

To : See Distribution

From : H. Derrien, A. Schofield and P. Johnston

Subject : CINDA Blocking and Cleaning-up

3rd November, 1976

The present memo. concerns the letters from C. Dunford (20th August), H. Goldstein (7th September), H.D. Lemmel (7th September) and the MEMO 4-C 1/92 from N.E. Holden. Several questions need to be clarified concerning different actions put on the Centres for CINDA from the last 4-C meetings, especially, cleaning-up and blocking. An important effort has been achieved in this direction and the main volume 76/77 is certainly better than the main volume 75. But it is obvious that a perfect version of CINDA is far from being achieved, and a large amount of work remains to be done. Also, several "mistakes" have been made during the blocking and cleaning-up, partly, because no rules have been established for these operations. We must consider the 76/77 main volume as an initial step : a definitive volume will be published in 1978. Eighteen months remain in which to complete the work, but before continuing, it now seems obvious that some general rules must be defined in order to avoid too many criticisms from those responsible for CINDA in each Centre (although we must confess that there are very few criticisms on the part of CINDA users!).

The following points concerning CINDA will be discussed, and for each of them we will endeavour to propose some general rules :

I - Existing mistakes,

II - Blocking,

III - "no book" flag and

IV - Laboratory indexing.

I - EXISTING MISTAKES IN CINDA

I.A We are rather distressed at the fact that once again Regions 1 & 3 are exchanging correspondance taking examples from Region 2 entries. We find it hard to believe that errors are not as easy to detect in Regions 1, 3 and 4! In point of fact, each page of CINDA contains several errors : references without author name, references without comment, author name without . and + , lines without references, double references, etc.; all mistakes, perhaps minor, but which are deviations from the CINDA coding rules. The frequency of these kinds of errors is the same in each region, and can be as high as 30% in some pages in CINDA (for instance, page 223 of the 76/77 main volume). More serious errors are those concerning the energy range, the

Distribution : S. Pearlstein
J.J. Schmidt
V. Manokhin

Attree
Lemmel
Okamoto
Lorenz
Lammer
Lesser
Schweer
Smith
Yaplumbian

laboratory code, the reference code, the quantity, the incorrect isotope (for RES, LDL, etc...), the blocking, etc. Unfortunately most of these mistakes are not apparent and could be corrected only by the authors themselves or by the CINDA users. But, for the errors which are evident, the correction is a matter of doing the cleaning-up in the Centres; their existence is also evidence of the fact that the cleaning-up of CINDA is far from being complete.

I.B Rules proposed for the correction of the mistakes

(1) It is impossible and not essential to correct all the minor errors : a part of them should be corrected during the completion of the blocking work. To avoid such errors in the future, the CINDA indexers should be more careful and pay more attention to the existing CINDA coding rules;

(2) An effort should be made :
  (a) to enhance the feedback from the CINDA users;
  (b) to encourage authors to criticise the blocking of their own experiments (such an effort has been made for two countries in the CCDN service area, but the efficiency was very poor or nil);

(3) The correction of the apparent and severe errors should have a high priority in the Centres and should be done, in parallel with the blocking operation, by a systematic scanning of CINDA. The deadline for this operation should be spring 1978.

(4) Furthermore, those responsible for CINDA corrections in each Centre should be aware that errors can also occur during the blocking and correction operation itself; all feedback from the CCDN, after a CINDA operation, should be examined very carefully to ensure that the correction and blocking have been made correctly. Some evident errors are directly corrected by the CCDN staff; but the latter cannot ensure that all these kinds of errors could be corrected by them.

II BLOCKING OPERATION

II.A The blocking work done at the CCDN for the 76/77 main volume was of a rather general nature, i.e., a block corresponds to a set of experiments with (of course) the same ZAQ-LAB, done by the same team of experimentalists working on the same subject for a given purpose. The work was carried out by a single person (three months full time) under great pressure in order to fullfill an action issued by the 4-C. The reason for adopting a rather large blocking procedure was that it was the only way to realise, in a short space of time, a reasonable blocking for the majority of the experiments in Area 2. But, we do not claim that our method of blocking is definitive. The undertaking of a general cleaning-up with the deadline of spring 1978 has already begun at the CCDN. It is evident that some important blocks will be reconsidered and all the errors in the actual blocking will be corrected. Do not forget that the CCDN can block only "après coup" and not while compiling because of its Reader network, and entries arriving at the last moment cannot be blocked. This, for instance, explains the "oddity" spotted by H. Goldstein.

II.B Before discussing rules for the blocking, we will give an example of the complexity of this problem. We choose the block Pu-239 RES SAC (page 1787, main volume 76/77) which one of us knows very well. This block contains 19 lines including 2 EXFOR index lines. It is the result of the analysis of 3 types of measurements :

(1) Total cross-section from H. Derrien,
(2) Fission cross-section from J. Blons,
(3) Scattering cross-section from J. Trochon.

The data for each line of the block are related to the 3 measurements; the main reference is the thesis FRNC-TH-487 which contains a review of all experiments. The first total cross-section has been measured in 1964 and the last in 1970 (see the block Pu-239 TOT SAC and EXFOR 20445 which contains 17685 data points). There is a smooth progression in the measurements and in the improvement of the experimental conditions; it is difficult and undesirable to divide this work into partial blocks, because there is a very strong correlation between each measurement.

II.C Finding some precise rules for CINDA blocking is quite difficult : no agreement can be established without a concensus on the purpose of the blocking. One obvious reason for blocking is to avoid a scattering of information on the "same" experiment over one or several pages of a CINDA book or listing. If we consider only this purpose, a limit of the blocking concept should be obtained through the appearance in CINDA of a LAB sort inside each quantity, and this kind of blocking could be done automatically by program. Another purpose of the blocking, which is the opposite of the first, is to put only "equivalent" (= giving essentially the same information) publications in the same block. This kind of blocking was done, in the old CINDA book, in extra lines for a given quantity. Between these 2 extreme interpretations, there are several others which differ by the degree of correlation between the references in a block; it is, then, possible to define 3 types of correlation leading to the following definitions of the block :

(a) Different aspects of the same measurement

For instance, one article gives mainly the experimental results; a second describes the experimental method in more detail; in a third article more emphasis is put on the theoretical interpretation of the results; that corresponds to the description of " "a single experimental block".

(b) Historical development of the measurement of one quantity

All kinds of correlations as preliminary work, improvement of technique, extension of the energy range, change in the detector, new technique, new machine, etc.....; that is the description of a 'composite experimental block' which is more closely related to the history of a given quantity measurement in a given laboratory by the same team of experimentalists.

(c) Different measurements leading to the determination of the same quantity

The blocks for the quantities like RES or LDL must fall into this category because the determination of these parameters needs the contribution of several experiments (for instance, RES is often a mixing of results from total, fission, capture, scattering etc....).

Obviously the two abovementioned extreme interpretations must be rejected. We are inclined to adopt the "composite experimental block", which effectively consists of a)+b)+c), and which also means that, in our philosophy, the aim of the experiment (i.e. the measurement of a certain quantity) is predominant on the way the experiment is performed.

Since some arbitrariness is inevitable, we accept that some errors (provided that they are not "blatant") could appear in a "Composite experimental block". We understand that a general consensus on this point is very difficult because the decision cannot be based on very precise considerations.

Here again, our preference is mainly dictated by the expectations of users, or at least, by what we believe they should expect. One of the aims of the blocking operation is also to reduce the number of publications a user has to consult in order to find the information he requires. If we accept the principle that a new measurement of the same quantity by the same authors at the same laboratory using essentially the same apparatus aims at an improvement of the earlier experiment, it is reasonable to expect that the publication which refers to the most recent measurements carries enough information for the majority of the users. If the new articles are blocked together with the old ones, the users will have the possibility of restricting consultation to the most recent publication, whereas if each phase of this "macro-experiment" is blocked separately he will feel obliged to examine each of them.

A certain discretion should be left to the person responsible in each Centre in the blocking operation as he is in the best position to decide what type of blocking should be made for some complicated case. He also has the possibility of consulting the authors (although the latter are rarely disposed to provide feedback), or the data compilers in the Centres (who know the history of a experiment when compiling a set of experimental data). However, some general guide lines must be established. We propose the following :

II.D General guide lines for blocking

(a) The blocking should not be more restrictive than a "single experimental Block" as defined in II.C.a;

(b) A "composite experimental block" is allowed for sets of experiments which are sufficiently correlated : the degree of correlation is left to the discretion of the indexer, but not beyond a certain limit;

(c) limit of correlation in a "composite experimental block" : if the block corresponds to a long interval of time, no large gap should be allowed in this interval, i.e., the experiment should be continuously "going on" during this time.

(d) Exception to the precedent rule : if one experiment done in 1975 superseded the same experiment done in 1965, both experiments should appear in the same block with appropriate comments, although the gap is very large.

(e) In the same block, several EXFOR lines could appear for the following reasons :

(i) part of the results have been published in a preliminary paper, compiled and transmitted in EXFOR. The work has been completed, published later and another EXFOR tape is transmitted ;

(ii) there are two complementary "single experiments" in a "composite block" and the experimental data have been compiled at different times. However, the EXFOR compilers should avoid using new EXFOR "accession numbers" when adding tables to a work which has already been transmitted and they should make sure that all the relevant litterature is in the corresponding CINDA block ;

(f) Multiple entries for the same quantity from the same article should be in the same block, unless there is no correlation between these entries. However, many multiple entries should be avoided in future by using appropriate comments ;

(g) Review papers and evaluations should not appear in an experimental block, unless they have been made by the principal author in the block with particular emphasis on the experiment or the set of experiments in the block.

III. - THE "NO BOOK" FLAG

It is not possible to separate the "no book" flag problem from the blocking problem. It is a means of deciding what should appear in the CINDA book when the blocking has been done. At the present time, there are only a few "no book" flags within the Area 2 CINDA references. The assignment of the "no book" flag will be made for the CCDN area within the continuing cleaning-up planned for 1976-1978. This flag will be assigned only to those references concerning :

- abstracts,
- "preliminary", "to be done", "planned", etc.
- most of the progress reports, and
- some "superseded".

In so doing, we will cancel approximately 20% of the references.

At first sight, it seems normal to assign a "no book" flag to all "superseded" references. But, we do not agree with the philosophy which seems to be adopted in Area 1 concerning the comment "superseded" ; there is an exageration in the use of this comment. What is superseded in an article is often the experimental results (i.e., the results of the analysis of the raw data) and not other important information such as the experimental method or the method of analysis. Thus, it is dangerous to assign the comment "superseded" and nothing else and to systematically assign a "no book" flag. It would indeed be exageration to reduce each block to the main entry and its EXFOR index line in the CINDA book! In using such a method, there is a risk of dissatisfying many CINDA users who could complain of a lack of information.

Apart from the above-mentioned references, the aim of which is solely to state clearly what is going on, all other publications should appear in the book. This should be the case, in particular, for the so-called "equivalent publications" - typically a laboratory report, a conference paper and an article in Nuclear Physics - which contain essentially the same information and which should, nonetheless, all be cited. The only way to assure a user that they are equivalent is, in fact, to reference them all under the same block accompanied by a pertinent comment. Any omission may, in fact, be considered by the user to be an oversight on the part of the CINDA indexer. It is also a fact that there are few requests from CINDA users for CINDA retrievals : the users know only of the book and all omissions in the book would be considered as a missing reference.

We propose the following guide-lines for "no book flag" and "superseded" :

1. When assigning the comment "superseded" to a reference, the CINDA indexer should bear in mind that :

(a) sometimes only a part of the article, for instance, tables of values, is superseded by a more recent article which, often, refers to the precedent article as a description of the experiment;

(b) the notion of "superseded" depends on the need of the users; for instance, for an evaluator nothing should be superseded when comparing sets of data from the same group of physicists or from another group.

2. The "no book" flag should be assigned only to the abstract and to all references stating that an experiment is planned, to be done, or in progress (mainly from progress reports) which are really superseded references.

Now we will give some examples of (i) blocking, and
(ii) superseded references,
which seem to be exagerated and, sometimes erroneous

EXAMPLE I page 1604 - U-235 NF ORL

There are 3 references in the block, all of them superseded ; one of the references is superseded by a 1966 measurement the reference of which is not given ; another reference is apparently superseded by WASH 1053, p.58 which does not appear as a line in the block (for the reason that it is a reference for ALFA measurements also superseded and "no book" flagged.) The first reference, NSE 23 45, from Gerard de Saussure is mainly an α measurement, which should appear in the comment. The 1966 measurement should appear in the block as the main reference. Where is the corresponding CINDA line? the report ORNL-TM-1804 in another block U-235 NF ORL (page 1605), which supersedes the 1966 Paris Conference? It seems, then, that the two blocks are sufficiently correlated to appear in a single block. Another remark : NSE 23 45 is cited as a reference in 66 PARIS for a detailed description of the experimental conditions; at least, this part of the article is not superseded by 66 PARIS.

EXAMPLE II page 1609 - U-235 NF ORL (first ORL block on the page)

We agree that this block could not be mixed with the blocks on pages 1604 and 1605 mentioned above ; the aim of the experiments is different : α measurement in the 65-67 experiment (machine RPI+ORNL, responsible G. de Saussure); fission measurement in the 71-72 measurements (machine ORNL, responsible R. Gwin). Nevertheless, the presence of the review paper of R. Peelle in the block is a proof that the Area 1 philosophy of blocking is not as restrictive as it is claimed. Another question : Why EXFOR 10267.014 appears in this block with a comment : ..11 PTS Pu239/U-235, and not in the second Pu-239 NF ORL block on page 1758?

These examples also show that the complexities of the blocking problem are not specific to Area 2. We realized that the team comprising G. de Saussure, R. Gwin and L. Weston have been measuring U-235, U-233 and Pu-239 alfa and fission cross-sections for years and are still working in these fields. So, the blocking problem for NF ORL is quite a difficult job.

EXAMPLE III page 1598 - U-235 SNG BNL

There are 8 references in the block : 7 of which are superseded, without comments. Is there really nothing original in 70 HELSINKI or 71 KNOXVILLE?

EXAMPLE IV page 1588 U-235 TOT BNW (last block on the page)

There are two references in the block, with a gap of 6 years, and one EXFOR index line :

PR/C 3 576 Feb.71 FOSTER + ...
NIM 36 1 Sept.65 Superseded
DATA EXFOR 10047 Aug.73 486 PTS, SIGMA

The NIM 36 article is given in PR/C 3 as a reference on the experimental method. However, there are no values of U-235 total cross-sections in NIM 36. Thus the comment should not be "superseded", but something like : DETAILS OF EXP. METHOD, NDG.

IV - LABORATORY INDEXING

It is obvious that, due to the blocking problem, the CINDA rule must be modified. We agree with the proposition of N.E. Holden : "For the case where scientists from different laboratories collaborate, the convention should be changed to : the laboratory where the measurement was performed is listed." Thus, the mistakes in Examples (1) and (2) of Charlie's letter could be avoided. Concerning these examples, it is worth mentioning that the reference EANDC(E)-89U contains no statement suggesting that the work has a connection with any laboratory other than Munich. It is only by making a retrospective comparison with the other references that it becomes apparent that the experiment used the facilities of IRK.

We would also like to mention that conflict between the laboratory of the first author and the laboratory where an experiment has been performed occasionally arises with EXFOR/NEUDADA lines in CINDA because the X-4 rules state more categorically that the laboratory where the experiment was performed should be used. Thus, the Holden rule should be a normalisation between CINDA and EXFOR.

However, there are a few borderline cases. For instance, let us consider the following : a visiting group uses the facilities of a laboratory and all the references relevant to CINDA originate from the "home" laboratory. In this case, we are tempted to propose an exception to the Holden rule, because the blocking in CINDA could be done under the "home" laboratory without ambiguity. But, more complicated situations could evolve, as is shown in the following description (which is hardly a caricature of some real situations!):

The physicist X from laboratory A (area 2), Y from B (area 1), Z from C (area 2) collaborate for an experiment using the facilities of laboratory D (area 1) the Head of which is the physicist W ; 6 articles or abstracts have been written to report on the experiment :

from X, Y and Z in an A progress report ;

from Y, X and Z in a B progress report ;

from Z, X and Y in a C progress report ;

from W in a D progress report ;

from X, Y and Z in Nuclear Physics ;

from Y, X and Z in Nuclear Science and Engineering.

Due to the fact that there is no connection between the CINDA indexers, a correct block will never be built for this experiment, if we accept an exception to the Holden rule.

Consequently, we propose that the Holden rule be generalized to all CINDA entries : "Whoever the authors are, the laboratory where the experiment was performed is listed."

V - CONCLUSION

There was a time when CINDA was acceptable as an ordered, but structureless, collection of information on neutron physics literature ; a minimum of simple coding rules allowed a rapid accumulation of coded information ; cleaning-up operations could reasonably be considered a clerical task. However, something had to be done to prevent CINDA from simply reflecting the actual proliferation of redundant litterature within its own field. The general feeling one may have after reading this review Memo. on the "blocking problem" in CINDA is that the strict observation of related coding rules will require, on the part of the coder, a degree of discipline almost beyond that which could be expected, let alone a good general knowledge of neutron physics, and a specialized knowledge in certain fields! A lot of formal cleaning-up still remains to be carried out urgently on CINDA as it now stands, whereas any work connected with "blocking" may not be considered by everyone to be as equally urgent, and moreover, requires specialized manpower. But, even at this stage, it seems worthwhile establishing a concensus on some rules which will contribute to making CINDA a very useful tool despite some imperfections.