**WP2010-22**

# Automatic import information from NSR to CINDA

V.Zerkin, IAEA-NDS, March 2010



## 1. CINDA today

For many different reasons, last few years compilation to CINDA has been practically stopped. In order to partially compensate this and keep CINDA-service running for traditional users, an automatic procedure importing information from EXFOR to CINDA was developed and regularly performed by NDS since 2006. This procedure provides updating of CINDA contents synchronizing it with modifications of EXFOR. Similar procedure extending CINDA by contents of NSR database is proposed and described in this paper.

### 1.1 CINDA relational database

CINDA relation database developed in the IAEA-NDS (2002) is based on so-called CINDA Master file (original text file in CINDA format). The database is oriented to practical usage via Internet and CD-ROM users and includes also information from dictionary system and other extensions, programmatically done by software extracting information from other databases, namely:

- Initials of the first author from EXFOR
- Full list of authors from NSR
- Title of article from NSR and EXFOR
- KeyNo from NSR (to provide Web link CINDA-NSR)
- DOI from NSR (to provide link to Web journals)
- Patterns from a special Dictionary (to provide link to Web journals)

So, CINDA relational database is not only index to reaction data – it has features of real bibliographical database (although they are not complete).

CINDA-Relational is used for the most utilities:

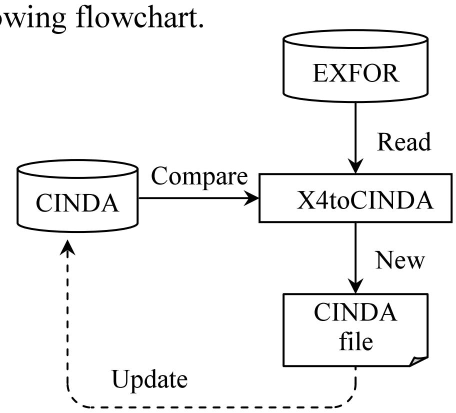
- comparison to EXFOR in the procedure of automatic import new information from EXFOR to CINDA

- comparison CINDA – EXFOR – NSR to find experimental data which are not yet included to EXFOR (SG30, 2008)
- matching CINDA with NSR references to import bibliographical data and DOI
- production of CINDA backup (Master file)

## 1.2. Automatic import from EXFOR

Current CINDA library contains "manual" part (contents of CINDA as of 2005 converted to the new format plus seven human compiled updates) and "automatic" part, regularly imported from EXFOR (always completely replaced every time). The library is maintained centrally in the IAEA[1]: full Master file and all exchange files are available for nuclear data centers on NDS Web site.

Generally, algorithm of importing procedure from EXFOR to CINDA can be described as following flowchart.




CINDA in new format and EXFOR have very compatible coding for all information (reaction, references), common system of dictionaries and similar scope of the databases. Algorithm includes comparison of imported information with existing data in CINDA database to avoid duplication.

# 2. Import from NSR: problems and possible solutions

Only recently NSR has become Nuclear Science References. From the beginning, NSR was Nuclear Structure References, and all (subject area, contents and coding of information) of NSR was very different from reaction databases (CINDA/EXFOR).

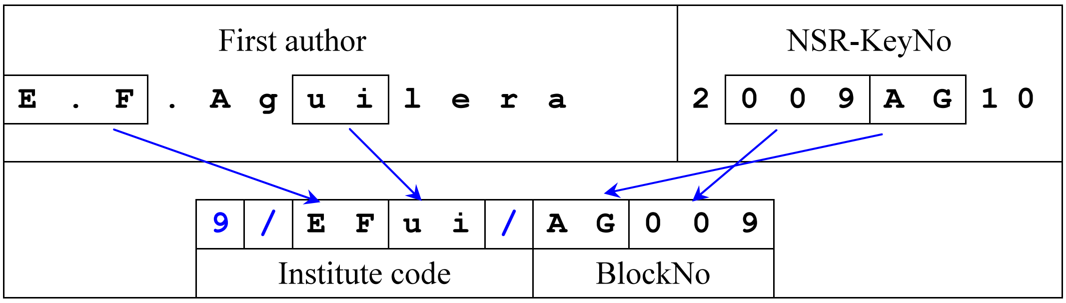
## 2.1. Blocking: Lab and Block-No

Both EXFOR and CINDA have so-called "blocking" - system, defining that data are produced using similar methods/equipment/conditions by the same group of authors. The decision to unite data is made by compiler. Technically, grouping in EXFOR is implemented via ENTRY, in CINDA – via combination [Reaction, Lab, Block-No].

[1] Last automatic updates performed in NDS: April, December 2009 and March 2010.

NSR does not have either Lab in its coding or Blocks – every reference is completely independent. How to provide blocking of data imported from NSR?
It can be done on the basis of assumption:
"Data published by the same author in the same year can be grouped to one block"
Of course, when data will be processed by CINDA software, Reaction-code will be considered.

In order to provide implementation of this assumption, the following method can be used:




So, Lab (Institute-code) is completely artificial: "**9**" is artificial area-code, "**/**" are used to differentiate country-code and lab-code from really existing codes. Note: if we build real table Author-Lab, then it could be difficult to fulfil our assumption.

## 2.2. Coding of reaction

NSR uses ENSDF coding of reactions, nuclides and particles using separate values of target, reaction, quantity connected by so-called selectors. In principal, this coding can be converted to EXFOR/CINDA using simple algorithms and dictionaries (tables of correspondence). Because scope of NSR and CINDA is very different, some tricks should be used to make conversion reasonable, so may be new codes should be introduced to CINDA. For example, NSR has quantity DSIGMA, CINDA has differential cross sections: DA, DE, DAE; probably new CINDA quantity DCS should be introduced.

| NSR reaction coding | CINDA reaction coding | |
|---|---|---|
| NSR coding: `T:27AL;A. R:(N,2N);A.`<br>`N:26AL;A. M:SIGMA;A.`<br>Equivalent: `27AL(N,2N),SIGMA` | `13 27 N,2N` | `CS` |
| `24MG(6LI,6LI),SIGMA` | `12 24 Li-6,EL` | `CS` |
| `12C(P,P'),DSIGMA` | `6 12 P,INL` | `DCS` |
| `58NI(A,PA),SIGMA` | `28 58 A,P+A` | `CS` |
| `249CF(13C,XNYP),SIGMA` | `98249 C-13,xN+yP` | `CS` |
| `64ZN(16O,XNYPZA),DSIGMA` | `30 64 O-16,xN+yP+zA` | `DCS` |
| `12C(24MG,216O),DSIGMA` | `6 12 Mg-24,2O-16` | `DCS` |
| `116SN(G,N),RESONANCE` | `50116 G,N` | `RP` |
| `235U(N,F),FISSION` | `92235 N,F` | `MFQ` |

Some of codes are not allowed in EXFOR now: [xN+yP]

### 2.2. Coding of references

NSR has well defined coding of journals only; having a correspondence table, conversion can be easily done. Conversion private communications and thesis's from NSR to CINDA is trivial. Building correspondence between reports and conferences is more difficult task, because NSR seems quite flexible rules of coding these types of references; it will obviously need additional dictionary.

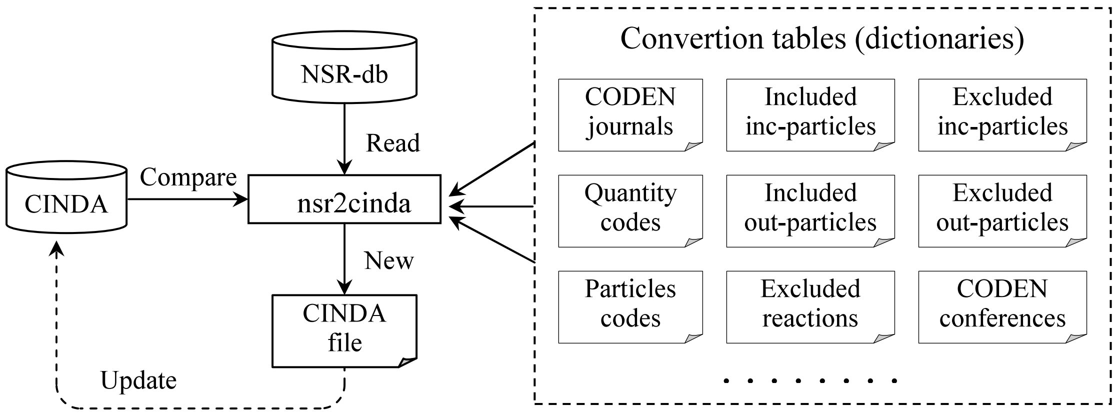
### 2.3. Incident energy

Although NSR usually has compiled information about incident energy for publications related to nuclear reactions, this information is not introduced into the NSR database, produced by NNDC (now only preliminary processing of energy is done).

## 3. Automatic import from NSR: general approach, algorithm

First of all, we have to accept the assumption, that information imported from NSR to CINDA will not be as accurate as information converted from old CINDA and imported from EXFOR. Although retrieval system built on this extended database would help to find some information which does not exists in EXFOR and old CINDA, it can also produce some irrelevant references ("garbage") in respond to the user's request.

Portion of unsure information imported from NSR can be regulated by the tables of correspondence-conversion (configuration files), or some limitations implemented in the algorithm of the importing program. For example, we can filter imported information by: only journals, or journals having Web links; selected projectiles (neutron, proton, etc.), by type of work ("Measured", "Calculated"), by incident energy (e.g. <1GeV), or by any combination of these parameters. Configuration of these parameters will actually define the scope of extended CINDA, therefore they should be discussed (or at least understood) before principal decision is made; of course, some filtering can be done on the level of retrieval system during "request-search-selection" process by end-user.



Important: program uses NSR relational database, produced by NNDC and supplied to the IAEA.

"Compare" on the flow chart has two parts:

1) Search NSR-KeyNo existing in CINDA (this is result of another program searching NSR references corresponding to CINDA references)
2) Search if CINDA has line(s) with references with given author/target/year of publication. If found, program assumes that given NSR reference is already in CINDA (at least, similar works were observer by compilers), and it is not included to output CINDA file.

# 4. Test run

Although, many things can (and should) be improved, the basic program for automatic import from NSR to CINDA is ready.

Applied limitations were not strict:

- stype="M" (Measured)
- subject="SIGMA", "DSIGMA", "RESONANCE", "FISSION"
- excluded particles: MU, PI+, PI-, PI, K-, K+, X-RAY, P-BAR, N-BAR, S+, S-

Summary result*:

- References found in NSR: 46345
- Found correspondence in CINDA (ignored): 14944 (32%);
- Added to CINDA (+% to existing CINDA, including imported from EXFOR):
  - References: 31401 (+68%), from journals:19821, Phys.Rev.C: 3431
  - Lines: 91987 (+21% to existing CINDA)
  - Blocks: 87887 (+45% to existing CINDA)

  Average lines per block (blocking rate) =1.05 (in existing CINDA =2.26)
- Run-2: test treatment of energy; blocking: works of the same author within 3 years
  - References with Energy found: 8616
  - Converted to Energy range: 6957 (22% of added references=31401)
  - Blocking rate =1.09

*Note. Addition looks huge (except for neutrons: +5% only); so, may be some types of data should be filtered out. In order to do this, some analysis could be done probably by experts deeply understanding needs of users' community.

Statistics: number of blocks sorted by incident particles, see in Appendix-1.

Web interface: http://nds121.iaea.org/exfor2/cinda.htm

Usage of new lines regulated by check-box: ☑ *Include lines imported from NSR*

# 5. Concluding remarks

After discussion of the scope of data imported from NSR, finalizing of algorithm and dictionaries, a regular procedure extending CINDA database by NSR contents can be established.

Import from NSR is probably one of "last attempts" to save CINDA, or at least make it more useful for some groups of users (e.g. EXFOR compilers). CINDA can continue to exist as "secondary" database - index of the contents of other databases with data (EXFOR, ENDF) and bibliographical information (NSR) until time of a revision of nuclear reaction bibliographical data will come - either in frameworks of EXFOR, NSR or as separate database…

**Appendix-1.**

| Extension CINDA by NSR (blocks) | | | | |
|---|---|---|---|---|
| projectile | before | after | plus | +percent |
| N | 150481 | 158751 | 8270 | 5 |
| P | 14294 | 32233 | 17939 | 125 |
| 0 | 7334 | 7334 | 0 | 0 |
| G | 6699 | 12918 | 6219 | 92 |
| A | 4592 | 14431 | 9839 | 214 |
| D | 4134 | 11510 | 7376 | 178 |
| HE3 | 1528 | 7759 | 6231 | 407 |
| C-12 | 796 | 3747 | 2951 | 370 |
| O-16 | 453 | 5124 | 4671 | 1031 |
| T | 443 | 1967 | 1524 | 344 |
| Li-6 | 411 | 2023 | 1612 | 392 |
| Li-7 | 376 | 1478 | 1102 | 293 |
| N-14 | 348 | 1600 | 1252 | 359 |
| O-18 | 155 | 1255 | 1100 | 709 |
| He-6 | 143 | 264 | 121 | 84 |
| Ar-40 | 138 | 1547 | 1409 | 1021 |
| Be-9 | 116 | 505 | 389 | 335 |
| Ne-20 | 115 | 1026 | 911 | 792 |
| Si-28 | 111 | 589 | 478 | 430 |
| E | 100 | 3581 | 3481 | 3481 |
| F-19 | 94 | 592 | 498 | 529 |
| Ne-22 | 93 | 458 | 365 | 392 |
| B-11 | 86 | 534 | 448 | 520 |
| C-13 | 67 | 624 | 557 | 831 |
| Li-8 | 53 | 95 | 42 | 79 |
| B-10 | 53 | 321 | 268 | 505 |
| Be-11 | 52 | 122 | 70 | 134 |
| Be-7 | 45 | 91 | 46 | 102 |
| U-238 | 43 | 326 | 283 | 658 |
| He-8 | 43 | 79 | 36 | 83 |
| Li-11 | 38 | 93 | 55 | 144 |
| Fe-56 | 38 | 202 | 164 | 431 |
| Mg-24 | 37 | 241 | 204 | 551 |
| Cu-63 | 36 | 126 | 90 | 250 |
| S-32 | 35 | 939 | 904 | 2582 |
| Ca-40 | 33 | 326 | 293 | 887 |
| Ni-58 | 32 | 499 | 467 | 1459 |
| Ca-48 | 30 | 341 | 311 | 1036 |
| Xe-132 | 28 | 83 | 55 | 196 |
| Kr-84 | 27 | 215 | 188 | 696 |
| B-8 | 27 | 94 | 67 | 248 |
| F-17 | 27 | 47 | 20 | 74 |
| Si-30 | 25 | 111 | 86 | 344 |
| Li-9 | 25 | 56 | 31 | 124 |
| Be-10 | 25 | 38 | 13 | 52 |
| F-18 | 23 | 35 | 12 | 52 |

| projectile | before | after | plus | +% |
|---|---|---|---|---|
| Ar-36 | 21 | 132 | 111 | 528 |
| Be-14 | 19 | 38 | 19 | 100 |
| Cl-35 | 19 | 353 | 334 | 1757 |
| C-14 | 18 | 265 | 247 | 1372 |
| N-15 | 18 | 201 | 183 | 1016 |
| Pb-208 | 18 | 180 | 162 | 900 |
| Be-12 | 17 | 33 | 16 | 94 |
| Mg-26 | 17 | 56 | 39 | 229 |
| Ni-64 | 16 | 246 | 230 | 1437 |
| Se-82 | 16 | 31 | 15 | 93 |
| C-11 | 15 | 34 | 19 | 126 |
| C-15 | 14 | 35 | 21 | 150 |
| C-10 | 14 | 21 | 7 | 50 |
| Al-27 | 13 | 93 | 80 | 615 |
| C-19 | 13 | 27 | 14 | 107 |
| Kr-86 | 12 | 207 | 195 | 1625 |
| O-17 | 12 | 231 | 219 | 1825 |
| Ti-48 | 12 | 90 | 78 | 650 |
| B-15 | 11 | 15 | 4 | 36 |
| Cl-37 | 10 | 117 | 107 | 1070 |
| Ni-57 | 10 | 10 | 0 | 0 |
| N-13 | 9 | 33 | 24 | 266 |
| O-14 | 9 | 23 | 14 | 155 |
| Na-21 | 9 | 23 | 14 | 155 |
| Ge-76 | 8 | 19 | 11 | 137 |
| B-17 | 8 | 24 | 16 | 200 |
| O-15 | 8 | 12 | 4 | 50 |
| C-16 | 8 | 20 | 12 | 150 |
| Ne-26 | 8 | 14 | 6 | 75 |
| C-9 | 8 | 24 | 16 | 200 |
| Fe-52 | 8 | 9 | 1 | 12 |
| Mg-32 | 8 | 15 | 7 | 87 |
| C-17 | 8 | 33 | 25 | 312 |
| Fe-58 | 7 | 88 | 81 | 1157 |
| C-18 | 7 | 13 | 6 | 85 |
| N-17 | 7 | 14 | 7 | 100 |
| O-22 | 6 | 23 | 17 | 283 |
| Mn-51 | 6 | 6 | 0 | 0 |
| B-14 | 6 | 15 | 9 | 150 |
| Fe-51 | 6 | 6 | 0 | 0 |
| Fe-50 | 6 | 7 | 1 | 16 |
| Ni-60 | 6 | 63 | 57 | 950 |
| F-24 | 6 | 14 | 8 | 133 |
| N-12 | 6 | 7 | 1 | 16 |
| Ni-56 | 6 | 31 | 25 | 416 |
| Mg-30 | 6 | 13 | 7 | 116 |