

Full EXFOR in C4 format: delivery to WPEC group SG-30

V.Zerkin, IAEA-NDS, 19/09/2008

Objectives and Tasks of NRDC start with the statement:

"The primary goal of the Network is the dissemination of nuclear reaction data and associated documentation to users."

Objectives of WPEC group SG30:

"Make EXFOR an easy accessible and correct database, available in computational format."

Although NRDC activity starts with definition of data format, organization of data exchange, data search and compilation, final aim of NRDC is to deliver data to users. And here users of NRDC have some problems because EXFOR designed as a universal format for data exchange between nuclear data centres and appear to be very complex for ordinary users using traditionally Fortran language in their programs. Access to the data in simplified formats via retrieval systems providing by several data centres satisfies most of users, but not all. There is not so large but important group of users – professionals working in data evaluation field and constructing comprehensive software packages (e.g. Talis, Empire) for massive data evaluation. These users organized WPEC working group SG30 and requested access to entire database but not in EXFOR format – in some computational format easy for programming.

This is a short summary what was done at the IAEA-NDS to deliver EXFOR data to SG30 community:

- 1) computational format and way of storage was selected, discussed and agreed (extended C4), see Appendix
- 2) software retrieving data from EXFOR and running conversion procedures was written
- 3) entire EXFOR database was converted to extended C4 format
- 4) full EXFOR in C4 is regularly delivered to SG30
- 5) full EXFOR backup – one Entry in one file – also regularly delivered
- 6) special FTP site is organized: <http://www-nds.iaea.org/x4toc4-master/>
- 7) Dates of releases:
 1. 2007-Jun 28
 2. 2007-Sep-04
 3. 2007-Sep-26
 4. 2007-Oct-23
 5. 2008-Mar-14
 6. 2008-Apr-29
 7. 2008-May-07
 8. 2008-Jun-10
 9. 2008-Jul-03
 10. 2008-Sep-18

Statistics of the last release			
# File	Size	Unzipped	Contents
1 C4-2008-09-18.zip	72Mb	896Mb	Full-C4 + EXFOR14A.DAT
	Files: 2		
2 X4-2008-09-18.zip	106Mb	464Mb	Full-X4 stored in files by ENTRY
	Files: 18187		in directory structure:
	Dir: 232		ENTRY:A1234 -> X4A11/A12/A1234.x4
#Summary of the contents:		EXFOR	C4+
Last updated:		11-Sep-2008	18-Sep-2008
Total Number of ENTRY:		18,187	
Number of ENTRY with data:		17,488	11,885 (68%)
Number of SUBENT with data:		118,314	
Number of Datasets:		131,054	63,029 (48%)
Number of Datalines		10,678,854	6,820,296 (64%)

Full C4 file is already used for several purposes by SG30 members. One of them is testing the contents of EXFOR by sophisticated procedures aiming to find and correct errors and improve EXFOR database, see:

"NRDC action the List of EXFOR Outliers", Naohiko Otsuka, WP2008-03, http://www-nds.iaea.org/nrdc/nrdc_2008/working/wp2008-03.pdf).

Contents of README file on IAEA-NDS site for SG30

<http://www-nds.iaea.org/x4toc4-master/>

Nuclear Data Section (NDS)
 Department of Nuclear Sciences and Applications
 International Atomic Energy Agency (IAEA)
 Wagramer Strasse 5, P.O.Box 100
 A-1400 Vienna, Austria
 Tel: (+43 1) 2600-21714; Fax: (+43 1) 26007

Full EXFOR in C4 format
 + Full EXFOR by Entries
 Created 8-May-2007 by Viktor Zerkin, e-mail: V.Zerkin@iaea.org
 Last updated: 19-September-2008

Contents:

1. C4-YYYY-MM-DD.zip (size:~72Mb) contains:
 - 1) C4-YYYY-MM-DD.xc4 (size:~900Mb):
 full EXFOR (as of YYYY/MM/DD in the IAEA-NDS) in extended C4 format*
 - 2) EXFOR14A.DAT (as of YYYY/MM/DD)
2. X4-YYYY-MM-DD.zip (size:~106Mb) contains:
 X4all: full EXFOR in a directory structure - one ENTRY in one file
 Total: 232 subdirectories, 18187 files, size:~500Mb
3. dev/ contains: several version of development
4. history.txt
5. readme.txt

Questions and Answers.

1.Q: What is extended C4 format?

A: C4 with identification information:

- a) Identification information is given as comment starting with #.
- b) Information is sorted by Entry-Subentry-Pointer and organized as follow:

```
#C4REQUEST
#ENTRY
...ENTRY Information: reference, title, full list of authors,...
#DATASET: SUBENTRY-Number+Pointer
...DATASET Information: EXFOR-Reaction, MF, MT,...
#DATA
...DATA: C4 lines as is in pure C4 file
#/DATA
....
#/DATASET
....
#/ENTRY
....
#/C4REQUEST
```

2.Q: How to update software reading C4 to be able to use extended C4?

A: Add to your code ignoring lines starting with "#";
or use "filter" - utility, which will read extended C4 and write plain C4

3.Q: Why full EXFOR database is presented in one C4 file (not by smaller parts)?

- A: One C4 file containing all EXFOR data (although it is huge ~1Gb) seems to be preferable, because:
- a) user decides how to organize data for his/her application - can easy write software to split full C4-file to parts convenient for his applications and store them in appropriate form, e.g.:
 - 1) index file + directory structure sorting data by EXFOR numbers,
 - 2) index file + directory structure sorting data by target/reaction,
 - 3) ENDF-like directory structure: ZAProjectile/ZAMaterial/MF/MT
 - 4) database
- etc. (full freedom: NDS does not dictate the method of data storage)

b) no need to provide software for data access (which can be different for different applications).

4.Q: Is there any alternative way to get EXFOR data in C4 format?

A: NDS provides two "standard" ways/methods to get EXFOR data in C4 format:

a) via Web EXFOR retrieval system:

<http://www-nds.iaea.org/exfor/>

b) using non-interactive stand-alone Java-utility retrieving data from EXFOR database (MySQL/MS-Access on CD-ROM or remotely), which can be called through external script by any application (as it is done for Empire and EndVer).

IAEA-NDS CD-ROMs:

- "EXFOR-CINDA for Applications" for Linux/Windows/Mac (MySQL)
- "EndVer/GUI and EXFOR-CINDA for Applications" (Lin/Win/Mac,MySQL)
- "EXFOR-CINDA Retrieval system for Windows" (MS-Access)

5.Q: What is the meaning of the fields in the lines #C4REQUEST and #/C4REQUEST ?

A: These lines mark begin/end of an extended C4-file:

a) #C4REQUEST N1 N2 N3

N1 - date of request (date when this C4-file was created)

N2 - time, when request started

N3 - date of last update of EXFOR database,
from which data were retrieved (source database)

b) #/C4REQUEST N1 N2 N3

N1 - number of Entries in this file (start with #ENTRY)

N2 - number of Datasets in this file (start with #DATASET)

N3 - total number of datasets in the source EXFOR database

-End-

Example of file in extended C4 format

Appendix.2

```

#C4REQUEST 20080704 121521 20080610
#ENTRY C0774
#AUTHOR1 C.H.Johnson+
#YEAR 1968
#INSTITUTE (1USAORL)
#TITLE The 89Y(p,n)89Zr cross section near the first two
#+ analogue resonances
#AUTHOR(S) C.H.Johnson, R.L.Kernell, S.Ramavataram
#REF-CODE (J,NP/A,107,21,1968)
#REFERENCE Jour. Nuclear Physics, Section A
#+ Vol.107, p.21, 1968
#DATASETS 1
#
#DATASET C0774002
#DATE 20010404
#REACTION 39-Y-89(P,N)40-ZR-89,,SIG
#PROJ 1001
#TARG 39089
#MF 3
#MT 4
#C4BEGIN [ 1001 39089 3 4 ]
#DATA 59
# Prj Targ M MF MT PXC Energy dEnergy Data dData Cos/LO dCos/LO ELV/HL dELV/HL I78
#----->o<----->ooo<----->----->----->----->----->----->----->
1001 39089 3 4 3660000. 4.6600-4 2.2000-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3670000. 5.3500-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3690000. 5.2800-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3690000. 5.7800-4 1.9000-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3710000. 6.0800-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3720000. 6.3700-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3730000. 7.0300-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3750000. 7.5700-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3780000. 7.4800-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3830000. 9.3100-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3880000. 9.3100-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3930000. 1.1100-3 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 3980000. 1.2300-3 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4030000. 1.3300-3 5.3200-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4080000. 1.5200-3 6.0800-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4140000. 1.6100-3 6.4400-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4170000. 1.7100-3 6.8400-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4190000. 1.7900-3 7.1600-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4220000. 1.7800-3 7.1200-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4230000. 1.8400-3 7.3600-5 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4300000. 3.0100-3 1.2040-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4360000. 3.3200-3 1.3280-4 C.H. JOHNSON,ET.AL. (68) C0774 2
1001 39089 3 4 4410000. 3.6200-3 1.4480-4 C.H. JOHNSON,ET.AL. (68) C0774 2

```

1001 39089	3	4	4470000.	3.6400-3	1.4560-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4520000.	4.0500-3	1.6200-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4580000.	4.1600-3	1.6640-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4630000.	4.4500-3	1.7800-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4630000.	4.3500-3	1.7400-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4680000.	4.5900-3	1.8360-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4750000.	4.8500-3	1.9400-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4890000.	5.6400-3	2.2560-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4910000.	5.7900-3	2.3160-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4940000.	5.9400-3	2.3760-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4940000.	6.1200-3	2.4480-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	4970000.	6.4700-3	2.5880-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5000000.	7.8700-3	3.1480-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5030000.	8.6500-3	3.4600-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5060000.	6.5900-3	2.6360-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5080000.	6.2500-3	2.5000-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5110000.	6.3800-3	2.5520-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5140000.	6.6400-3	2.6560-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5170000.	6.6400-3	2.6560-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5200000.	6.7100-3	2.6840-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5230000.	6.8900-3	2.7560-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5260000.	6.9700-3	2.7880-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5290000.	7.1000-3	2.8400-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5320000.	7.4300-3	2.9720-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5350000.	7.3100-3	2.9240-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5380000.	7.5300-3	3.0120-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5410000.	7.8000-3	3.1200-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5440000.	7.8700-3	3.1480-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5470000.	8.0700-3	3.2280-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5500000.	8.1500-3	3.2600-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5530000.	8.4700-3	3.3880-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5590000.	8.6700-3	3.4680-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5650000.	8.9500-3	3.5800-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5730000.	9.3500-3	3.7400-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5780000.	9.5700-3	3.8280-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
1001 39089	3	4	5840000.	0.010200	4.0800-4	C. H. JOHNSON, ET. AL.	(68)	C0774	2
#/DATA	59								
#/DATASET									
#/ENTRY	1								
#									
#									
#/C4REQUEST	1	1	1	0	59				