From Paper to Data: GSYS and the Digital Transformation

M.Odsuren, National University of Mongolia
Co-worker: N.Otsuka, L.Vrapcenjak, IAEA

Vienna, Austria
5 December 2024

What are the Steps of Digitization

The first significant step in preserving data from paper to digital format occurred in the mid-20th century with the development of early technologies designed to convert physical documents into digital form.

1. **Scanning Technology (1930s-1950s)**: The use of light-based systems to create digital images of paper documents.
2. **Optical Character Recognition (1960s)**: OCR software was developed to convert scanned images into machine-readable text.
3. **Microfilm and Microfiche (1970s)**: Analog methods for preserving documents in a compact, easy-to-store format.
4. **Digital Libraries and Cloud Storage (1990s-Present)**: The rise of digital preservation via cloud computing and OCR software to archive and manage vast amounts of data.

These steps were crucial in making data accessible and easier to store, preserve, and manage over time, laying the foundation for today's digital archival systems.

Nuclear Science and Technology

The **International Nuclear Information System (INIS)** is a specialized database established by the **International Atomic Energy Agency (IAEA)** in **1969**. Its primary goal is to serve as a global information resource **for nuclear science and technology**. INIS collects, stores, and disseminates nuclear-related information from various scientific, academic, and governmental sources.

- **INIS, IAEA** is a prime example of preserving data from paper-based documents to digital formats. It has transformed the way nuclear research, technical data, and regulatory documents are stored and accessed.
  By digitizing historical materials and continually updating with new content, INIS ensures that vital nuclear information is preserved for future generations, supporting research and policy development on a global scale.

Specialized Libraries, Databases

INIS is unique **in its specific focus on nuclear science and technology**, there are several similar specialized libraries and digital repositories that focus on preserving and disseminating scientific and technical data in various fields, such as medicine, engineering, health, physics, and agriculture

1. **PubMed (National Library of Medicine):** Biomedical and life sciences research.
2. **IEEE Xplore:** Electrical engineering, computer science, and electronics
3. **World Health Organization (WHO) Global Health Library:** Global health information.
4. **ScienceDirect (Elsevier):** Scientific, technical, and medical research.
5. **UNESCO's Digital Library:** Cultural heritage, education, and sustainable development.
6. **ArXiv:** Physics, mathematics, computer science, and other scientific fields.
7. **CERN Document Server (CDS):** High-energy physics, CERN-related research.
8. **NASA Astrophysics Data System (ADS):** Astronomy and astrophysics.
9. **FAO Corporate Document Repository:** Agriculture and food security.
10. **AGRIS (International Information System for the Agricultural Sciences and Technology):** Agricultural sciences and technology.

Data Digitization in EXFOR

The EXFOR database was established in the 1960s by the IAEA (International Atomic Energy Agency) and associated organizations.

Digitization Methodology:

Digitization involves two primary tasks:

- converting experimental results into formats understandable to computers
- organizing the data for easy retrieval and analysis

For EXFOR, a specific format (the “EXFOR format”) was developed to ensure consistency and reliability in data conversion.

The data are digitized into tables, categorizing each physical quantity separately.

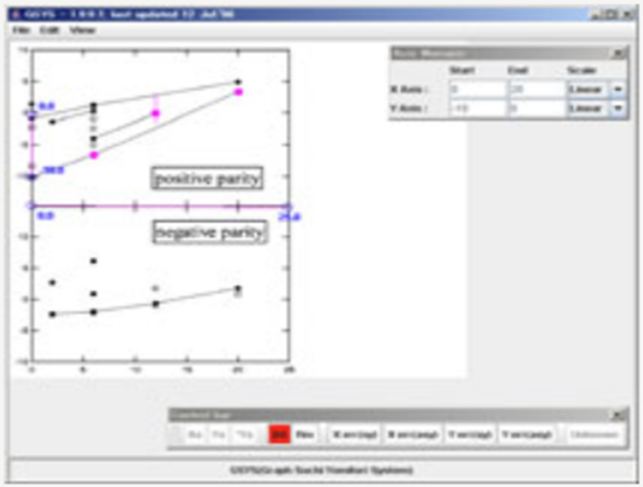
Data Digitization in EXFOR

D8070	J.Galin+,J,PR/C,9,1126,1974	Curve*
D8069	J.Galin+,J,PR/C,9,1113,1974	Curve+
D8071	G.Bizard+,J,NP/A,285,461,1977	Curve+
D8077	J.C.Steckmeyer+,J,NP/A,500,372,1989	Curve*
D8073	R.Billerey+,J,PRL,47,639,1981	Curve*

(*: Digitization was done by Lidija Vrapcenjak. +: Digitization was done by me.)

Hokkaido University Nuclear Reaction Data Centre (JCPRG)

Graph Digitizing System (GSYS, SyGRD)



Latest version
GSYS 2.4 (May 16, 2022)

Development version
GSYS 2.5.27 (Oct. 7, 2022)

Choose a different version?
GSYS Ver. 2.2 (Dec. 31, 2006)
GSYS Ver.2 (Aug. 02, 2006)
GSYS Ver.1 (Feb. 02, 2005)
SyGRD Ver.2 (Oct. 12, 2001)

Please use following address, if you have any questions or request.
gsys@jcprg.org

Feel free to inquire anything (web service, contribution to databases etc.):
Nuclear Reaction Data Centre, Faculty of Science, Hokkaido University
060-0810 Sapporo, Japan
TEL +81(JPN)-11-706-3723 / FAX +81(JPN)-11-706-3724

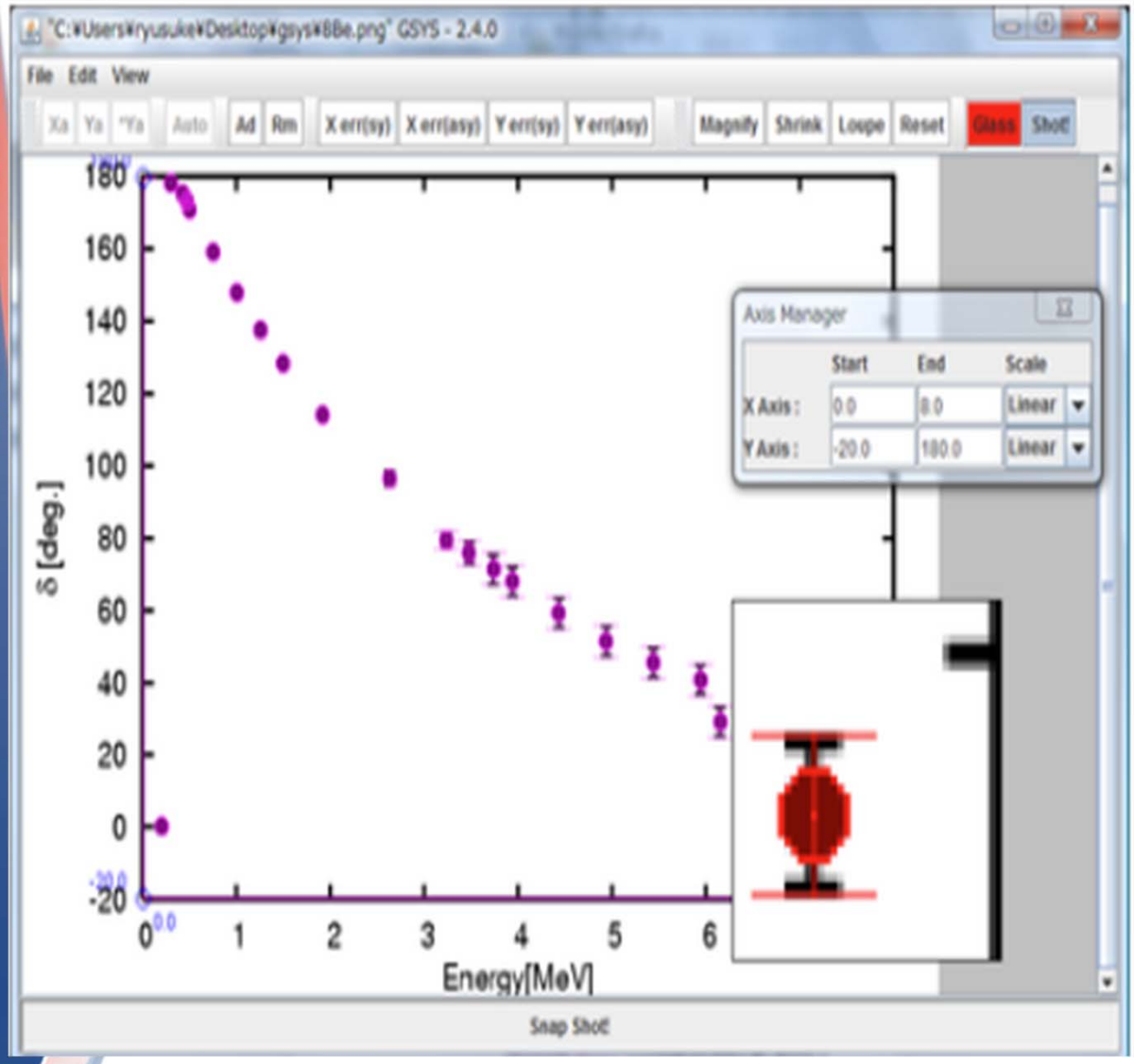
Japan Charged-Particle Nuclear Reaction Data Group (JCPRG)

Graph Suchi Yomitori System (GSYS2.4)

GSYS2.4

GSYS is a software to digitize data points on the figure in a form of graphical image (printed matter, image file,...).

This software is also useful to confirm whether the numerical value is correct or not.



Release Version: GSYS 2.4.9
(May 16, 2022)

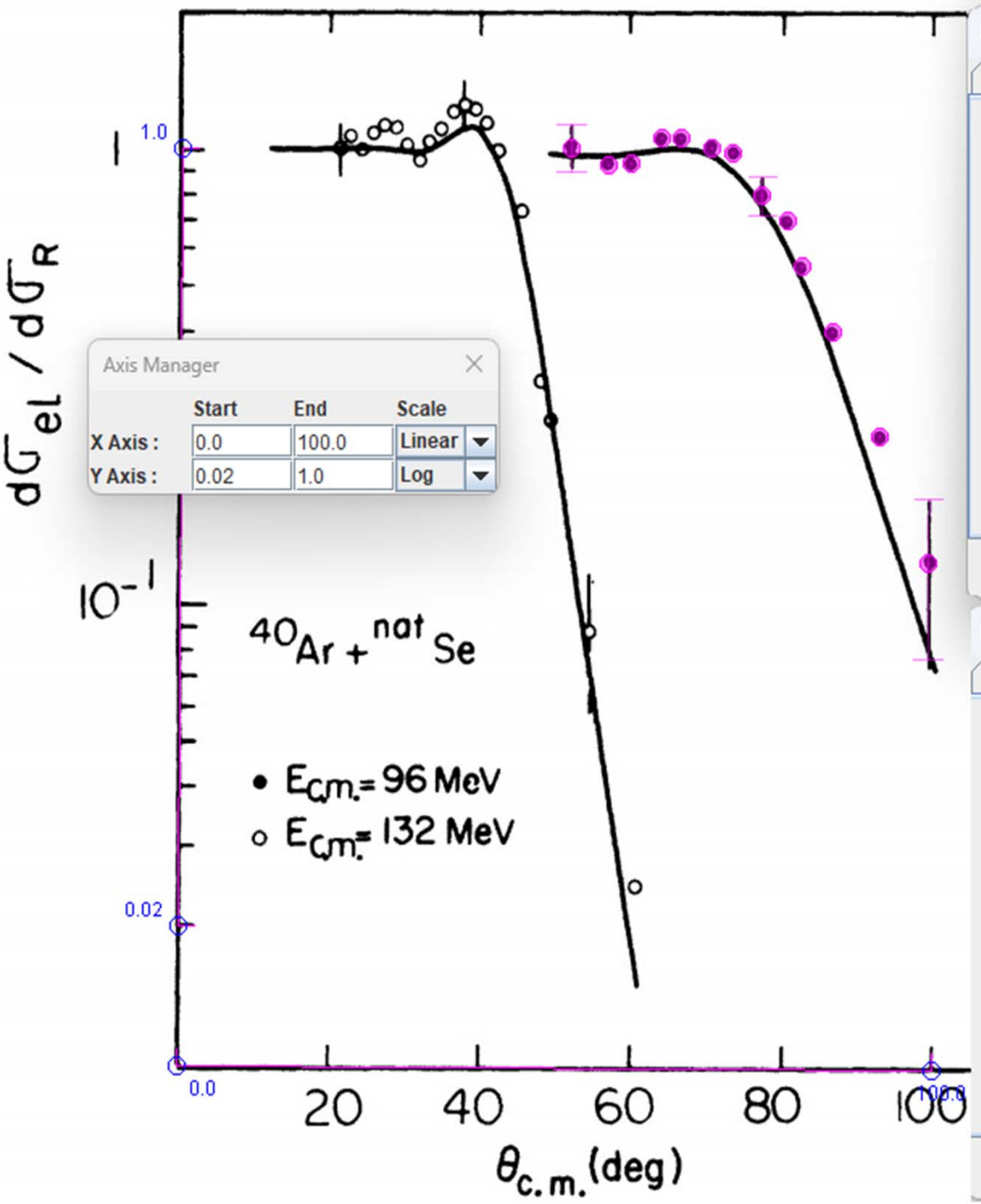
Download (jar, exe)
Download (for EXFOR compilers) (jar, exe)
Download (for NRDF compilers) (jar, exe)

Release notes
Manual for Gsys2.4 (Oct. 28. 2012)
See CP-E/145
Sample figure (contributed by K.Arai)
Sample figure 2 (made by ryusuke)

D8070

```
                D8070    20241001                                    D80700000001
SUBENT        D8070001   20241001                                    D80700100001
BIB                9         16                                      D80700100002
TITLE      Study of charged particles emitted from 117Te compound    D80700100003
           nuclei. II.Comparison between 40Ar+77Se and 14N+103Rh     D80700100004
           reactions and determination of critical angular           D80700100005
           momenta                                                   D80700100006
AUTHOR     (J.Galin, B.Gatty, D.Guerreau, C.Rousset,                 D80700100007
            U.C.Schlotthauer-Voos, X.Tarrago)                        D80700100008
INSTITUTE  (2FR PAR)                                                 D80700100009
REFERENCE  (J,PR/C,9,1126,1974)                                      D80700100010
SAMPLE     - Chemical-form of target is element.                     D80700100011
           - Physical-form of target is solid.                       D80700100012
           - Target-thickness is 1.64 mg/cm2.                        D80700100013
           - Target is self supported.                               D80700100014
FACILITY   (CYCLO,2FR PAR)                                           D80700100015
DETECTOR   (TELES,SIBAR,SILI)                                        D80700100016
REL-REF    (M,D8069001,J.Galin+,J,PR/C,9,1113,1974)                  D80700100017
HISTORY    (20240924C) M.Odsuren, L.Vrapcenjak                       D80700100018
ENDBIB            16          0                                      D80700100019
NOCOMMON           0          0                                      D80700100020
ENDSUBENT         19          0                                      D80700199999
SUBENT        D8070002   20241001                                    D80700200001
BIB                3          3                                      D80700200002
REACTION   (45-RH-103(7-N-14,EL)45-RH-103,,DA,,RTH)                  D80700200003
ANG-SEC    (ANG-CM,45-RH-103)                                        D80700200004
STATUS     (TABLE,,J.Galin+,J,PR/C,9,1126,1974) Figure 1             D80700200005
```

D8070 (Figure 3)



Axis Manager

	Start	End	Scale
X Axis :	0.0	100.0	Linear
Y Axis :	0.02	1.0	Log

Properties

Color & Points | Error & Axis | Format

Format of Output/Input Data: EXFOR (Standard / NRDF / EXFOR)

Fix "Log" scale output format : Not fix

Fix "Linear" scale output format : Not fix

Apply | Cancel | Default | Close

Properties

Color & Points | Error & Axis | Format

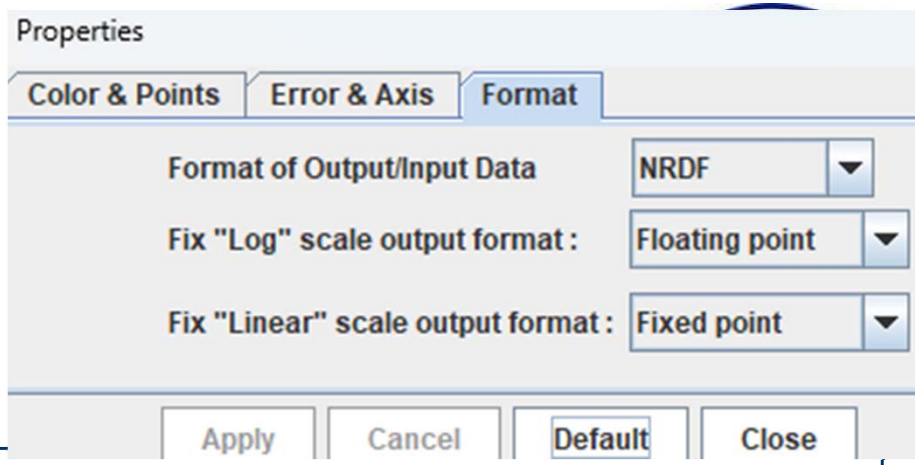
Format of Output/Input Data: Standard

Fix "Log" scale output format : Floating point

Fix "Linear" scale output format : Fixed point

Apply | Cancel | Default | Close

D8070 (Figure 3)



- Formatting is important! -> ?

# x	y	dy
51.64	1.02E+00	1.21E-01
56.49	9.34E-01	0.00E+00
59.54	9.42E-01	0.00E+00
63.52	1.07E+00	0.00E+00
66.09	1.07E+00	0.00E+00
70.16	1.02E+00	0.00E+00
72.98	1.00E+00	0.00E+00
76.93	8.05E-01	7.87E-02
80.23	7.11E-01	0.00E+00
82.30	5.65E-01	0.00E+00
86.27	4.05E-01	0.00E+00
92.61	2.40E-01	0.00E+00
99.23	1.27E-01	4.84E-02

# x	y	+-dy
51.64	1.02E+00	+-1.21E-01
56.49	9.34E-01	+-NEGLIGIBLE
59.54	9.42E-01	+-NEGLIGIBLE
63.52	1.07E+00	+-NEGLIGIBLE
66.09	1.07E+00	+-NEGLIGIBLE
70.16	1.02E+00	+-NEGLIGIBLE
72.98	1.00E+00	+-NEGLIGIBLE
76.93	8.05E-01	+-7.87E-02
80.23	7.11E-01	+-NEGLIGIBLE
82.30	5.65E-01	+-NEGLIGIBLE
86.27	4.05E-01	+-NEGLIGIBLE
92.61	2.40E-01	+-NEGLIGIBLE
99.23	1.27E-01	+-4.84E-02

D8070

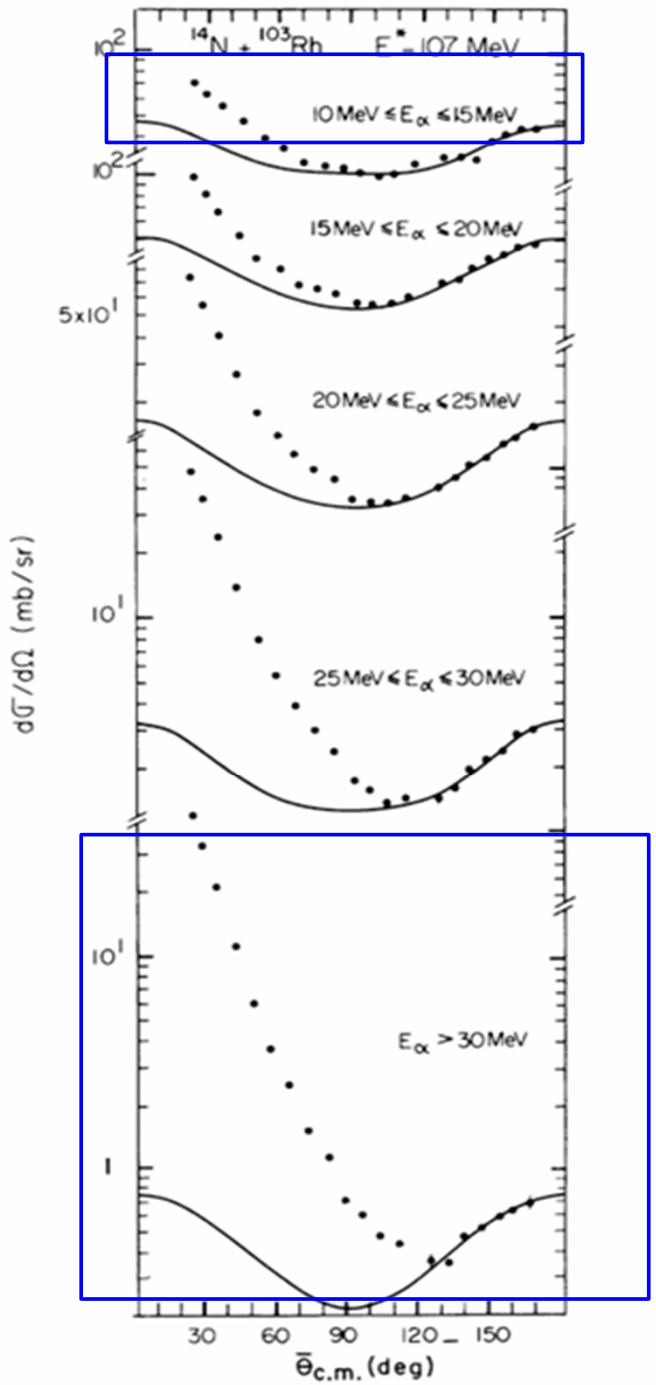
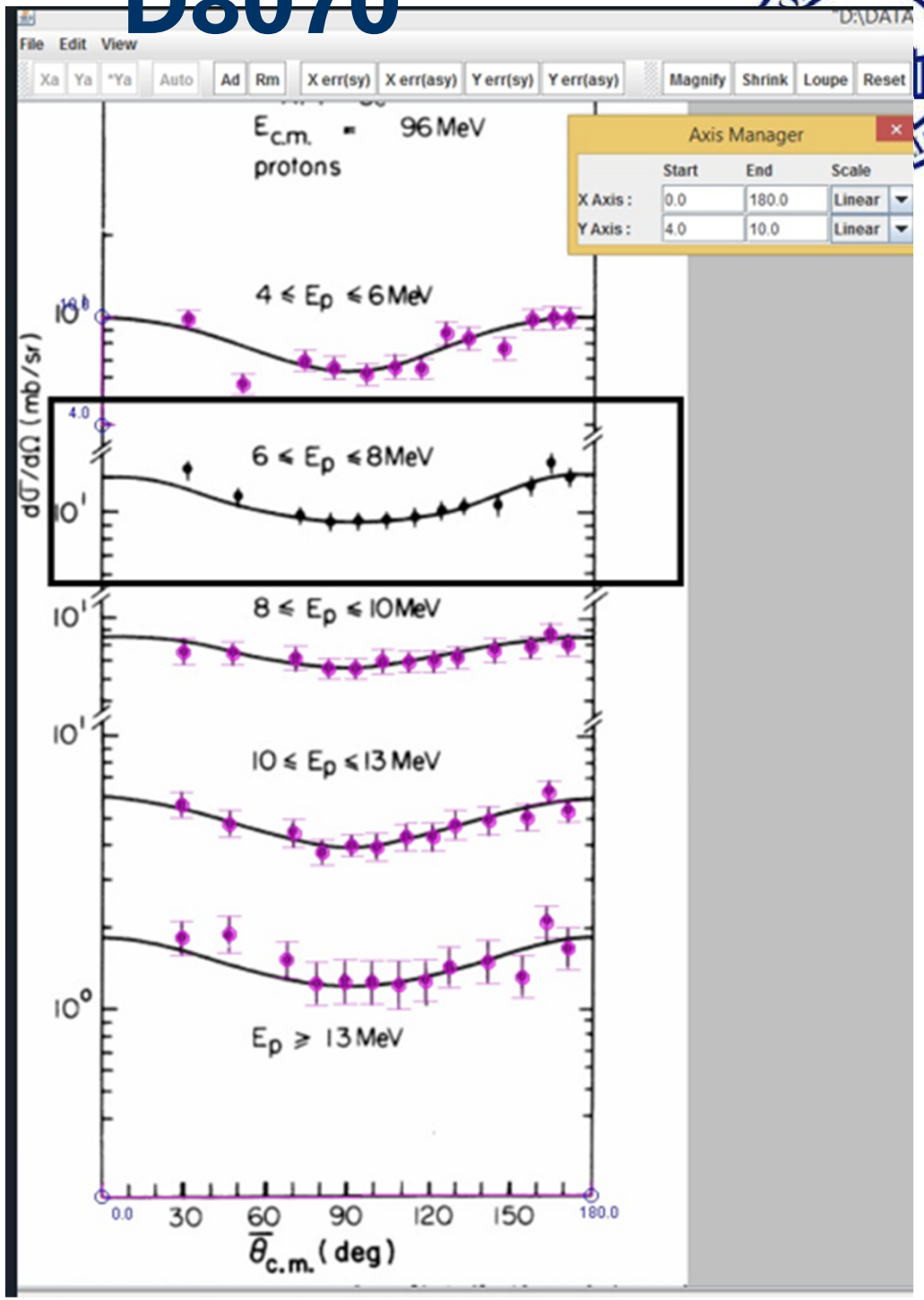


FIG. 7. ^{14}N-induced proton and α-particle angular distributions at the higher excitation energy integrated over small ranges of the particles' kinetic energy. The solid lines indicate the contribution of the compound-nuclear reaction; for details, see text.



D8077

Edit:

d8077

edit

Convert:

d8077

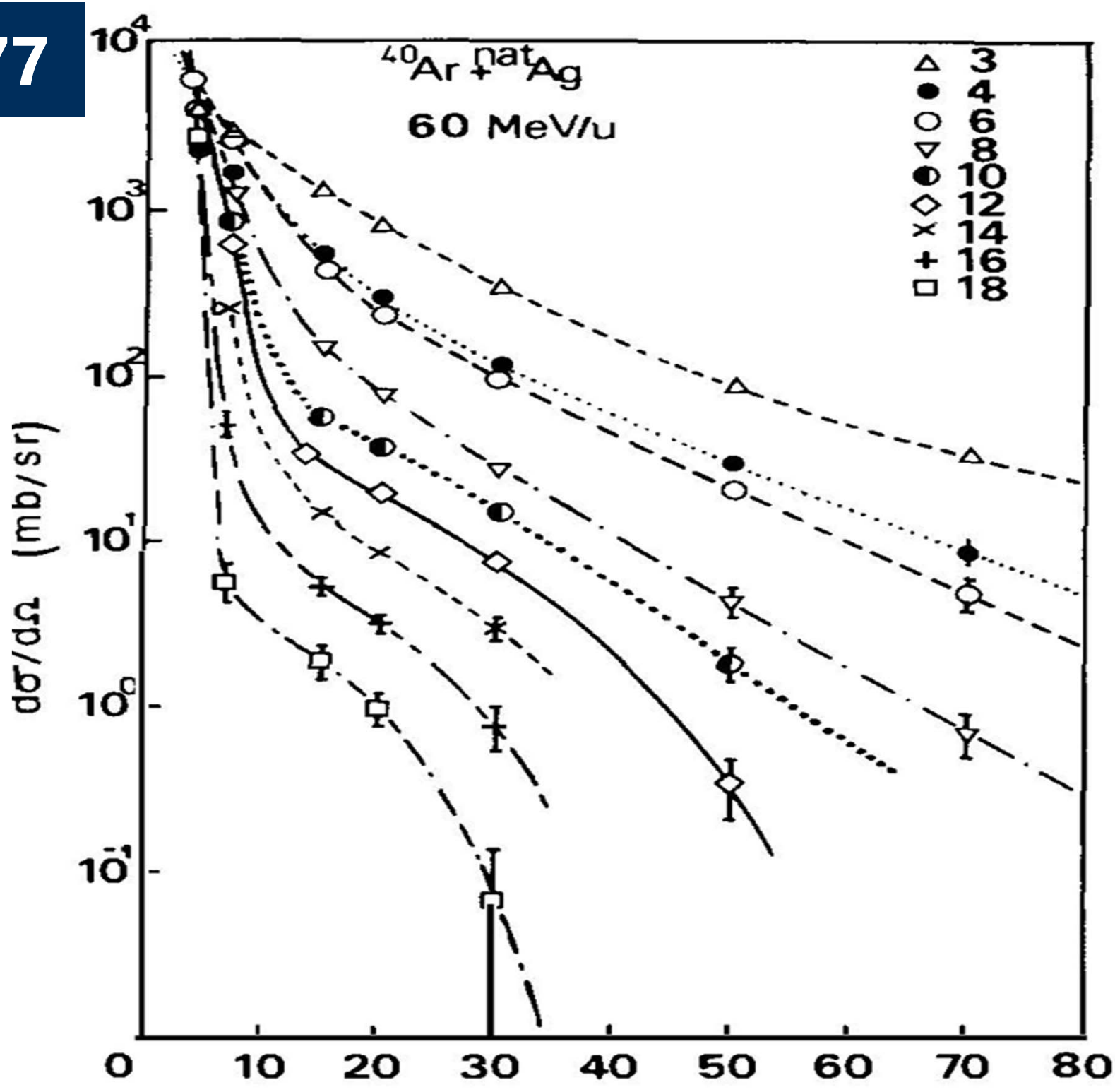
conv

NRDF
w/o data
CHEN
EXFOR
w/o data
CHEX
JANIS
EXFOR+
Graph

Bib
Data 0A
Data 0B
Data 0X
Data 1

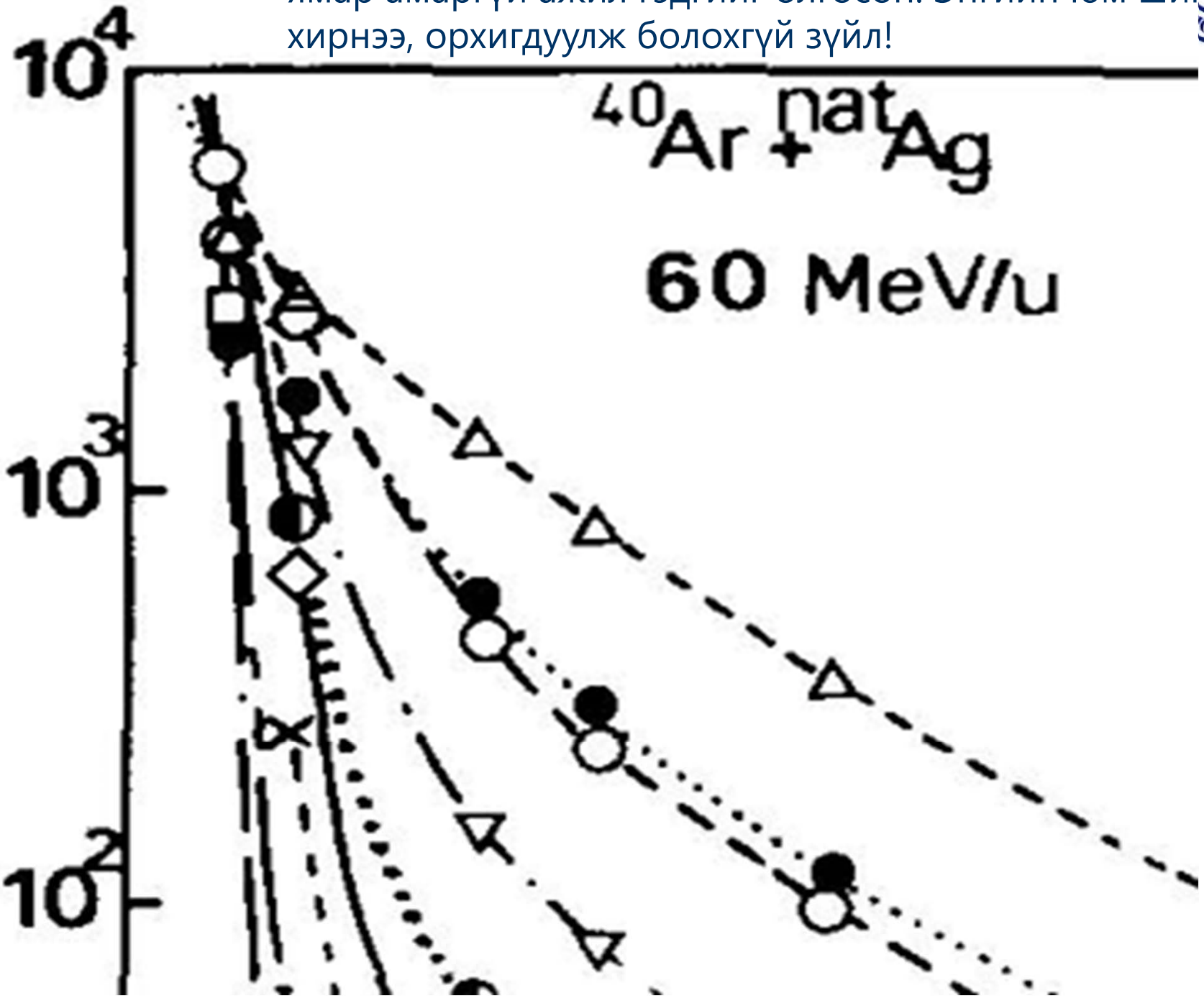
```
ENTRY          D8077   20241025                                  D80700000001
SUBENT        D8077001 20241025                                  D80700100001
BIB                11        18                                  D80700100002
TITLE      Transfer reactions and sequential decays of the       D80700100003
           projectile-like fragments in the 60MeV/nucleon        D80700100004
           40Ar+NatAg, 197Au reactions                           D80700100005
AUTHOR     (J.C.Steckmeyer, G.Bizard, R.Brou, P.Eudes,           D80700100006
            J.L.Laville, J.B.Natowitz, J.P.Patry, B.Tamain,      D80700100007
            A.Thiphagne, H.Doubre, A.Peghaire, J.Peter, E.Rosato, D80700100008
            J.C.Adloff, A.Kamili, G.Rudolf, F.Scheibling,        D80700100009
            F.Guilbault, C.Lebrun, F.Hanappe)                    D80700100010
INSTITUTE  (2FR CAE,2FR GAN,2FR STR,2FR NTE,2BLGBRU)             D80700100011
REFERENCE  (J,NP/A,500,372,1989)                                 D80700100012
SAMPLE     - Chemical-form of target is element.                 D80700100013
           - Physical-form of target is solid.                   D80700100014
METHOD     (EDE)                                                 D80700100015
FACILITY   (ACCEL,2FR GAN)                                       D80700100016
DETECTOR   (TELES,SOLST)                                         D80700100017
ERR-ANALYS (DATA-ERR) No information on source of uncertainties  D80700100018
STATUS     (TABLE,,J.C.Steckmeyer,J,NP/A,500,372,1989) Figure 3  D80700100019
HISTORY    (20241021C) M.Odsuren, L.Vrapcenjak                   D80700100020
ENDBIB             18          0                                 D80700100021
COMMON              1          3                                 D80700100022
EN                                                               D80700100023
MEV/A                                                            D80700100024
 60.                                                             D80700100025
ENDCOMMON           3          0                                 D80700100026
ENDSUBENT          25          0                                 D80700199999
```

D8077


$^{40}Ar + ^{nat}Ag$
60 MeV/u
3
4
6
8
10
12
14
16
18
dσ/dΩ (mb/sr)
10^4
10^3
10^2
10^1
10^0
10^{-1}
0
10
20
30
40
50
60
70
80

D8077

Муруй дээрээ биш цэгүүд будлиан үүсгээд цөмийн физик ч надад ийм бхад өөр хүмүүс компайл хийх ямар амаргүй ажил гэдгийг олгосон. Энгийн юм шиг хирнээ, орхигдуулж болохгүй зүйл!



D8069

Edit:
d8069
edit

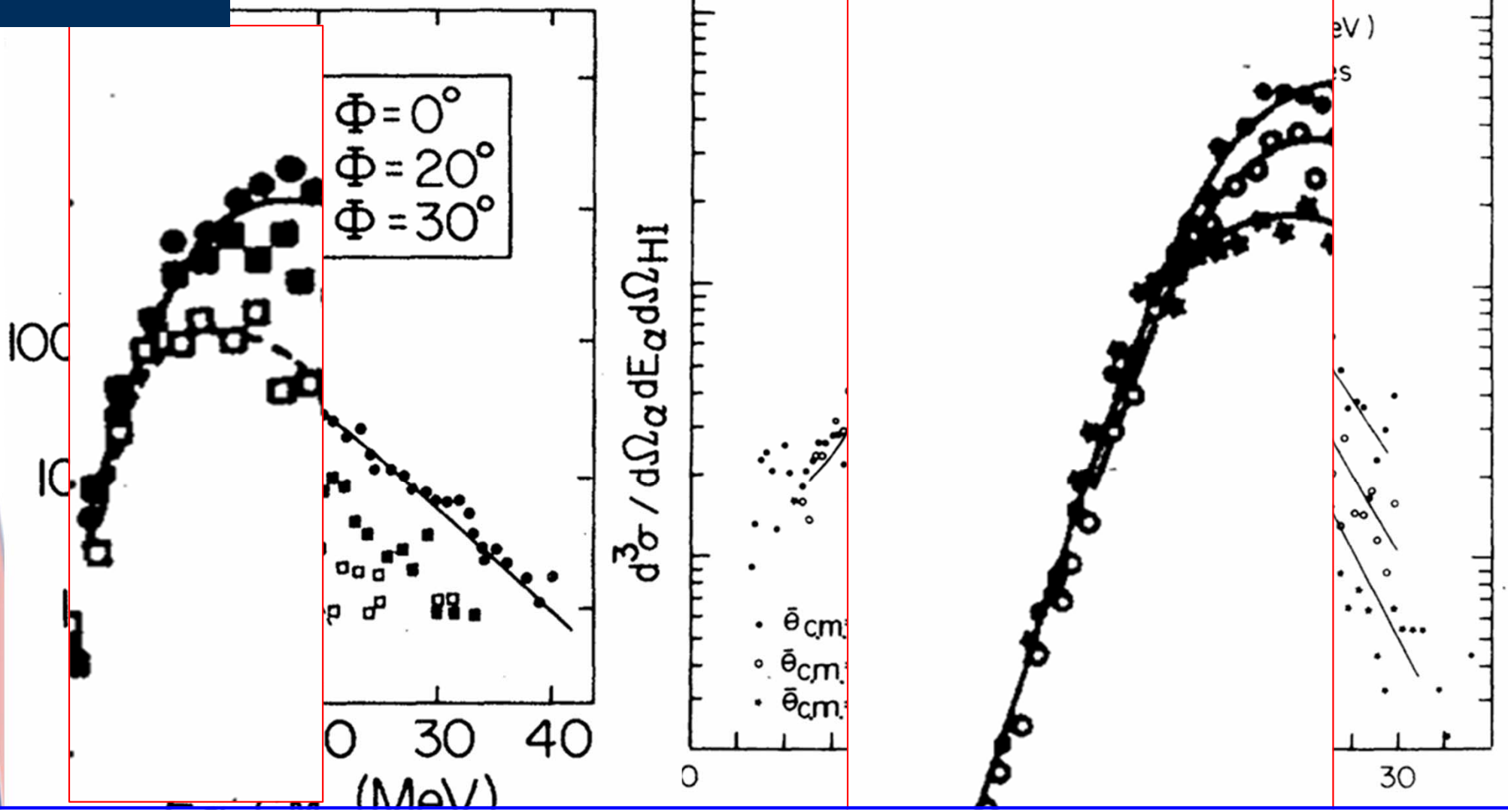
Convert:
d8069
conv

NRDF w/o data
CHEN
EXFOR w/o data
CHEX
JANIS
EXFOR+ Graph

Bib
Data 0A
Data 0B
Data 0X
Data 1
Data 2
Data 3

```
ENTRY            D8069   20240919                                D806900000001
SUBENT        D8069001   20240919                                D806900100001
BIB                 12         27                                D806900100002
TITLE       Study of charged particles emitted from 117Te compound D806900100003
            nuclei. I. Argon-induced reactions                   D806900100004
AUTHOR      (J.Galin, B.Gatty, D.Guerreau, C.Rousset,            D806900100005
             U.C.Schlotthauer-Voos, X.Tarrago)                   D806900100006
INSTITUTE   (2FR PAR)                                            D806900100007
REFERENCE   (J,PR/C,9,1113,1974)                                 D806900100008
INC-SOURCE  Beam intensity is 20 nA.                             D806900100009
SAMPLE      - Chemical-form of target is element.                D806900100010
            - Physical-form of target is solid.                  D806900100011
            - Target-thickness: Au (300 ug/cm2) - Se (1 mg/cm2) - D806900100012
              Au (800 ug/cm2). Au-Se-Au sandwich target.         D806900100013
            - Backing is gold.                                   D806900100014
            - Backing-thickness: 800 ug/cm2                      D806900100015
METHOD      (EDE)                                                D806900100016
FACILITY    (CYCLO,2FR PAR) ALICE                                D806900100017
DETECTOR    (TELES,SIBAR,SILI)                                   D806900100018
ERR-ANALYS  (ERR-1,10.,15.) Absolute cross section uncertainty of D806900100019
                            10% (higher energy) and 15% (lower   D806900100020
                            energy). (10-15%)                    D806900100021
ADD-RES     - Absolute cross sections integrated over angles at  D806900100022
              E(CM)=132 MeV for natSe and 77Se targets of emission D806900100023
              lithium are <1 mb. For the 77Se target at E(CM)=96 D806900100024
              MeV for emission of lithium is <0.1 mb.            D806900100025
            - Fig 12 cannot be digitized because it is not       D806900100026
              possible to distinguish between open and solid     D806900100027
              symbols.                                           D806900100028
HISTORY     (20240815C) M.Odsuren, N.Otsuka                      D806900100029
```

D8069



REACTION	(34-SE-0(18-AR-40,X)2-HE-4,,DA/DE)	D806900600003
	some symbols between 10 and 13 MeV are not suitable	D806900600004
	for digitization and not compiled.	D806900600005
ANG-SEC	(ANG-CM,A)	D806900600006
EN-SEC	(E-CM,A)	D806900600007
STATUS	(CURVE,,J.Galin+,J,PR/C,9,1113,1974) Figure 8	D806900600008

Fig12

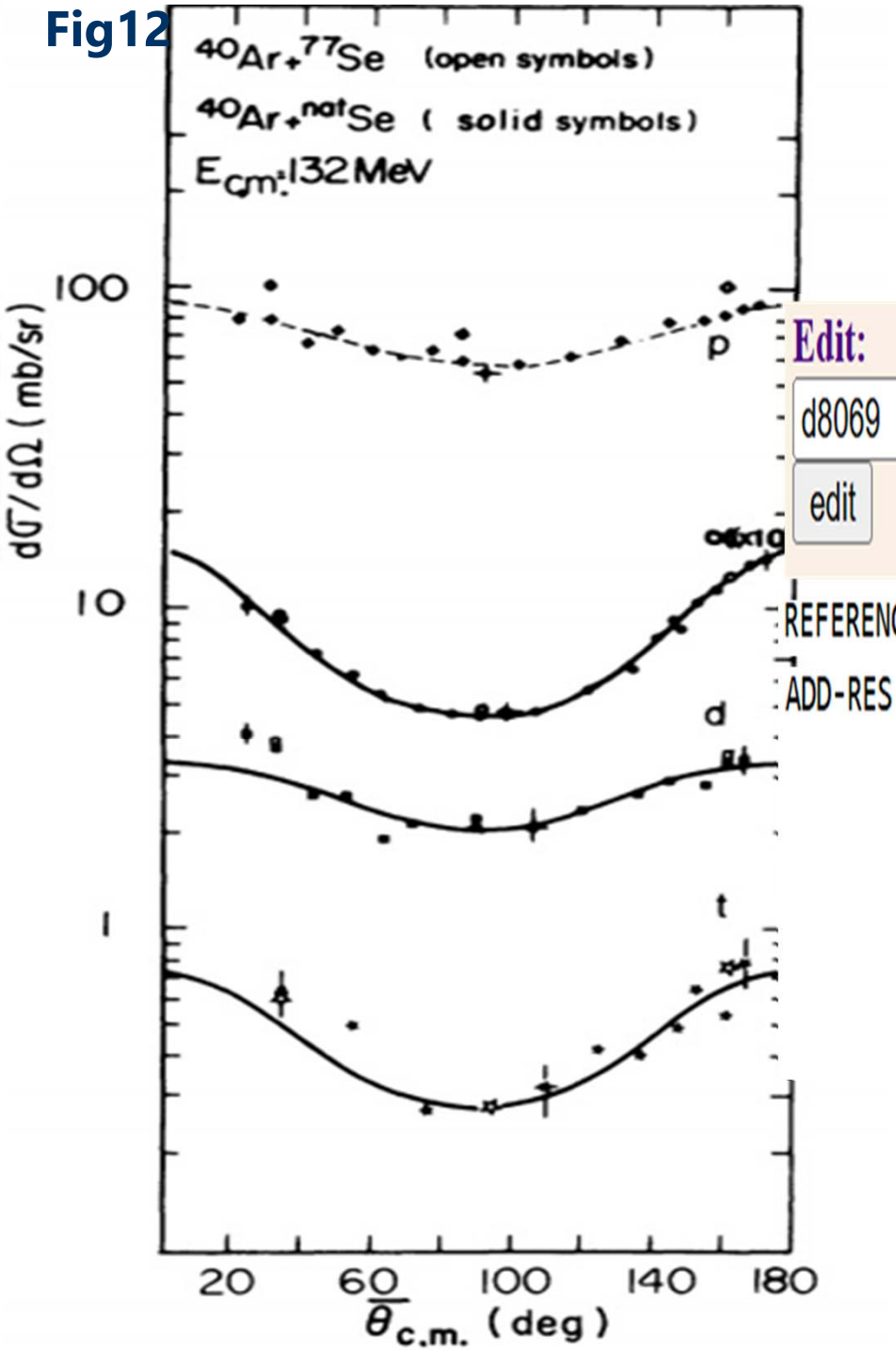


FIG. 12. Angular distributions of pr
les integrated over the total spectra a

D8069

Edit:

d8069

edit

REFERENCE	(J,PR/C,9,1113,1974)	D806900100008
ADD-RES	- Absolute cross sections integrated over angles at	D806900100022
	E(CM)=132 MeV for natSe and 77Se targets of emission	D806900100023
	lithium are <1 mb. For the 77Se target at E(CM)=96	D806900100024
	MeV for emission of lithium is <0.1 mb.	D806900100025
	- Fig 12 cannot be digitized because it is not	D806900100026
	possible to distinguish between open and solid	D806900100027
	symbols.	D806900100028

Edit:
d8071
edit

Convert:
d8071
conv

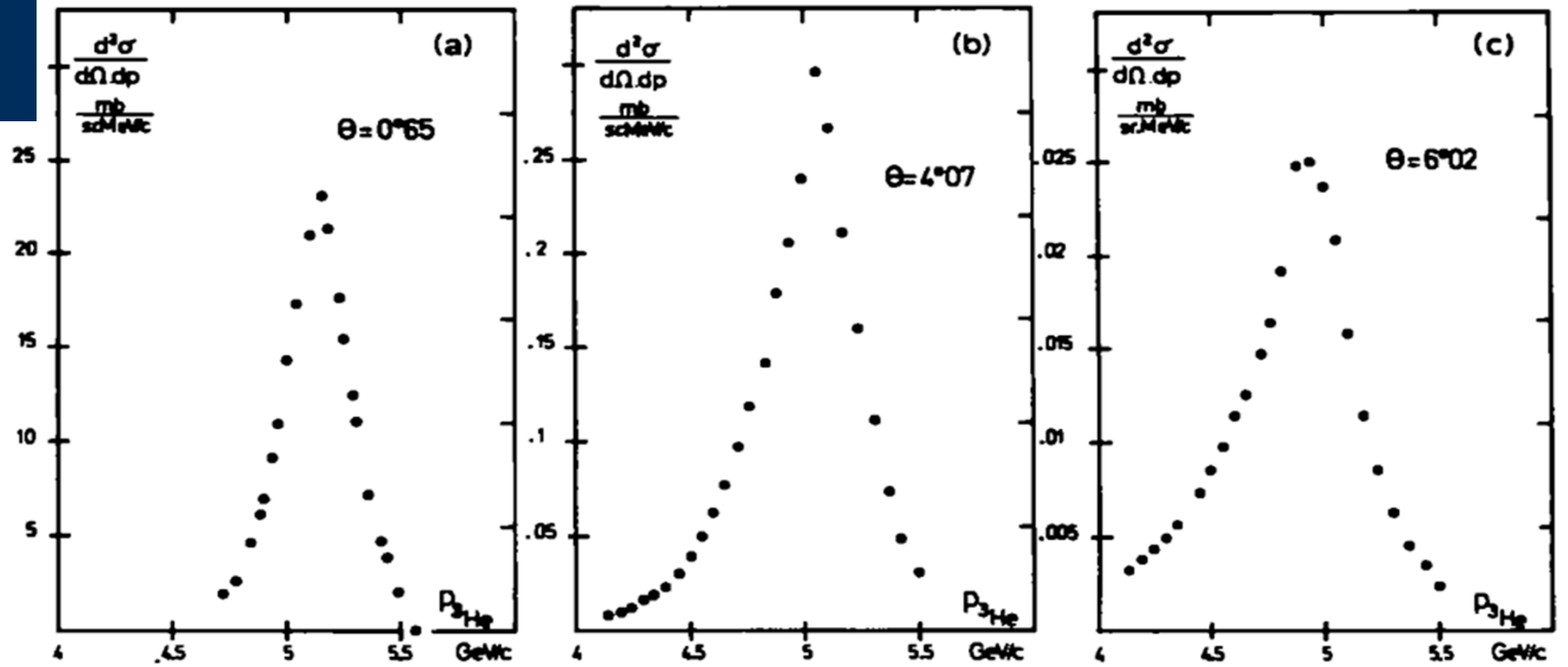
NRDF
w/o data
CHEN
EXFOR
w/o data
CHEX
JANIS
EXFOR+
Graph

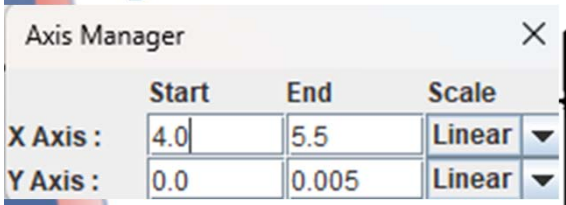
Bib
Data 0A
Data 0B
Data 0X
Data 1
Data 2
Data 3
Data 4

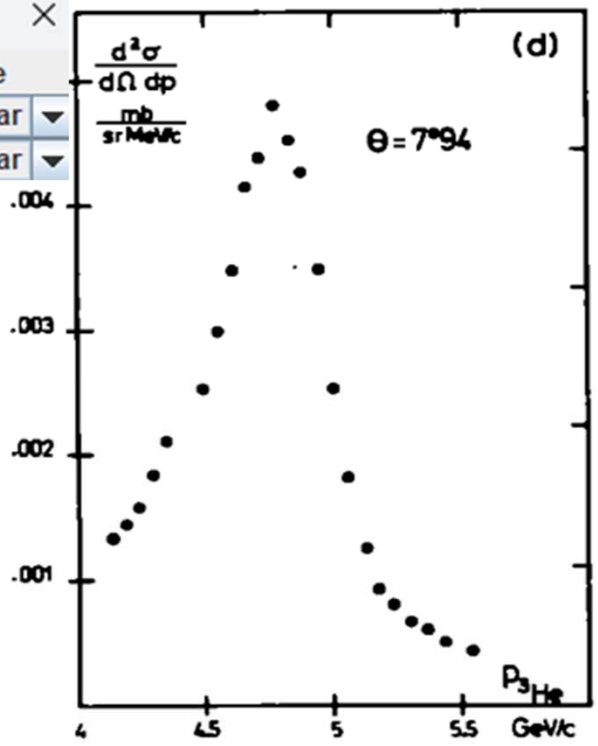
D8071

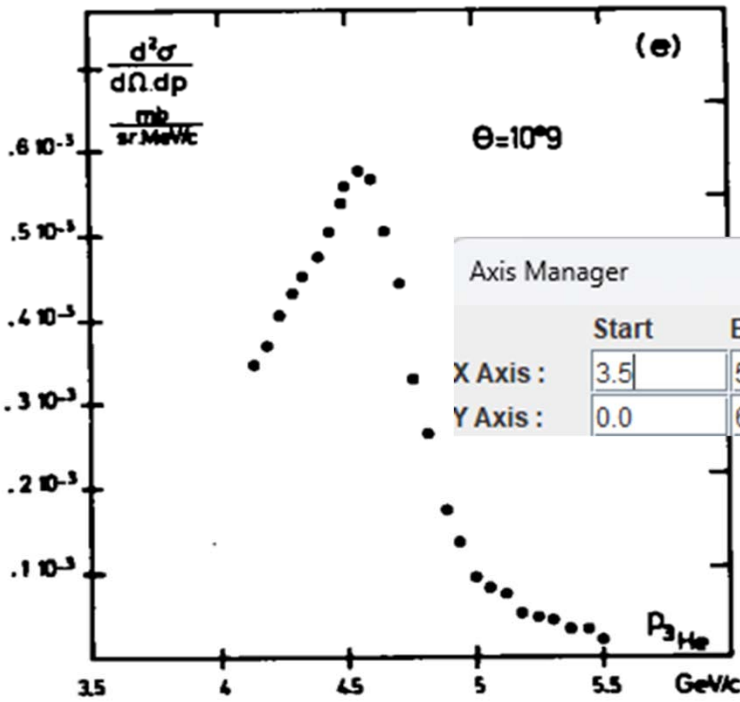
```
            D8071   20240917                                    D80710000001
          D8071001  20240917                                    D80710010001
                11        17                                    D80710010002
TITLE     3He production in 4He fragmentation on protons at 6.85 D80710010003
          GeV/c                                                 D80710010004
AUTHOR    (G.Bizard, C.Le Brun, J.Berger, J.Duflo, L.Goldzahl,  D80710010005
           F.Plouin, J.Oostens, M.Van Den Bossche, L.Vu Hai,    D80710010006
           F.L.Fabbri, P.Picozza, L.Satta)                      D80710010007
INSTITUTE (2FR CAE,2FR SAT,2ITYFRA)                             D80710010008
REFERENCE (J,NP/A,285,461,1977)                                 D80710010009
MONITOR   (6-C-12(A,X)6-C-11,,SIG)                              D80710010010
INC-SOURCE Beam intensity is 2.0E+10 PPS.                       D80710010011
SAMPLE    - Chemical-form of target is element.                 D80710010012
          - Physical-form of target is solid.                   D80710010013
          - Target-thickness: 5.9 cm long liquid hydrogen.      D80710010014
METHOD    (TOF)                                                 D80710010015
FACILITY  (SYNCH,2FR SAT)                                       D80710010016
DETECTOR  (MAGSP,SCIN)                                          D80710010017
          (IOCH)                                                D80710010018
HISTORY   (20240811C) M.Odsuren, N.Otsuka                       D80710010019
ENDBIB              17          0                               D80710010020
COMMON               1          3                               D80710010021
MOM                                                             D80710010022
GEV/C                                                           D80710010023
 6.85                                                           D80710010024
ENDCOMMON            3          0                               D80710010025
ENDSUBENT           24          0                               D80710199999
SUBENT       D8071002   20240917                                D80710200001
BIB                  3          3                               D80710200002
REACTION  (1-H-1(A,X)2-HE-3,,DA/DP)                             D80710200003
ANG-SEC   (ANG,2-HE-3)                                          D80710200004
STATUS    (CURVE,,G.Bizard+,J,NP/A,285,461,1977) Figure 2 (a-e) D80710200005
```

D8071


(a)
θ = 0°65
(b)
θ = 4°07
(c)
θ = 6°02
$\frac{d^2\sigma}{d\Omega.dp}$
mb / sr.MeV/c
$p_{^3He}$
GeV/c


Axis Manager
Start
End
Scale
X Axis :
4.0
5.5
Linear
Y Axis :
0.0
0.005
Linear


(d)
θ = 7°94
$\frac{d^2\sigma}{d\Omega\,dp}$
mb / sr.MeV/c
$p_{^3He}$
GeV/c


(e)
θ = 10°9
$\frac{d^2\sigma}{d\Omega.dp}$
mb / sr.MeV/c
$p_{^3He}$
GeV/c

Axis Manager
Start
End
Scale
X Axis :
3.5
5.5
Linear
Y Axis :
0.0
6.0E-4
Linear

Data Digitization in EXFOR

Automation and the GSYS System:
Digitizing this kind of information is labor-intensive, requiring skilled personnel to ensure accuracy.

The GSYS system was developed to streamline and automate this process for EXFOR, allowing older articles and experimental results to be digitized efficiently and accurately.

GSYS minimizes errors and speeds up the data conversion process, making it easier to capture and organize historical data.

Future Directions:
With the increasing demand for accessible data, the EXFOR community is exploring the application of AI to improve data processing, analysis, and usability.

Thank you for your attention!