Recent work on ENSDF: OODB, machine learning, APIs

Adam Hayes

on behalf of the

National Nuclear Data Center
Brookhaven National Laboratory

GNDS
November 2020

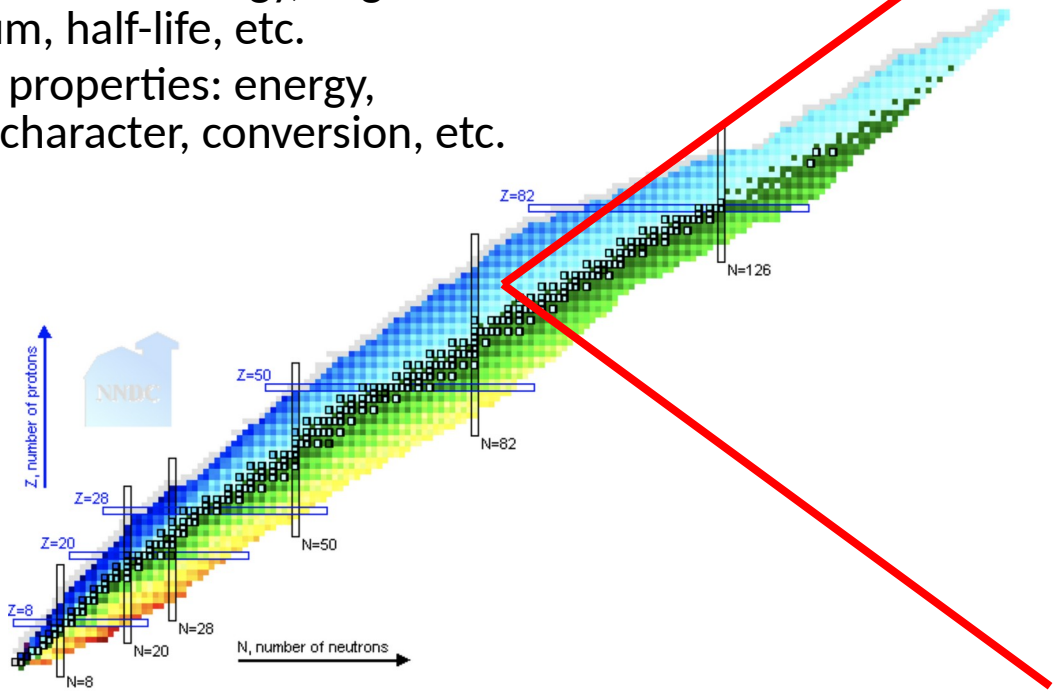
Outline

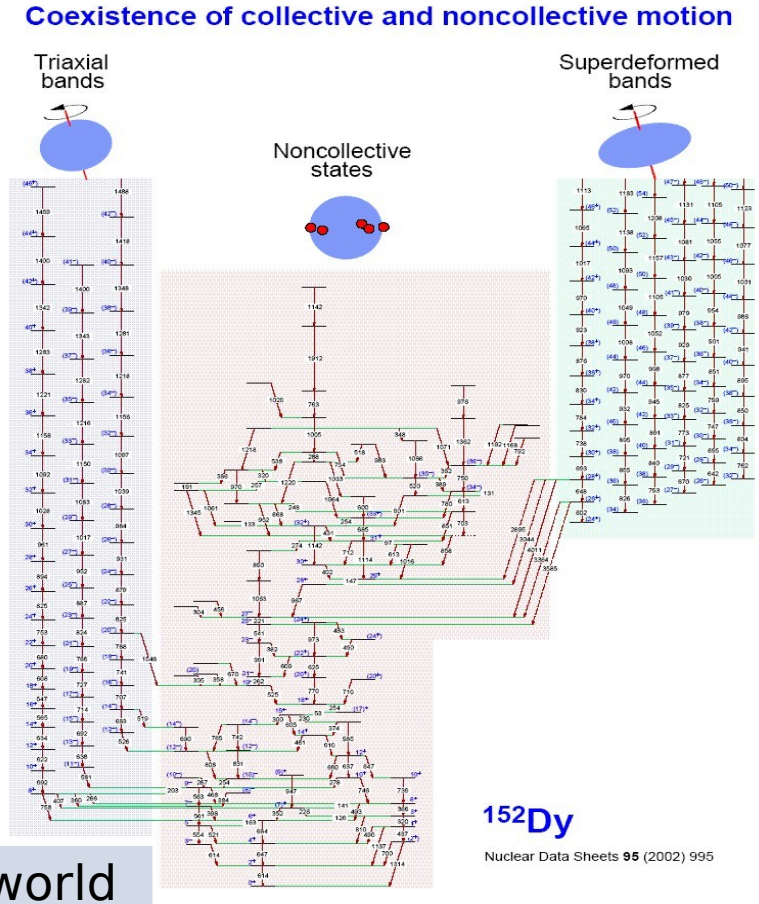
- ENSDF now
- Rethinking ENSDF and XUNDL
- New since May GNDS meeting
- Status
- Questions for
  the downstream ENSDF user

Evaluated Nuclear Structure Data Files (ENSDF)

For each of 3,300 nuclei:

- Level properties : energy, angular momentum, half-life, etc.
- Radiation properties: energy, intensity, character, conversion, etc.





It is Unique: Only Nuclear Database of this kind in the world

It is Complete: **All** nuclei and **all** level and radiation properties

It is Versatile: Feeds back into both basic and applied sciences

National Nuclear Data Center

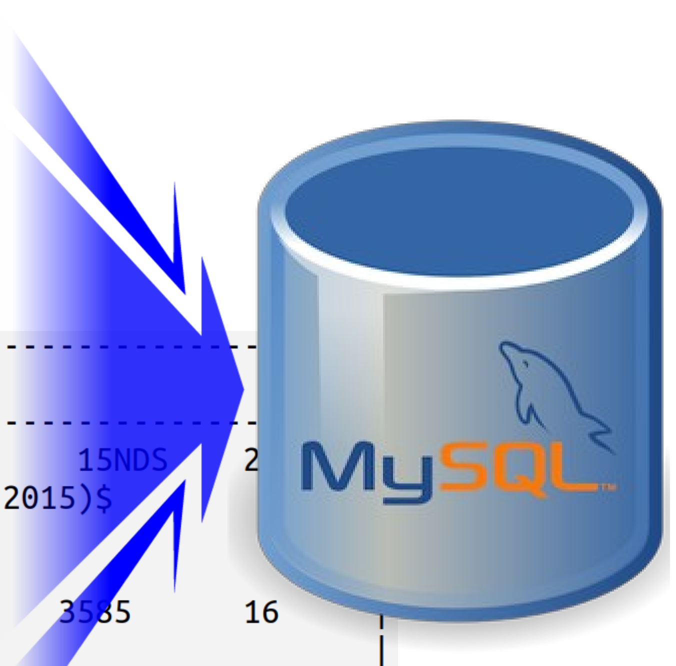
Slide: E.A. McCutchan

A.B. Hayes

Present state of ENSDF and XUNDL

rec_id	dsid	ds_type	nucid
238092001	ADOPTED LEVELS, GAMMAS	ADOPTED	238U
238092002	238PA B- DECAY	DECAY	238U
238092003	238U IT DECAY (280 NS)	DECAY	238U
238092004	242PU A DECAY	DECAY	238U
238092005	236U(T,P)	REACTI	238U

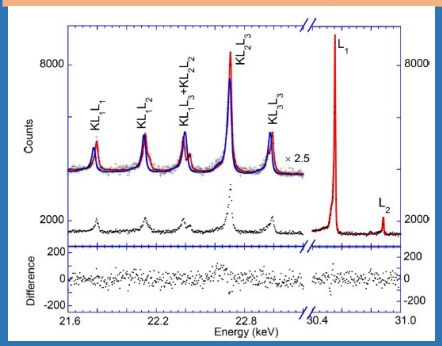
rec_id	line_text
238092002	238U 238PA B- DECAY 15NDS 2
238092002	238U H TYP=FUL$AUT=E. BROWNE, J. K. TULI$CIT=NDS 127, 191 (2015)$
238092002	238U 2 H CUT=1-Jun-2014$
238092002	238U D Slightly modified by E. Browne (7/24, 2014).
238092002	238PA P 0.0 (3-) 2.28 M 10 3585 16
238092002	238U N 1.0
238092002	238U c The Gamow-Teller |b-decay strength function has been calculated in
238092002	238U 2c 1978Iz04 and 1979KlZT.



Rethinking ENSDF and XUNDL

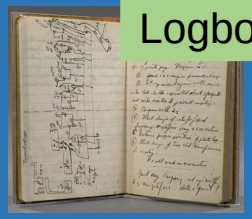
ENSDF now:
Discrete levels
Discrete radiation

X-rays and Auger electrons

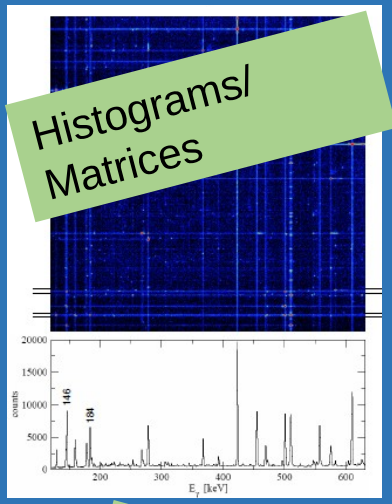


M. Alotiby et al., J. Elec. Spect. 232, 73 (2019)

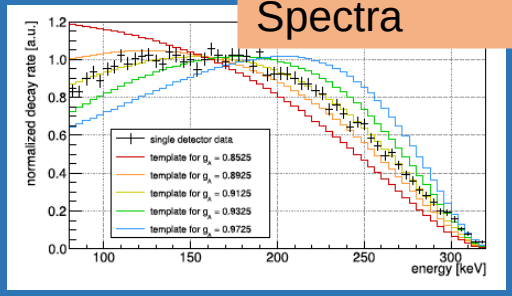
Logbook



Histograms/ Matrices



Continuous Spectra



Cobra, Phys. Lett. B 800, 135092 (2020)

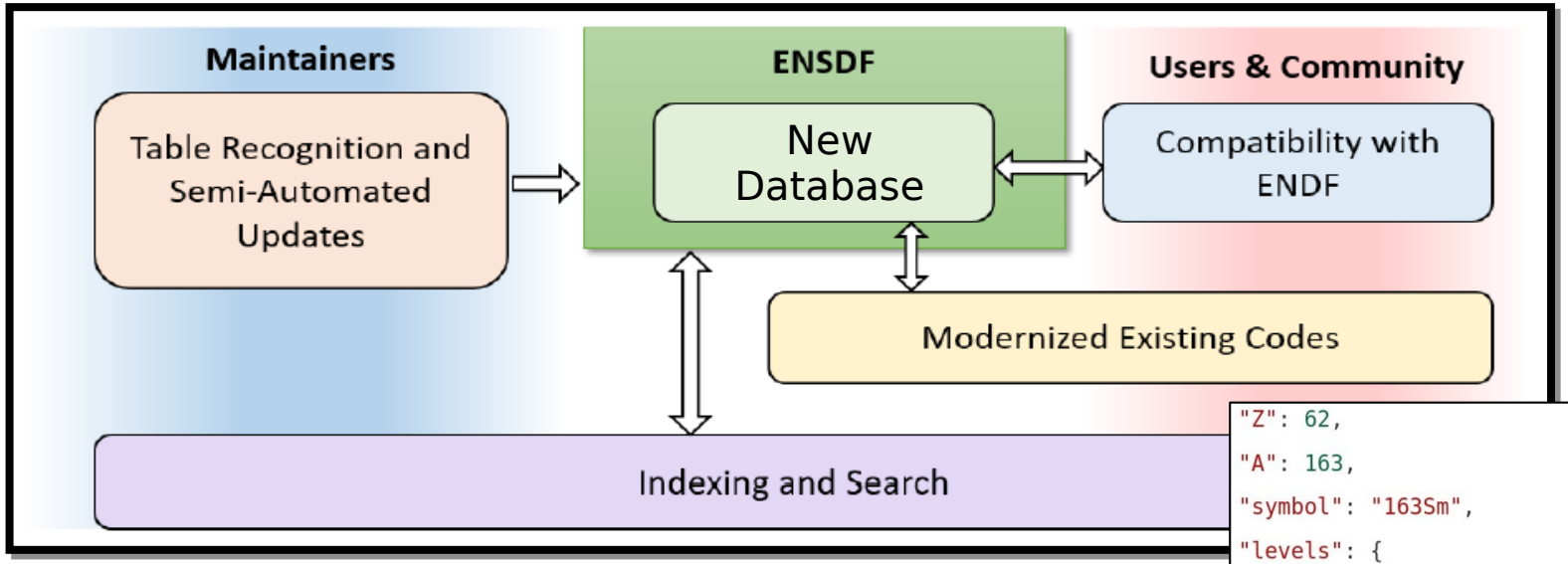
Raw Event Data

```
Event 20
Det 37
Det 61
Event 20
Det 04
Det 24   E 615
Det 46   E 339
Event 20585
Det 17   E 761
Det 20   E 186
Det 59   E 615
```

ENSDF modernization project

Received 3 years of funding in July 2020 through Nuclear Data Interagency Working Group – FOA LAB 19-2114

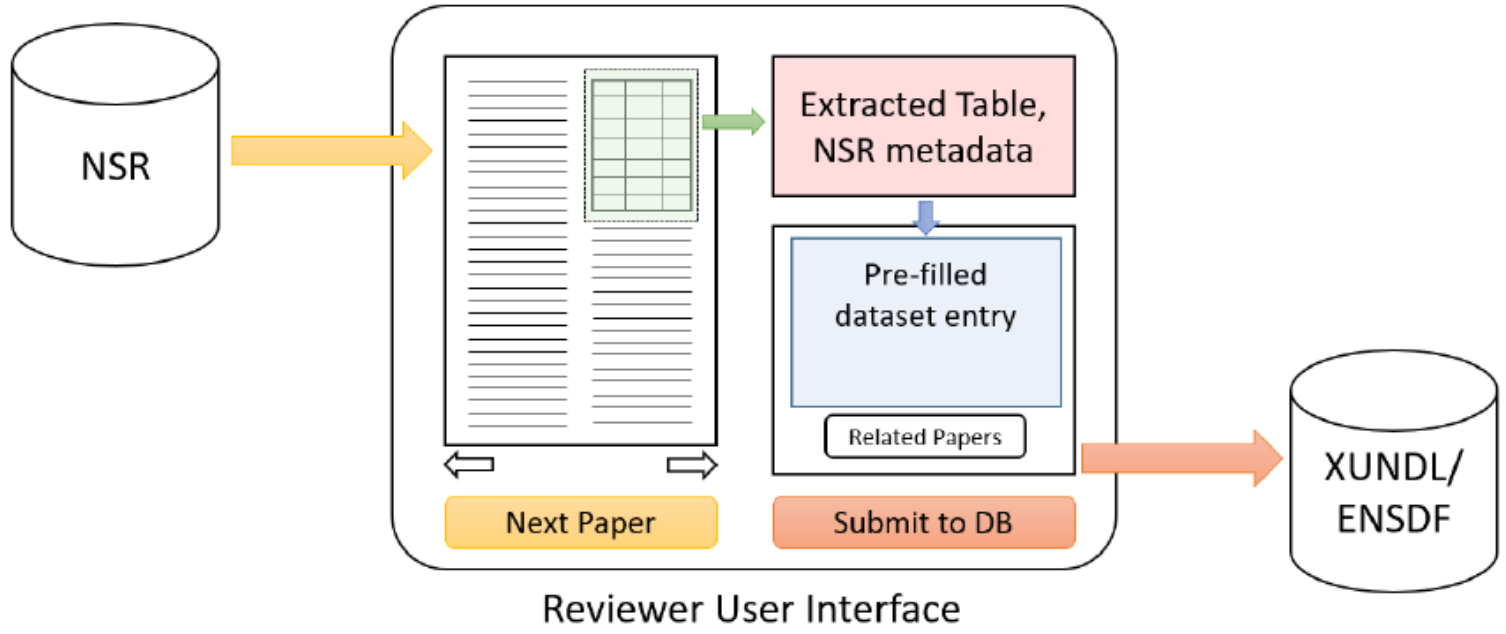
Collaboration with BNL (NNDC and CSI), ANL, LLNL – 3 year proposal




```
"Z": 62,
"A": 163,
"symbol": "163Sm",
"levels": {
    "GS": {
        "tHalf": {
            "published": {},
            "evaluations": {
                "2017, 04, 01, 00, 00, 00, 000000": {…}
            },
            "measurements": {}
        },
```


1) Develop a new Object-Oriented Database for ENSDF
2) Modernize existing codes used with ENSDF
3) Develop streamlined publication-to-ENSDF software
4) Apply machine learning to increase efficiency of evaluations
5) Ensure compatibility with end users

Make the most of the evaluator's effort



- OCR for extraction of metadata and numerical data
- Automatic population of database

- Reviewer interface for side-by-side comparison
- Consideration by scientist still essential

Table extraction using Machine Learning

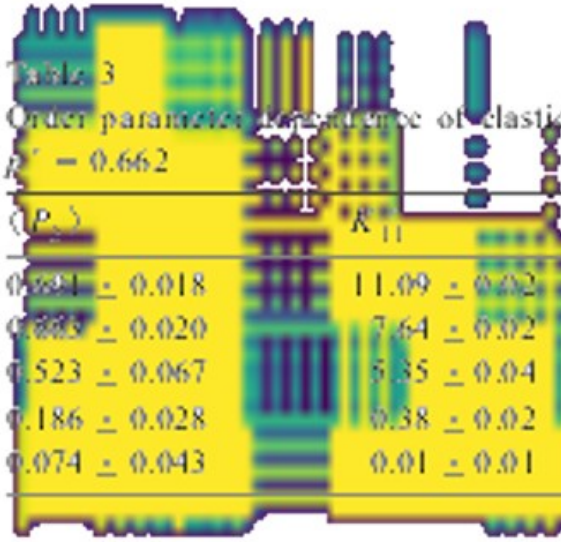
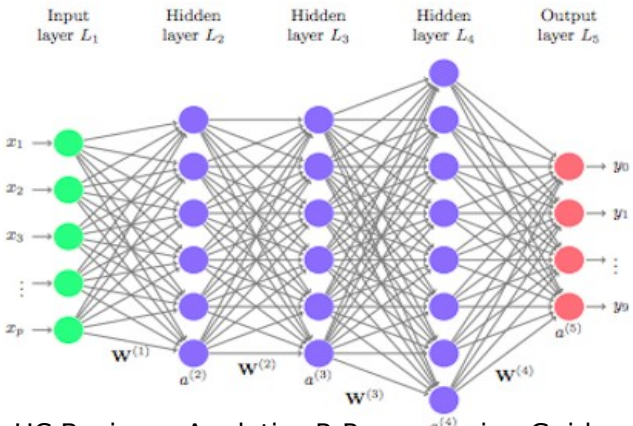


Image: Calos Soto, BNL's Computer Science Initiative

- Developed a preliminary schema for "raw" table data
- CSI evaluating deep-learning models, including TableNet https://arxiv.org/abs/2001.01469
  - Example of segmentation before tuning thresholding
  - Operating on a table image
- Preparing for training on both PDF and LaTeX

```
{"levelEnergy" : 171.7
{"initialJpi"  : "3/2-"
{"gammaEnergy" : 171.7
{"A0"          : 6
{"RDCO"        : 0.65
```

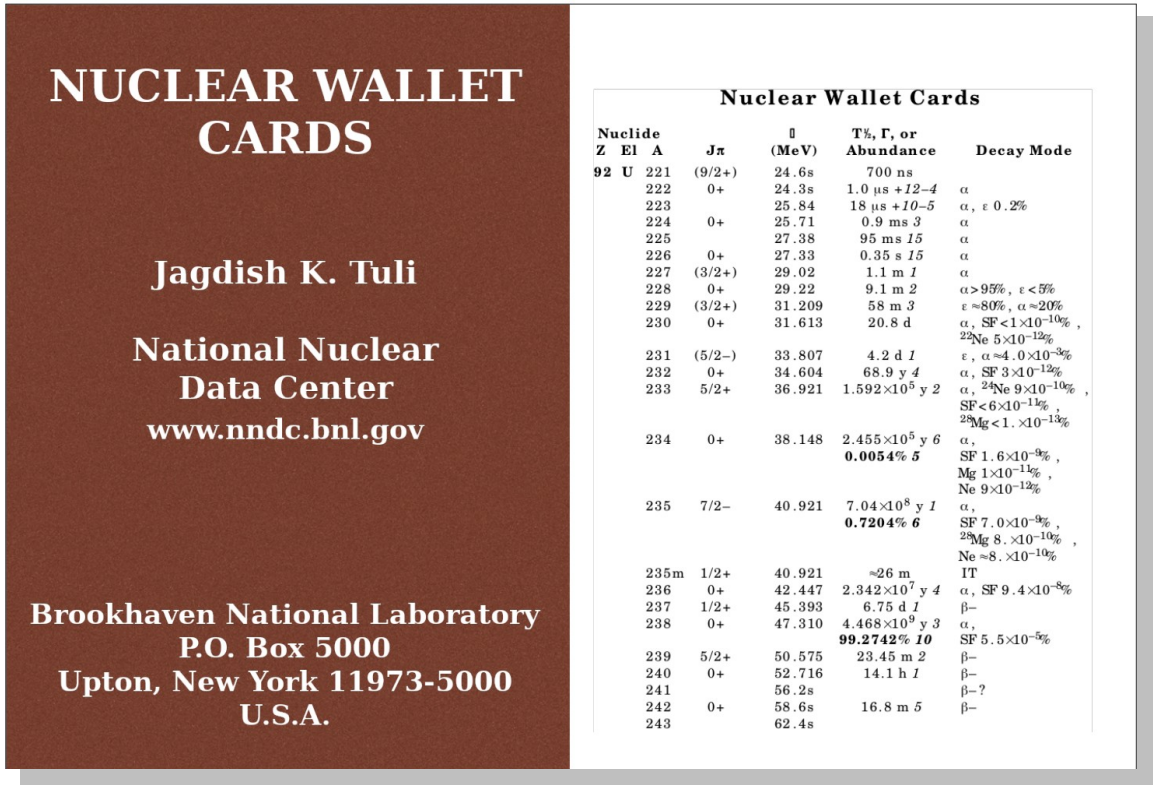
Training UI



UC Business Analytics R Programming Guide

Successful object-oriented DB
TEST CASE: Nuclear Wallet Cards

- Basic structure and decay data for ground and long-lived states
- Currently, production involves unnecessary repetition
- Average 6.5 years between editions



NUCLEAR WALLET CARDS

Jagdish K. Tuli

National Nuclear Data Center
www.nndc.bnl.gov

Brookhaven National Laboratory
P.O. Box 5000
Upton, New York 11973-5000
U.S.A.

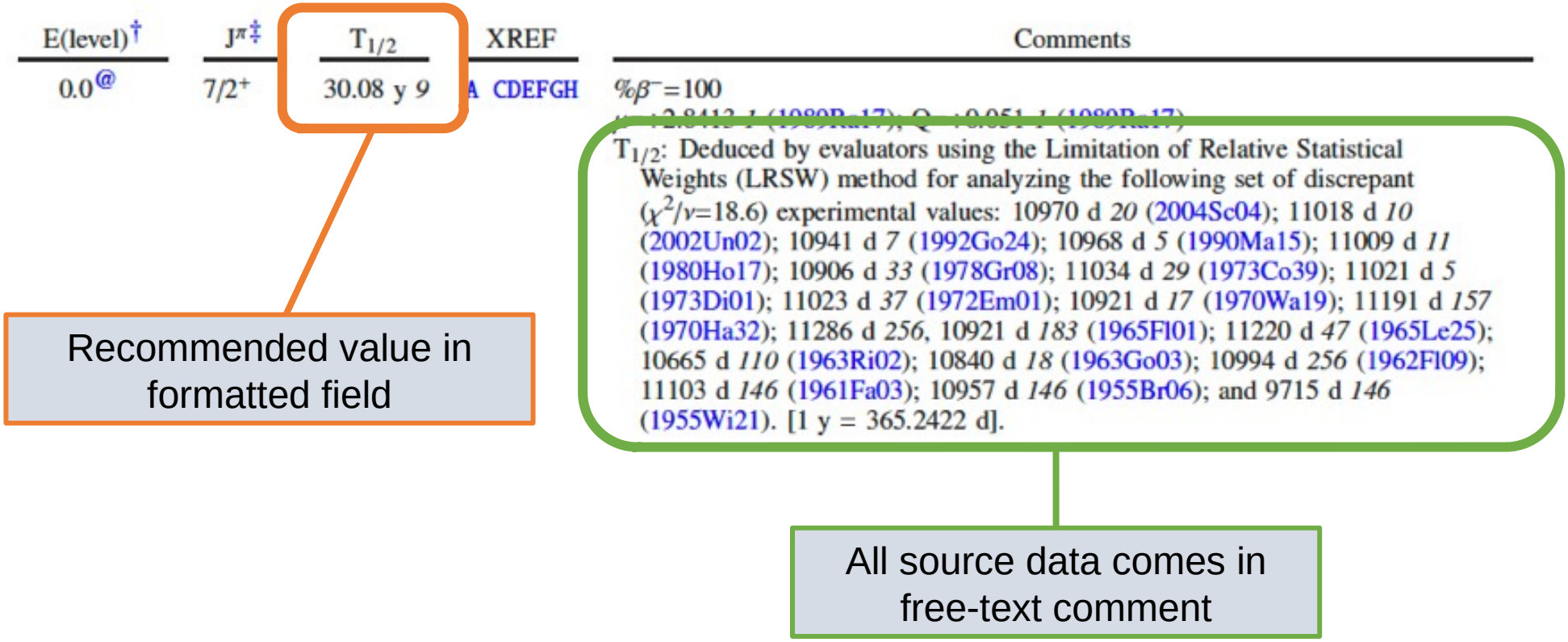
Nuclear Wallet Cards

Nuclide Z	El	A	Jπ	(MeV)	T½, Γ, or Abundance	Decay Mode
92	U	221	(9/2+)	24.6s	700 ns	
		222	0+	24.3s	1.0 μs *+12–4*	α
		223		25.84	18 μs *+10–5*	α, ε 0.2%
		224	0+	25.71	0.9 ms *3*	α
		225		27.38	95 ms *15*	α
		226	0+	27.33	0.35 s *15*	α
		227	(3/2+)	29.02	1.1 m *1*	α
		228	0+	29.22	9.1 m *2*	α>95%, ε<5%
		229	(3/2+)	31.209	58 m *3*	ε≈80%, α≈20%
		230	0+	31.613	20.8 d	α, SF<1×10^{-10}%, ^{22}Ne 5×10^{-12}%
		231	(5/2–)	33.807	4.2 d *1*	ε, α≈4.0×10^{-3}%
		232	0+	34.604	68.9 y *4*	α, SF 3×10^{-12}%
		233	5/2+	36.921	1.592×10^{5} y *2*	α, ^{24}Ne 9×10^{-10}%, SF<6×10^{-11}%, ^{28}Mg<1.×10^{-13}%
		234	0+	38.148	2.455×10^{5} y *6* **0.0054% *5***	α, SF 1.6×10^{-9}%, Mg 1×10^{-11}%, Ne 9×10^{-12}%
		235	7/2–	40.921	7.04×10^{8} y *1* **0.7204% *6***	α, SF 7.0×10^{-9}%, ^{28}Mg 8.×10^{-10}%, Ne≈8.×10^{-10}%
		235m	1/2+	40.921	≈26 m	IT
		236	0+	42.447	2.342×10^{7} y *4*	α, SF 9.4×10^{-8}%
		237	1/2+	45.393	6.75 d *1*	β–
		238	0+	47.310	4.468×10^{9} y *3* **99.2742% *10***	α, SF 5.5×10^{-5}%
		239	5/2+	50.575	23.45 m *2*	β–
		240	0+	52.716	14.1 h *1*	β–
		241		56.2s		β–?
		242	0+	58.6s	16.8 m *5*	β–
		243		62.4s		

- CouchDB + User Interface
- Eliminates reptition for evaluators
- History of evaluator's work
  - Methods
  - Data considered
  - Data used
  - Comments
- Uniformity of calculation methods
- Freeze publication versions
- Reduce publication time
- Test case for larger projects (e.g. binary data easily added)

Wallet card updates

The half-life of ^{137}Cs has been measured 20 times



E(level)†	Jπ‡	$T_{1/2}$	XREF	Comments
0.0@	7/2+	30.08 y 9	A CDEFGH	%β^-=100

$T_{1/2}$: Deduced by evaluators using the Limitation of Relative Statistical Weights (LRSW) method for analyzing the following set of discrepant (χ^2/ν=18.6) experimental values: 10970 d *20* (2004Sc04); 11018 d *10* (2002Un02); 10941 d *7* (1992Go24); 10968 d *5* (1990Ma15); 11009 d *11* (1980Ho17); 10906 d *33* (1978Gr08); 11034 d *29* (1973Co39); 11021 d *5* (1973Di01); 11023 d *37* (1972Em01); 10921 d *17* (1970Wa19); 11191 d *157* (1970Ha32); 11286 d *256*, 10921 d *183* (1965Fl01); 11220 d *47* (1965Le25); 10665 d *110* (1963Ri02); 10840 d *18* (1963Go03); 10994 d *256* (1962Fl09); 11103 d *146* (1961Fa03); 10957 d *146* (1955Br06); and 9715 d *146* (1955Wi21). [1 y = 365.2422 d].

- Evaluator needs to repeat manual lookups and calculations
- Repetitive, inefficient and error prone
- Almost impossible to implement automation or AI

Nuclear Wallet Cards interface with evaluator's history

Video: Web interface
by Ben Shu

- ~3000 nuclides
- In use at NNDC
- Will be used for next version of WC

Status

- July 2020: McCutchan *et al*. received 3 years of funding for major ENSDF database overhaul & machine learning
- Started redesign of database and internal discussions on user interface
- Beginning work with Brookhaven's Computer Science Initiative (CSI)
  - OCR and table recognition on LaTeX, images of PDF files
  - UI to streamline human training of learning model
- Working on prototype evaluators' UI (led by Ben Shu, NNDC).
- Designing the schema for a full ENSDF nuclide entry
  - Organization of measured values, references, historical evaluator notes, *etc*.
  - Defining very uniform human-readable structure and key names (intuitive to the developer, obvious to readers)
  - Streamline evaluation process
- Beginning work on ENSDF API.

Parting questions

- Generally,

  In what other ways can we exploit new **database technology** and related **software & APIs** to improve database currency?

- For the downstream user:
  - What features would be helpful (e.g. for ENDF) in a general-purpose API?

    - Level energies
    - Gamma energies
    - …

- What ENSDF content would benefit ENDF?
  - Specific data
  - Views (indexes)
- Thinking broadly, what new features would be beneficial?

END