



Modernizing Workflow for NSR Updates

Benjamin Shu

October 2024



Current State: Data Model

Limited by its adherence to “exchange” format:

- Entries need to be pre-processed before use in code

```
<KEYNO    >2002RI06
<HISTORY  >A20020708
<CODEN    >JOUR NUPAB...
<REFERENCE>Nucl.Phys. A...
<AUTHORS  >M.V.Ricciardi...
<TITLE    >Experimental Study...
<KEYWORDS>NUCLEAR REACTIONS...
<SELECTRS>T:1H;A. ...
```

Current State: Data Model

Limited by its adherence to “exchange” format:

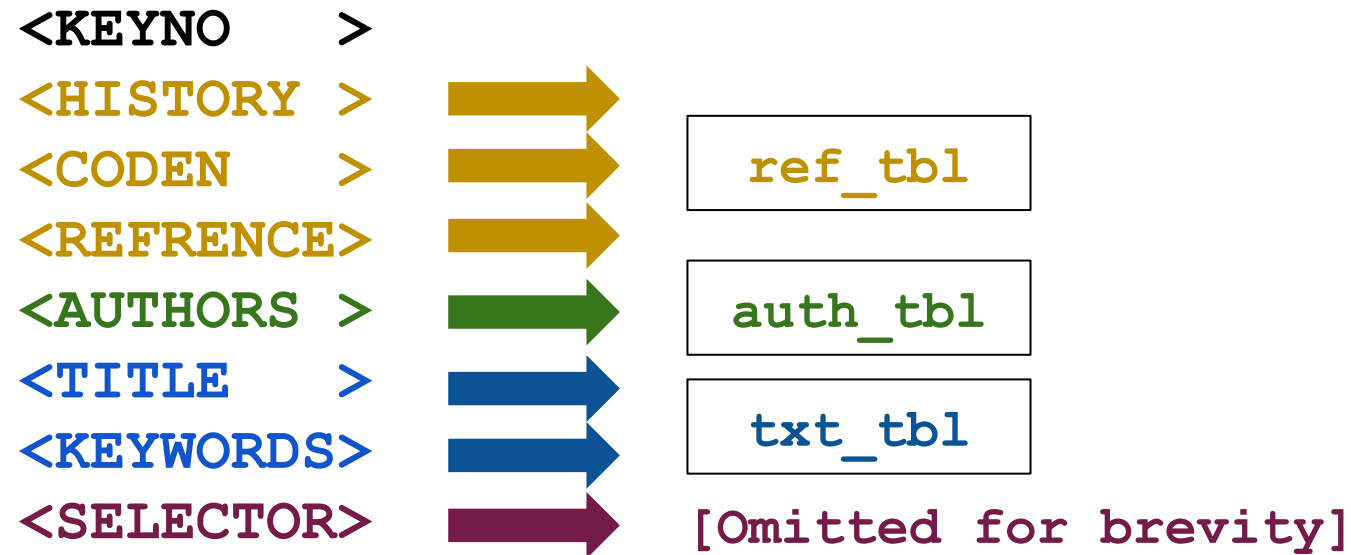
- Entries need to be pre-processed before use in code

<KEYNO >	2002RI06	Unique ID
<HISTORY >	A20020708	Entry/modification dates
<CODEN >	JOUR NUPAB...	Source type, journal
<REFERENCE>	Nucl.Phys. A...	Citation information
<AUTHORS >	M.V.Ricciardi...	List of authors
<TITLE >	Experimental Study...	Formatted title
<KEYWORDS>	NUCLEAR REACTIONS...	Sentences about content
<SELECTRS>	T:1H;A. ...	Structured search keys

Current State: Data Model

Limited by its adherence to “exchange” format:

- Text files need to be synchronized with SQL tables



Current State: Data Model

Limited by its adherence to “exchange” format:

- Custom parsing required for LaTeX-like syntax

Title	First half-life measurement of a low-lying isomer in ^{37}Si
Keywords	<ul style="list-style-type: none">• RADIOACTIVITY $^{37, 38}\text{Al}(\beta^-)$, (β^-n)[from $^{48}\text{Ca}(^9\text{Be}, X)$, $E=140$ MeV/ nucleon]; measured E_γ, I_γ, I_β, $\gamma\gamma$-coin, $\beta\gamma$-coin; deduced $T_{1/2}$. ^{37}Si; deduced levels, J, π, $T_{1/2}$ of isomeric states at 68- and 156-keV, $B(M1)$, matrix elements for low-lying ground-state M1 transitions. $^{36, 38}\text{Si}$; deduced E_γ, I_γ, transition intensities. Comparison to other experimental data and shell model calculations using the SDPF-MU, SDPF-U-SI, and FSU interactions. CeBr_3 based implantation detector with position-sensitive photomultiplier tube surrounded by 16 segmented Ge detectors (SeGa) and 15 LaBr_3 detectors at Coupled Cyclotron Facility (NSCL, MSU).

<https://www.nndc.bnl.gov/nsr/?search-type=keynumber&key=2023OG04>

Current State: Data Model

Limited by its adherence to “exchange” format:

- Custom parsing required for LaTeX-like syntax

```
<KEYWORDS>RADIOACTIVITY  $\{+37\}, \{+38\}Al(|b{\+-}) , (|b{\+-}n)$  [from  $\{+48\}Ca(\{+9\}Be, X)$ ,  $E=140$  MeV/nucleon]; measured  $E|g, I|g, I|b, |g|g\text{-coin}, |b|g\text{-coin}$ ; deduced  $T\{-1/2\}$ .  $\{+37\}Si$ ; deduced levels,  $J, |p, T\{-1/2\}$  of isomeric states at 68- and 156-keV,  $B(M1)$ , matrix elements for low-lying ground-state  $M1$  transitions.  $\{+36\}, \{+38\}Si$ ; deduced  $E|g, I|g$ , transition intensities. Comparison to other experimental data and shell model calculations using the SDPF-MU, SDPF-U-SI, and FSU interactions.  $CeBr\{-3\}$  based implantation detector with position-sensitive photomultiplier tube surrounded by 16 segmented Ge detectors (SeGa) and 15  $LaBr\{-3\}$  detectors at Coupled Cyclotron Facility (NSCL, MSU)
```

<https://www.nndc.bnl.gov/nsr/?search-type=keynumber&key=2023OG04>

Current State: Data Model

Limited by its adherence to “exchange” format:

- No way to uniquely identify author names
 - No space for ORCID (or other fields)

```
<AUTHORS >T.Redpath, P.Gueye,  
T.Baumann, B.A.Brown,  
A.Cunningham, P.A.DeYoung,  
N.Frank, C.R.Hoffman,  
A.N.Kuchera, B.Monteagudo  
Godoy, C.Persch, A.Revel,  
W.F.Rogers, M.Thoennessen,  
J.A.Tostevin, D.Votaw
```

```
MariaDB [nsr_test]> SHOW COLUMNS IN auth_both;  
+-----+-----+-----+-----+-----+-----+  
| Field | Type          | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+-----+  
| akey  | int(11)       | NO   | MUL | NULL    |       |  
| aname | varchar(32)   | NO   | MUL | NULL    |       |  
| a1    | char(1)       | NO   |     |         |       |  
| a2    | char(1)       | NO   |     |         |       |  
+-----+-----+-----+-----+-----+-----+  
4 rows in set (0.001 sec)
```

Current State: Database Hosting


Only exists as a MariaDB database instance


- Limited ability to track updates
 - *Who* made the last update?
 - *What* did they change?

At time of writing,
this information is
NOT being tracked



Aug 28, 2024

 Updated json/entries (again) after adding code to check for XUNDL datasets.
Benjamin Shu authored 2 weeks ago

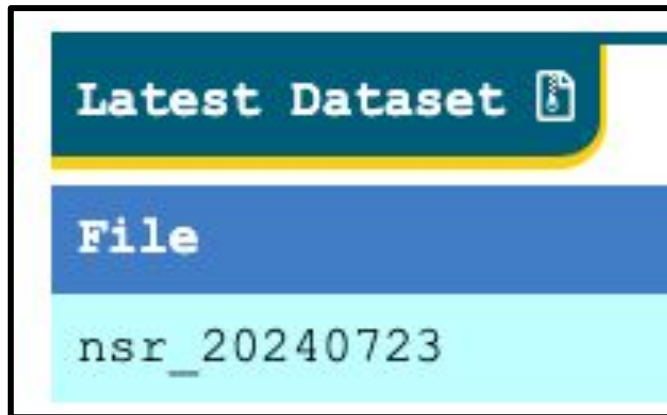
 Updated json/entries after adding code to store EXFOR keys.
Benjamin Shu authored 2 weeks ago

Changelog of updates to NSR

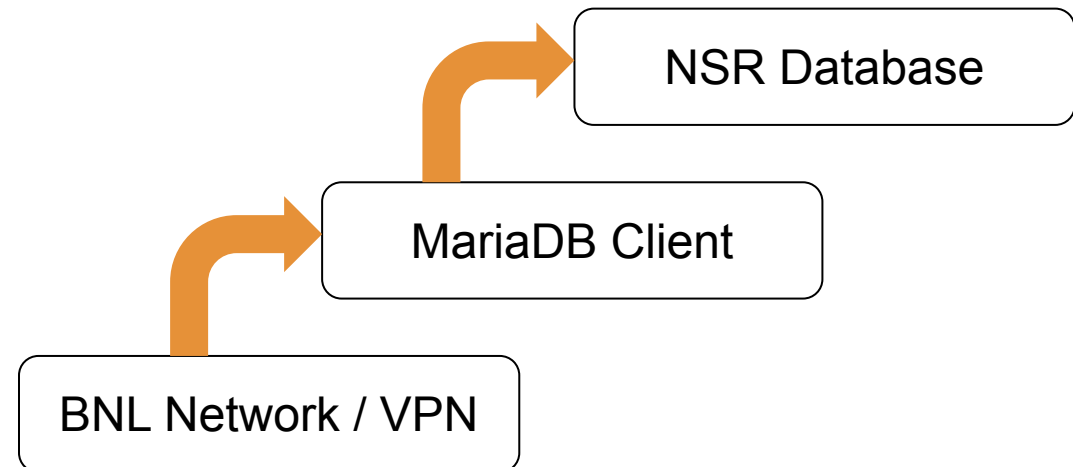
Current State: Database Hosting

Only exists as a MariaDB database instance

- No offline version for developers
 - Latest archive from 2024-07-23
 - Data files only accessible through SQL client or website



<https://www.nndc.bnl.gov/nsrarchivals/>



Current State: Database Hosting

Only exists as a MariaDB database instance

- Table schema cannot be changed easily

```
MariaDB [nsr_test]> SHOW TABLES;
+-----+
| Tables_in_nsr_test |
+-----+
| auth_alias          |
| auth_both           |
| auth_dict           |
| auth_tbl            |
| az_tbl              |
| coden_dict          |
| ensdf_link          |
| exch_tbl            |
| nuc_dict            |
| nuc_tbl             |
| part_dict           |
```

```
  reac_conv
  reac_dict
  reac_out_dict
  reac_tbl
  ref_tbl
  sub_conv
  sub_dict
  sub_tbl
  tempalias
  topic_tbl
  txt_tbl
+-----+
22 rows in set (0.000 sec)
```

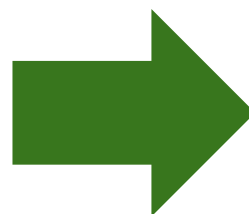
Step 0: JSON Data Model

[Working]

Based on existing JSON structure used in website

- Preserve text from original “exchange” tags:
 - <AUTHOR>, <TITLE>, <REFERENCE>, etc.

<KEYNO	>	2002RI06
<HISTORY	>	A20020708
<CODEN	>	JOUR NUPAB...



```
{
  "name": "KEYNO",
  "content": "2002RI06"
},
{
  "name": "HISTORY",
  "content": "A20020708"
},
{
  "name": "CODEN",
  "content": "JOUR NUPAB 701 156c"
},
}
```

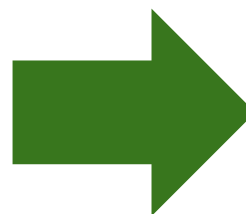
Step 0: JSON Data Model

[Working]

Based on existing JSON structure used in website

- Parse NSR free text as JSON key-value pairs

```
{
  "name": "KEYNO",
  "content": "2002RI06"
},
{
  "name": "HISTORY",
  "content": "A20020708"
},
{
  "name": "CODEN",
  "content": "JOUR NUPAB 701 156c"
},
}
```



```
"key_number": "2002RI06",
"publication_year": 2002,
"entry_type": "A",
"entry_date": "2002-07-08",
"modification_date": "2002-07-08",
"coden_text": "JOUR NUPAB 701 156c",
"reference_type": "JOURNAL",
"journal_code": "NUPAB",
"index_information": "701 156c",
```

Step 0: JSON Data Model

[Working]

Based on existing JSON structure used in website

- Add fields for data not stored or made explicit

```
{
  "type": "SINGLE_AUTHOR",
  "text": "M.V.Ricciardi",
  "initials": [
    "M",
    "V"
  ],
  "last_name": "Ricciardi"
},
```

```
"nuclide_selectors": [
  {
    "sub_type": "TARGET",
    "z": 1,
    "n": 0,
    "a": 1,
    "element_name": "Hydrogen",
    "element_symbol": "H"
  }
],
```

```
"is_primary": true,
"pdf_available": false,
"xundl_available": false,
"exfor_keys": []
```

Step 1: Preserving Existing Data

[Working]

Conversion of original exchange files into JSON

- **nsr-pipeline** project - reusable programs for:
 - Downloading exchange files from SQL database
 - Downloading additional information not stored in text files
 - Combining data and printing into new JSON schema

```
Collecting entries: 100% [=====] 249699/249699 (0:09:49 / 0:00:00)

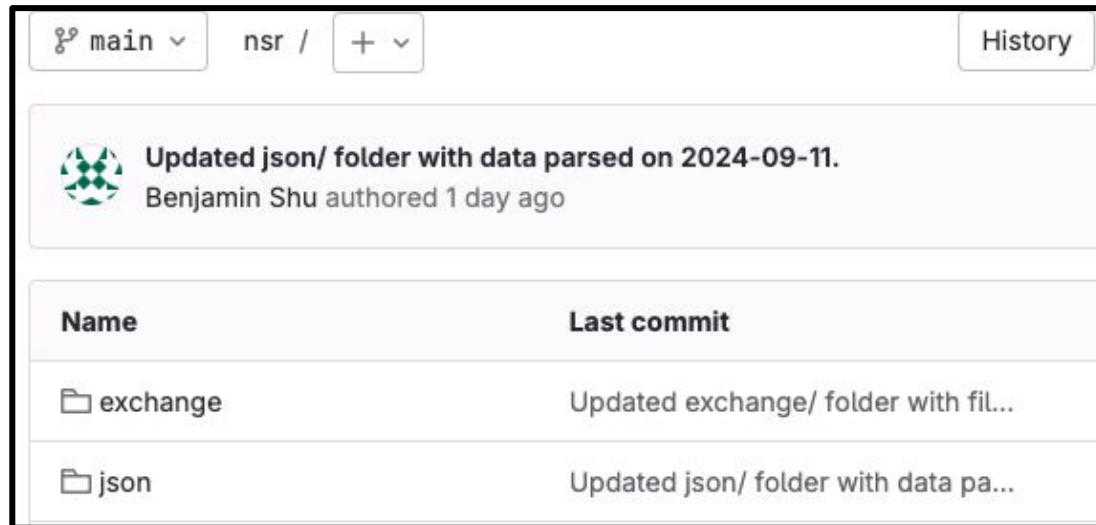
BUILD SUCCESSFUL in 10m 21s
20 actionable tasks: 8 executed, 12 up-to-date
[bshu@development3 combine-exchange-and-ref_tbl]$
```

Step 1: Preserving Existing Data

[Working]

Conversion of original exchange files into JSON

- Output files stored in GitLab under **databases/nsr**
 - Updating data no longer requires persistent connection
 - Possible public access? (if allowed by BNL)



The screenshot shows the GitLab interface for the 'nsr' repository. At the top, there are navigation elements: 'main' with a dropdown arrow, 'nsr /', a '+' button with a dropdown arrow, and a 'History' button. Below this is a commit message: 'Updated json/ folder with data parsed on 2024-09-11. Benjamin Shu authored 1 day ago'. Below the commit message is a table with two columns: 'Name' and 'Last commit'. The table lists two folders: 'exchange' and 'json', each with a corresponding commit message.

Name	Last commit
exchange	Updated exchange/ folder with fil...
json	Updated json/ folder with data pa...

```
databases/nsr
exchange/
...
json/
tags/
documents/
entries/
```

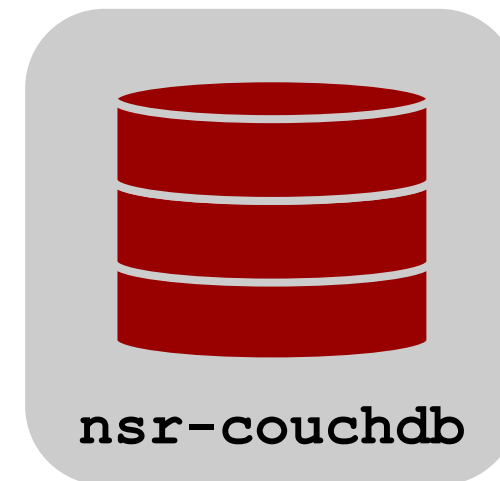
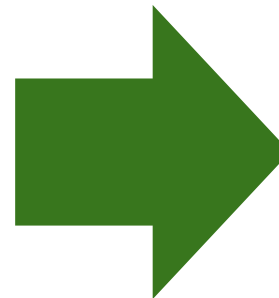
Step 2: Containerized Databases

[Working]

Use JSON files to build a CouchDB Docker image

- Build container with reusable Python scripts
 - Inserts JSON files and `.js` for database views
 - Can be restarted/rebuilt if necessary
 - Can be redeployed in case of server failure

```
databases/nsr/  
design_documents/  
Dockerfile  
load.sh
```

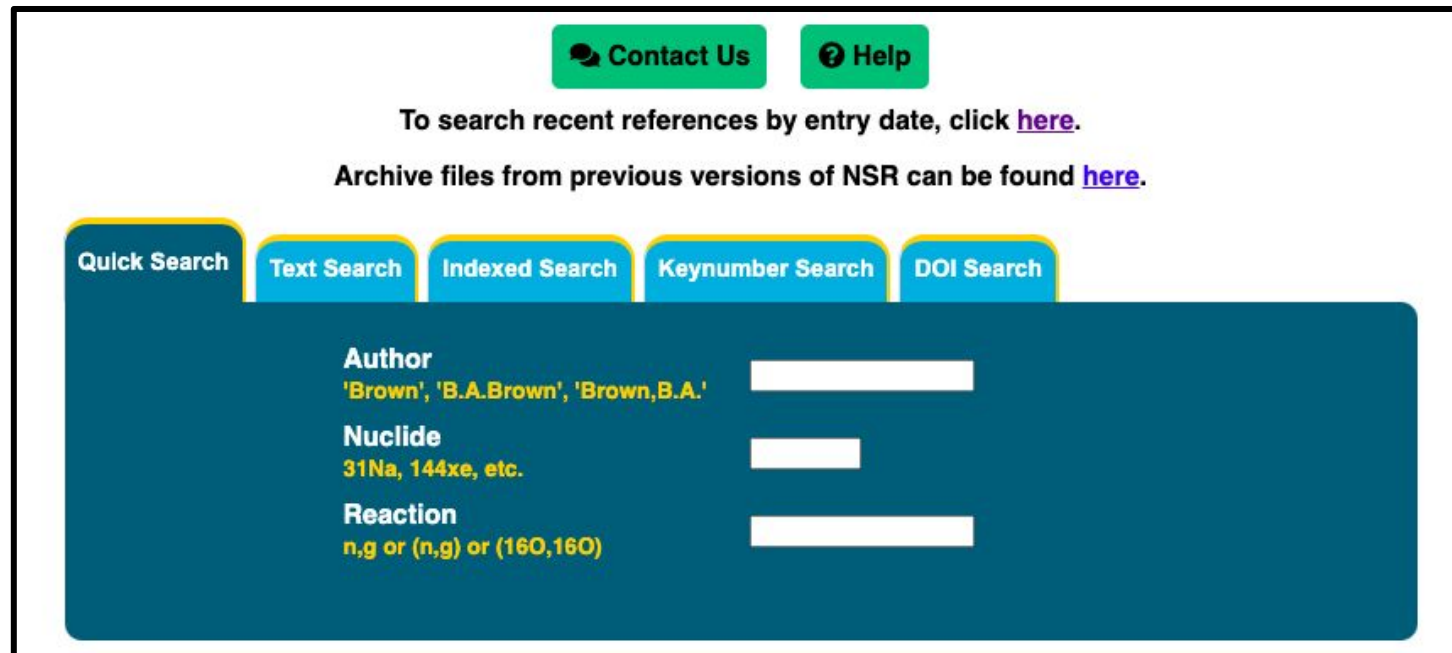


Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

- Current focus: preserving existing features



The screenshot shows a web application interface with a white background. At the top, there are two green buttons: "Contact Us" with a speech bubble icon and "Help" with a question mark icon. Below these buttons, there is a line of text: "To search recent references by entry date, click [here](#)." followed by another line: "Archive files from previous versions of NSR can be found [here](#)." Below the text, there are five blue buttons with white text: "Quick Search", "Text Search", "Indexed Search", "Keynumber Search", and "DOI Search". Below the buttons is a dark blue search form with three input fields. The first field is labeled "Author" and has the example text "'Brown', 'B.A.Brown', 'Brown,B.A.'" below it. The second field is labeled "Nuclide" and has the example text "31Na, 144xe, etc." below it. The third field is labeled "Reaction" and has the example text "n,g or (n,g) or (16O,16O)" below it. Each input field has a white rectangular box for text entry.

Step 3: Web Development

[In Progress]

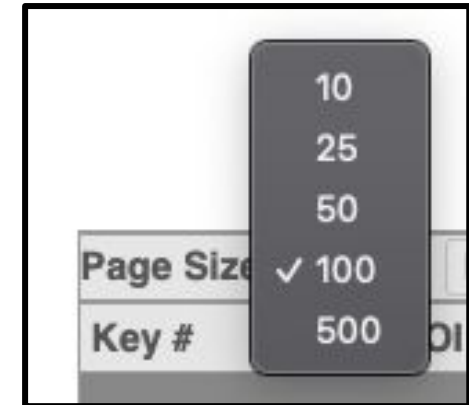
Update webapp to pull data from CouchDB image

- Latest updates:
 - Remote pagination for search results

• start-date: '1980-01-01'
• end-date: '2024-09-13'
• keyword-option: 'all'
• journal-option: 'all'
• source-option: 'all'
• page: '1'
• size: '100'

249699 results found!

Page Size	100	First	Prev	1	2	3	4	5	Next	Last
Key #	DOI + Links	Authors	Reference							
2024ZU01	10.1103/PhysRevC.109.014319	L.Zurek S.K Bogner	Phys.Rev. C 109, 014319 (2024)							



<https://www.nndc.bnl.gov/nsr/recent-references.html?start-date=1980-01-01&end-date=2024-09-13>

Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

- Latest updates:
 - Bookmarkable URLs & query parameters

1)

```
• search-type: 'quick'  
• author: 'b.a.brown'  
• nuclide: '31Na'  
• page: '1'  
• size: '100'  
  
5 results found!
```

2)

```
• search-type: 'quick'  
• reaction: 'n,g'  
• page: '1'  
• size: '100'  
  
6706 results found!
```

3)

```
• search-type: 'quick'  
• nuclide: '40Ca'  
• reaction: 'n,g'  
• page: '1'  
• size: '100'  
  
20 results found!
```

- 1) <https://www.nndc.bnl.gov/nsr/?search-type=quick&author=b.a.brown&nuclide=31Na>
- 2) <https://www.nndc.bnl.gov/nsr/?search-type=quick&reaction=n,g>
- 3) <https://www.nndc.bnl.gov/nsr/?search-type=quick&nuclide=40Ca&reaction=n,g>

Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

- Latest updates:
 - “Browse” pages no longer open in new tab

Browse Nuclides

Search By Z (Required)

Element:
Gallium ▾

Proton # (Z):
[] → []

Proton # (Z) ▲	Neutron # (N) ▲	Atomic Ma... ▲	Element ▲
31			Gallium
31	23	54	Gallium
31	24	55	Gallium
31	25	56	Gallium
31	26	57	Gallium
31	27	58	Gallium
31	28	59	Gallium
31	29	60	Gallium
31	30	61	Gallium
31	31	62	Gallium

Page Size 10 ▾ First Prev 1 2 3 4 5 Next Last

Step 4: Distribution, Collaboration

[In Progress]

Goal: Make NSR as accessible as possible

- Changes have already enabled collaboration with INSPIRE
 - “INSPIRE now links to over 100,000 records in the NSR.”



<https://blog.inspirehep.net/2024/09/inspire-and-the-nuclear-science-references-database/>

Step 4: Distribution, Collaboration

[In Progress]

Goal: Make NSR as accessible as possible

- Possible future data model updates
 - ORCID's for authors?

```
{  
  "type": "SINGLE_AUTHOR",  
  "text": "S.F.Mughabghab",  
  "initials": ["S", "F"],  
  "last_name": "Mughabghab",  
  "orcid": "INSERT ORCID HERE"  
}
```

The JSON schema makes it easy to add this field.

Finding and filling in ORCID's will be much harder.

Step 4: Distribution, Collaboration

[In Progress]

Goal: Make NSR as accessible as possible

- Possible future software
 - JSON editor for entries?
 - Offline versions of code tables for nuclides, subjects, reactions, etc.
- Contact us:
<https://www.nndc.bnl.gov/contact-us/>
- Email: bshu@bnl.gov

End Note: Crawlers/Bots

~37% of retrievals from FY2024 came from 11 IP addresses

- Most identified as bots belonging to Amazon/Google
- Current NSR uses Google Analytics, which filters out known crawlers

