Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Modernizing Workflow for NSR Updates

Benjamin Shu

October 2024

@BrookhavenLab

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...
<REFRENCE>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:

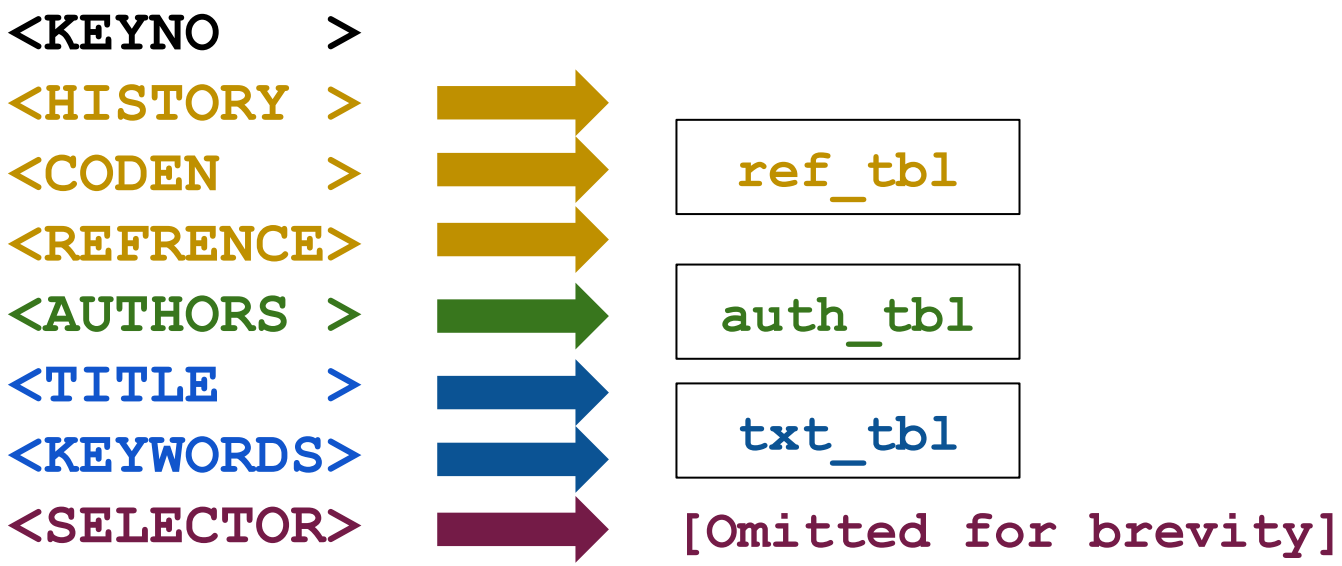
- Entries need to be pre-processed before use in code

```
<KEYNO   >    2002RI06                 Unique ID
<HISTORY >    A20020708                Entry/modification dates
<CODEN   >    JOUR NUPAB...            Source type, journal
<REFRENCE>    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	• RADIOACTIVITY ^{37,38}Al(β^-), (β^-n)[from ^{48}Ca(^{9}Be, X), E=140 MeV/ nucleon]; measured Eγ, Iγ, Iβ, γγ-coin, βγ-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}Si; deduced Eγ, Iγ, 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-coin, |b|g-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}S
i; deduced E|g, I|g, transition intensities. Comparison to other experimental da
ta and shell model calculations using the SDPF-MU, SDPF-U-SI, and FSU interactio
ns. CeBr{-3} based implantation detector with position-sensitive photomultiplier
 tube surrounded by 16 segmented Ge detectors (SeGa) and 15 LaBr{-3} detectors a
t Coupled Cyclotron Facility (NSCL, MSU)
```

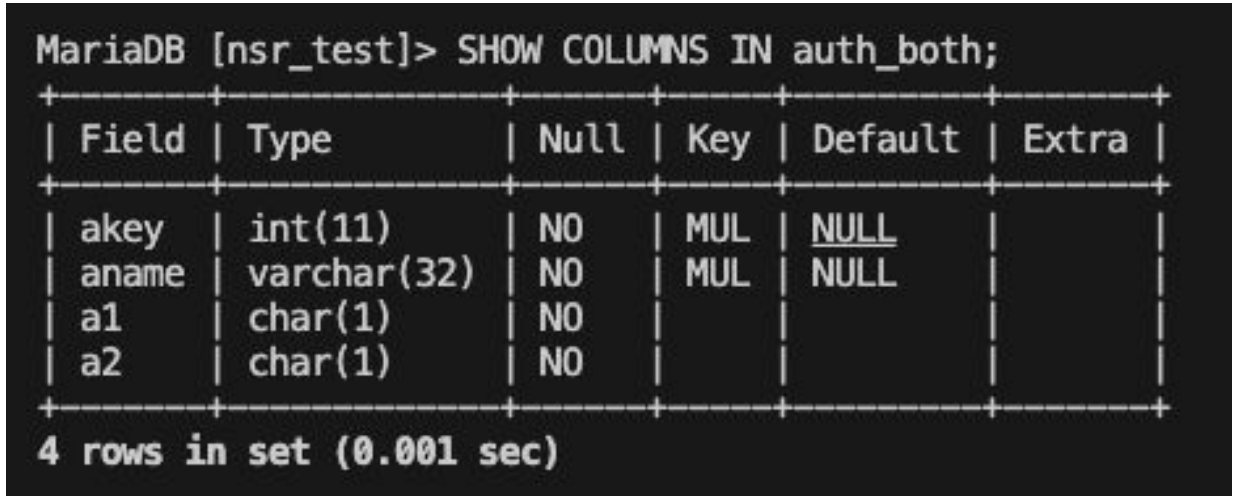
https://www.nndc.bnl.gov/nsr/?search-type=keynumber&key=2023OG04

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
```



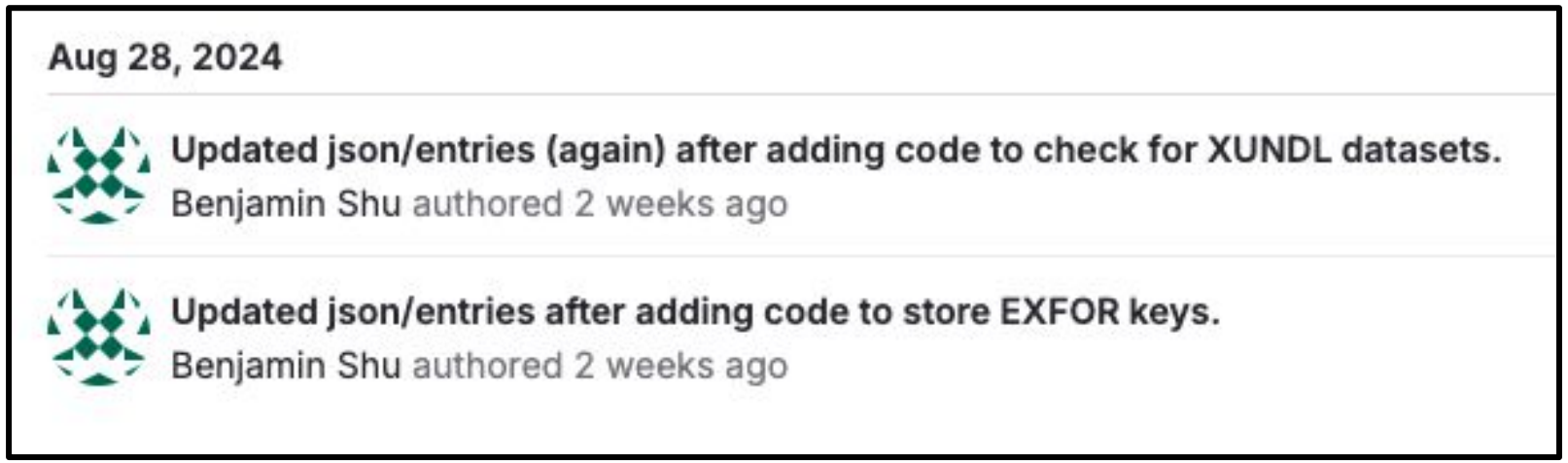
```
MariaDB [nsr_test]> SHOW COLUMNS IN auth_both;
+--------+--------------+------+-----+---------+-------+
| Field  | Type         | Null | Key | Default | Extra |
+--------+--------------+------+-----+---------+-------+
| akey   | int(11)      | NO   | MUL | NULL    |       |
| aname  | varchar(32)  | 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

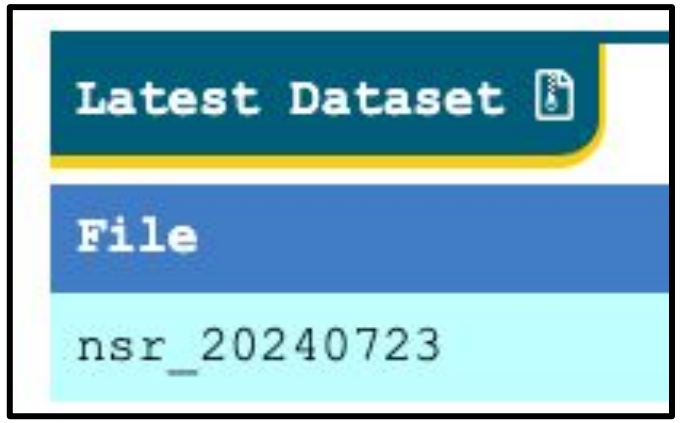


Changelog of updates to NSR

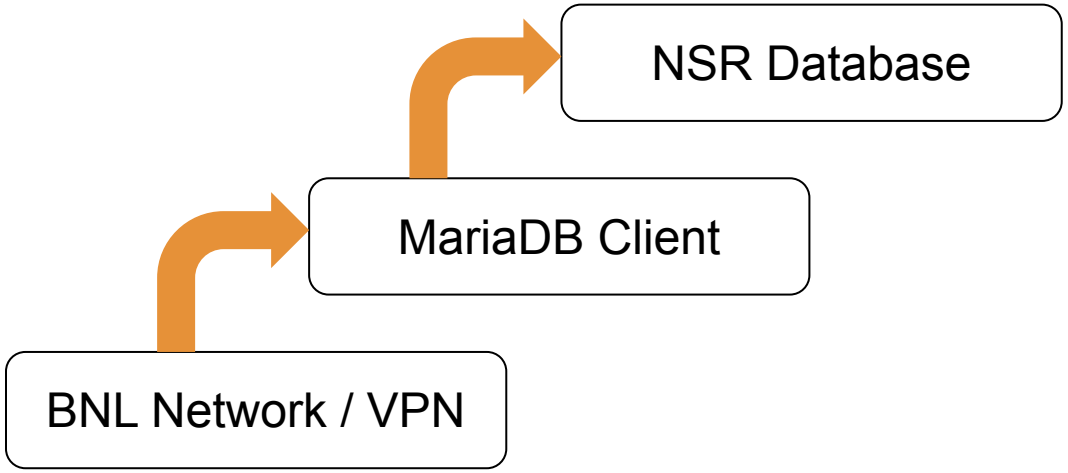
Current State: Database Hosting

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)
```

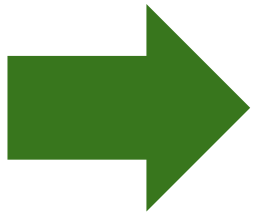
Step 0: JSON Data Model

[Working]

Based on existing JSON structure used in website

- Preserve text from original “exchange” tags:
  - <AUTHOR>, <TITLE>, <REFRENCE>, 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

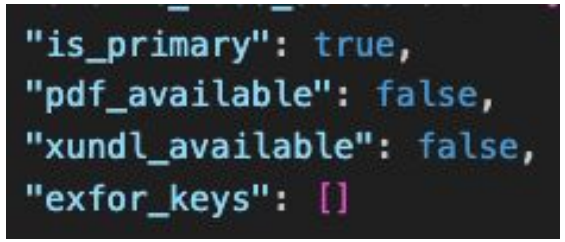
- Parse NSR free text as JSON key-value pairs

Step 0: JSON Data Model

[Working]

Based on existing JSON structure used in website

- Add fields for data not stored or made explicit

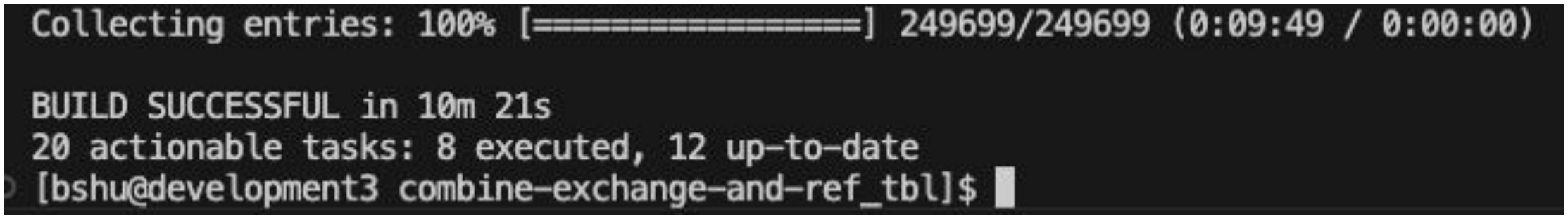


Step 1: Preserving Existing Data

[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

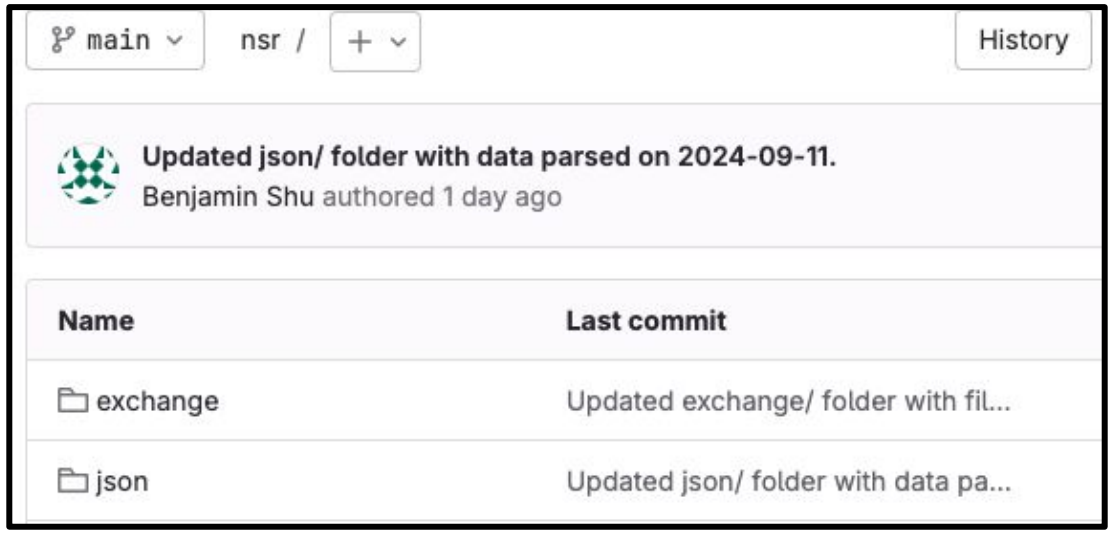


Step 1: Preserving Existing Data

[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)




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

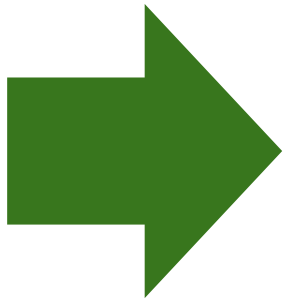
Step 2: Containerized Databases

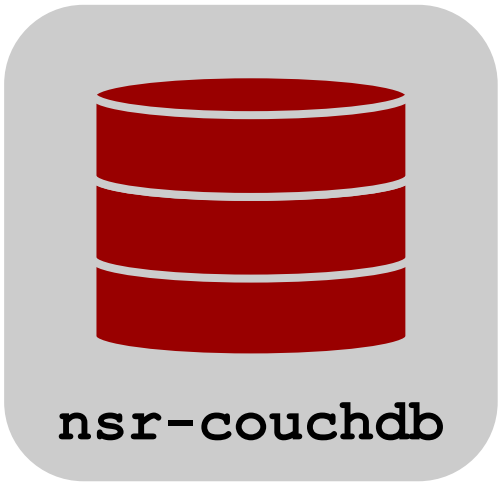
[Working]

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
```



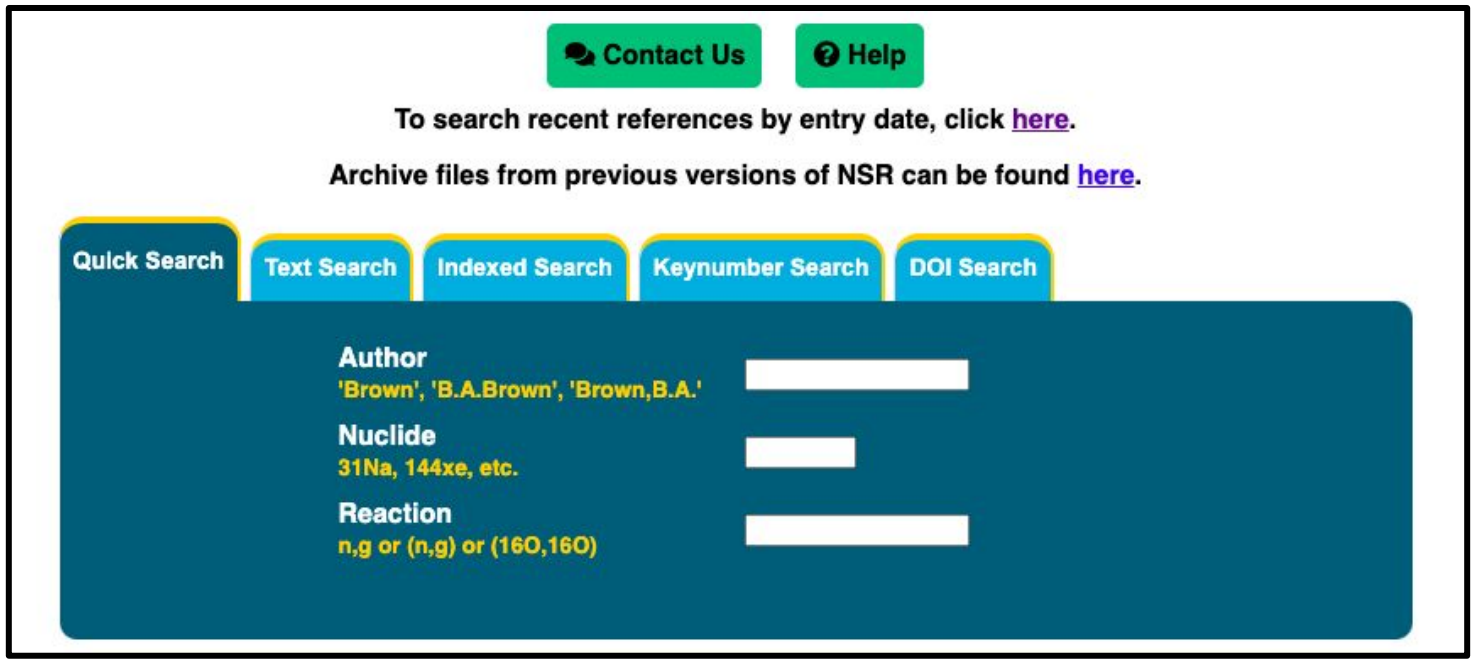


Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

- Current focus: preserving existing features

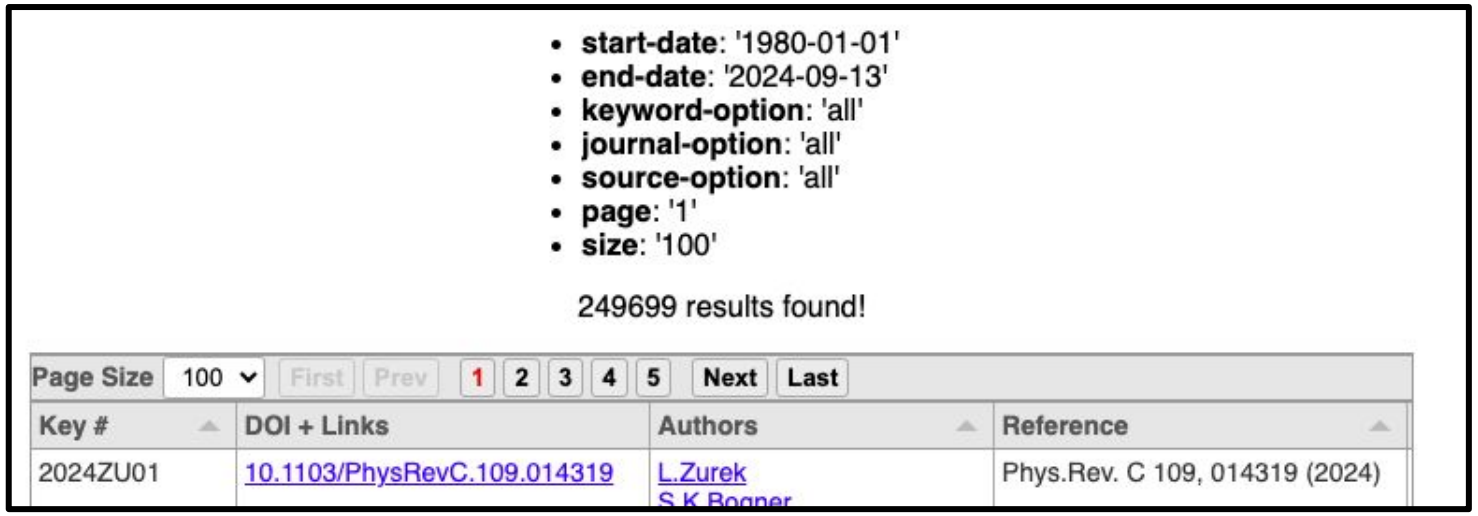


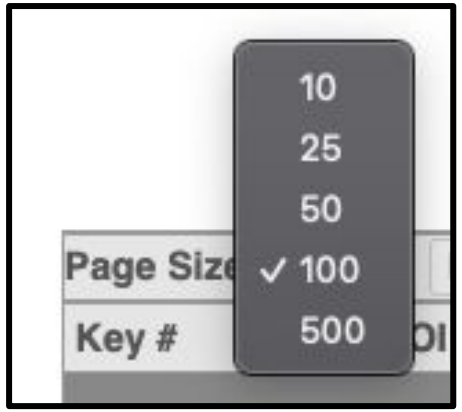
Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

- Latest updates:
  - Remote pagination for search results





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!

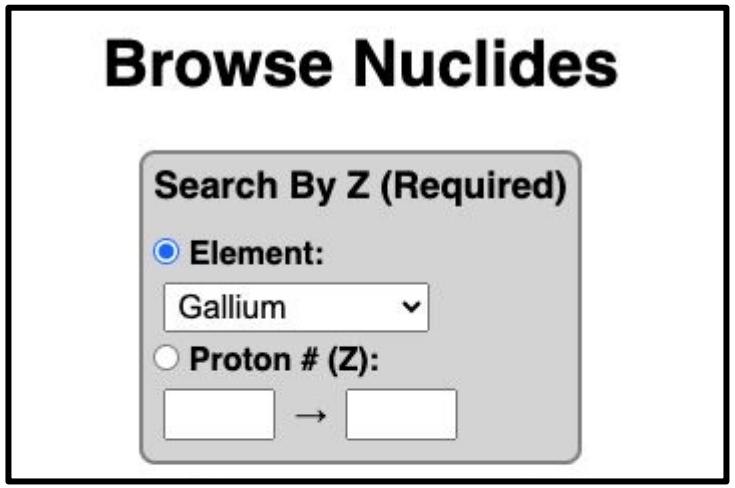
1) https://www.nndc.bnl.gov/nsr/?search-type=quick&author=b.a.brown&nuclide=31Na
2) https://www.nndc.bnl.gov/nsr/?search-type=quick&reaction=n,g
3) https://www.nndc.bnl.gov/nsr/?search-type=quick&nuclide=40Ca&reaction=n,g

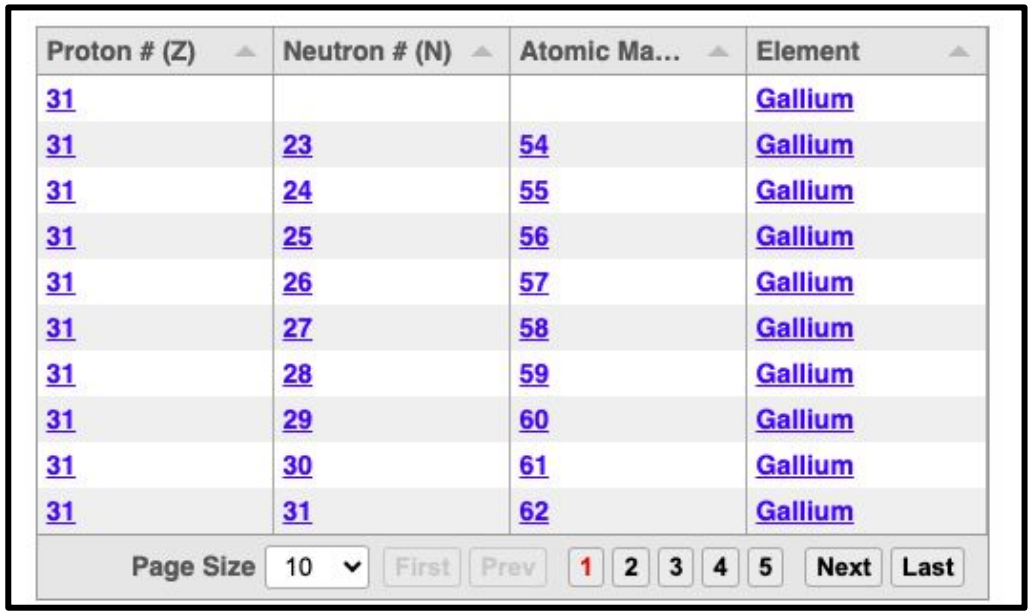
Step 3: Web Development

[In Progress]

Update webapp to pull data from CouchDB image

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





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

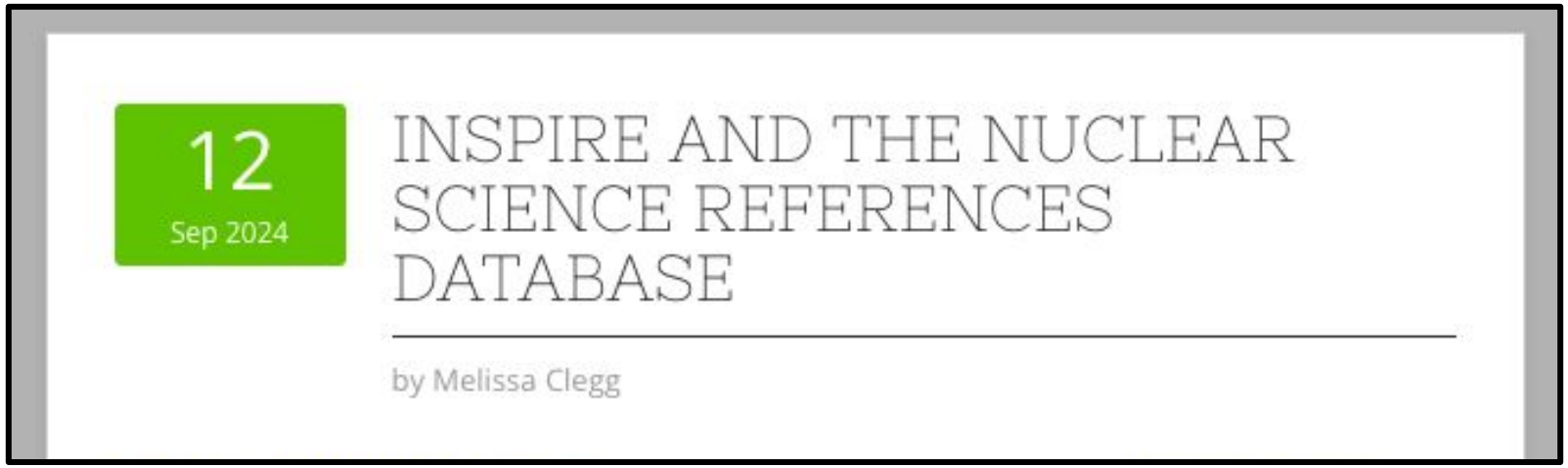
Page Size 10 First Prev 1 2 3 4 5 Next Last

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
  - ORCIDs 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 ORCIDs will be much harder.

Step 4: Distribution, Collaboration

[In Progress]

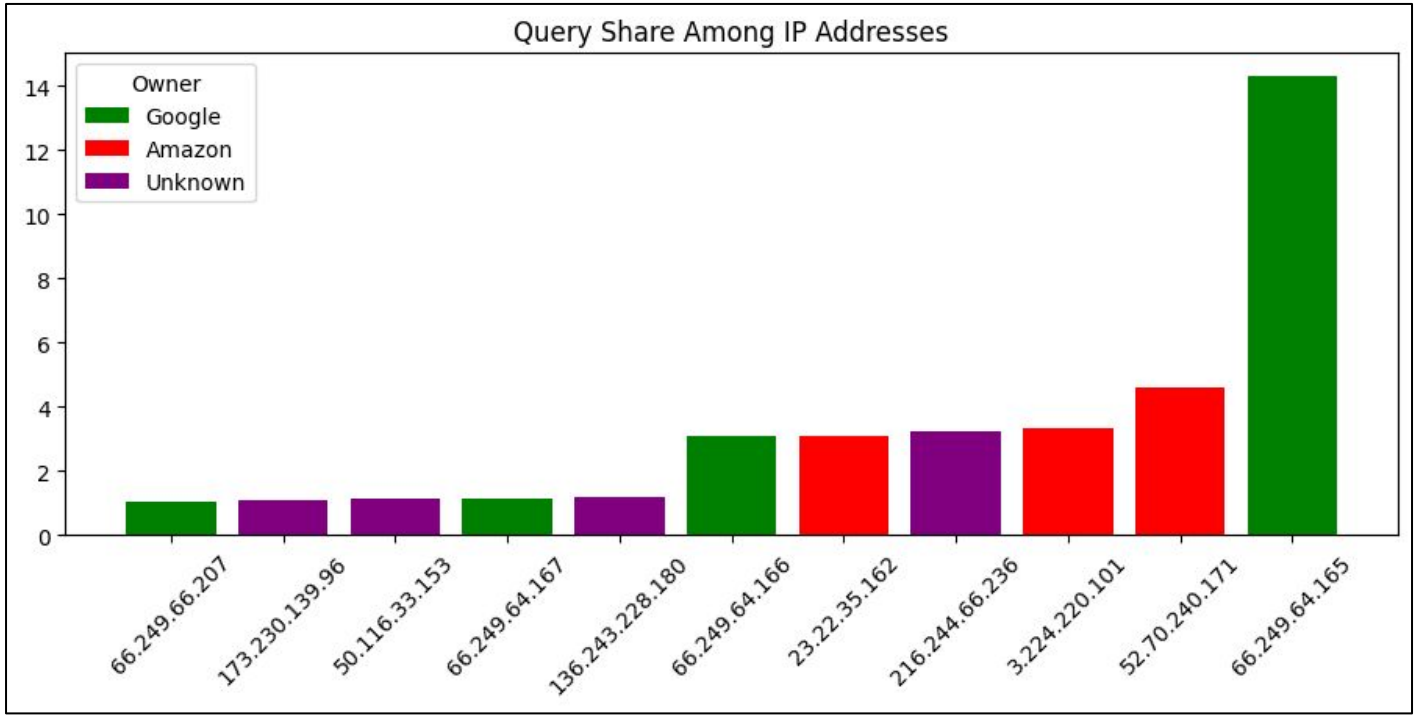
Goal: Make NSR as accessible as possible

- 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