Status of the ADVANCE CI/CD

David Brown, Ramon Arcilla, Rebecca Coles, Gustavo Nobre
National Nuclear Data Center

23 February 2023
NCSP TPR, Albuquerque, NM

@BrookhavenLab

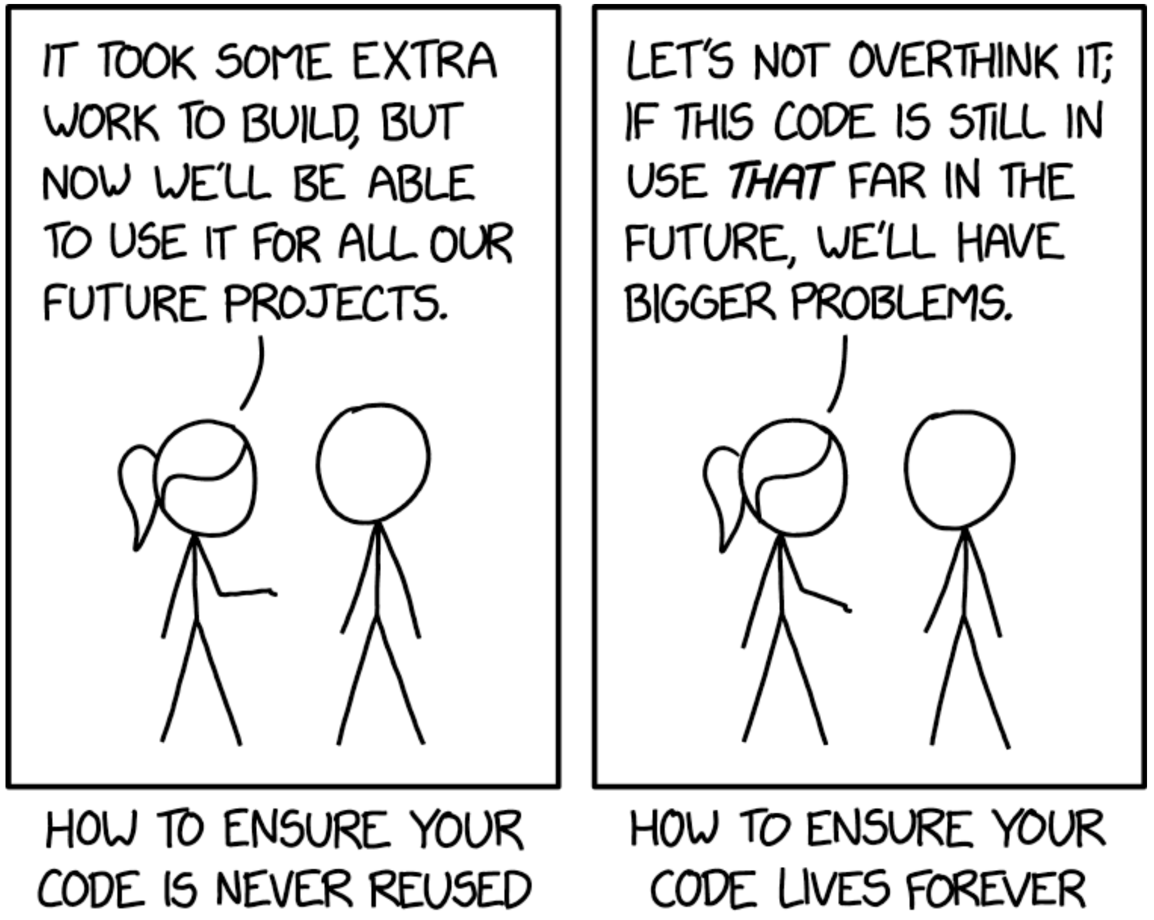
ADVANCE is the ENDF continuous integration system

- Stood up as quick-n-dirty hack to get ENDF/B-VII.1 out the door

  … in 2011

- Faithfully (if at times erratically) serving CSEWG since

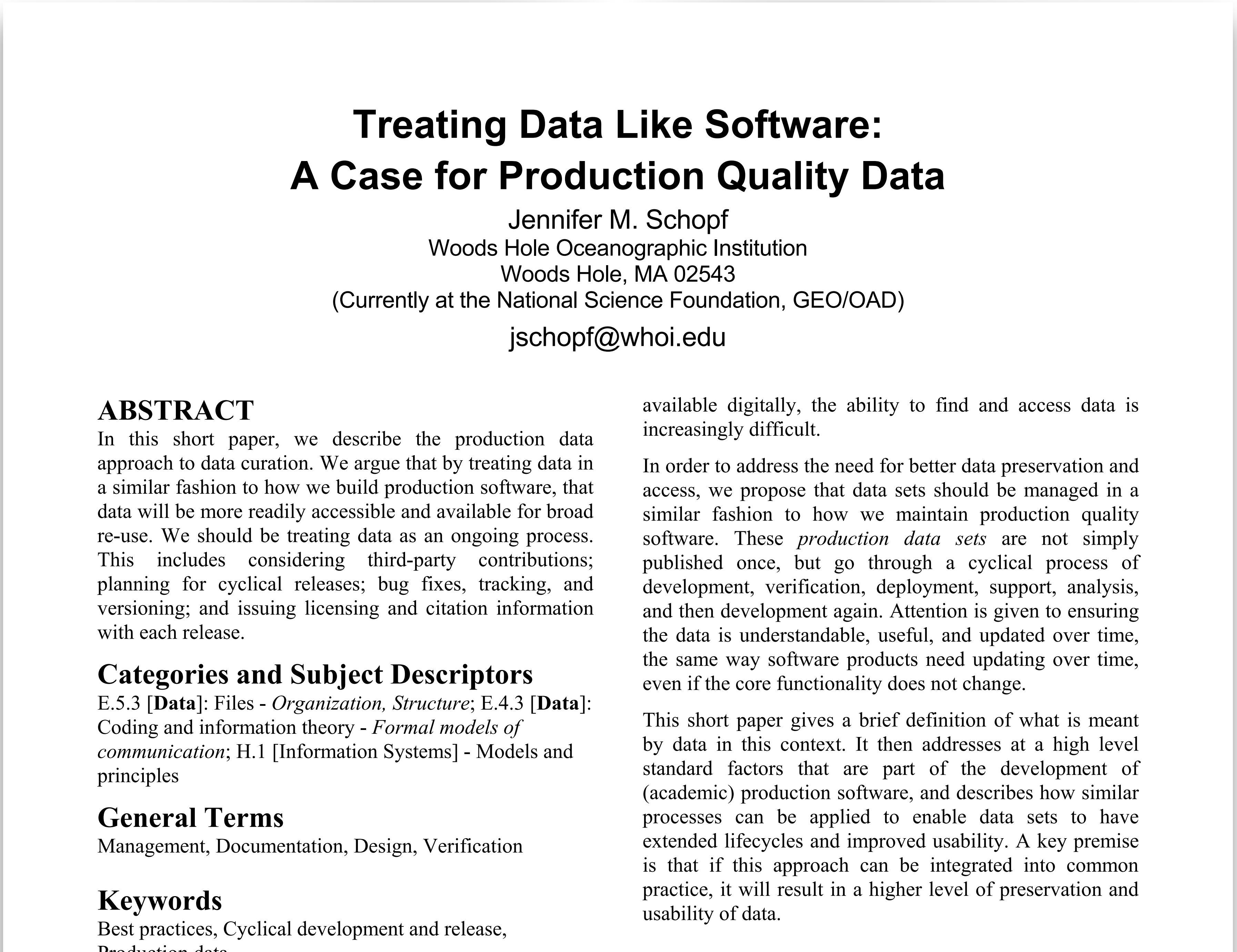
- All ENDF files ran through variety of physics & format checks as well as processing codes


IT TOOK SOME EXTRA WORK TO BUILD, BUT NOW WE'LL BE ABLE TO USE IT FOR ALL OUR FUTURE PROJECTS.
HOW TO ENSURE YOUR CODE IS NEVER REUSED
LET'S NOT OVERTHINK IT; IF THIS CODE IS STILL IN USE THAT FAR IN THE FUTURE, WE'LL HAVE BIGGER PROBLEMS.
HOW TO ENSURE YOUR CODE LIVES FOREVER

https://xkcd.com/2730

Brookhaven National Laboratory

Jennifer M. Schopf. 2012. Treating data like software: a case for production quality data. In Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (JCDL '12). Association for Computing Machinery, New York, NY, USA, 153–156. DOI:https://doi.org/10.1145/2232817.2232846



Treating Data Like Software: A Case for Production Quality Data

Jennifer M. Schopf
Woods Hole Oceanographic Institution
Woods Hole, MA 02543
(Currently at the National Science Foundation, GEO/OAD)
jschopf@whoi.edu

ABSTRACT

In this short paper, we describe the production data approach to data curation. We argue that by treating data in a similar fashion to how we build production software, that data will be more readily accessible and available for broad re-use. We should be treating data as an ongoing process. This includes considering third-party contributions; planning for cyclical releases; bug fixes, tracking, and versioning; and issuing licensing and citation information with each release.

Categories and Subject Descriptors

E.5.3 [**Data**]: Files - *Organization, Structure*; E.4.3 [**Data**]: Coding and information theory - *Formal models of communication*; H.1 [Information Systems] - Models and principles

General Terms

Management, Documentation, Design, Verification

Keywords

Best practices, Cyclical development and release, Production data

available digitally, the ability to find and access data is increasingly difficult.

In order to address the need for better data preservation and access, we propose that data sets should be managed in a similar fashion to how we maintain production quality software. These *production data sets* are not simply published once, but go through a cyclical process of development, verification, deployment, support, analysis, and then development again. Attention is given to ensuring the data is understandable, useful, and updated over time, the same way software products need updating over time, even if the core functionality does not change.

This short paper gives a brief definition of what is meant by data in this context. It then addresses at a high level standard factors that are part of the development of (academic) production software, and describes how similar processes can be applied to enable data sets to have extended lifecycles and improved usability. A key premise is that if this approach can be integrated into common practice, it will result in a higher level of preservation and usability of data.

Phase I testing automated for nearly 10 years

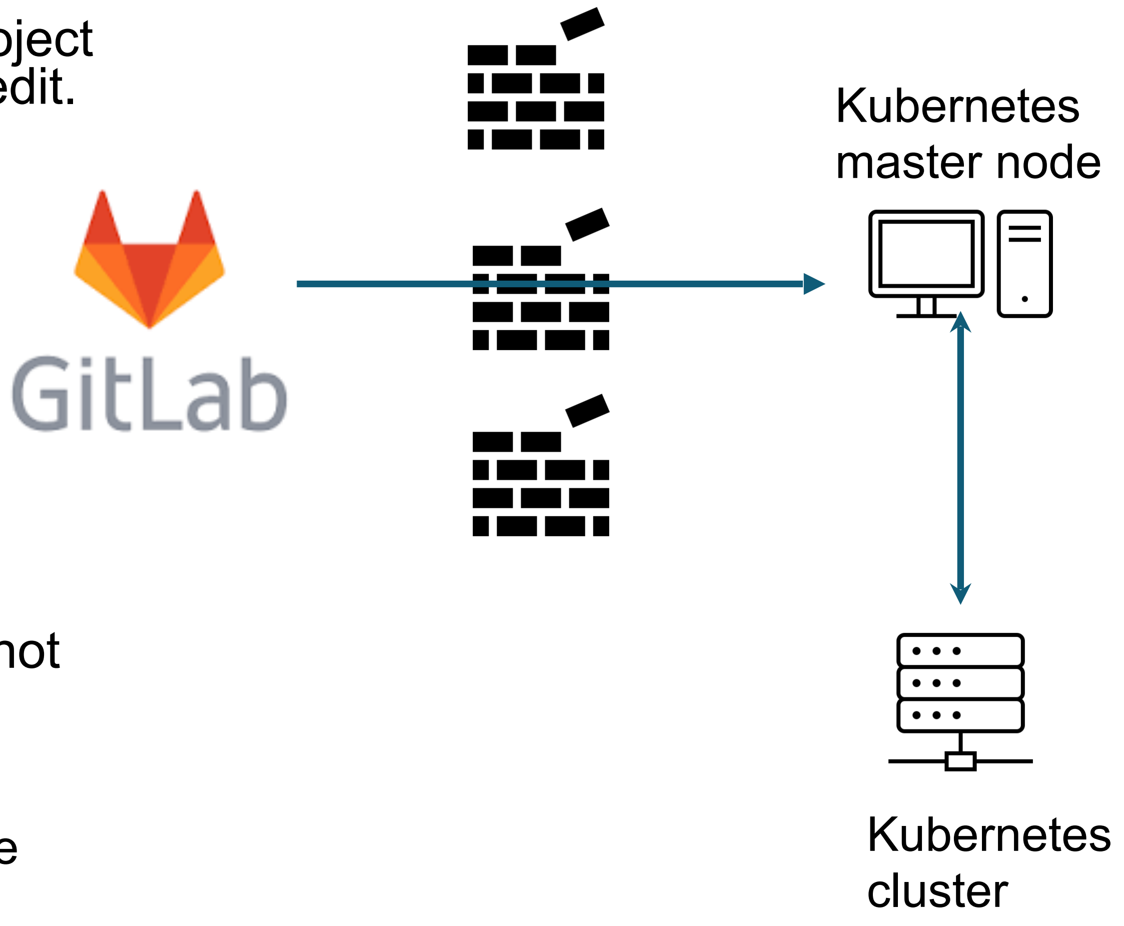
Can we automate reading build reports? What other tests can we automate?

Code	Test	pre-VII	Now
	File summary complete & correct	🤔	
STAN, STANEF, CHECKR, fudge	**ENDF format compliance**	🖥️	🖥️
FIZCON, fudge	**Mathematical correctness** (e.g. probabilities valid, covariances positive)	🖥️	🖥️
FIZCON, PSYCHE, fudge	**Physical correctness** (e.g. Q, thresholds, energy deposition/KERMA)	🖥️	🖥️
INTER, fudge (inter.py)	**Compute & check integral metrics** (e.g. RI, thermal cross sections, MACS)	🖥️	🖥️
fudge	**Completeness** (all outgoing particles, including gammas)	🤔	🖥️
ADVANCE	**Comparisons to microscopic experimental data** (EXFOR)	🤔	🖥️
	Assessment of application suitability (e.g. usable for fast reactors or spaceflight)	🤔	
	Reasonable (e.g. covariances, angular distributions)	🤔	
NEW fudge (grokres.py)	**Resonance quality** (missing resonances? widths realistic?)		🖥️
PREPRO, fudge, NJOY (not SCALE yet)	**Can process for user codes**		🖥️
	Is state of the art? Is best we can do?	🤔	

Brookhaven National Laboratory

Major reworking of ADVANCE innards

- Gitlab's Ci/CD configuration controlled by per-project YAML file anyone who is part of the project can edit.
- To prevent accidental (or intentional?) troubles, must execute checking codes in container
- Gitlab, the Kubernetes master node (development2) and the Kubernetes worker node (ADVANCE2 server) are all behind the BNL FireWall
- Reverse proxy server (outside the BNL FireWall) secures access from the Internet to GitLab.
- Two configuration requirements Gitlab.com had not anticipated:
    - Our cluster must live behind firewall
    - Legacy checking codes need temp space
- Gitlab and BNL/ITD cyber security engineers worked with us to resolve issues.



Brookhaven National Laboratory

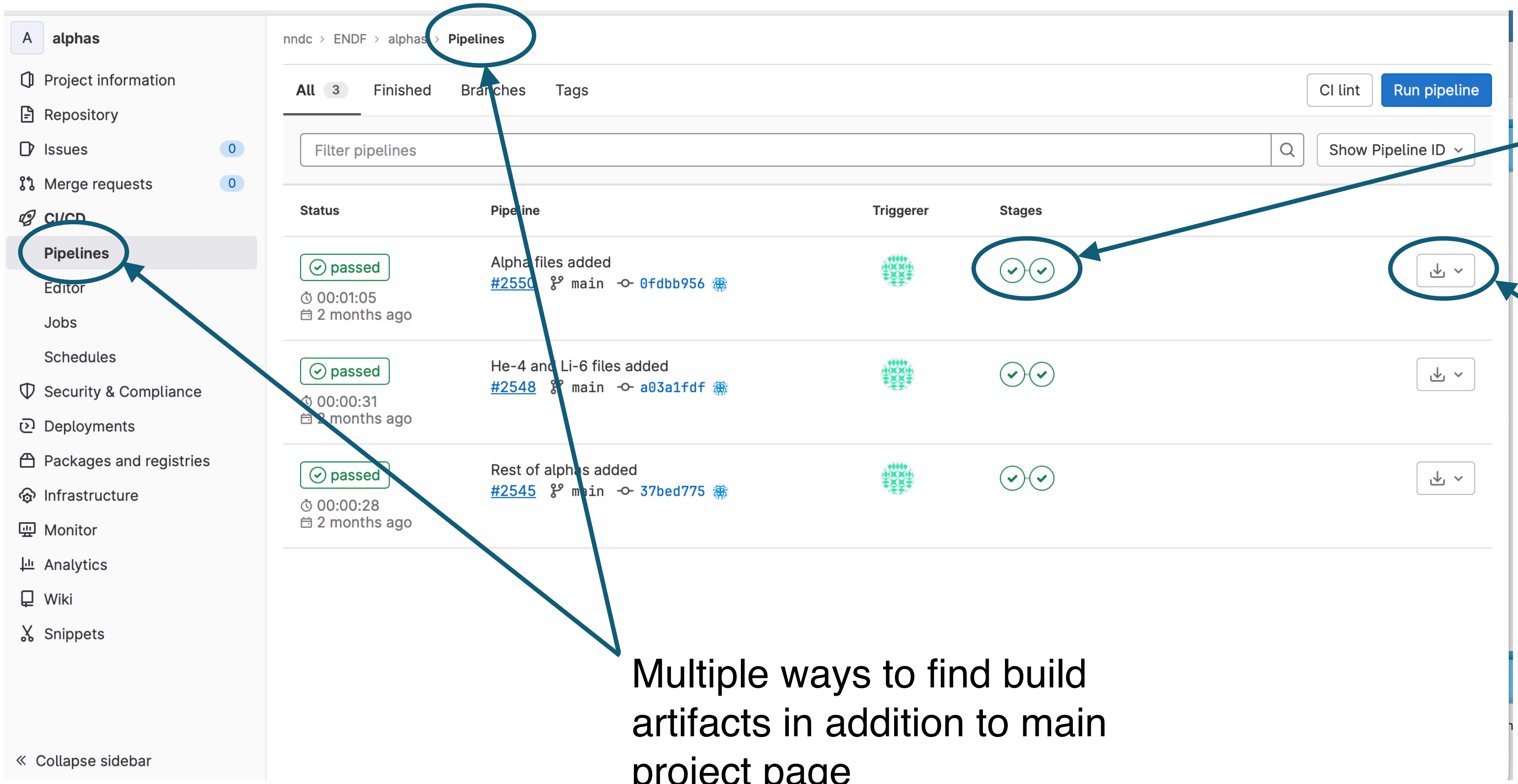
Simpler reports

- Reports will be per-commit, on any git branch (but probably restricted to review branches for now)
- They must be light weight, but not compromise content
- Solution:
  - Summary markdown, per code, replacing website
  - Any important build artifacts made by code (xsdir, ace files, …)
  - Pictures & lists of bugs in summary markdown

Thanks to:

- Ramon Arcilla (BNL-NNDC) for fighting the fight with GitLab and
- Rebecca Coles (BNL-NNP) for creating the simplified reports

Executes on every commit on every branch



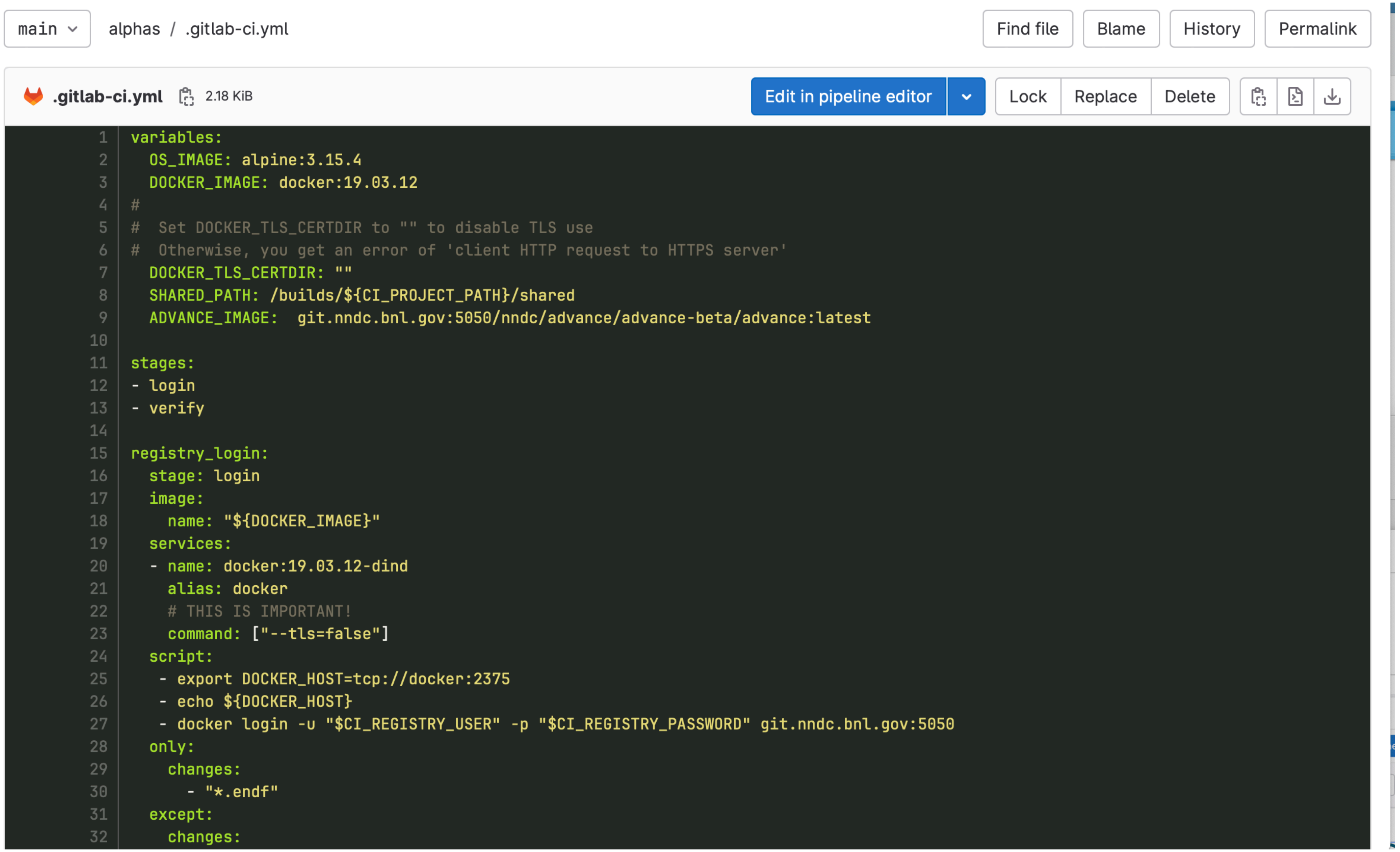
Status	Pipeline	Triggerer	Stages
passed 00:01:05 2 months ago	Alpha files added #2550 main 0fdbb956		✓ ✓
passed 00:00:31 2 months ago	He-4 and Li-6 files added #2548 main a03a1fdf		✓ ✓
passed 00:00:28 2 months ago	Rest of alphas added #2545 main 37bed775		✓ ✓

It passed!

artifacts.zip

Multiple ways to find build artifacts in addition to main project page

Execution controlled by a YAML file



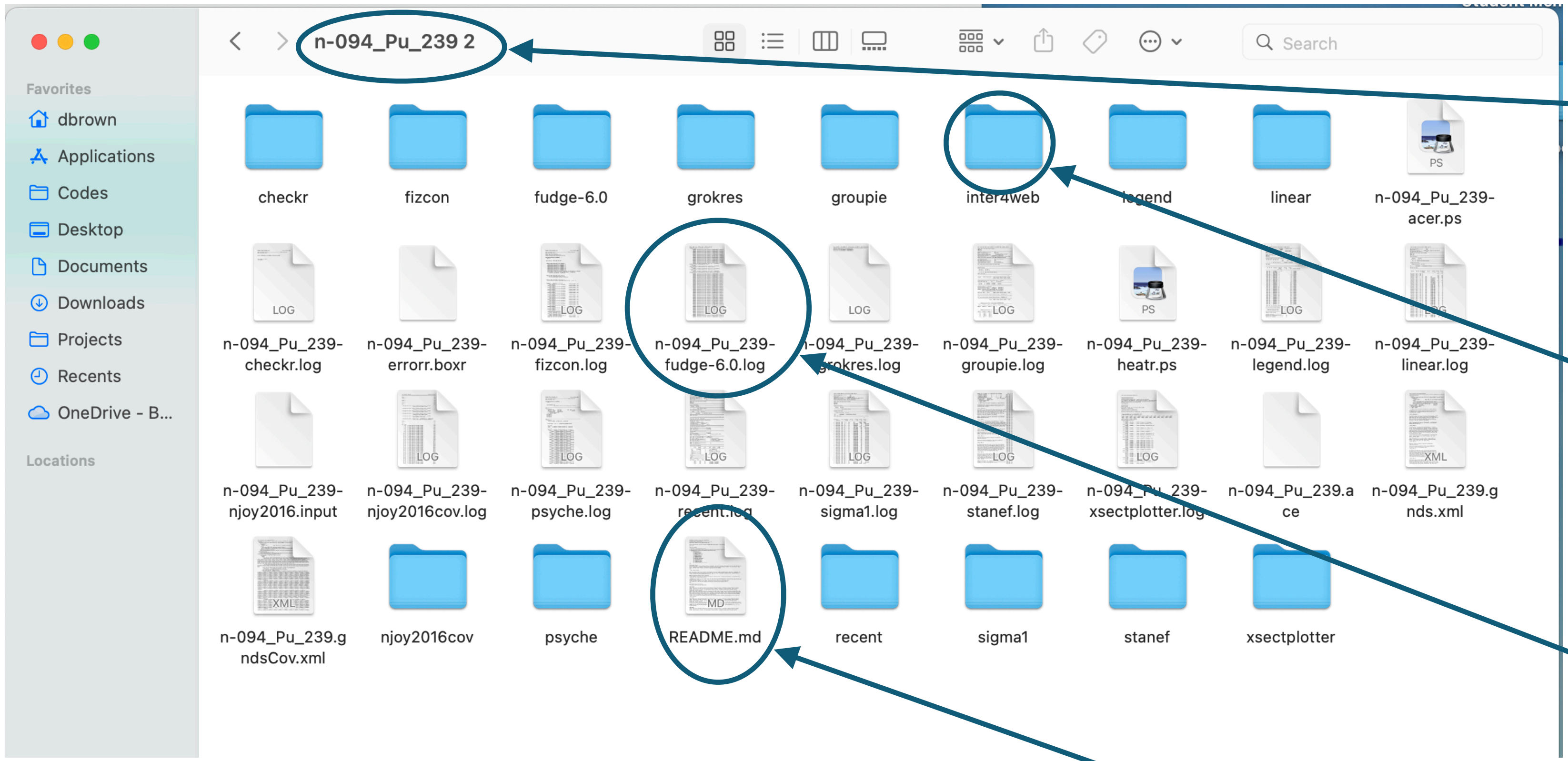
```yaml
variables:
  OS_IMAGE: alpine:3.15.4
  DOCKER_IMAGE: docker:19.03.12
#
#  Set DOCKER_TLS_CERTDIR to "" to disable TLS use
#  Otherwise, you get an error of 'client HTTP request to HTTPS server'
  DOCKER_TLS_CERTDIR: ""
  SHARED_PATH: /builds/${CI_PROJECT_PATH}/shared
  ADVANCE_IMAGE:  git.nndc.bnl.gov:5050/nndc/advance/advance-beta/advance:latest

stages:
- login
- verify

registry_login:
  stage: login
  image:
    name: "${DOCKER_IMAGE}"
  services:
  - name: docker:19.03.12-dind
    alias: docker
    # THIS IS IMPORTANT!
    command: ["--tls=false"]
  script:
   - export DOCKER_HOST=tcp://docker:2375
   - echo ${DOCKER_HOST}
   - docker login -u "$CI_REGISTRY_USER" -p "$CI_REGISTRY_PASSWORD" git.nndc.bnl.gov:5050
  only:
    changes:
      - "*.endf"
  except:
    changes:
```

Brool
National

An unpacked artifacts.zip file



checkr
fizcon
fudge-6.0
grokres
groupie
inter4web
legend
linear
n-094_Pu_239-acer.ps
n-094_Pu_239-checkr.log
n-094_Pu_239-errorr.boxr
n-094_Pu_239-fizcon.log
n-094_Pu_239-fudge-6.0.log
n-094_Pu_239-grokres.log
n-094_Pu_239-groupie.log
n-094_Pu_239-heatr.ps
n-094_Pu_239-legend.log
n-094_Pu_239-linear.log
n-094_Pu_239-njoy2016.input
n-094_Pu_239-njoy2016cov.log
n-094_Pu_239-psyche.log
n-094_Pu_239-recent.log
n-094_Pu_239-sigma1.log
n-094_Pu_239-stanef.log
n-094_Pu_239-xsectplotter.log
n-094_Pu_239.ace
n-094_Pu_239.gnds.xml
n-094_Pu_239.gndsCov.xml
njoy2016cov
psyche
README.md
recent
sigma1
stanef
xsectplotter

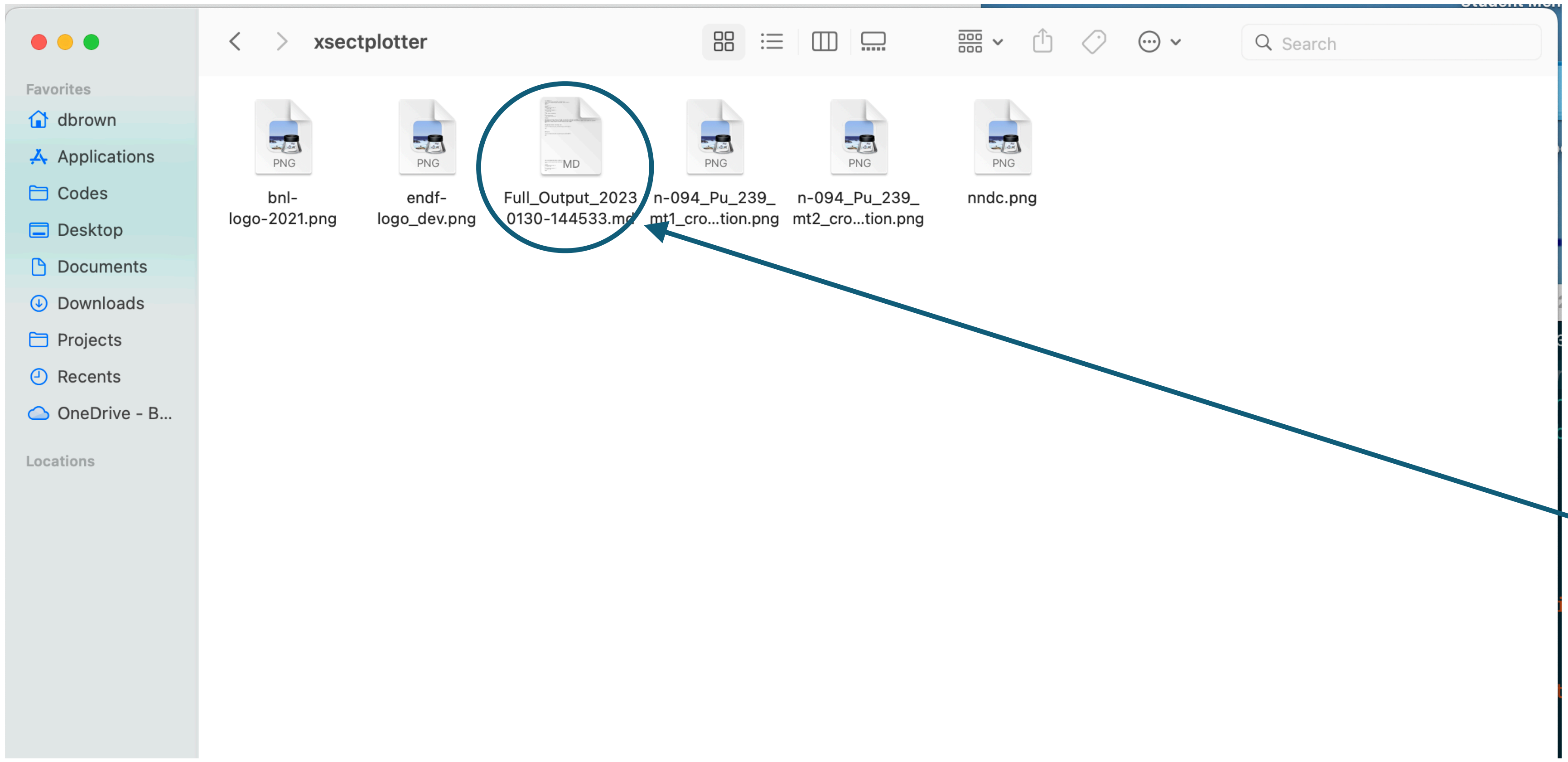
Each changed evaluation gets its own directory

Full build artifact for each code in own directory

stdout stream from code

Helpful instructions

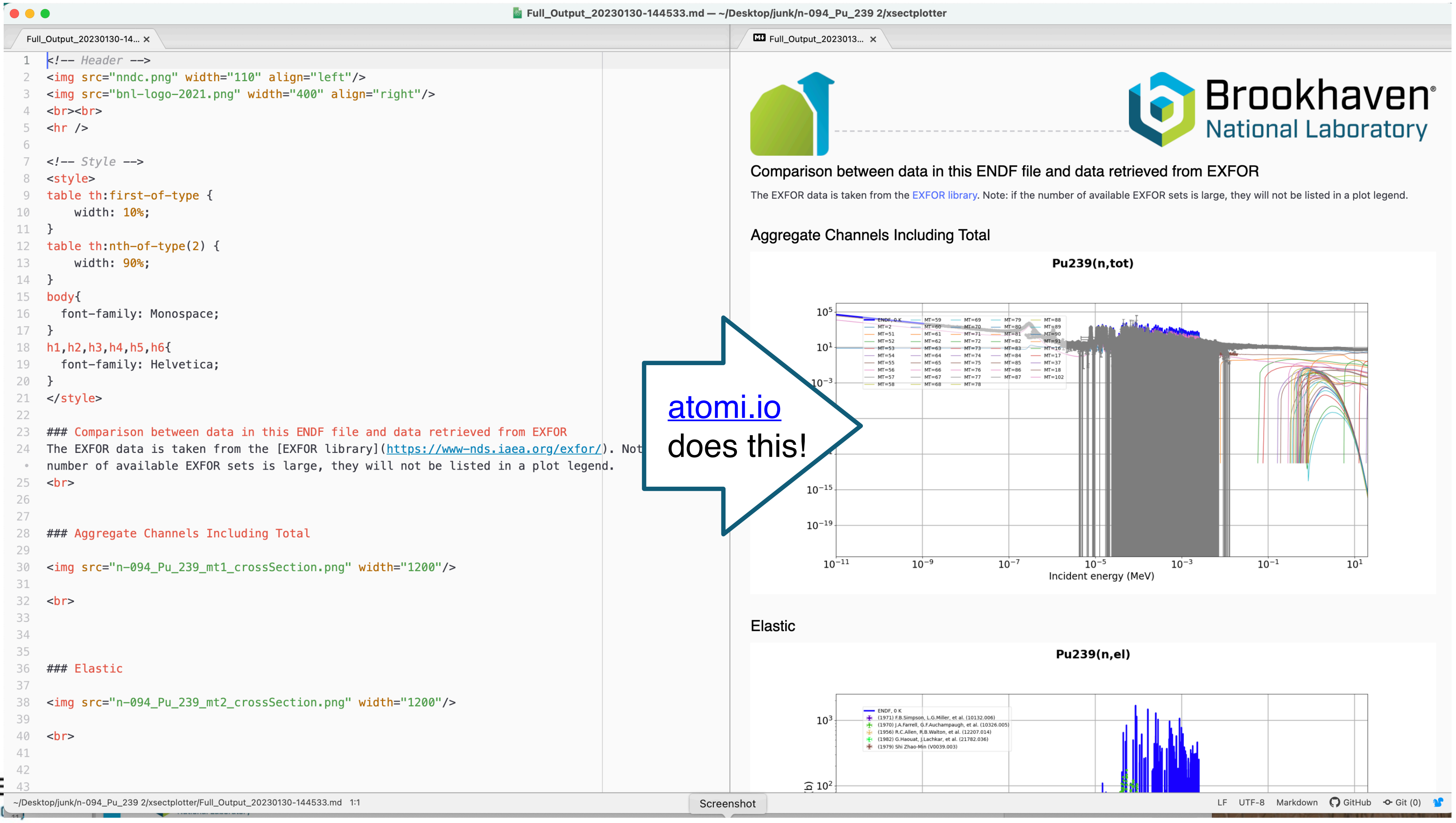
Sample directory



Open this. “Full_output_…”

Brookhaven National Laboratory

Output is in markdown and can be viewed in many text editors and web browsers


Full_Output_20230130-144533.md — ~/Desktop/junk/n-094_Pu_239 2/xsectplotter
Full_Output_20230130-14...
Full_Output_2023013...
Screenshot
LF UTF-8 Markdown GitHub Git (0)
Pu239(n,tot)
Incident energy (MeV)
Pu239(n,el)

```
<!-- Header -->
<img src="nndc.png" width="110" align="left"/>
<img src="bnl-logo-2021.png" width="400" align="right"/>
<br><br>
<hr />

<!-- Style -->
<style>
table th:first-of-type {
    width: 10%;
}
table th:nth-of-type(2) {
    width: 90%;
}
body{
  font-family: Monospace;
}
h1,h2,h3,h4,h5,h6{
  font-family: Helvetica;
}
</style>

### Comparison between data in this ENDF file and data retrieved from EXFOR
The EXFOR data is taken from the [EXFOR library](https://www-nds.iaea.org/exfor/). Not
number of available EXFOR sets is large, they will not be listed in a plot legend.
<br>


### Aggregate Channels Including Total

<img src="n-094_Pu_239_mt1_crossSection.png" width="1200"/>

<br>



### Elastic

<img src="n-094_Pu_239_mt2_crossSection.png" width="1200"/>

<br>
```

~/Desktop/junk/n-094_Pu_239 2/xsectplotter/Full_Output_20230130-144533.md 1:1

Brookhaven National Laboratory

Comparison between data in this ENDF file and data retrieved from EXFOR

The EXFOR data is taken from the EXFOR library. Note: if the number of available EXFOR sets is large, they will not be listed in a plot legend.

Aggregate Channels Including Total

Elastic

atomi.io does this!

Where are we now

- Final testing of full system on alphas sub library (it is small)
- Will run only on review branches for now
- Once live, will adjust review forms to point to the correct artifact URL

- We hope (expect?) you all will have many comments about the build reports
    - Good comments: dbrown@bnl.gov
    - Bad comments: gnobre@bnl.gov