NEA
NUCLEAR ENERGY AGENCY

NEA perspectives

WANDA
12 FEB 2025

Michael FLEMING
Head of the Data Bank
michael.fleming@oecd-nea.org

US membership in the Data Bank

- Since 1977 the US has been an observer to the Data Bank
- Bilateral agreements between NEA / DB and DOE / RSICC treated exchange of static packages
- Negotiations over 2024 proposed partial assessed in-kind contributions approved by all NEA countries and by OECD Council on 16 DEC 2024
- US has joined the Data Bank by notification to the OECD Secretary-General

Prior to 2025

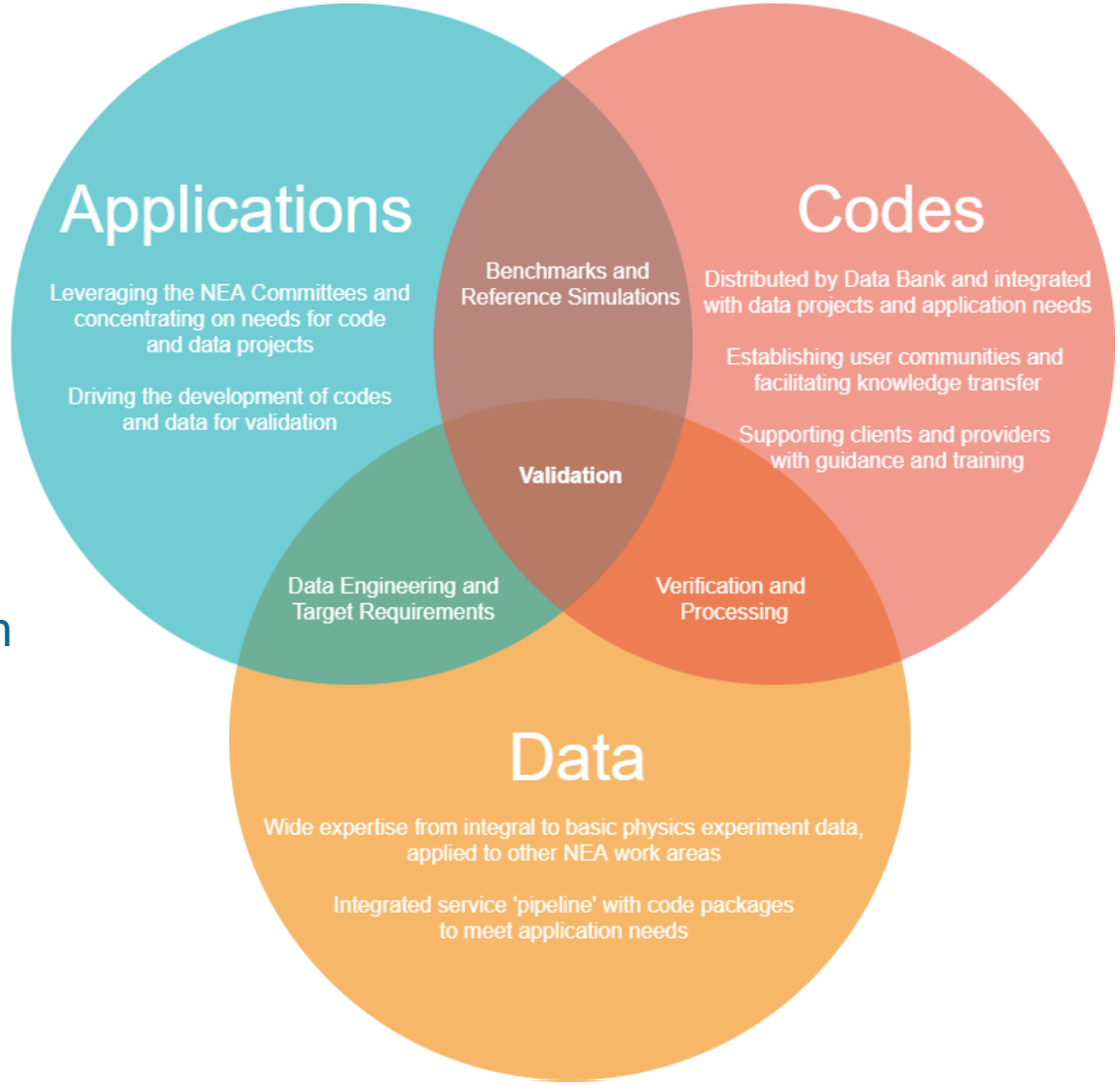
NEA Data Bank Programme of Work

- Data Bank has three Programme of Work areas:
  - 1.1 Computer Program Services
  - 1.2 Nuclear Data Services
  - 1.3 Nuclear Knowledge Management

- 'One-stop shop' provides opportunity for cross-programmer added value services/products

- IT managed by Data Bank Head since Q3 2023 with major overhaul to corporate and other services

- Embraced the use of GitLab and related systems to collaboratively develop resources, disseminate and create integrated products



Inventory of information / data / knowledge

Nuclear Data

- EXFOR contributions in NRDC Areas 2 / O
- JEFF Nuclear Data Library and Co-ordination Group
- Graphical applications including JANIS alongside others for benchmarks (DICE, IDAT, DATIF, NDaST)

Experimental Data

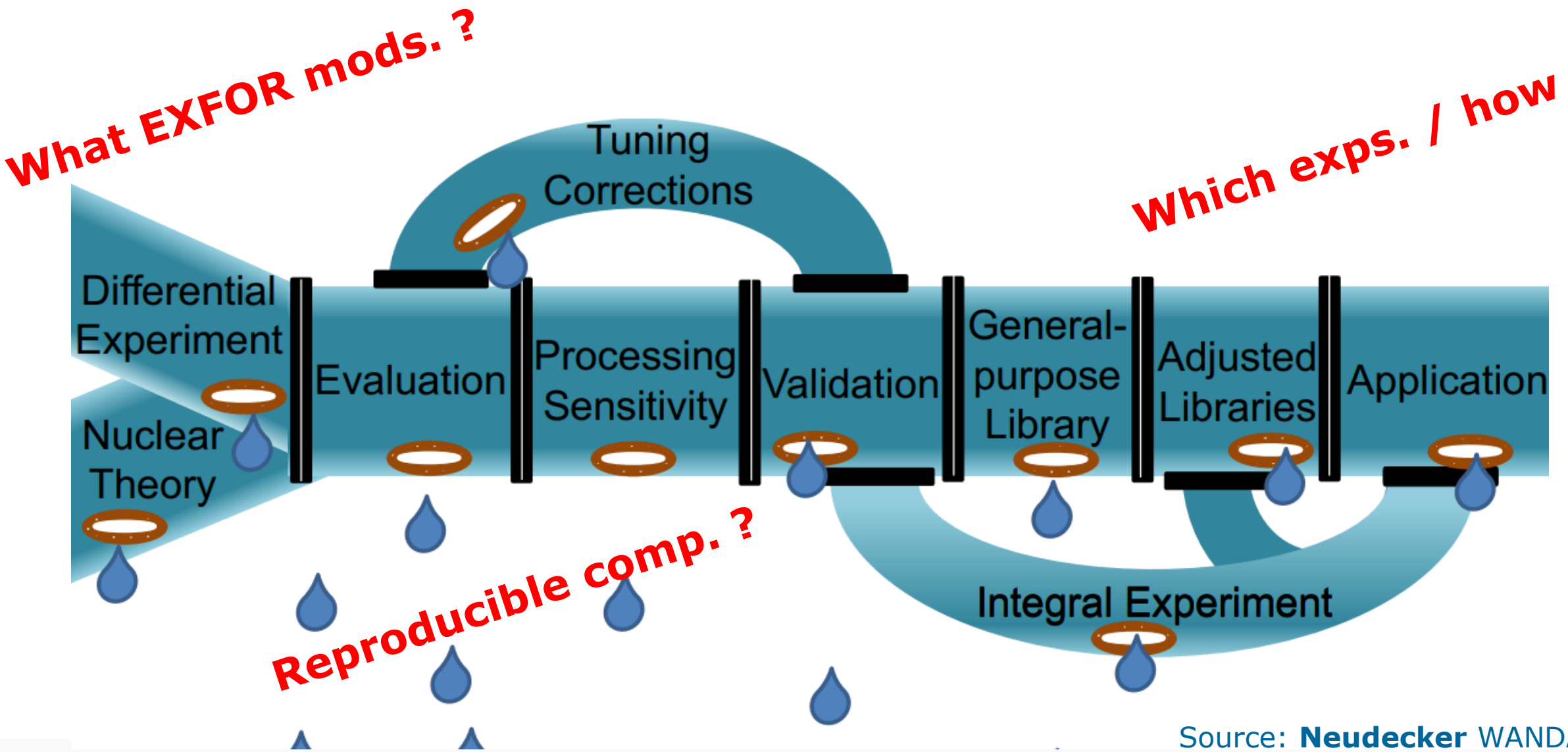
- NSC integral experiments / benchmarks (criticality, shielding, reactor physics, fuel, etc)
- NEA Joint Project datasets (e.g. Halden, FIDES, safety projects, etc)

Computer Codes

- Catalogue of access-controlled software (Serpent-2, Fispact-II, KRAKEN, Tripoli-4.12, PHITS-3, PENELOPE, FIFRELIN, etc)
- Source code with full build / processing / deployment mechanisms to containerised environment

Integration – these resources must be assembled with implicit knowledge

Missing knowledge / reproducibility



Source: **Neudecker** WANDA 2020

Challenges with information / knowledge loss

- "Evaluated EXFOR" is created locally by virtually all users, duplicating effort and often not preserved as durable knowledge / data.
- Evaluators create data files locally through processes and with software / scripts (even hand modification) that cannot be (practically) reproduced.
- Experiment selection and all relevant materials for validation is not preserved and/or disseminated, resulting in duplication and loss of knowledge.
- Adjustments are performed locally and not well documented and/or communicated. Creates misunderstanding with users (c.f. U8 inelastic note by LLNL in WANDA-25).
- Multiple drivers for this including legacy technology, job security, IP / information control.
- "Rule of 2" is not always possible and is not necessarily ideal.

A non-exhaustive wish list

Differential data

- Promptly compiled EXFOR(-like?) data with purely machine-readable formats
- Augmented uncertainty information to include un(der)-reported data (c.f. Denise's talk)

Evaluation

- Truly reproducible evaluation resources leveraging VC / containerisation – complete with documentation including experiment down selection, optimization, etc.

Validation / processing

- QA computational models with peer-review and other input resources coupled with containerised pipelines built from source in version control.
- Downstream V&V pipelines to users (blind if IP issues) with automated (or prompt) feedback.

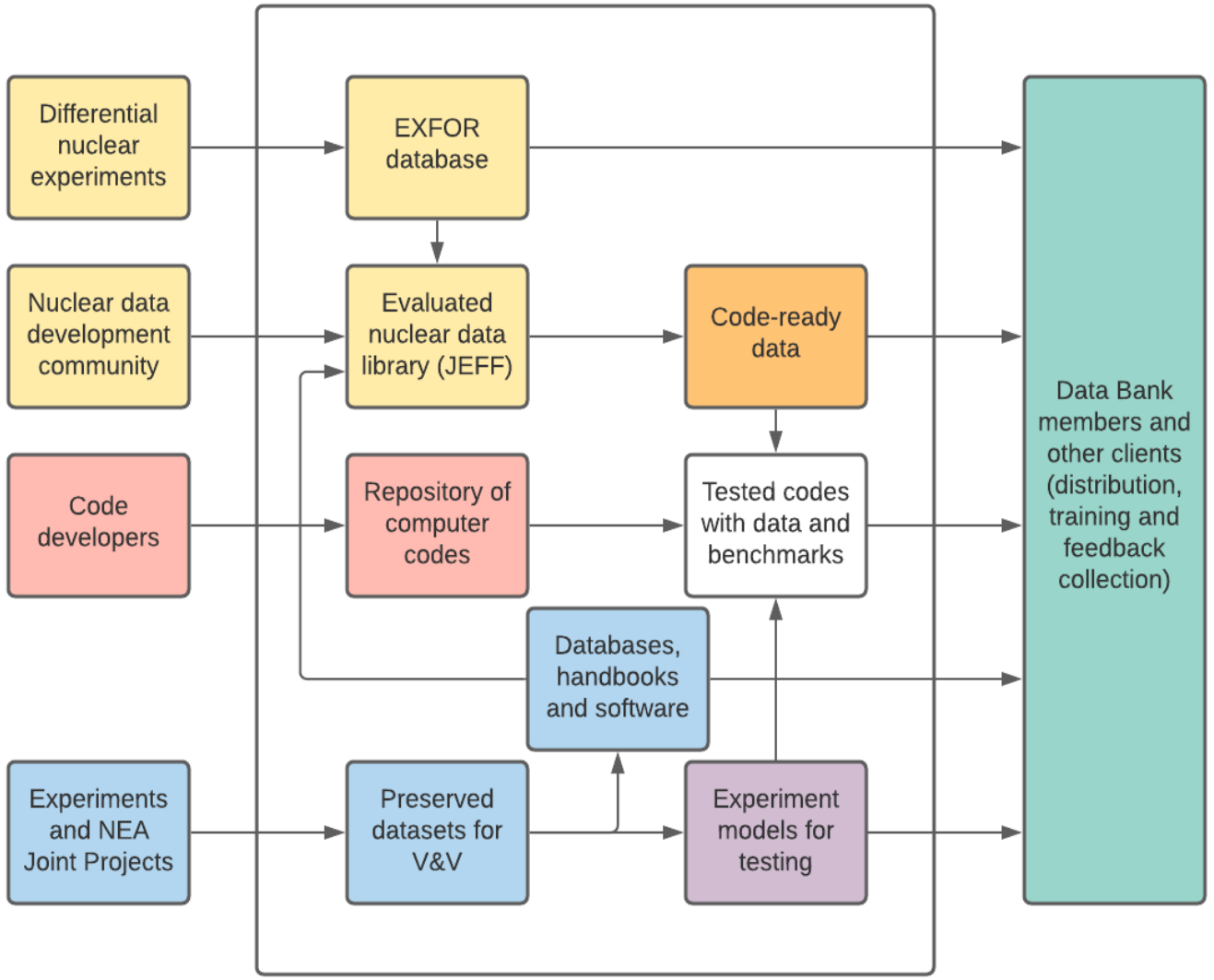
Adjustment

- Example adjustment processes with experiment selection, covariances, executed through containerised pipelines from source and w/documentation on choices – ready for users to modify!
- Database of sensitivity data (allowing 'blind') preferencing openly reproducible data.

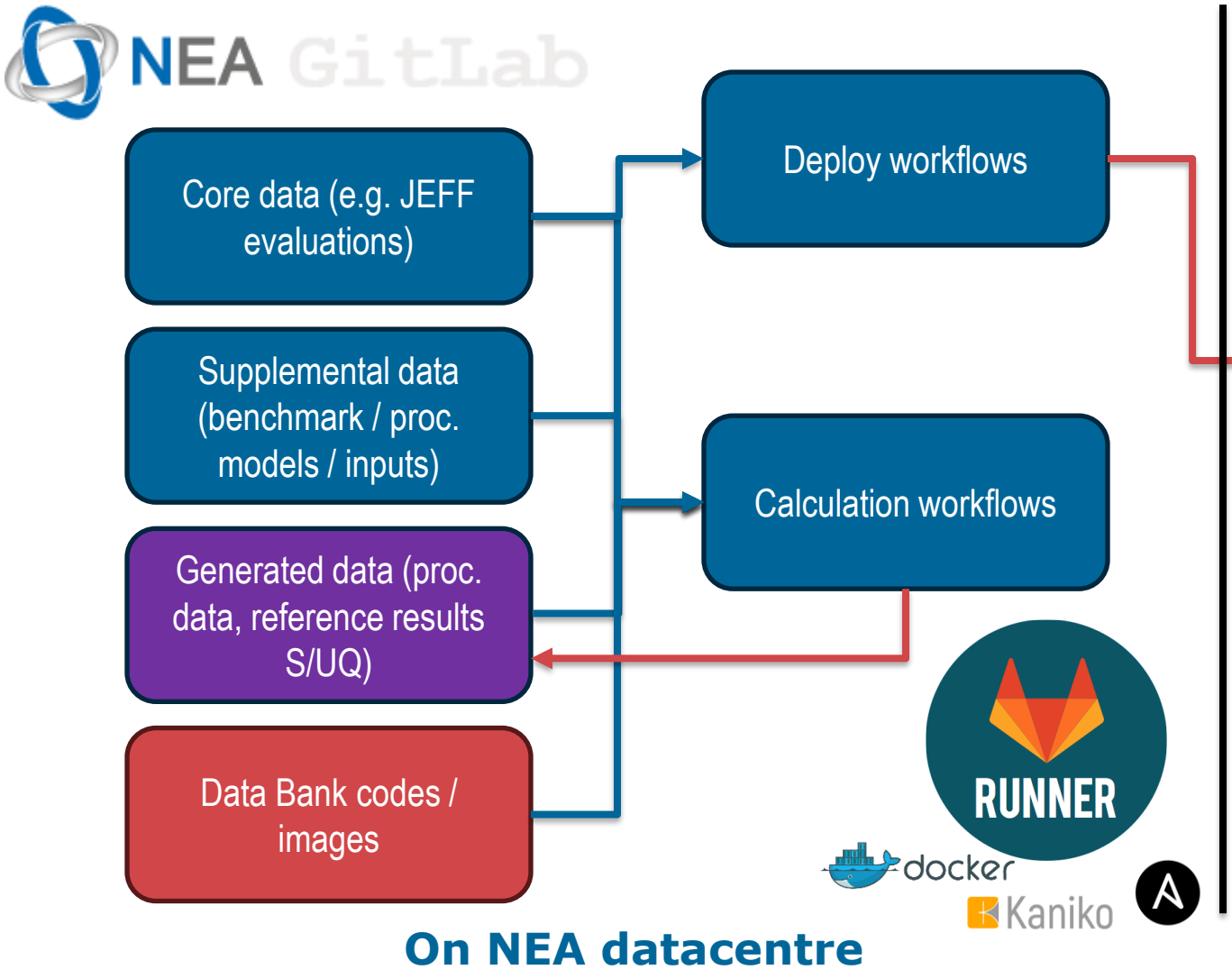
(Scientific) data management model

- Rigorous version control and project management systems
- Cross-project integration and automated workflow management.
- Integrated containerisation for precise environment replication (e.g. Docker).
- A suite of supplemental tools to track issues, submit/review changes, store process outputs and publish to closed/open webpages, and more.
- Contributors and partners directly tap into live physics, software and benchmark experts and **continuously-updated data**.

All content is securely stored, can be referenced, accessed and reused.



(Scientific) data management model



RDM / RDBs with API

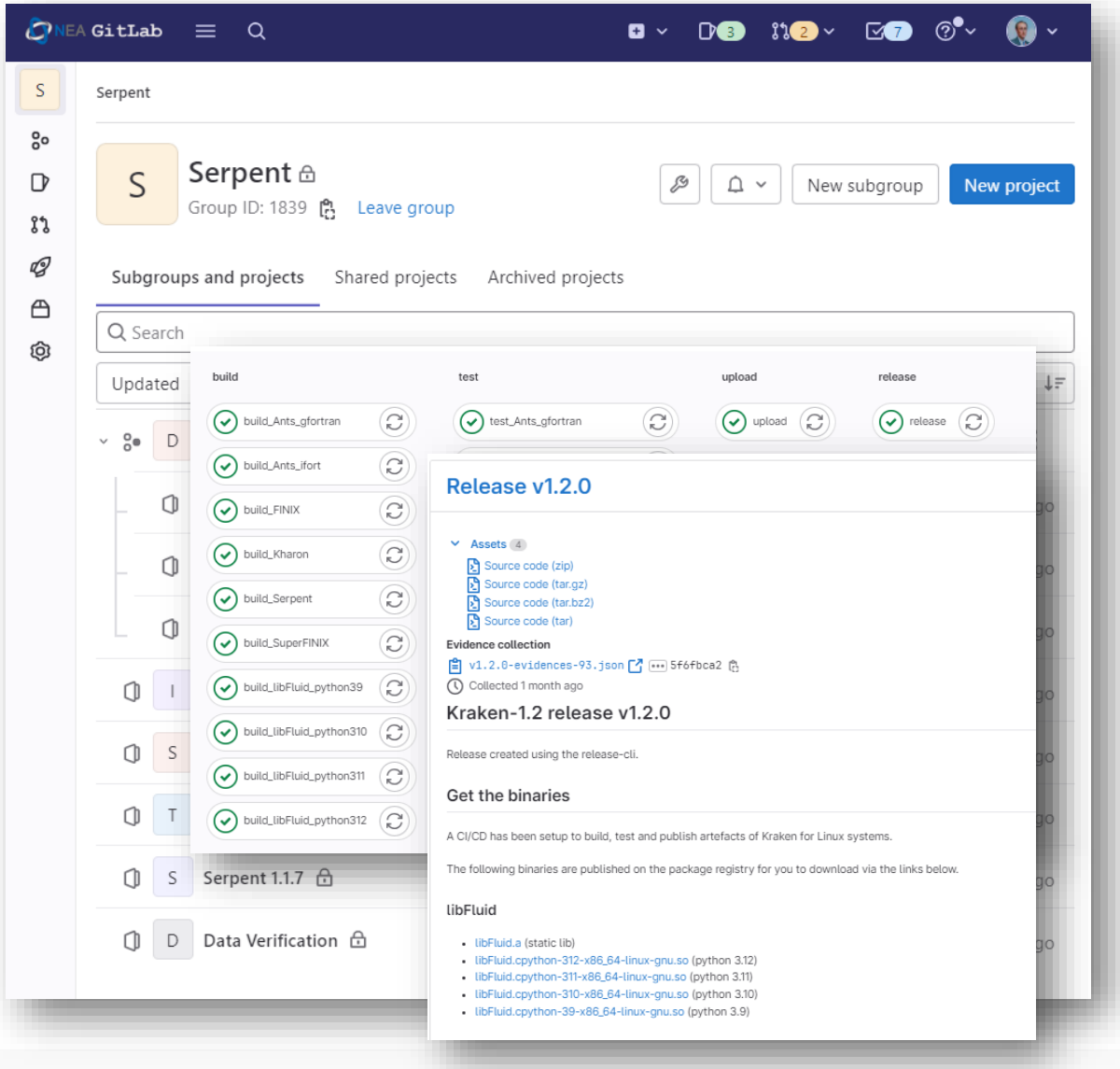
Public data

on NEA corporate cloud

- Databases publicly accessible through graphical interfaces (JANIS, DICE, etc.), directly and/or via API.
- All protected content only under EULAs and delegates on NEA datacentre.
- No manual calculations.

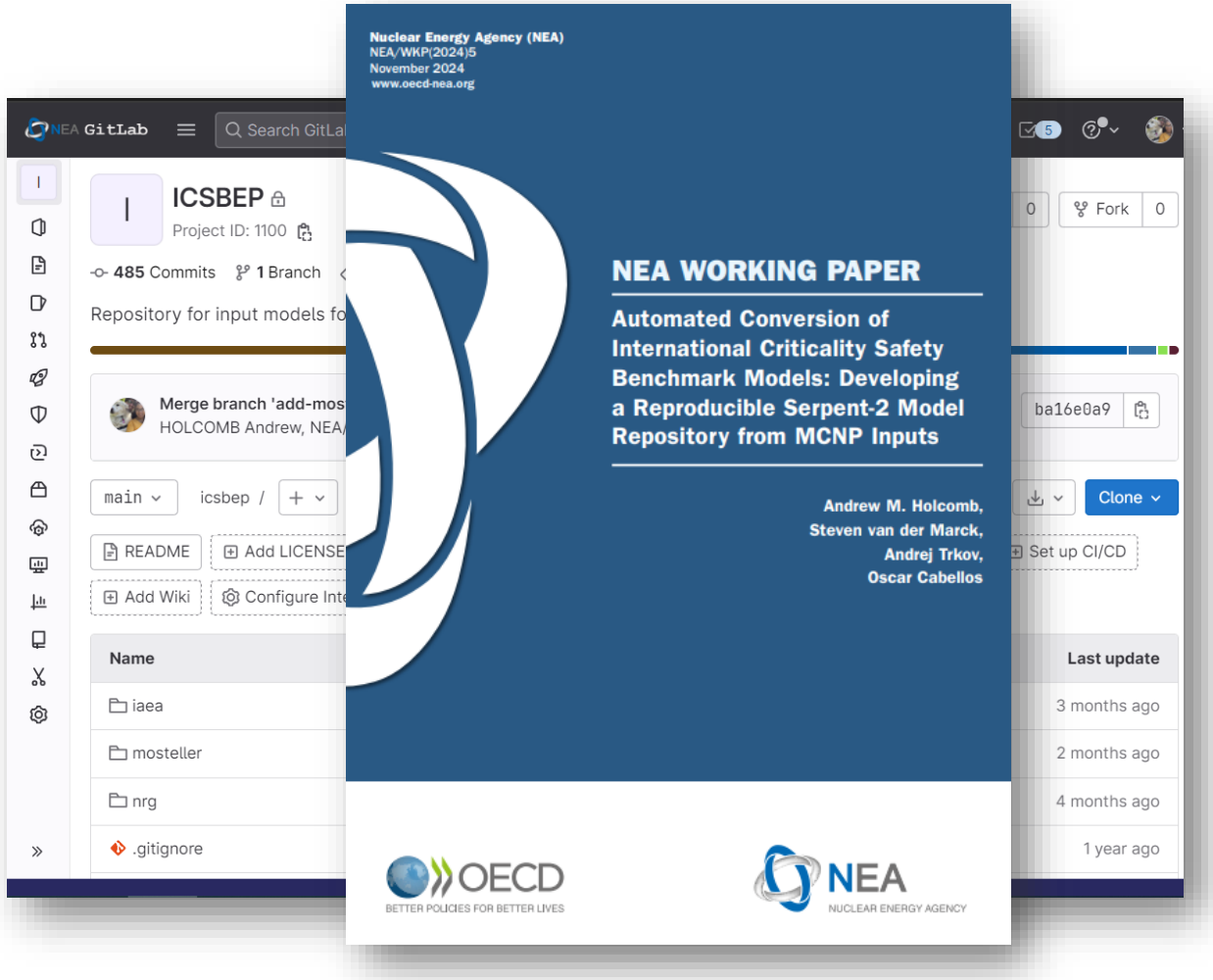
NEA GitLab-based services

- As previously reported, the **NEA GitLab is central** to Data Bank service model, hosting ~1800 licensed users and delegates.
- **It is growing** with new packages: TRIPOLI, PENELOPE, KRAKEN, HALDEN and updates to SERPENT, PHITS, FIFRELIN, FISPACT and more.
- Coupled with benchmark models for ICSBEP, with reference calculations run through open and automatic processes through GitLab.
- Support to help developers with Carpentries-trained trainers.



Benchmark computational repositories

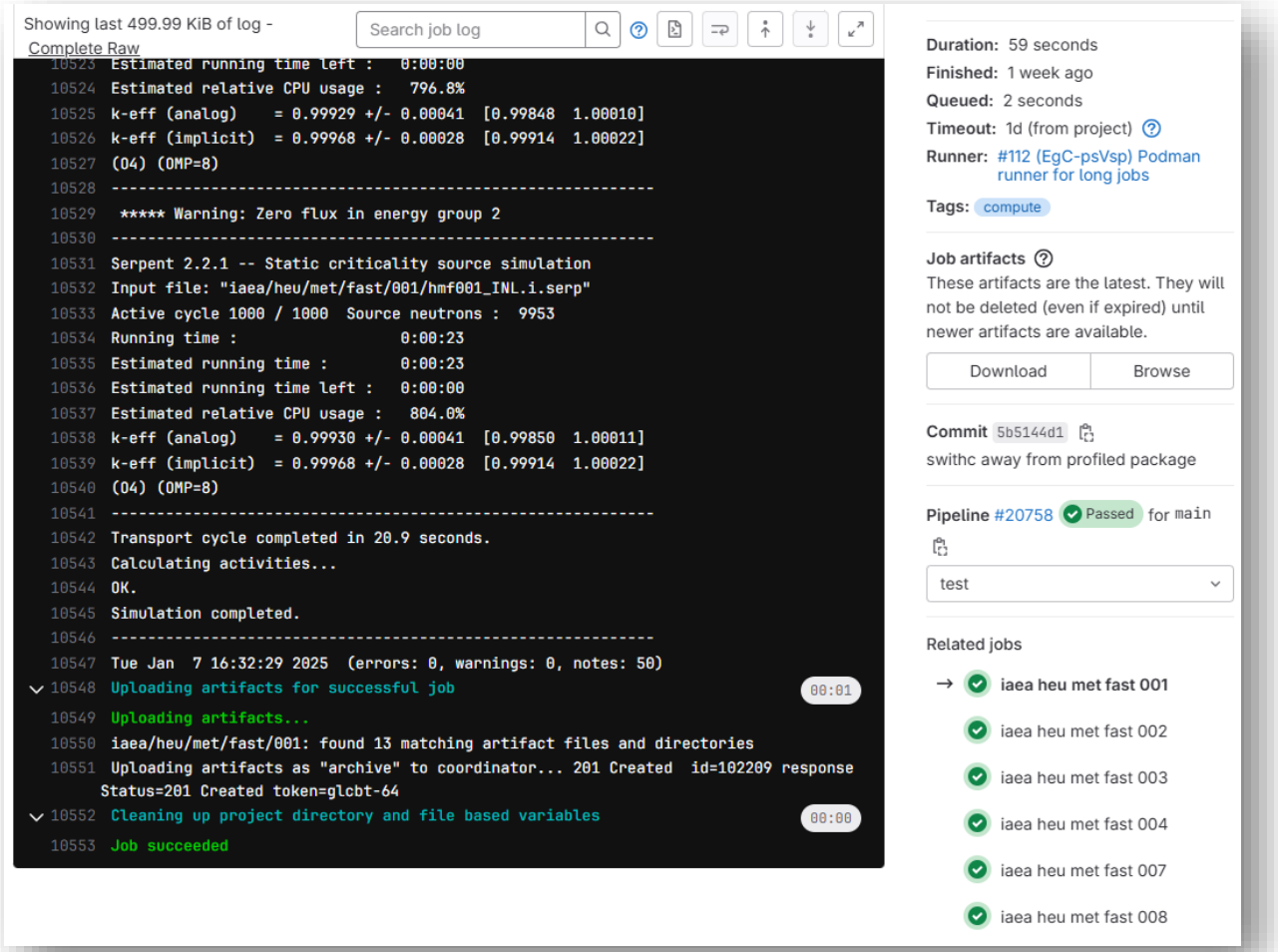
- Nuclear Science Committee benchmarks are international standards used around the world for V&V.
- The Handbooks are used by computational physicists, but their purpose is not to manage code-specific resources.
- Data Bank services now centralise repositories of benchmarks to:
  - Provide V&V resources for users
  - Provide testing frameworks for devs
  - Enable community-driven, controlled and licensed data sharing
  - Develop automated JEFF testing pipelines
  - Build a platform for future V&V collaboration



Source: **Holcomb** et al. NEA/WKP(2024)5

JEFF nuclear data library

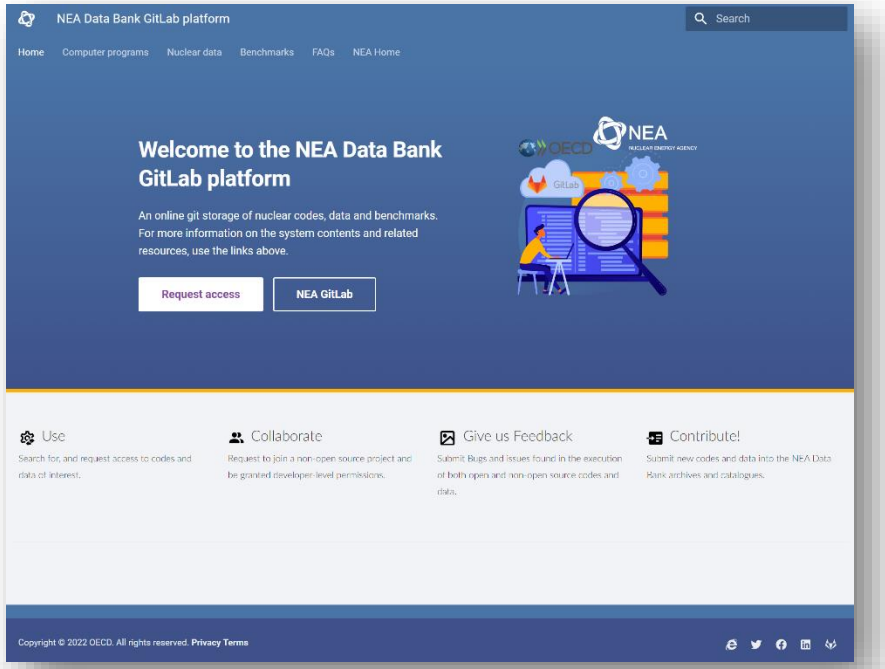
- JEFF-4 test libraries demonstrated improved performance with industry. Release in Q2 25.
- GitLab provides a completely traceable system for version control and automated task execution.
- JEFF development data used for direct calculations that are 100% FAIR Open Science.
- Data summarized for publication with full reference stack for (licensed) users.
- These will form the basis of JEFF-4 QA package documented on release.



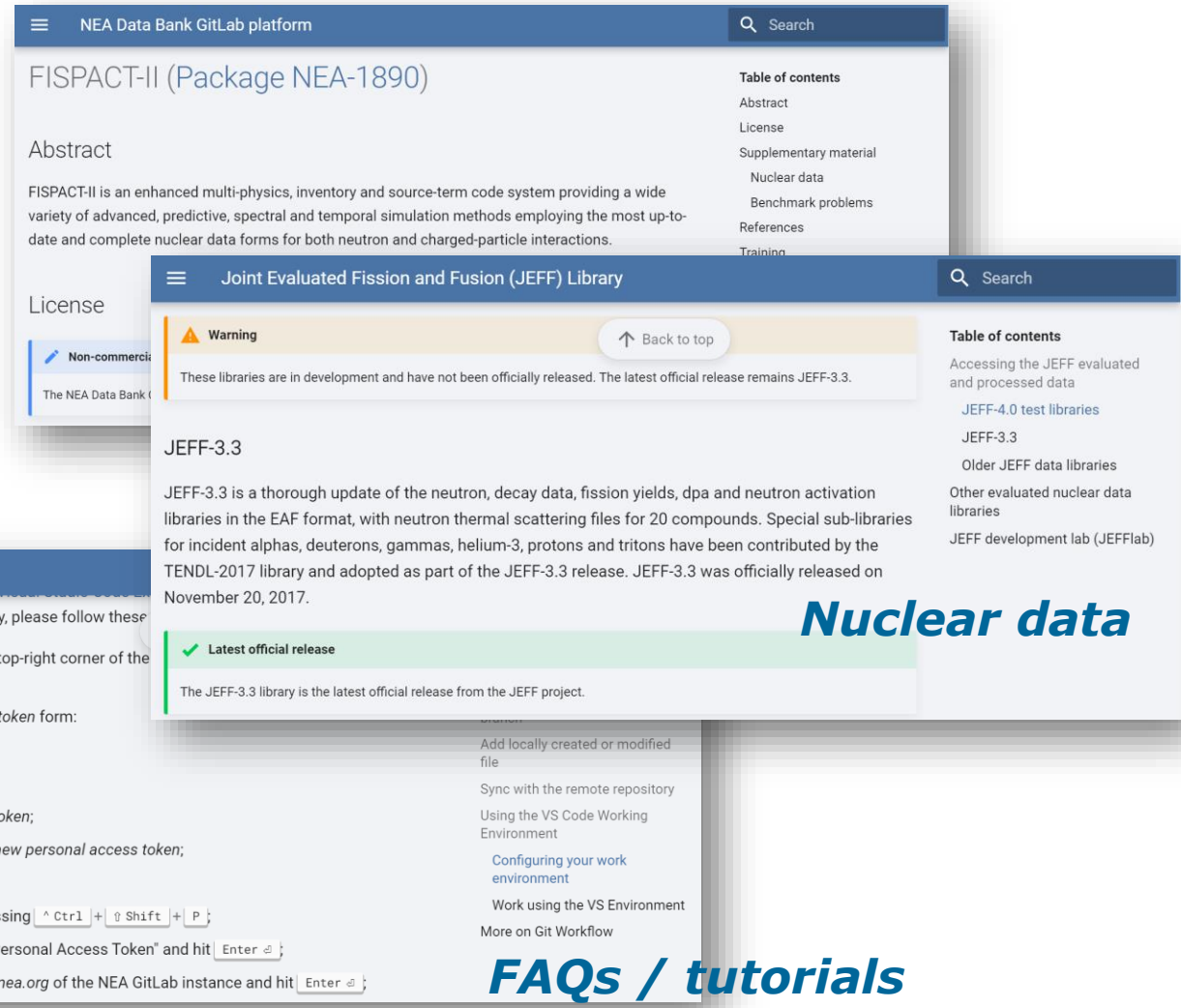
Interfaces for packages

One public landing page

- *'An interface for GitLab-hosted content'*
- All managed markdown within GitLab, effortless maintainability, content drafted with owners
- All content interlinked and easily searchable
- Will contain reference calculation content

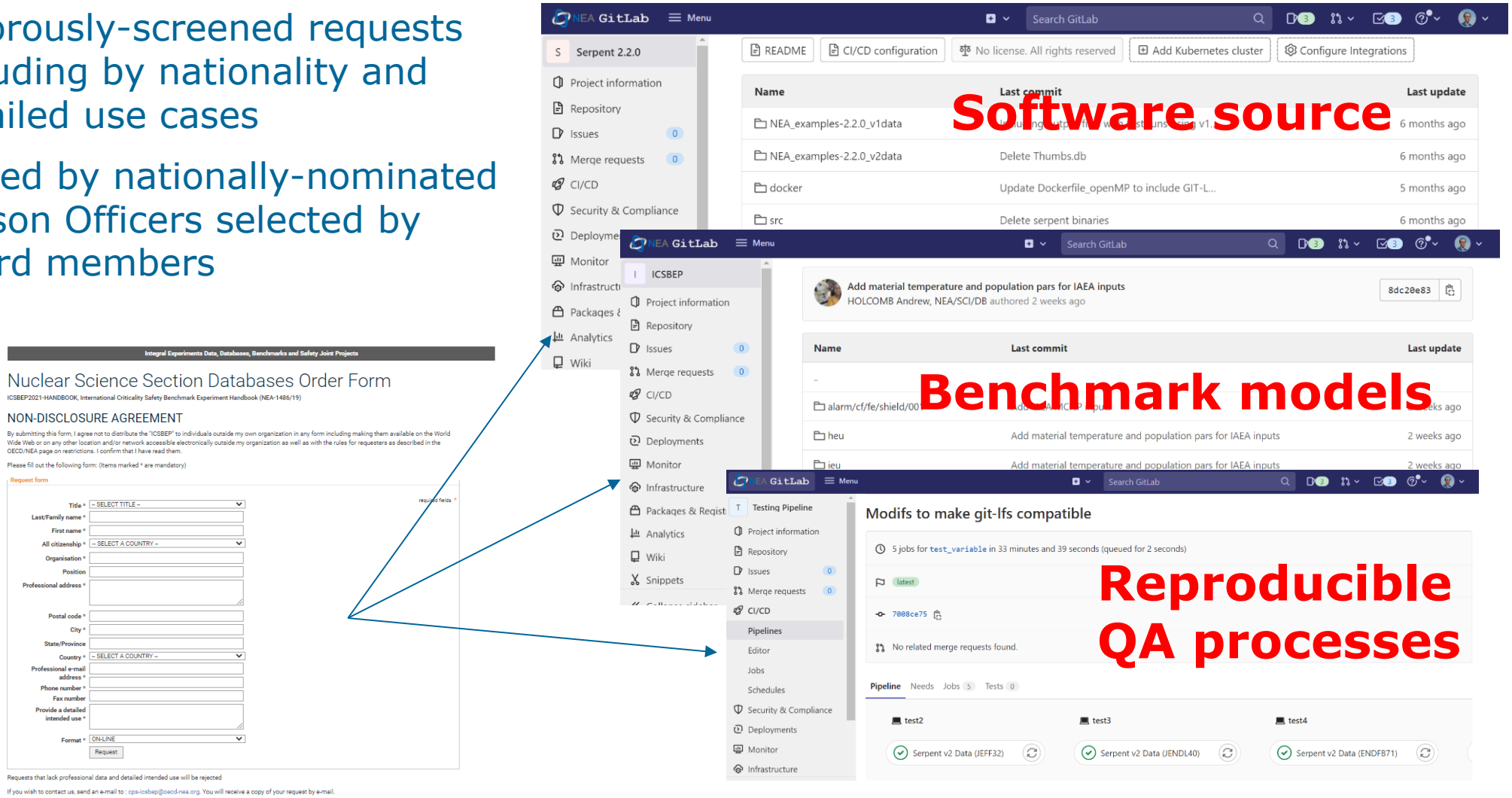


CPS packages (software + NSC + SAF)



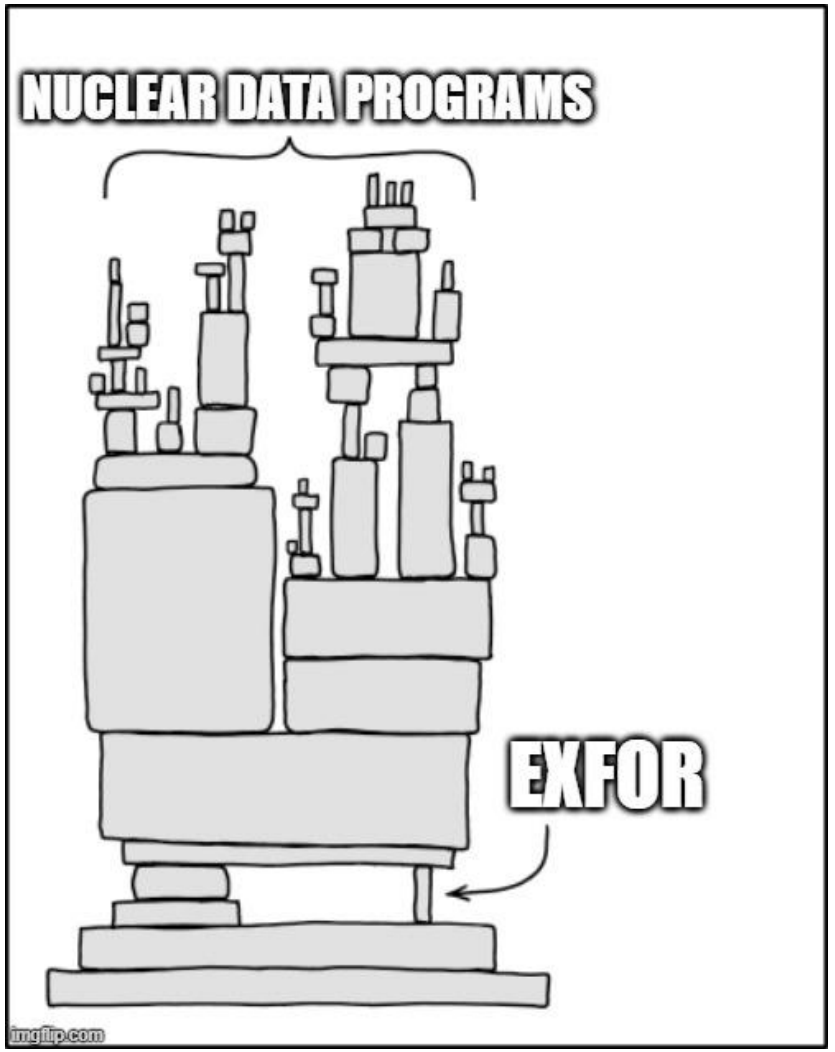
Access-controlled resources

- Rigorously-screened requests including by nationality and detailed use cases
- Vetted by nationally-nominated Liaison Officers selected by Board members



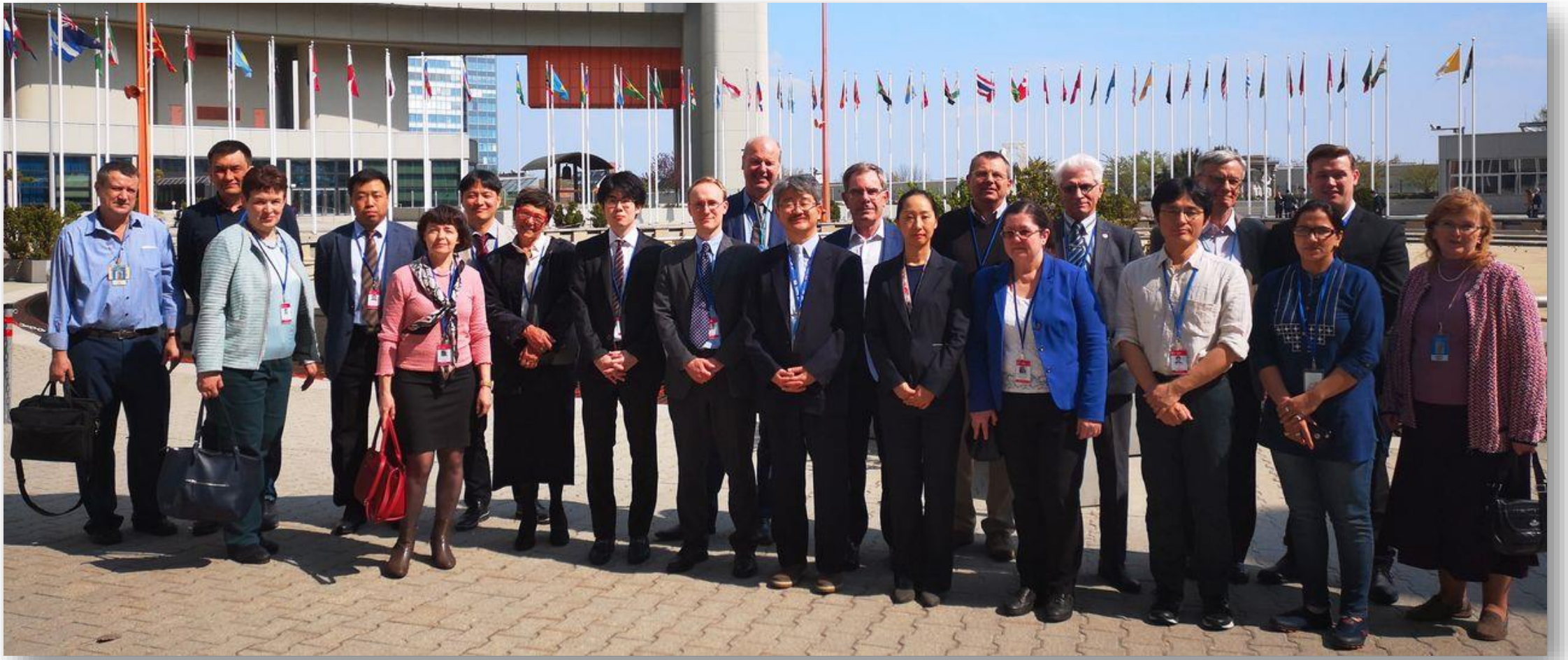
EXFOR – the foundation of nuclear data

- EXFOR is the essential differential measurement database used in virtually all nuclear data activities.
- Originally a US / NEA / USSR / IAEA project with four ‘core centres’ responsible for >90% of all content.
- Technology has transitioned over time with 2005 consolidation into one ‘master’ (EMF) maintained by IAEA
- Some progress made in EXFOR management but significant space for improvement remains:
  1. Do we want to do any AI/ML with differential data ?
  2. Do we want meaningful error analysis ?
  3. Do experimentalists want to write to EXFOR ?



Source: XKCD 2347

NRDC (EXFOR group)



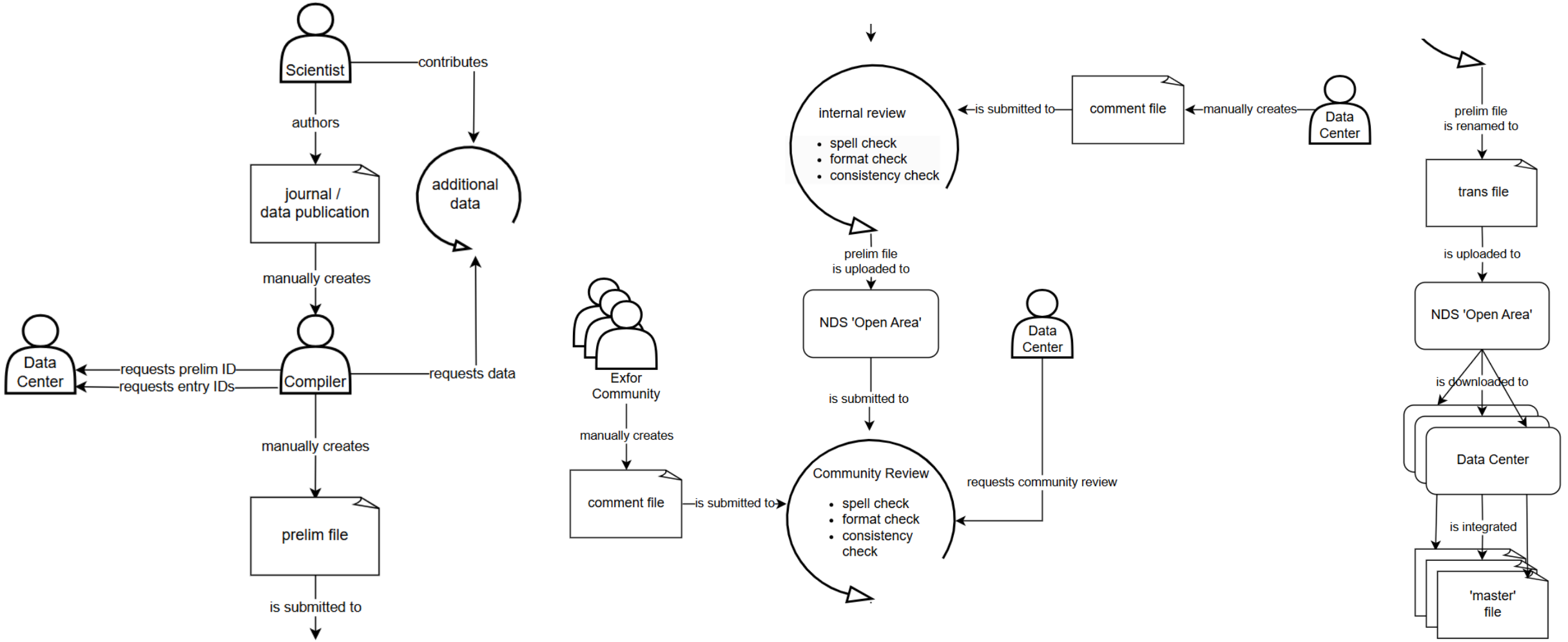
Source: https://www-nds.iaea.org/nrdc/

What is in EXFOR (actual EXFOR transmitted in 2024)

- Emailed text files with text file line patching system
- Extensive use of **free text** for content such as error analysis or incident particle spectra
- Version history with line of text inside the modified text file
- Version recovery *should* be possible
- Tech stack is not open / DevOps'ed

```
REACTION    (83-BI-209(N,X)0-NN-1,,DA/DE)
SAMPLE       Hollow cylinder of chemically pure bismuth
ERR-ANALYS   Sources of uncertainties at neutron energy ~ 1.7 MeV:
             Statistics of the counts for Bi                    1.5
             Correction for energy spread of neutron source    18.
             Correction for absorption and multiple
             scattering for Bi                                  5.
             Method of analysis                                 7.
             Statistics for np-scattering peak                  1.
             Correction for absorption and multiple
             scattering for np-scattering                       3.
             np scattering vross section                        1.
             Relative efficiency of detector                    4.
             Total uncertainty                                 21.
           (DATA-ERR)  Total uncertainty
STATUS     (TABLE,,G.A.Prokopets,J,SNP,32,19,1980) Table 1.
HISTORY    (19931021U)   New reference added
           (19981111U)  .Reference deleted
           (20070321U) ERR-1 was chanhed into ERR-S
           (20240916A) ERR-ANALYS was corrected.
                       ERR-T -> DATA-ERR, EN-RSL -> EN-RSL-HW
```

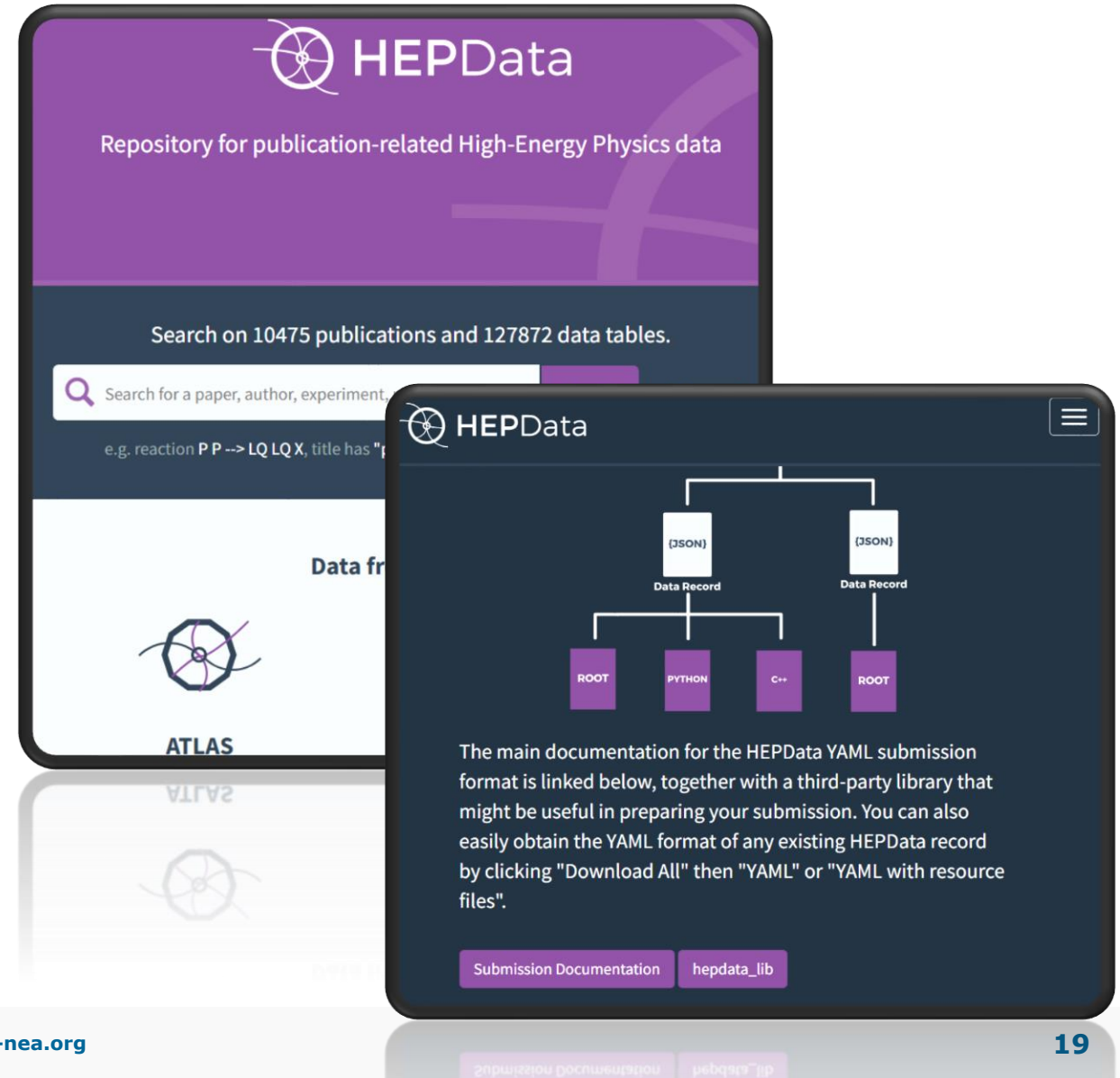
EXFOR actual workflow (for information)



Source: **FOLIGNO** NRDC 2022

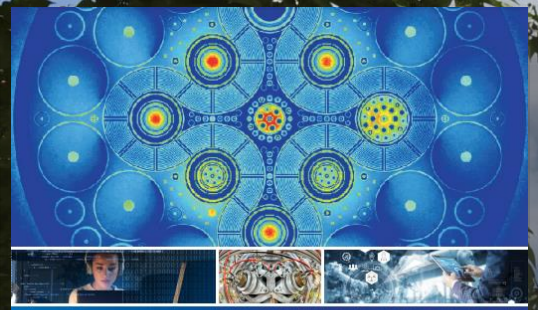
Where should we go ?

- Existing open-source technology stacks can be leveraged to provide the underlying service c.f. hepdata.net
- The experimental community should be directly engaged and should become contributors
- Decisions must come from stakeholders convened through open processes
- Existing formats, processes and status quo mentality must not dictate the future
- Maintain as much of the existing expertise from the NRDC as possible



Recap

- The US has joined the Data Bank in 2025, and more details will be communicated.
- Access to a wealth of resources in controlled, collaborative environments will be possible in 2025 with relevant nominations.
- JEFF-4 will be released in Q2 2025 and will include entirely FAIR Open Science reference results coupling NSC benchmarks with software and physics data.
- More progress is needed in key areas including:
  - Evaluation reproducibility [not discussed]
  - Adjustment methods, transparency and knowledge dissemination [WPEC SG 52]
  - Collaborating on benchmark computational resources [NEA TRGs / WPs]
  - Differential data / EXFOR [WPEC SG 54]



The NEA Data Bank

The NEA Data Bank acts a
participating countries w
computer codes and nuclear
of phenomena in the nuclear
providing the means to develo
uting them in accordance wi
ing and education activities t

OECD
BETTER POLICIES FOR BETTER LIVES

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

DATA BANK NEWSLETTER

ISSUE 1. OCTOBER 2022

Survey for the 2023 training courses

The NEA Data Bank offers training courses in several computer programs used in nuclear energy applications. Since 2020 these include both in-person and virtual training events that use an e-Learning platform to deliver course materials and hands-on tutorials. We are expanding the options for 2023 and would be grateful for your input on which courses should be prioritised and which topics covered. If you or any of your colleagues would be willing to participate, please complete the online survey here.

Upcoming events

The NEA Data Bank will organise four courses this autumn. Following the success of the first OpenMC training earlier this year, an online OpenMC course will take place on 24-27 October. Participants will gain expertise in this open source Monte Carlo neutron and photon transport simulation code.

A new edition of SCALE KENO-MAVRIC criticality safety and radiation shielding training will follow on 14-18 November at the NEA offices in Boulogne-Billancourt, France. A course on PHITS, a Japanese general-purpose Monte Carlo particle transport simulation code, will take place on 21-25 November in Paris. Please note that registration periods for these courses close in October 2022. You can find more information about these courses online here.

Course	Dates	Venue	Cost (EUR)
OpenMC	24-27 Oct	Virtual	450
SCALE-NCS/RS	14-18 Nov	NEA	2 300
PHITS	21-25 Nov	NEA	650

New e-Learning system and credentials

Moving to virtual training for computer programs introduced several challenges which were common to millions of students around the world. In 2022, the Data Bank launched a Canvas Learning Management System (LMS) to support virtual training events and coupled this with an Accredible-based certification system that allows students to receive verifiable recognition for their completed work.

The systems were tested in the first of a kind OpenMC training in April 2022 and rolled out in subsequent events for FISPACT-II, PENELOPE and the NEA International Radiation Protection School.

Message from the Head of the NEA Data Bank

After more than 10 years, we are resuming the NEA Data Bank newsletter to provide you with updates on our services, recent and upcoming events, new packages that are available to our members and other news items of interest.

Following two years with lockdowns, restricted travel and never-ending virtual meetings, the Data Bank Secretariat have been delighted to welcome you back to the OECD Conference Centre and our Boulogne-Billancourt offices, whether as delegates to official bodies, experts in our technical groups or students in training workshops. The Joint Evaluated Fusion and Fusion (JEFF) project has already hosted two physical full nuclear data weeks, we have resumed live training events and we welcomed our Management Board delegates to their first in-person meetings since 2019.

In the past months we have been working diligently to create improved and modern services that integrate our different work areas. We have implemented a new GitLab platform with custom on-site infrastructure, are moving to SharePoint Online for delegate and official body engagement and launched an e-Learning system. The team have been piloting these tools across our work areas and we are grateful to all our partners for their engagement and enthusiastic responses.

I sincerely hope to welcome you (again) to the NEA in the near future. If you have any questions, please reach out to us using the contact details provided below.

Michael Fleming

New NEA credentials for online and in-person courses offered through the Data Bank with projects launched between April – July 2022.

michael.fleming@oecd-nea.org
oecd-nea.org/databank

data@oecd-nea.org
programs@oecd-nea.org