Nuclear Energy Agency

NEA/MBDAV/DOC(2021)13

Unclassified

English - Or. English

4 January 2022

NUCLEAR ENERGY AGENCY
MANAGEMENT BOARD FOR THE DEVELOPMENT, APPLICATION AND VALIDATION OF NUCLEAR DATA AND CODES

Cancels & replaces the same document of 17 December 2021

The NEA Data Bank Task Force on Open Science and Nuclear Energy Research Data: Emerging Needs, Opportunities, and Possible Role for the NEA Data Bank "DB-OpeNER"

Main findings from stakeholders' interviews and NEA DB users' survey

Alice Dufresne
Email: alice.dufresne@oecd-nea.org
Tel: +33 (0)1 73 21 28 30

JT03487859

Contents

1. Background of the TF DB-OpeNER and its working methods

The on-going movement towards Open Science (OS) is leading the NEA Data Bank (NEA DB) to take a prospective look at its services and working methods, so as to foster a wider access to Research Data and the joint production of new data (in a broad sense). In September 2020, the NEA Management Board for the Development, Application and Validation of nuclear data and codes (MBDAV) established a one-year Task Force (TF) on Open Science and Nuclear Energy Research Data: Emerging Needs, Opportunities, and Possible Role for the NEA Data Bank, called "TF DB-OpeNER" with the goal of "reviewing" the NEA DB services in the context of OS and evolving users' expectations (See TF mandate in Appendix 1). The conclusions of the TF are expected to serve as a basis for proposing adjustments in the NEA DB missions over the 2023-2027 period.

1.1. Background on the NEA Data Bank

Since its inception, the NEA DB has been practising and promoting transparent and collaborative activities, much in the spirit of OS and Open Data (OD). Today "the Data Bank acts as an international centre of reference for its participating countries with respect to basic nuclear tools, […] [and] provides a direct service to its users by providing the means to develop, improve and validate these tools and by making them available as requested". The NEA DB carries out scientific activities related to computer codes and nuclear data and supports various parts of the NEA with its expertise. The NEA DB is implementing OS working approaches in the JEFF[1] collaboration and construction of collaborative workspace (currently the NEA DB is working on setting up a long-term locally-hosted GitLab system). The NEA DB also has the central mission to manage data in the long term (preservation and distribution), and thus plays a transverse role within the NEA. Most of the data and codes at the NEA DB are voluntary contributions made freely accessible to member countries on request. At the same time the NEA statutory elements, member country positions, project funding mechanisms, data/software ownership rights, publication rules, and limited resources provide arguments against adopting full-fledged OS practices indiscriminately as, at the NEA, most of the data are collected and become open only after reaching a consensus between member countries.

1.2. Task Force working approach

The TF adopted a three-step approach. As an initial step, the TF members decided to interview NEA DB services users, NEA database managers and OS practitioners in order to collect their anticipated needs in terms of research data and new services, as well as any recent experience in terms of building OS culture and implementation. As a second step, the TF team organised a workshop with the aim to share the main findings of individual interviews and surveys, and to foster further exchanges on specific issues raised during the interviews. The TF workshop held

[1] JEFF – The Joint Evaluated Fission and Fusion File https://www.oecd-nea.org/jcms/pl_20182/jeff

on May 27-28, 2021 gathered 61 participants from 15 countries, the European Commission (EC) and the International Atomic Energy Agency (IAEA), the French National Research Institute for Agriculture, Food and the Environment (INRAE), the Strasbourg Astronomical Observatory, the Japanese Research Data Utilization Forum (RDUF), the CERN (the European Organisation for Nuclear Research) and a staff member of the GitLab company. As a third step, the TF prepared a set of suggestions and recommendations relative to future NEA DB activities based on the main findings and outcomes of the workshop, for consideration by the MBDAV.

1.2.1. Preparation of the stakeholders' interviews

The TF members prepared a set of questions (see Appendix 2) to be addressed in the interviews, which covered four major topics:

- Modern data repositories
- Improved practices and collaborative work
- Opening access to research data in the nuclear energy field, including educational activities
- DB's possible role in connection with OS, e.g., providing an OS portal

The main goals of the interviews were to identify the needs of NEA DB users, to hear their views on the possible benefits/drawbacks of sharing information more openly in the nuclear energy community, to provide examples of modern working practices, to hear their needs for data management infrastructures in the context of OS, and to collect ideas/suggestions/proposals regarding the evolution of NEA DB services regarding OS.
The TF organised twelve interviews over three months (February-April 2021), the list of interviewees' organisations is given in Appendix 3. It includes software developers, database managers, and NEA DB services users. They represented European and governmental institutions, international and national research centres, academia and universities, regulatory bodies, private companies and NEA divisions. Each interviewee was free to focus on selected questions or topics in accordance with his/her area of expertise. The interviewees were encouraged to speak their mind and voice their opinion freely, so all replies were processed anonymously except for a few messages provided in § 2.1 below.

1.2.2. Survey of NEA DB Liaison Officers

In addition to the interviews, a short survey (See Appendix 4) was sent to NEA DB Liaison Officers. 32 respondents from 14 countries (including Belgium, Czech Republic, France, Germany, Greece, Hungary, Italy, Netherlands, Poland, Romania, Russia, Slovakia, Spain, UK) shared their views regarding possible evolutions of NEA DB services, and anticipated needs toward OS, and for new types of infrastructure and long-term evolution of the Data Bank. Several NEA DB member countries including Argentina, Belgium, Japan and Russia provided independent consolidated vision from groups of experts and representatives. The list of countries/organisations is given in Appendix 5.

1.3. Motivations for this document

The main findings derived from the stakeholders' interviews and DB users' survey were compiled and presented at the 27-28 May TF workshop. They triggered many reactions and interesting discussions (see the detailed summary record of the TF Workshop NEA/MBDAV/DOC(2021)11). This prompted the TF to record these findings in the present document, so that they can be shared more broadly.

2. Interviews' main findings

2.1. Open Science and the NEA DB

According to the European Open Science Cloud[2], Open Science can be defined as "an approach to the scientific process based on cooperative work and ways of disseminating knowledge, improving accessibility to and re-usability of research outputs by using digital technologies and collaborative tools". At its core, OS aims at "increasing research quality, boosting collaboration, speeding up the research process, making the assessment of research more transparent, promoting public access to scientific results, as well as introducing more people to academic research".

During the TF-led interviews, the general movement towards OS and its implications for our community were discussed. The TF collected the interviewees' opinion regarding OS, the status of OS in their country, and their views on possible applications of OS in the nuclear energy field. Interviewees were also encouraged to express their anticipated expectations towards the NEA DB and MBDAV but also the NEA in general with respect to adopting OS principles.

While some of the interviewees feel insufficiently informed about OS, there is a general perception that the nuclear community is lagging behind other communities in the ongoing OS/Open Access (OA) movement. Many countries have turned OS and FAIR (Findable, Accessible, Interoperable, Reusable) principles into laws, mandating open data policies (Data Management Plans (DMPs) etc.) for government-funded R&D organisations. OS principles are fairly easily implemented in basic research communities, universities, and government-funded organisations. They are more difficult to implement in nuclear organisations and industry. There are political, legal, export control, proprietary rights concerns, as well as other obstacles such as psychological and cultural barriers.

National policies towards OS differ across NEA member countries. In most European countries, EC-funded projects are bound to publish and release projects results openly. The European legislation is pushing forward the adoption of OS practices. As an example, the European Initiative Horizon 2020 supports opening data in the nuclear field. Argentina approved a national legislation related to OA and Institutional Repositories. As a consequence, the "Nuclea" Institutional Digital Repository of CNEA was created, and it adheres to the National System of Institutional Repositories. In Japan, some initiatives on OS were established in the past few years. The RDUF is successful in leading OS in the Japanese scientific community, and this includes collaboration with JAEA. However, in the nuclear energy field, there is an established practice of carefully judging what data can be made open-access and what data should remain restricted. In Russia, OS concepts have not been formalised so far and there is a shared view that considers these OS topics are only of interest to academia, universities and fundamental nuclear physics research. Despite this diversity of positions, TF members and interviewees recognise that there is a general drive toward the adoption of (more) Open Access (OA) policies. An important question is therefore to identify potential OS communities and forums where the DB could or should be involved for the benefit of its participating countries.

The various opinions of the interviewees with regard to OS reflect a diversity of interests and stakes. Representatives from academia/universities and publicly-funded laboratories/research centres tend to be more positive and proactive. It is easier for them to adhere to and adopt OS policies, not to mention that some of their research funds are granted under the conditions that their research outcomes are made accessible. Representatives from industry and regulatory bodies are more cautious or hesitant. While they are generally convinced that it is important to share research outcomes, they also highlight the need to clarify upfront what information can be

[2] See the glossary page https://eosc-portal.eu/glossary

shared, under what conditions, and emphasize that legitimate restrictions can apply for many reasons (IP rights, commercial interests, security concerns, etc.).
The interviewees and survey respondents recognised the valuable work undertaken by the NEA DB to make data more easily accessible, and also to develop a platform for improving collaboration around data and codes. They suggested that the following initiatives could be taken towards OS:

1. **To consolidate what the NEA DB is doing**, which is of high value and in the spirit of OS and OA. An example is the current TF on decay heat reference data, which is a very good initiative.
2. **To think of OS together with the other NEA divisions** as the OS cultural change and the implementation of new practices would be beneficial to all the working areas of the NEA.
3. **To demonstrate the benefit of OS practices through projects**, and these demonstrations will contribute to a better understanding of the OS benefits within the nuclear community.
4. **To assess to what extent arising nuclear needs could be better addressed by OS initiatives**, and the expected benefits, such as more industry (stakeholders) engagement.
5. **To prepare an information booklet on OS for the purpose of helping the various NEA stakeholders** understand the underlying concepts and benefits.

Among the interviewees, only two were experienced OS practitioners. Their recommendations were to initiate deeper actions, at the NEA level, to establish and promote OS and OS culture. Francoise Genova, who is an emeritus astronomer at Strasbourg Astronomical Observatory – one of the founding institutions of the disciplinary international data sharing framework, the so-called astronomical Virtual Observatory – highlighted:

> *"The first goal for the NEA DB should be to develop and promote OS culture in the nuclear community by fostering OS/OA practices and providing an OS infrastructure. OS is about practices and its adoption implies a cultural change".*
> *"OS should not be adopted only at the NEA DB, but also at contributing organisations. A cultural change is needed, which involves understanding the big pieces making up the system. Researchers, developers, other actors, everybody should understand what is and means OS in their discipline. A global engagement is needed, it is not only thinking about databases, but also about involving the users and training them".*

Alberto Mengoni (CERN/ENEA), who is the Italian representative at the G7 Working Group on Open Science, recommended to consider the same approach as the one already developed in European Research Centres.

> *"The NEA DB could consider a similar approach as CERN and the Computer Physics Community, which support developers by peer-reviewing open-source codes before distribution".*
> *"The NEA DB could also promote the establishments of agreements for jointly developing and sharing data and codes".*
> *"The NEA DB could suggest ways of helping open data providers and open-source code developers get more recognition."*

Many NEA DB member countries consider that the NEA and its Data Bank would be well-placed for promoting OS in the nuclear energy field. They supported this DB-OpeNER TF initiative and gave the following inputs:

> "The active participation of international organisations such as NEA to promote OS among its member countries is essential and should also be done among non-members. The activities should be focused on training in OS context, tools

for FAIR Data Management Plan and functionalities of OS repositories."

"The NEA DB could take a leading role at the NEA in assessing how OS initiatives could attract more industry (stakeholders) involvement."

"OS principles should be broadly adopted, as they are essential for sustainable progress in technology and science."

"If the NEA and NEA DB lead the OS movement in the nuclear energy field, NEA member countries will be more supportive for OS projects and initiatives as they view NEA initiatives as flagship programmes for member countries."

"The NEA DB could help in providing guidance regarding opening and sharing data. Another important aspect of long-term data preservation is the connection of different document management systems into one general system."

"OS concept in general has a goal to enlarge the capability of publishing research results in different OA formats."

"The NEA DB could be a "Centre of Excellence" for the integrated Verification, Validation, Benchmarking of Data Evaluation and Software Validation."

"The NEA DB could provide a standardised High Performance Computing (HPC) platform, which should guarantee suitability of codes hosted at NEA, plus all produced data within any benchmark/report at NEA."

"It would be good for the NEA DB to provide a working environment such as GitLab to host joint work as it would facilitate collaborations among various groups, including universities."

"RDUF expects extending the activities in regard to OS globally and considers the NEA as a key actor regarding global OS."

"The NEA DB could provide support to Universities which are the key stakeholders for OS, OD and OSS (Open-Source Software). Professors, students, postgraduate students, and young researchers should be involved in such projects since they are the future generation. They can be "test approvers" in developing OSS and in testing some benchmarks."

"Sharing data and codes improves their recognition. The NEA DB should foster this by negotiating with the authors/owners."

2.2. The NEA DB evolution in the context of OS – Data, databases and related software

All interviewees and survey respondents recognised the importance of the NEA DB services, and expressed a high level of satisfaction with them. They also expect that the NEA DB will continue providing free and (un)restricted access to its basic nuclear data files and related software: JEFF evaluated nuclear data file, EXFOR experimental data, JANIS visual software, processing codes, experimental benchmark databases and related software: ICSBEP[3] (DICE), IRPHE[4] (IDAT);

[3] ICBEP, the International Criticality Safety Benchmark Evaluation Project (https://www.oecd-nea.org/jcms/pl_24498/international-criticality-safety-benchmark-evaluation-project-icsbep)

[4] IRPhE, the International Handbook of Evaluated Reactor Physics Benchmark Experiments (https://www.oecd-nea.org/jcms/pl_20279/international-handbook-of-evaluated-reactor-physics-benchmark-experiments-irphe)

DICE[5], IDAT[6] SINBAD[7], SFCOMPO[8], NDaST[9] and nuclear data sensitivity profiles and other data.

The interviewees and respondents highlighted the following needs regarding data and databases, which are a call for either improved or new services, while stressing the importance of implementing dedicated platform(s)/repository(ies) to store data and databases.

First, there are needs for a stable, supported and QA-controlled JEFF library with nuclear data evaluation fully documented, with access not only to the basic (JEFF) data files, but also to the code-specific application libraries derived from those basic files. For the Verification and Validation (V&V) of nuclear data, the recommendation is to consider developing open-source code validation suites including the description of ICSBEP and other experiments. The results would be open, even when the software used is not open-source.

Second, users would like to have access to more benchmark-quality data including benchmarks to support safety analyses. There are increasing needs in new high-fidelity benchmarks and measurements to support multi-physics (MP) simulations. As complexity increases, the issue is not only to provide access to these data but also to get the help from specialists for properly using such data in connection with computer models.

Third, in view of the increasing data volumes being produced, there are needs in implementing efficient “search, find and select” functions in the NEA databases. In addition, these databases would become more user-friendly if they could include additional information such as summary descriptions, metadata, limitations, etc. They could also integrate code inputs corresponding to benchmark calculation models (for various codes, that would be linked as well), link the data with the relevant NEA reports and/or application examples, ancillary information, etc. The NEA DB could help the users find the relevant information and provide recommendations on how to use some data, e.g. to validate nuclear data using NEA DB codes. Last, users also expect to have clearer guidelines on the terms of use of the data, i.e. to know “what is allowed to do with the data”.

Fourth, storing data over long timescales always raises issues regarding the long-term maintenance and ageing of data and databases. Indeed, as an example, ICSBEP contains many instances of old data whose original authors/suppliers are no longer active. The NEA DB could exchange with the owners or custodians of legacy data to see what (raw) data would be worth curating and sharing with others NEA (DB) member countries. Assistance of the NEA DB would be welcome in identifying appropriate standards for preserving information/data properly, and how to check/use archives by future generations.

Fifth, the larger issue of saving and preserving valuable data was highlighted and interviewees were eager to know how the NEA DB/NEA plans to prepare the storage of and access to large volumes of shared data. The NEA DB has already launched the implementation of a locally hosted GitLab-platform to streamline its services, and will continue acting proactively in this field.

Sixth, it was also suggested that the NEA DB could act as a facilitator for utilities and other “data owners” to convert their data into “something useful” for the broader nuclear energy community. There are untapped reservoirs of data, in particular in the areas of Nuclear Power Plant (NPP) operation, including their decommissioning, that could be seeds for the making of additional reactor benchmarks, curated operational records, etc. Some organisations, the owners of these data, are willing to share these data but lack the manpower to do it.

---

[5] IDAT, the International Reactor Physics Handbook Database and Analysis Tool (https://www.oecd-nea.org/jcms/pl_20296/international-reactor-physics-handbook-database-and-analysis-tool-idat)

[6] DICE, the Database for ICSBEP (https://www.oecd-nea.org/jcms/pl_20293/database-for-icsbep-dice)

[7] SINBAD, the Shielding Integral Benchmark Archive and Database (https://www.oecd-nea.org/jcms/pl_32139/shielding-integral-benchmark-archive-and-database-sinbad)

[8] SFCOMPO, the Spent Fuel Isotopic Composition (https://www.oecd-nea.org/jcms/pl_21515/sfcompo-2-0-spent-fuel-isotopic-composition)

[9] NDaST, the Nuclear Data Sensitivity Tool (https://www.oecd-nea.org/jcms/pl_32450/nuclear-data-sensitivity-tool-ndast)

Seventh, it was made clear by the majority of interviewees that a crucial issue of OD and OSS discussion is their reliability and the related curation and quality control issues. Making the data open and accessible is only one aspect of the problem. Another key aspect is to turn them into useful information. Expert knowledge and work are required to curate these data and prepare benchmark-quality data.
Finally, based on these discussions, the interviewees/respondents made the following suggestions for the future activities of the NEA DB regarding data and datasets:

- **For all types of data (from nuclear data library to experimental benchmark), the NEA DB should take actions to address reliability and quality control issues,** such as:
    - Commit to providing validated and trusted data using appropriate criteria and tests;
    - Organise the provision of data curation services (reviewing and/or evaluation) through appropriate mandated bodies;
    - Provide the infrastructure to host the data (e.g. GitLab based), and include the reviewing work into it (including the definition of metadata and format).
- **For legacy and raw data, the NEA DB could take actions to improve their reuse and sharing,** such as:
    - Organise the review of available raw data and collaborate with owners/custodians to identify what data would be worth preserving, curating and sharing with other NEA DB member countries (a case in point is the CSNI Senior Export Group on Key Datasets Preservation);
    - Establish guidelines for data owners on how to convert their data into “something useful” for the broader community.
- **To answer users’ needs, the NEA DB could provide recommendations on how to use data**, such as:
    - How to test and validate nuclear data using NEA DB codes;
    - How to use experimental and benchmark data in the area of multi-physics modelling and simulations.
- **For the increasing volumes of data to be managed and the related needed infrastructure, the NEA DB could work in the following directions:**
    - Participate in NEA Joint Research Projects (JRP) at their initial stage for discussing data management plans (access during and after the project, including after the non-disclosure period and the data licensing), including discussions on the content and format expected by the users, and on data quality control, formatting and reporting. The NEA DB is expected to play a role in advocating more open approaches, highlighting their benefits to convince JRPs partners to adopt them. The NEA DB could take the lead in recommending a template Data Management Plan (DMP) framework for NEA JRPs;
    - Provide expertise to contributing organisations on data management practices, taking developers’ and users’ needs into account;
    - Consider hosting services associated with OS data management (platform integration, digital preservation) thanks to technical cooperation with interested stakeholders to understand their needs and identify common problems and leading projects;
    - Cooperate with other international initiatives and existing platforms in OS such as e.g. CERN Zenodo project, EERA activities on open science and digitalisation, IAEA initiatives, ongoing H2020 projects such as ENTENTE or ORIENT-NM, etc.

2.3. The NEA DB evolution in the context of OS – Computer software

All interviewees and survey respondents highlighted the importance of the NEA DB Computer Program Services (CPS) and expressed their high level of satisfaction with the current service offering. It is important to have in mind that most DB services' users are "consumers" of data/codes, with a small percentage of them contributing with their own data/codes.
The interviewees and respondents noted the following.
First, they confirmed their needs for an open access to computer codes, i.e., source codes, as well as binary files or beta versions in special open repositories. They expect to find these resources on a NEA DB platform. Such a platform would be a way for OSS developers (and OD providers) to get more recognition and international visibility. The NEA DB distributed OSS packages should include the relevant documentation, instructions, examples on how to use them, and (link to) related publications. In addition, there are many small code developments by students that the NEA DB could consider hosting when judged mature enough to be reused. The NEA DB would thus give them international visibility once they passed some review.
Second, academia and universities are the main stakeholders when it comes to OD and OSS. Getting access to OSS is very important for addressing academic research and training needs. They should be the targeted audience to organise joint activities around OD and OSS. The NEA DB could help universities to get access to OSS, and to find the "best code" for a particular application. Such codes would allow "learning by simulating" and would thus complement traditional Education and Training (E&T) methods based on books, manuals, etc.
Third, quality control is the crucial issue for OSS and associated data. In order to foster OSS development and validation, the NEA DB could consider a similar approach as CERN and the Computer Physics Community, which supports developers by peer-reviewing OSS before distribution. The NEA DB has already started to work on code validation suites to support V&V of nuclear data including OSS such as Open MC, NJOY, and FRENDY. This practice should be extended to other (un)restricted software and go beyond nuclear data V&V; and the outcomes of these activities should be made openly available. The NEA DB could also organise training that would go beyond the use of a specific software and include dedicated sessions on V&V and benchmarking practices and/or specific modelling issues.
Fourth, software license is an important topic for both restricted and open-source software, in particular when supporting their dissemination and re-use. It is important that OSS original authors chose and make clear what licence applies to their software. Indeed, there are several types of open-source licenses, each having its specificities, such as e.g., the LGPL (GNU Lesser General Public License) or the Creative Commons licenses, as CC-BY-NC-ND (Attribution Non-Commercial No-Derivative works). The NEA DB should pay attention to this point when attracting OSS and could even encourage the adoption of licenses which are most suited for e.g., the update or improvement of a code through the integration of users' contributions while acknowledging the contributors and preserving the open-source nature of the software.
Fifth, the NEA DB has already launched the implementation of a GitLab platform hosted on NEA servers. GitLab is one of the main platforms for collaborative software development and is designed to host and facilitate collaborative work. This platform could facilitate the sharing of data and codes, and also of the different steps and processes undertaken to improve/test/curate them. The needs for repositories and other related infrastructures would require more detailed discussions with code developers, users and Information Technology (IT) experts in order to find the best technical solution. This point should be accounted for when developing a strategy and selecting a platform for hosting OSS, the work done in OS on TRUST (Transparency, Responsibility, User community, Sustainability, Technology)[10] repositories would provide helpful insights in this respect. In parallel, the GitLab platform can become a means of fostering

[10] For a definition, see e.g. the Research Data alliance contribution https://www.rd-alliance.org/trust-principles-trustworthy-data-repositories-%E2%80%93-update

collaborations among code developers and advanced users, and help the management of users' communities. If OSS developers (and OD providers) find this co-operation beneficial, they will contribute more, with additional data and codes. Private companies may also be willing to contribute to such collaborative endeavours fostered by the NEA DB, including OSS development and OD evaluation as long as they see benefits for their clients. One interviewee even suggested that the NEA DB might take the initiative to foster and lead the joint development of an open-source Monte Carlo code similar to MCNP by the NEA DB countries. The survey respondents were very positive regarding the implementation of a NEA-hosted GitLab platform and the organisation of collaboration around computer codes. As soon as the GitLab platform is implemented, some research institutes expressed their willingness to contribute computer codes (both open and restricted ones), experimental data and benchmark for collaboration activities, and more publications. Universities may contribute with experimental data and physics models embedded in suitable simulation tools, computer codes and share their experience by organising training courses.
Last, many respondents stressed the importance of training courses, they expect that this activity will be increased and extended for open-source codes. They also mentioned the efficiency of workshops and forums, such as e.g., the Serpent code forum which is very valuable to users.
In short, the interviewees/respondents made the following suggestions for the future activities of the NEA DB regarding computer codes:

- **To modernise its infrastructure and services, the NEA DB should continue its efforts to implement a collaborative platform that would allow the following:**
    - Share and preserve nuclear computer codes, including additional re-usable information on testing and working operating systems.
    - Host open-source code development as a first step. OSS produced at/by the NEA can be hosted at the NEA DB platform. As a second step, to start collecting and integrating software developments done by university students. Other organisations may share their small computer codes, as they can be beneficial, especially for small countries/small institutions, which may lack work force.
    - Host collaborative work on restricted software, with the appropriate licensing framework to handle IP rights.
- **To provide more guidance to its users, the NEA DB should expand its work on quality assurance:**
    - Provide clear guidelines to code developers on the requirements to have their software distributed by the NEA DB, including guidelines on the tests to pass;
    - Provide clear information to code users regarding the tests passed, and, when possible, give them access to the GitLab pipeline;
    - Develop V&V pipelines allowing users to access partial code validation suites, including (un)restricted software, and these pipelines should link software with other NEA "products" such as e.g. ICSBEP;
    - Propose a scheme for implementing a peer-reviewing workflow for new code developments and sharing;
    - Provide more user-oriented information on codes in general.
- **To strengthen its educational activities, the NEA DB should continue organising training courses, and include OSS and the use of NEA developed tools (such as e.g., DICE) in its offer**
- **To facilitate the use of OSS and OD in international cooperation, the NEA DB should also:**
    - Help OSS developers in the licensing process of their software, as many OSS licenses exist;

- Collaborate with other international initiative such as e.g., the ONCORE[11] initiative.

3. Conclusions

The TF-led interviews and surveys highlighted a **very high level of users' satisfaction with the current NEA DB activities and services** on nuclear data, computer codes, training courses, access to experimental benchmark databases and joint-project outcomes. **It is expected that these services will continue to be provided by the NEA DB, as many users depend on them.**

3.1. Expression of needs and desired improvements

In general, **users expect more data and codes to become accessible, in a form that facilitates their re-use.** This means that the available data should be as comprehensive as possible, including all necessary ancillary information to facilitate their re-use (meta-data, documentation, use-cases, benchmarks, input files for modelling). Computer programs are expected to be available as source codes, and not just as binary codes.
It is suggested that **the NEA DB should collect information on the actual end-use made of the distributed data and software, and to make that information broadly available.** This would facilitate the creation of users' communities, the improvement of data and codes, and the identification of unanswered needs.

3.2. Suggestions for new activities

Many interviewees would welcome **the development of an NEA DB-hosted modern web-based collaboration platform, so that contributions from various groups can be efficiently and seamlessly integrated, providing several benefits.** Even in situations of undisclosed proprietary data and codes, it is felt that some organisations may be willing to make subsets of their data or parts of their codes available in a DB-hosted repository for education and training purposes.
As it is anticipated that digitalisation will inevitably lead to larger volumes of data being generated in the coming years, **it is essential that the NEA DB properly ensure that the data and codes have efficient search-retrieve-update browsing features**. Building on its expertise with databases and on the lessons learnt with legacy data, **the NEA DB could take a leading role to help contributing organisations and the NEA JRPs adopting data management best practices**.
As several NEA DB stakeholders feel insufficiently informed about OS, they expressed strong interest in having **the NEA DB act as a « middleperson » between them and the OS community and provide NEA DB stakeholders with more information on OS**. In many ways, the NEA DB has a long history of promoting open access to data and codes. It is expected to continue fostering the emergence of more OA data and OS codes, while preserving the added value of testing and quality-controlling all the data and codes it distributes.

[11] The Open-source Nuclear Codes for Reactor Analysis (ONCORE) initiative is an IAEA-facilitated international collaboration framework.

3.3. Take-away messages from the interviews, survey and followed Workshop

It was clear from the workshop discussions that **the central issue is neither the data and codes nor the infrastructure provided by the DB, but the way our community works together**. The experience of other communities having adopted OS practices can help our community transition to more efficient collaborative processes. Although the OS model is no silver bullet, it can be a game-changer, as some organisations have already experienced. **Expected benefits are increased reactivity and efficiency, reduced development costs, easier integration, improved quality and confidence in the results, increase stakeholders' satisfaction, and higher visibility of the outcomes.**

The NEA DB has been fostering open information exchanges for decades, thereby being an OS/OA precursor for the nuclear energy community. It has therefore *de facto* legitimacy to spearhead a transition in our collaborative model and practices. If such a transition is decided, it will require a mind-set change in our community, as well-established practices have taken root after so many years of valued DB services. A natural resistance to change is to be expected.

This transition will likely not succeed unless it builds momentum on needs-driven activities and projects (JEFF file, open-source code developments …), with all the stakeholders involved and motivated. The NEA DB and the NEA in general should act as catalysts for implementing OS practices in connection with these projects. Top-down and bottom-up incentives will have to be provided. Special consideration should be given to involving universities and students, so as to prepare the next generation of stakeholders.

4. Appendix 1: TF Mandate

MBDAV Task Force DB-OpeNER "Open Science and Nuclear Energy Research Data: Emerging Needs, Opportunities, and Possible Role for the NEA Data Bank"

"Nuclear Energy Research Data" encompasses a wide scope of content: data, software, reports, etc. This scope reflects the diversity of content preserved and distributed by the NEA Data Bank.

Background and objectives:

Over the past three decades or so, scientific activities have been highly impacted by both the digitisation and the increasing amount of data that research generates. This evolution triggers questions regarding the efficient long-term (re)use of the data. This and other related issues are part of the on-going movement referred to as Open Science or Science 2.0. Open Science invites us to rethink our ways of working with potential benefits such as efficiency gains for R&D (higher impact of the collective work and shared Research Data), wider access to Research Data, easier re-use and accrued confidence in the Research Data. This rethinking goes along with a corresponding transition towards modern infrastructures for data management. The NEA Data Bank, which was created under the Science 1.0 environment, should therefore reassess its missions and services in order to adapt to such a rapid change and to new expectations.

As decided during its seventh meeting, September session, held on 21-22 September, the MBDAV is establishing a Task Force on "Open Science and Nuclear Energy Research Data: Emerging Needs, Opportunities, and Possible Role for the NEA Data Bank", called "DB-OpeNER". This task force aims at collecting participating countries' needs and expectations in order to develop motivated lines of actions and a consensual vision. The outcome of the Task Force is expected to serve as a basis for proposing adjustments in the Data Bank missions.

Specifically, DB-OpeNER has the following goals:

1. Collect and summarise relevant Open Science background information, also recent feedback and lessons learnt from actual transitions to open science practices and infrastructure;
2. Gather anticipated needs and operational feedback from various specialists involved in nuclear energy related databases, especially database curators, nuclear code developers, and other stakeholders;
3. Organise a stakeholders' workshop to share their views and expectations on the main challenges of nuclear energy research data (in the context of Open Science), on modern working practices, and on the needed infrastructures (at organisation level, country level and international level);
4. Draft high-level objectives for the NEA Data Bank in the area of Nuclear Energy Research Data as input to upcoming MBDAV discussions in the context of the 2023-2027 NEA Data Bank strategic plan preparation.

The MBDAV is scheduled 7-8 June 2021 so a special MBDAV session will be called in the fall 2021 for the in-depth discussion on Open-Science and the drafting of Strategic Statements for 2023-2027.

Task Force Members:

- Mr Robert JACQMIN (France, TF leader)

- Mr Raul BARRACHINA (Argentina)
- Mr Helmut LEEB (Austria)
- Mr Arjan PLOMPEN (European Comission)
- Mr Antti RINTALA (Finland)
- Ms Françoise GENOVA (France)
- Mr Massimo SEPIELLI (Italy)
- Mr Marco SUMINI (Italy)
- Mr Yoshio SUZUKI (Japan)
- Mr Osamu IWAMOTO (Japan)
- Mr Georgy TIKHOMIROV (Russia)
- Mr Cheuk LAU (Sweden)
- NEA Secretariat, Data Bank: Ms Alice DUFRESNE, Ms Elena POPLAVSKAIA

Draft work plan and main deliverables:

1. **2020 Q4: Background information**

a) Prepare a common bibliography: list of selected documents and key information

b) Collect and share* country specific relevant information

c) Draft a short note on the current Open Science environment, including country/organisation initiatives on Open Science (policy statements, definition/implementations of data management plan, etc.) and identification of key stakeholders and experts in the Open Science field

d) Prepare list of interviews, along with a list of questions/points for discussion

*Online meeting: end of November-early December.

2. **2021 Q1: Interviews**

Draft schedule and guidelines early 2021.

Interviews to be held between January and mid-March. Format: 2-3 TF members + interviewee + NEA secretariat. For each interview, the expected outcomes are a short summary record highlighting key messages, needs and suggestions; and additional resources (persons, documents) – identification of potential speakers/topics for the workshop.

Halfway: status meeting (~mid February).

Mid-March: finalise interview material as input for the workshop.

3. **2021 Q2: Workshop**

To be scheduled before the next MBDAV meeting (7-8 June 2021). The workshop will most likely be held online (two 4 hour-sessions).

Potential date for the workshop: end of May (24-28) – early June (1-2). Workshop to be announced early April.

4. **2021 Q3-Q4: Post-workshop actions**

Collect reactions from participants (through 1-hour discussion with TF leader and NEA secretariat) and prepare a short presentation for the MBDAV June meeting.

Over the summer, draft the outcomes of the workshop; include possible missing topics and background document for the fall MBDAV meeting discussion.

In-depth discussion should include possible short-term actions in connection with on-going NEA Data Bank activities and proposals for the next strategic plan 2023-2027 (on-going).

Action	**Date**	**Deliverable**
Organise a first TF meeting before the end of 2020	10 December 2020	Agreement on the TF scope, working approaches and work plan.
Interview of relevant stakeholders to prepare the workshop and provide inputs to the TF	February-April 2021	12 interviews organised (see Appendices 3 and 4). A survey was distributed to MBDAV delegates and to NEA DB CPS liaison officers (LOs). 32 LOs answered, along with four collective feedback from DB participating countries (see Appendices 5 and 6).
Organise a second TF meeting to summarise the outcomes of the interviews and prepare the workshop	March, 29, 2021	Preliminary findings of interviews were shared. Decision on the topics for discussions in the upcoming workshop, identifications of main speakers and of participants.
Organise a workshop	27-28 May 2021	See Appendix 7 (Workshop agenda) and the summary record, published NEA/MBDAV/DOC(2021)11
Report to MBDAV meeting	7 June 2021 and 11 October 2021	Document summarising the main proposals NEA/MBDAV/DOC(2021)12 Presentations by the TF Leader, Robert Jacqmin.
Extract main messages for the MBDAV	December 2021	This document.

5. Appendix 2: Selected references

OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD (2007), https://doi.org/10.1787/9789264034020-en-fr

«The digital disruption of science – Governments and scientists toward an 'Open Science'», Antoine Maire, 2019, https://www.nap.edu/read/5504/chapter/2#10

Progress on Open Science: Towards a Shared Research Knowledge System, Final Report of the Open Science Policy Platform (2020), https://doi.org/10.2777/00139

Business models for sustainable research data repositories, OECD Science, Technology and Industry Policy Papers n°47, OECD (December 2017), https://doi.org/10.1787/23074957

Digital platforms for facilitating access to research infrastructures, OECD Science, Technology and Industry Policy Papers n°49, OECD (December 2017), https://doi.org/10.1787/8288d208-en

6. Appendix 3: TF interview guidelines

NEA Data Bank Task Force on Open Science and Nuclear Energy Research Data Guidelines for interviewing Data Bank service users database managers

February 8, 2021

Background

The on-going movement towards Open Science (OS) is leading the NEA Data Bank to take a prospective look at its services and working methods, so as to foster a wider access to Research Data and the joint production of new data (in a broad sense). The DB-OPENer Task Force was recently created with the goal of "reviewing" the NEA DB services in the context of OS and evolving users' expectations. As an initial step, the TF members have decided to interview NEA DB service users and NEA database managers in order to collect their anticipated needs in terms of research data and new services.

Here below is a list of issues and questions that should be addressed during these interviews. Topics of relevance include: (1) modern repositories, (2) improved practices and joint work, (3) opening/broadening of research data in the nuclear energy field, including educational activities, (4) the DB as an OS portal. "Research data" is to be understood as covering all types of scientific and research data relating to nuclear energy, including experiments, benchmarks, evaluated files, computer codes and related input/output files, publications... The TF seeks positions and viewpoints representative of broad communities.

Objective of the interview:

to identify the needs of the NEA DB users, to share the interviewee's vision on benefits open information sharing can bring into nuclear energy fields and co-operation, to provide the examples of modern working practices, needed infrastructures data in the context of OS and the ideas/suggestions/proposals for evolving the NEA DB services regarding OS; all these inputs will be used as materials for preparing the Stakeholder's Workshop.

Distribution of the guidelines for interview

These guidelines will be distributed among the interviewees identified by the TF DB-OPENer participants. As proposed at the TF DB-OPENer kick-off meeting the list of interviewees may include the NEA DB users, database managers, open source code developer and users, OS practitioners, the representatives of the NEA divisions/NEA STCs.

Schedule of interview and participants

All interviews should be scheduled in February-March 2021 based on the availability of the interviewees and TF participants. The proposed format of interview includes 2-3 TF participants, the interviewee and the NEA DB secretariat. All interviews will be organised remotely via ZOOM/WebEx link provided by the NEA DB secretariat. If due to some reasons the identified expert-interviewees cannot participate in the interview, the TF participant can distribute these guidelines with questionnaire among those interviewees and edit the answers in accordance with expected outcomes.

Expected outcomes

A summary record of the interview should be prepared including a brief information on the interviewee (name of the interviewee, name of organisation, contact information, his/her background/expertise); answers to the key questions presented in the guidelines, including the needs of users and proposals for evolving the NEA DB services regarding OS.

The questionnaire survey

I. Introduction and brief description of the area of work/research/expertise of the interviewee

1. Please specify the area of your work/research/expertise

II. General questions regarding Open Science

1. What is your opinion regarding the on-going movement to Open Science?
2. Could you provide a brief information on the status of Open Science in your country, if there is a support of OS on national level in your country, for example, a national law that obliges researchers to share their results in an Open Science environment?
3. Did you have discussions on Open Science in your organisation?
   - Does your organisation participate at Open Science initiatives?
4. Do you know exactly about what is intended with terms as Open Science, open source, open data, open access, open education?
5. Do you think that Open Science, open source, open data concepts can be smoothly applied to the nuclear field or you see potential hinders in this application?
6. If you find potential hinders, what of them you could highlight – if you wish please add comments to the term you select :
   a. Property
   b. Copyright
   c. Misuse
   d. Malevolent use
   e. Unreliability
   f. Others (please, specify).
7. In case you are positive to application of Open Science concept to nuclear, what do you expect from NEA/MBDAV to foster this approach?
   a. Indicate and suggest to users open data and open-source codes
   b. Dispatch directly open-source codes and open data for free
   c. Trade-off with data owners to acquire and make open former reserved / confidential data / codes
   g. Other (please, specify)
8. Do you think that criteria should be defined for data or source code to be open? If yes do you have criteria to propose?

III. Questions to NEA DB service users – The focus should be on needs, *not* on solutions. All data and services relating to nuclear energy are within the TF scope.

1. What DB data and services do you use? (Examples are JEFF, ICSBEP, SCALE…)
2. What difficulties/limitations do you face in accessing/using those data from the NEA DB? Are you satisfied with the level of information you get with the data, the format in which the data are provided to you?
3. Are there other research data and related services that your organization could immediately make valuable use of, which you would expect the NEA DB to provide but cannot get from it?
4. Beyond traditional data, what additional information or new services (such as joint work) do you anticipate your research organization will need in the coming 10+ years?
5. Suggestions for consideration by the TF

IV. Questions to experienced database custodians or repository practitioners/managers – The focus should be on lessons learnt from managing databases, interfacing with data providers and data users.

1. What are the specific databases that you oversee and type of information stored?

2. What type of infrastructure is used for these databases, what are its limitations and expected evolutions?
3. How is the database filled and managed?
4. What are the typical updating cycles?
5. Do you monitor database usage and how?
6. Who are the users, how many are they, what do they do with the data?
7. How do you interact with your users, and how can they provide feedback?'
8. Are there joint collaborative activities revolving around the supply or use of data?
9. In broad terms, what is the feedback collected from users?
10. What evolutions have you observed in usage and users' expectations over the past 10 years?
11. Is your organisation using or planning to use GitHub, GitLab, SourceForge or any equivalent platform?
12. Suggestions for consideration by the TF

V. Questions to open-source code developers and users

1. What is your opinion regarding development of open-source codes?
2. Can you provide some valuable insights regarding the development of open source codes in collaborative hosting facilities?
3. What kind of challenges do you foresee?
4. If you are an open-source code developer, would you like to submit the open source code to the NEA/DB?
5. Suggestions for consideration by the TF

VI. Stakeholder's Workshop

1. Could you please identify the key stakeholders and experts in the Open Science field, whose participation at Stakeholder's Workshop is very important from the view of preparing the recommendations to MBDAV

VII. Any other question would you like to discuss additionally in the context of this TF

7. Appendix 4: List of interviewees

Interviewee, organisation, country	Field of activity and expertise
Dr Luca Fiorito, SCK CEN, Belgium	Research Centre, researcher
Prof. Alain Hébert, Ecole Polytechnique, Montreal, Canada	University, code developer
Joint inverview of **Martin Elias, RAWRA, Czech Republic** **Dr József I. Fekete, PURAM Hungary**	Radioactive Waste, Information Data & Knowledge Management, database managers
Dr Manuel GARCIA, KIT, Germany	Research Centre, researcher
Dr Alberto Mengoni, CERN/EU, Italy	EU, G7 WG on Open Science
Dr Vladimir Artisuyk, ROSATOM/Russia	State corporation, scientific advisor
Dr Michael Österlund, Uppsala University, Sweden	University, researcher
Dr Jose Ignacio Marquez Damian, ESS, Sweden	European Spallation Source, researcher
Dr Philippe Pinard, Oxford Instruments, UK	Private Company, code developer
Dr Ian Hill, OECD/NEA, Nuclear Science	NSC, reactor physics and criticality safety, database interfaces
Dr Didier Jacquemain, OECD/NEA, Nuclear Safety	CSNI, severe accidents, exp. data preservation
Dr Luminita Grancea, OECD/NEA, Nuclear Technology and Economics	Nuclear Technology and Economics, policy analyst

8. Appendix 5: TF survey sent to nominated Liaison Officers

A quick DB-OpeNER survey

1. **Background**

Please specify your work/research area and the typical use that you make of the NEA DB services.

Also, please specify the contributions that your organisation makes to the NEA DB activities.

2. **Open Science**

Are you or is your organisation involved in the on-going movement on Open Science, open access, open research data?

If so, can you suggest specific OS areas in which the DB could be involved, to the benefit of its member organisations?

3. **Anticipated needs**

Faced with the rapid growth in the production rate and use of research data, some OS groups anticipate that new services and infrastructures will be necessary in conjunction with open access policies, in order to derive the full benefits from the information produced. Within the realm of your activities on nuclear energy, what specific needs do you anticipate in connection with this global trend?

4. **Evolution of the NEA DB services in a long-term perspective**

Are there new or improved services that your organisation would expect the NEA DB to provide in the future in order to meet your needs, complementing the current portfolio of DB services covering nuclear data, computer programs, training courses & workshops?

Please specify the expected services and the corresponding added value for your organisation

5. **Possible contributions of your (or another) organisation to future NEA DB activities**

Would your organisation be willing to contribute to joint collaborative activities using a NEA-hosted integration platform?

If so, please elaborate:

More generally, in the spirit of the rapidly expanding OS movement, can you think of collaborative actions that the NEA DB could legitimately spearhead and host, and to which you would willingly contribute?

9. Appendix 6: List of participating countries/organisations in the LO survey

Survey respondents	Represented organisations
32 NEA DB *Liaison Officers*, in Belgium, Czech Republic, France, Germany, Greece, Hungary, Italy, Netherlands, Poland, Romania, Russia, Slovakia, Spain, UK	Nuclear research institutes, R&D and industrial companies, government, regulator and TSO, universities and medical centres
Japan, Dr Yoshio Suzuki + 6 LOs	Nuclear research institute, Research Centre for Open Science and Data Platform Japan Agency for Marine-Earth Science and Technology
Argentina, Dr Raul Barrachina + 12 people	Nuclear research organisations, regulator, NPP operator, NPP builder, medical centres
Russia, Dr Georgy Tikhomirov + 6 people	Nuclear physics and engineering institute, nuclear safety institute
Belgium, Dr Gert Van den Eynde	Nuclear energy research institute

10. Appendix 7: Workshop Agenda (May 27-28, 2021, 13:00-17:00 CEST)

13:00 – Opening – 5' – NEA DB

13:05-14:50 – Day 1 – Session 1 – Data, databases, JEFF file, joint developments and NEA Data Bank related services – Moderator: Arjan Plompen, TF member

Introduction – 5', Session moderator

1.1 – 10' – Main lessons learned from the DB-OpeNER TF (Part 1), Robert Jacqmin, TF leader

1.2 – 30' – The JEFF4 collaboration challenges, Arjan Plompen, JEFF chair

1.3 – 30' – The CERN's experience in setting up large-scale collaborative data-driven research activities, Alberto Di Meglio, CERN

1.4 – Panel discussion, Q/A – 30' – Topics: Overcoming traditional barriers and legacy practices, engaging all stakeholders, assessing the value of legacy data,…

Conclusion – 5', Session moderator

--- 20' Break ---

15h10-17:00 – Day 1 – Session 2 – Computer codes, Open Source software developments, and NEA Data Bank related services – Moderator: Raul Barrachina, TF member

Introduction – 5', Session moderator

2.1 – 10' – Main lessons learned from the DB-OpeNER TF, outlining the users' needs and important specificities of the nuclear energy community (Part 2), Robert Jacqmin, TF leader

2.2 – 30' – A prospective view on computer program services, Luca Fiorito, SCK-CEN

2.3 – 30' – Fostering a culture of collaboration around shared challenges, Philippe Charrière, GitLab

2.4 – Panel discussion, Q/A – 30' – Topics: users' satisfaction, assuring some quality control, community engagement, OSS best practices,…

Conclusion – 5', Session moderator

13:00-14:50 – Day 2 – Session 3 – Possible role for the Data Bank in connection with the Open Science movement – Moderator: Françoise Genova, TF member

Summary of Day 1 and introduction – 5', Session moderator

3.1 – 10' – Main messages from the DB-OpeNER TF (Part 3), Robert Jacqmin, TF leader

3.2 – 15' – Experimenting Open Science and Open Access to data beyond basic principles: What can we learn from existing initiatives, Françoise Genova, TF member

3.3 – 15' – Research Data Utilization Forum (RDUF), Prof. Takaaki Aoki of Nagoya University the RDUF Planning Committee Chair.

3.4– 30' – Successfully implementing an open research data infrastructure at INRAE, Michel Bamouni, Dimitri Szabo, INRAE

3.5 – Panel discussion, Q/A – 30' – Topics: Involving academia, hosting direct integration services, fostering a cultural change (new model, levers)

Conclusion – 5', Session moderator

--- 20' Break ---

15:10-16:50 – Day 2 – Session 4 – Wrap-up discussions – Moderator: Robert Jacqmin, TF member

Introduction – 5', Session moderator

General discussion – 1h30' – Further exchange on the above topics, expert suggestions on "the best way forward", all participants

Conclusion – 5', Session moderator

16:20 – Closing – 10' – NEA DB

11. Appendix 8: Abbreviations

BY-NC-ND	Attribution Non-Commercial No-Derivative
CC	Creative Commons
CERN	European Organisation for Nuclear Research
CPS	Computer Program Services
DB	Data Bank
DB-OpeNER	Open Science and Nuclear Energy Research Data: Emerging Needs, Opportunities, and Possible Role for the NEA Data Bank
DMPs	Data Management Plans
EC	European Commission
EERA	European Energy Research Alliance
ENEA	Italian National Agency for New Technologies, Energy and Sustainable Economic Development
EXFOR	Experimental Nuclear Reaction Data
FAIR	Findable, Accessible, Interoperable, Reusable
HPC	High Performance Computing
IAEA	International Atomic Energy Agency
INRAE	French National Research Institute for Agriculture, Food and the Environment
IP	*Intellectual Property*
IT	Information Technology
JAEA	Japan Atomic Energy Agency
JEFF	Joint Evaluated Fission and Fusion File
JRP	Joint Research Project
LGPL	Lesser General Public License
MBDAV	Management Board for the Development, Application and Validation of nuclear data and codes
MP	Multi-physics
NEA	Nuclear Energy Agency
NEA DB	NEA Data Bank
NPP	Nuclear Power Plant
OA	Open Access
OD	Open Data
OECD	Organisation for Economic Co-operation and Development

ONCORE	Open-source Nuclear Codes for Reactor Analysis, IAEA initiative
OS	Open Science
OSS	Open-Source Software
RDUF	Japanese Research Data Utilization Forum
R&D	Research and Development
TF	Task Force
TRUST	Transparency, Responsibility, User community, Sustainability, Technology
V&V	Verification and Validation