Containerization and microservices for nuclear data

Georg Schnabel
g.schnabel [at] iaea.org

Nuclear Data Section
Division of Physical and Chemical Sciences NAPC
Department for Nuclear Sciences and Applications
IAEA, Vienna

Workshop for Applied Nuclear Data Activities
29 January 2021
Nuclear data pipeline

Compilation → Evaluation → Processing, Verification, Validation → Applications
Collaborative effort of experts
Technical overhead

Compilation → Cluster → Evaluation → Processing, Verification → Validation → Applications

Seattle Municipal Archives
Empower great minds
Automation

Compilation

Cluster

Examples:
BNL Advance
OECD NEA V&V system

Evaluation

Processing, Verification Validation

Applications
Microservices

Data from EXFOR DB

Evaluation

Model predictions

Q&A Service

Verification service
Microservices

Data from EXFOR DB

Q&A Service

Evaluation service

Model predictions

Verification service
Containerization

“A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.” (quoted from www.docker.com)

Open Container Initiative

The Open Container Initiative is an open governance structure for the express purpose of creating open industry standards around container formats and runtimes.
Evaluation with advanced statistical processing

In this work: parallel computing on ~80 CPU cores sufficient; statistical inference could benefit from use of GPU in the future (but not the bottleneck at present)

925 TALYS parameters considered
144 TALYS parameters adjusted
4322 experimental data points

[2009.00521] Conception and software implementation of a nuclear data evaluation pipeline (arxiv.org)
Results
Evaluation as pipeline

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Retrieval of relevant experimental data</td>
</tr>
<tr>
<td>2</td>
<td>Generation of predictions based on a reference calculation</td>
</tr>
<tr>
<td>3</td>
<td>Rule-based correction of experimental uncertainties</td>
</tr>
<tr>
<td>4</td>
<td>Correction of systematic experimental uncertainties using MLO</td>
</tr>
<tr>
<td>5</td>
<td>Evaluation of the Jacobian associated with the reference calculation</td>
</tr>
<tr>
<td>6</td>
<td>Setup of Gaussian processes for energy-dependent model parameters</td>
</tr>
<tr>
<td>7</td>
<td>Optimization of TALYS parameters using the LM algorithm</td>
</tr>
<tr>
<td>8</td>
<td>Calculation of a MVN approximation of the posterior pdf</td>
</tr>
<tr>
<td>9</td>
<td>Generation of representative random files</td>
</tr>
</tbody>
</table>
## Evaluation as pipeline

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Retrieval of relevant experimental data</td>
</tr>
<tr>
<td>2</td>
<td>Generation of predictions based on a reference calculation</td>
</tr>
<tr>
<td>3</td>
<td>Rule-based correction of experimental uncertainties</td>
</tr>
<tr>
<td>4</td>
<td>Correction of systematic experimental uncertainties using MLO</td>
</tr>
<tr>
<td>5</td>
<td>Evaluation of the Jacobian associated with the reference calculation</td>
</tr>
<tr>
<td>6</td>
<td>Setup of Gaussian processes for energy-dependent model parameters</td>
</tr>
<tr>
<td>7</td>
<td>Optimization of TALYS parameters using the LM algorithm</td>
</tr>
<tr>
<td>8</td>
<td>Calculation of a MVN approximation of the posterior pdf</td>
</tr>
<tr>
<td>9</td>
<td>Generation of representative random files</td>
</tr>
</tbody>
</table>
Implemented Containers (Prototypes)

EXFOR DB Container
- Database software

Pipeline Container
- Data retrieval
- Model fitting
- Writing output

Computation container
- Model and processing codes

Current limitations:
- Pipeline container specialized to Fe56,
- Communication with computation container coupled to R programming language
Advantages

• Composability
• Transparency
• Extensibility
• Reusability
• Transferability
• Reproducibility
• Collaboration

docker run -it -p 9090:8787 \
-v outdata:/home/username/eval-fe56/outdata \
-v talysResults:/home/username/talysResults \
-v /dev/shm:/dev/shm \
-e extUID=<UID> -e extGID=<GID> \
-e maxNumCPU=32 \
--name eval-fe56-cont eval-fe56-img test_eval
Nuclear data pipeline

Compilation → Cluster → Evaluation → Processing, Verification, Validation → Applications

Examples:
- BNL Advance
- OECD NEA V&V system
Conclusion

• Nuclear data pipeline requires diverse expertise and depends on a variety of codes
• Microservices for nuclear data enable human experts to write evaluation scripts concisely codifying what data should be used and what should be done with it
• Microservices as containers allow their easy shipping and reuse
• Prototypes of a nuclear database container and evaluation container available online
Outlook & Challenges

• Interaction with containers should be language agnostic, e.g., use widely employed data structures and protocols

• Microservices & data should be discoverable and usable, i.e., the conception that something is stored somewhere on a specific architecture becomes more and more irrelevant

• (Simple) Programmatic access to databases, model codes, statistical microservices and compute power will be an accelerator for AI/ML as enabling technology
Thank you for your attention!

Thanks to my collaborators:
Henrik Sjöstrand, Joachim Hansson, Dimitri Rochman, Arjan Koning, Roberto Capote
Links & References

• Pipeline paper

• Pipeline code
  https://github.com/gschnabel/eval-fe56

• EXFOR CouchDB database
  https://github.com/IAEA-NDS/exfor-couchdb-docker
Related work

- NEA NDS V&V System
  

- BNL ADVANCE (V&V System)
  