IAEA 60 Years
Atoms for Peace and Development

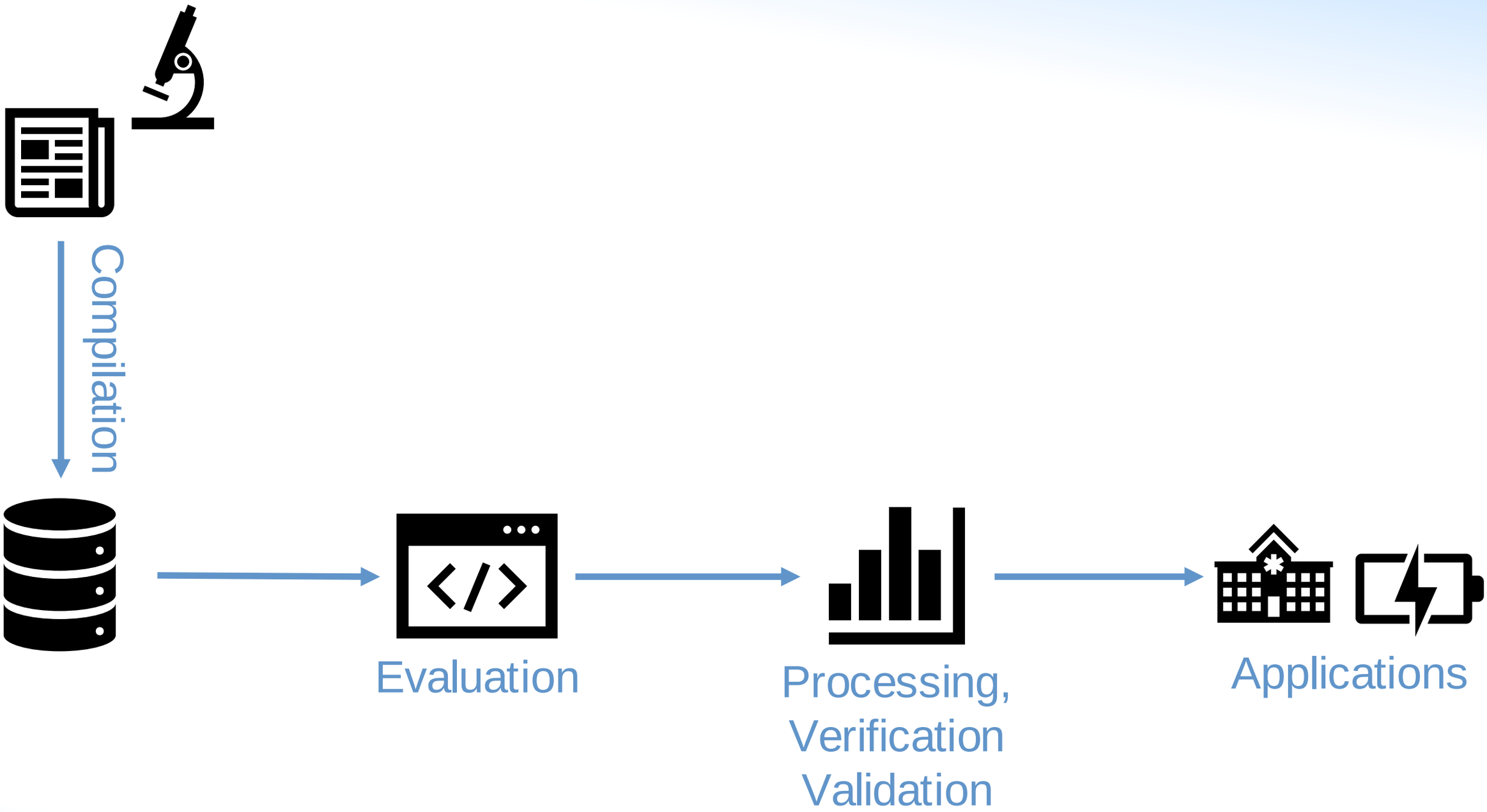
Containerization and microservices for nuclear data

Georg Schnabel
g.schnabel [at] iaea.org

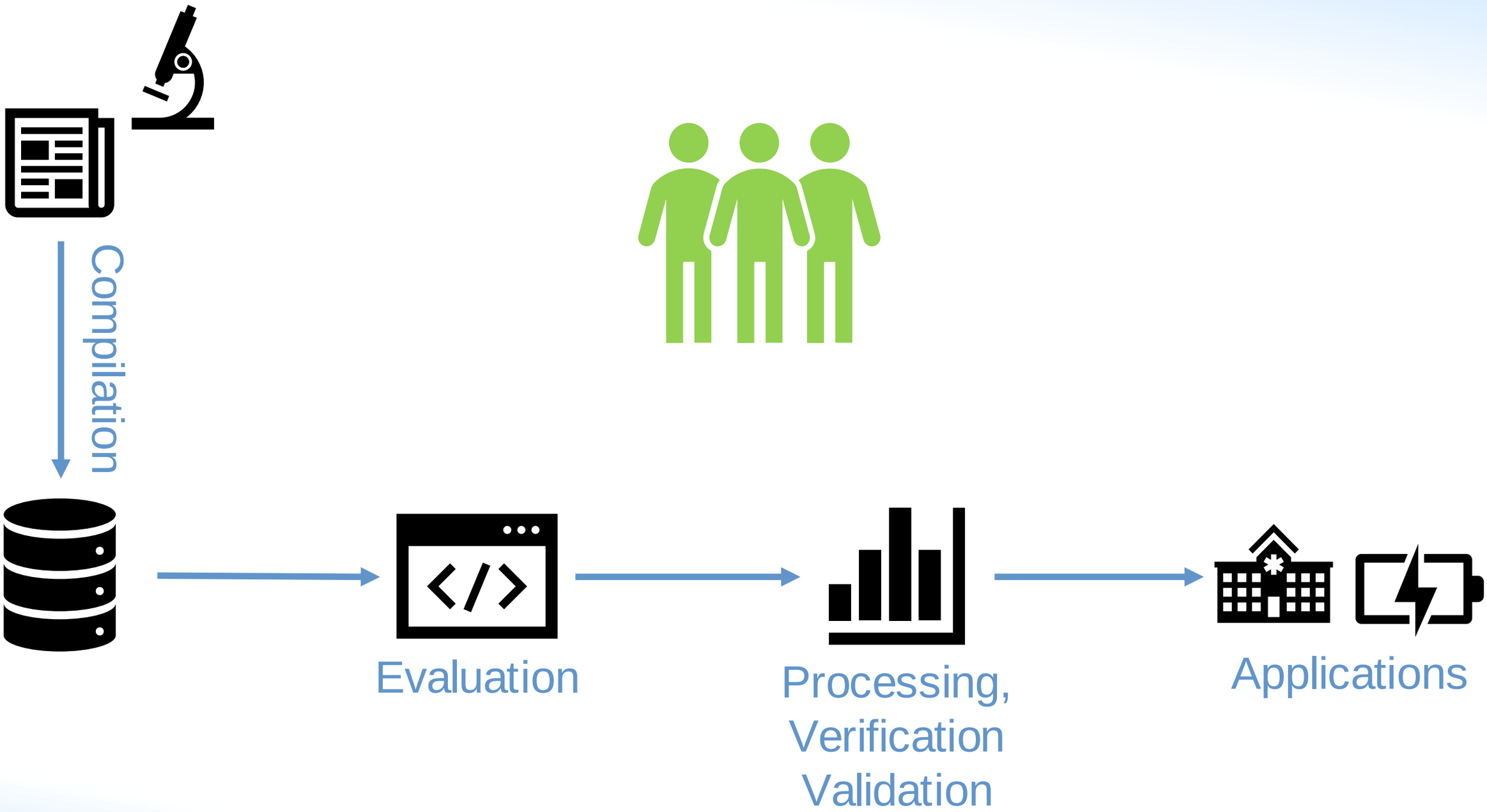
Nuclear Data Section
Division of Physical and Chemical Sciences NAPC
Department for Nuclear Sciences and Applications
IAEA, Vienna

Workshop for Applied Nuclear Data Activities
29 January 2021

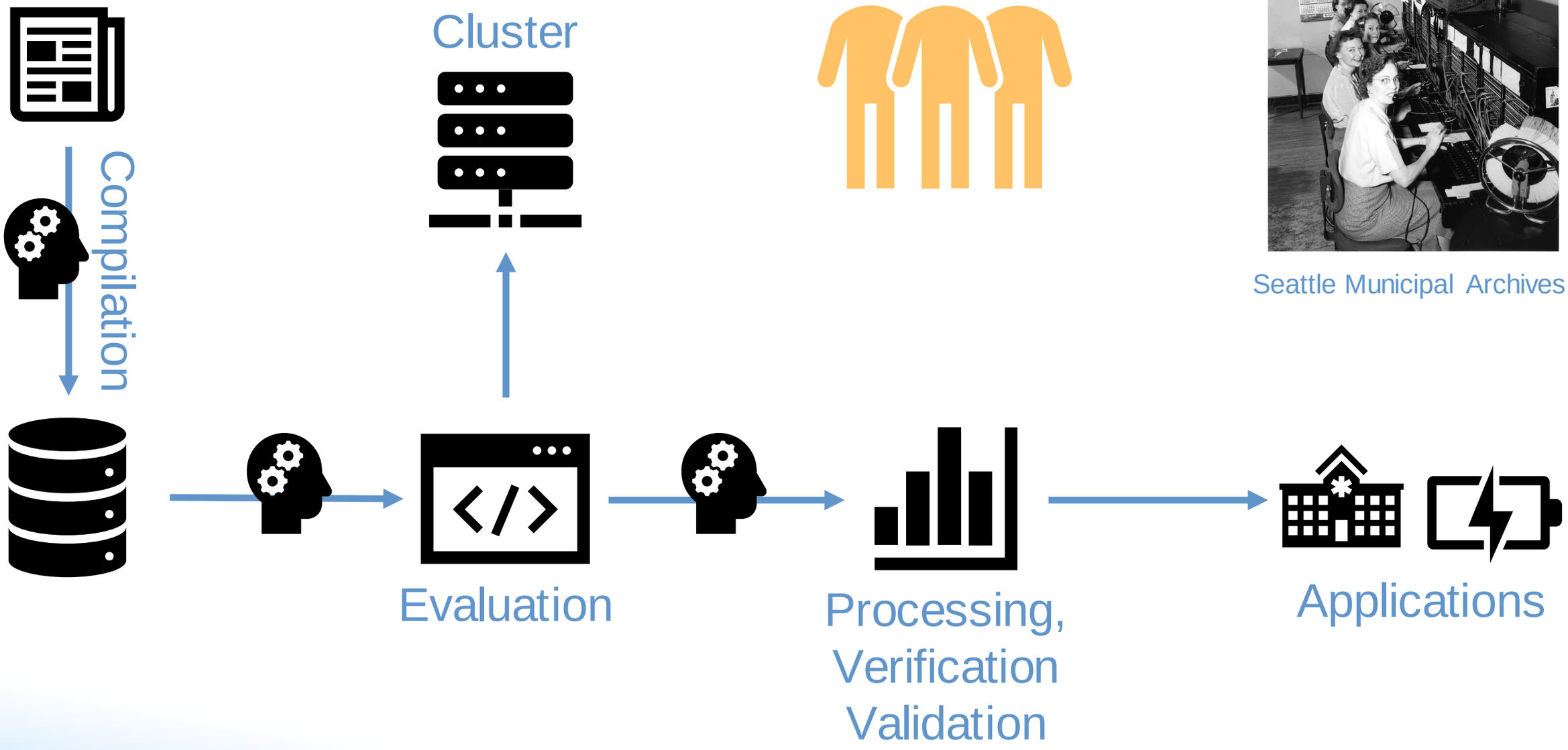
Nuclear data pipeline



Collaborative effort of experts



Technical overhead



Seattle Municipal Archives

Empower great minds

IAEA 60 Years
Atoms for Peace and Development

Seattle Municipal Archives

Cluster

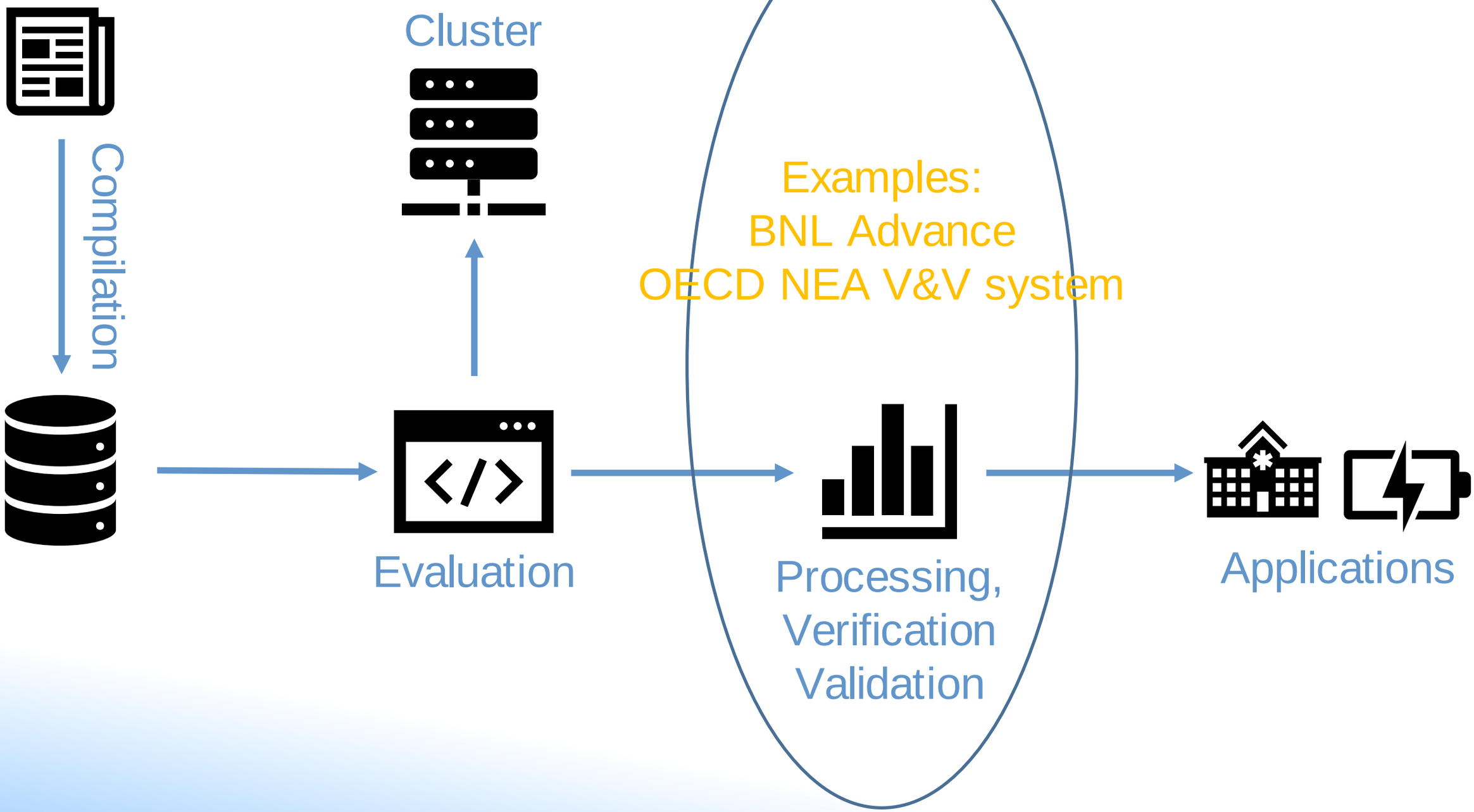
Compilation

Evaluation

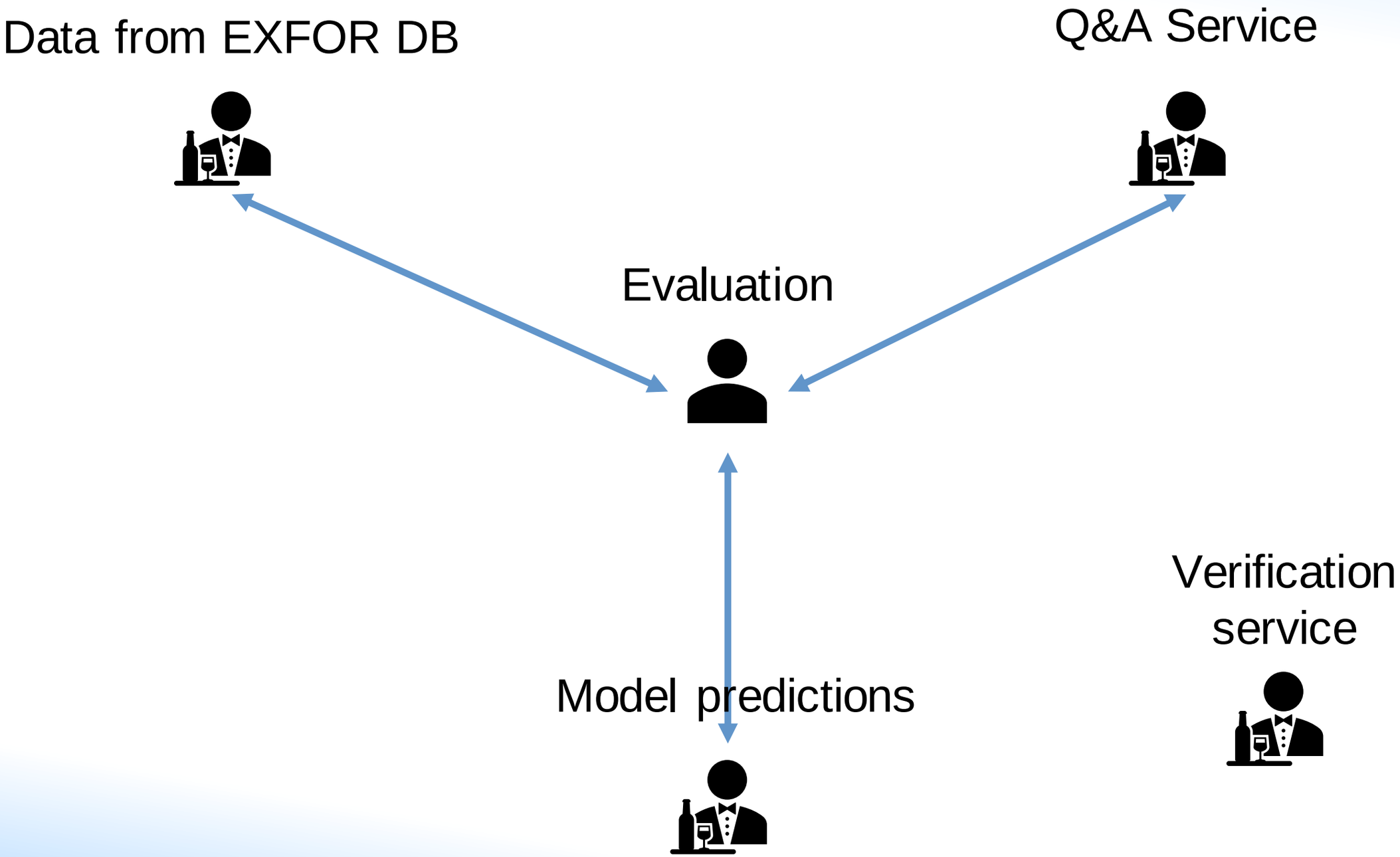
Processing,
Verification
Validation

Applications

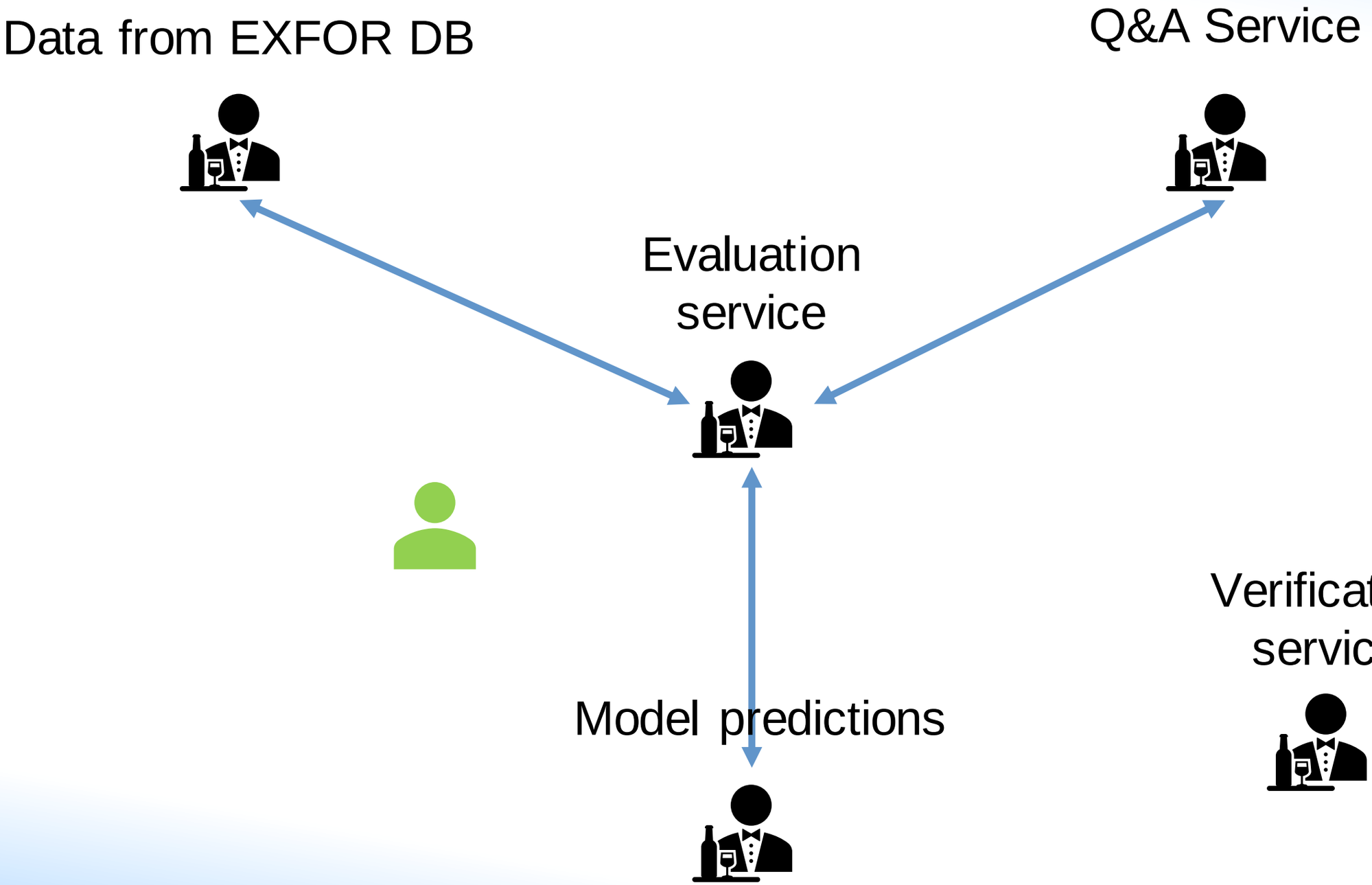
Automation


Compilation
Cluster
Examples:
BNL Advance
OECD NEA V&V system
Evaluation
Processing,
Verification
Validation
Applications

Microservices



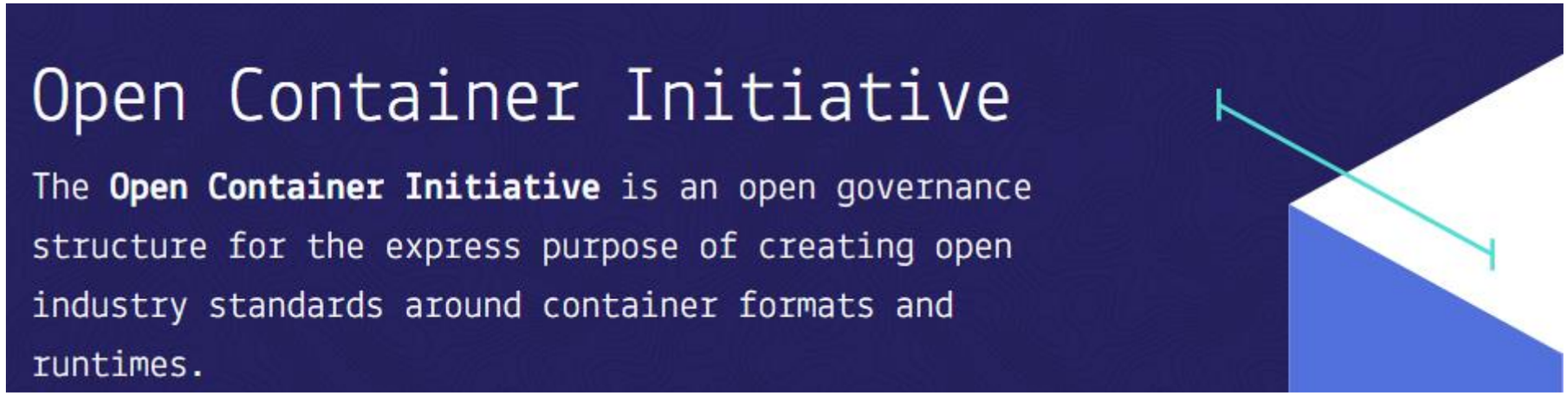
Microservices



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)

https://www.docker.com

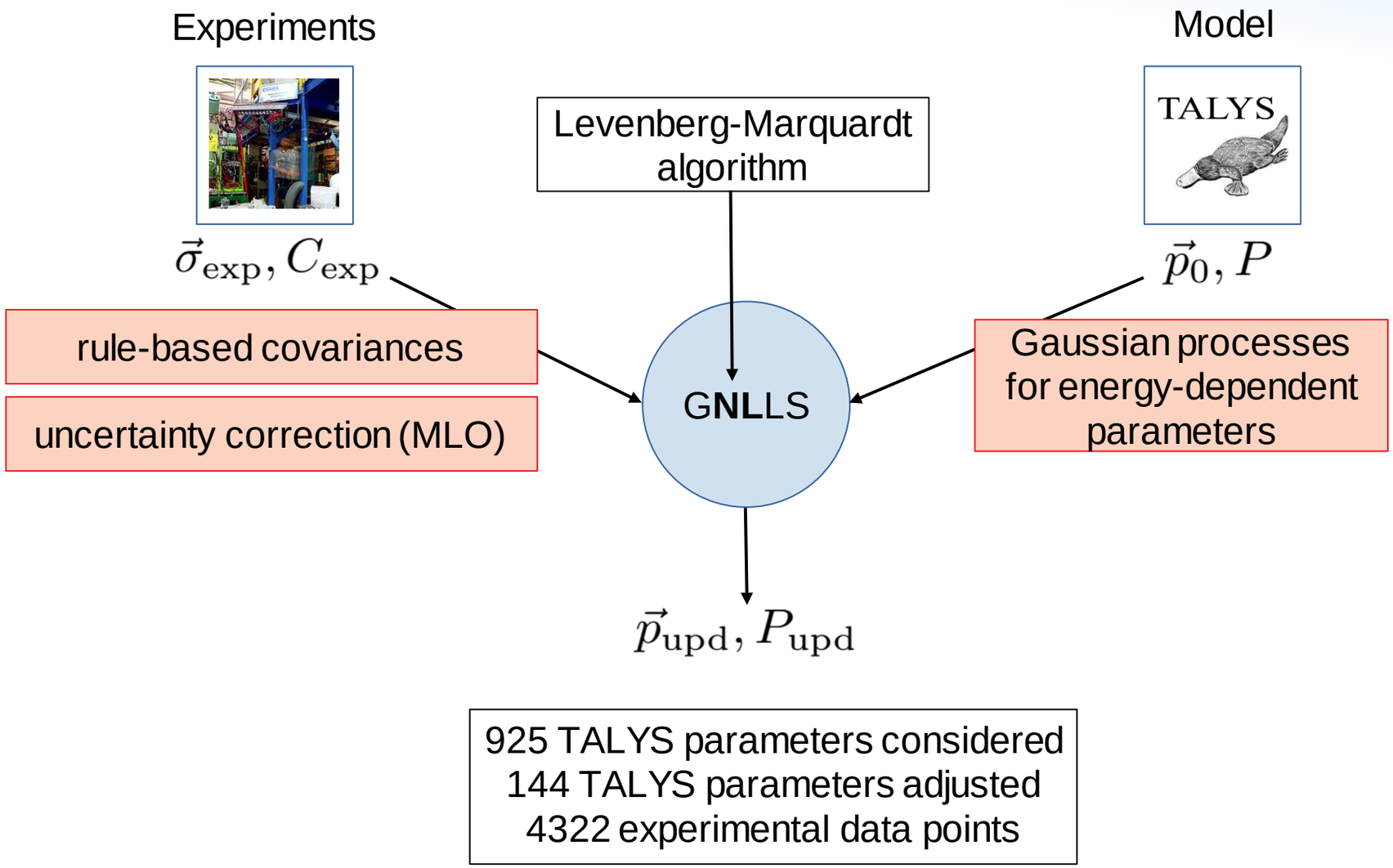


Sylabs.io
SINGULARITY
SIMPLE, FAST, SECURE

docker

podman

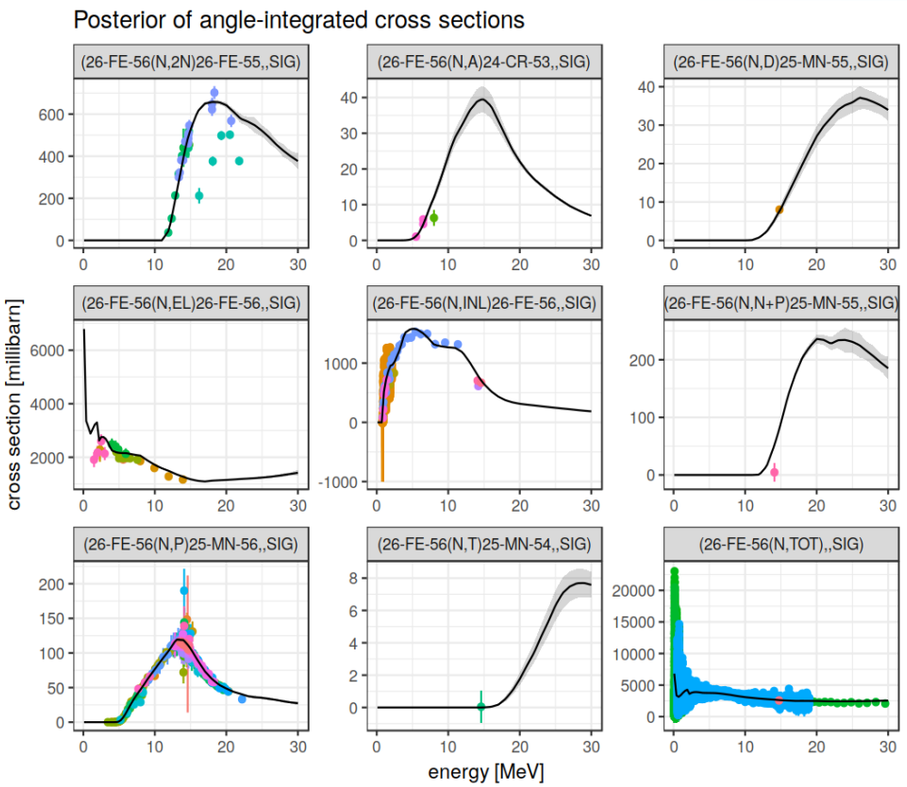
Evaluation with advanced statistical processing

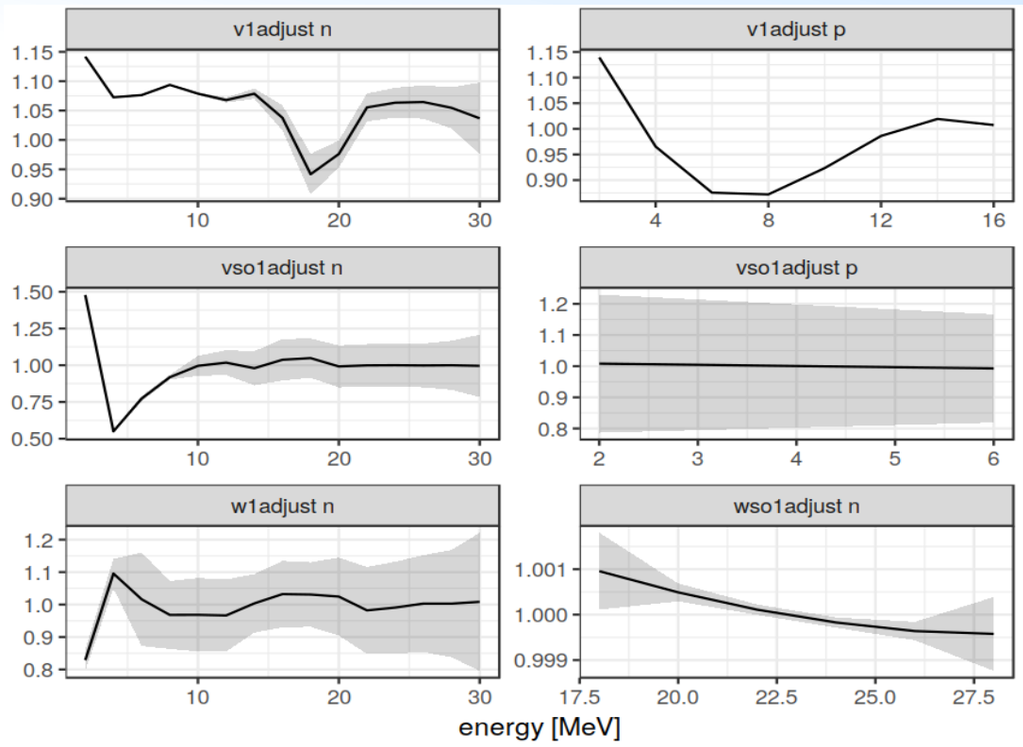


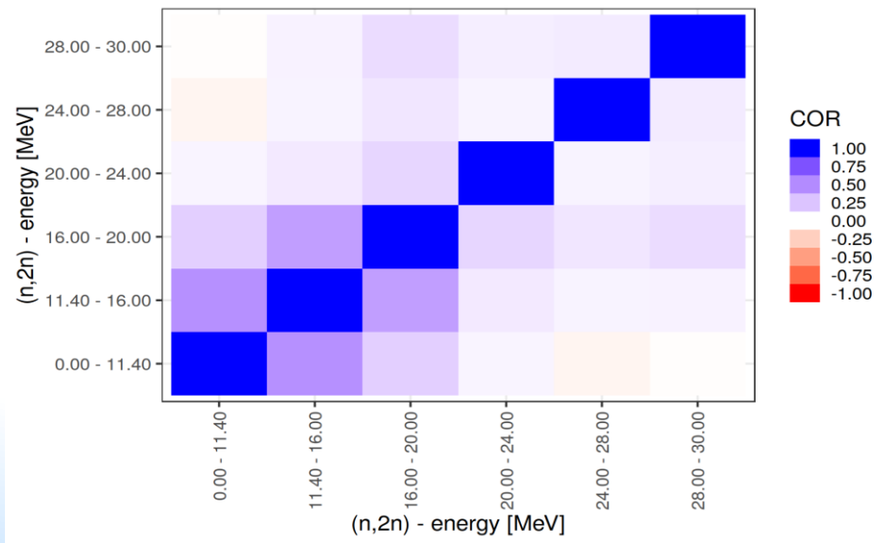
In this work: parallel computing on ~80 CPU cores sufficient; statistical inference could benefit from use of GPU in the future (but not the bottleneck at present)

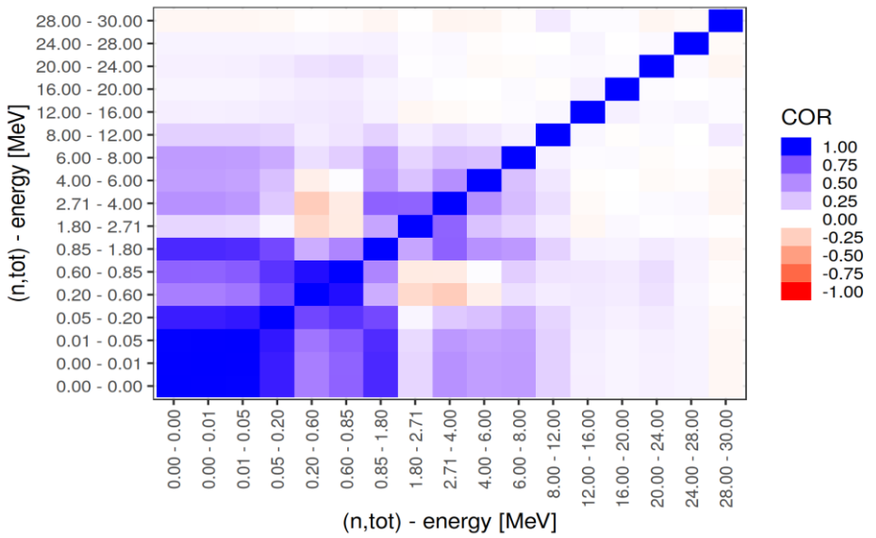
[2009.00521] Conception and software implementation of a nuclear data evaluation pipeline (arxiv.org)

Results


Posterior of angle-integrated cross sections
(26-FE-56(N,2N)26-FE-55,,SIG)
(26-FE-56(N,A)24-CR-53,,SIG)
(26-FE-56(N,D)25-MN-55,,SIG)
(26-FE-56(N,EL)26-FE-56,,SIG)
(26-FE-56(N,INL)26-FE-56,,SIG)
(26-FE-56(N,N+P)25-MN-55,,SIG)
(26-FE-56(N,P)25-MN-56,,SIG)
(26-FE-56(N,T)25-MN-54,,SIG)
(26-FE-56(N,TOT),,SIG)
cross section [millibarn]
energy [MeV]


v1adjust n
v1adjust p
vso1adjust n
vso1adjust p
w1adjust n
wso1adjust n
energy [MeV]


(n,2n) - energy [MeV]
(n,2n) - energy [MeV]
COR


(n,tot) - energy [MeV]
(n,tot) - energy [MeV]
COR

Evaluation as pipeline

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

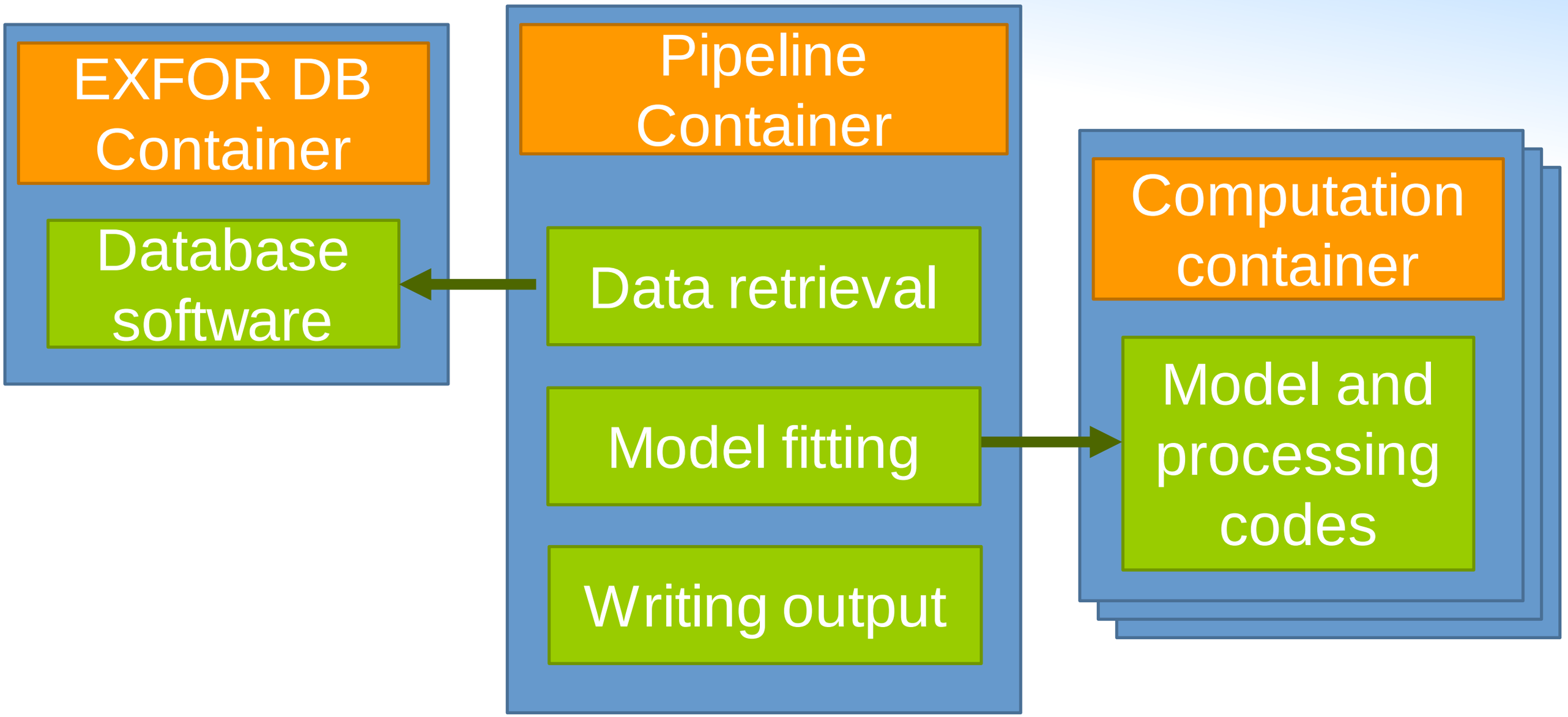
Evaluation as pipeline

Database interaction

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

Model calculation on computer cluster

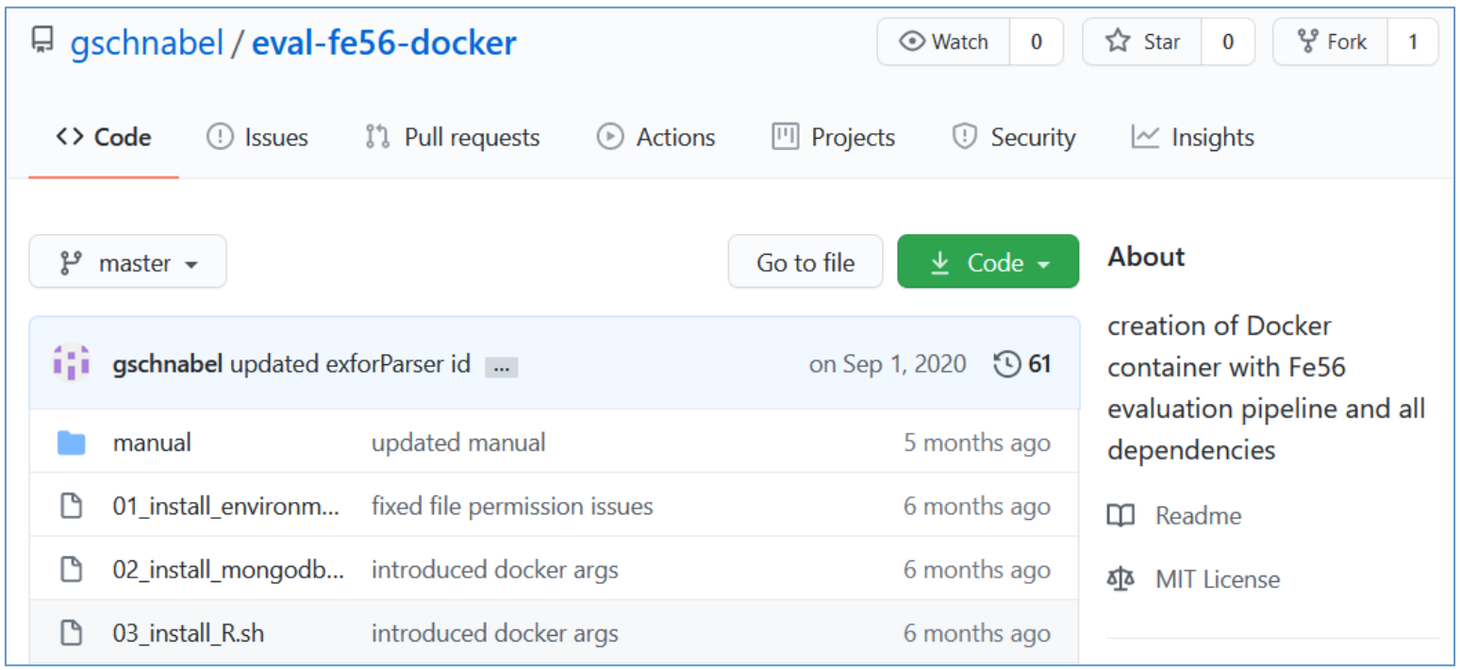
Implemented Containers (Prototypes)



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

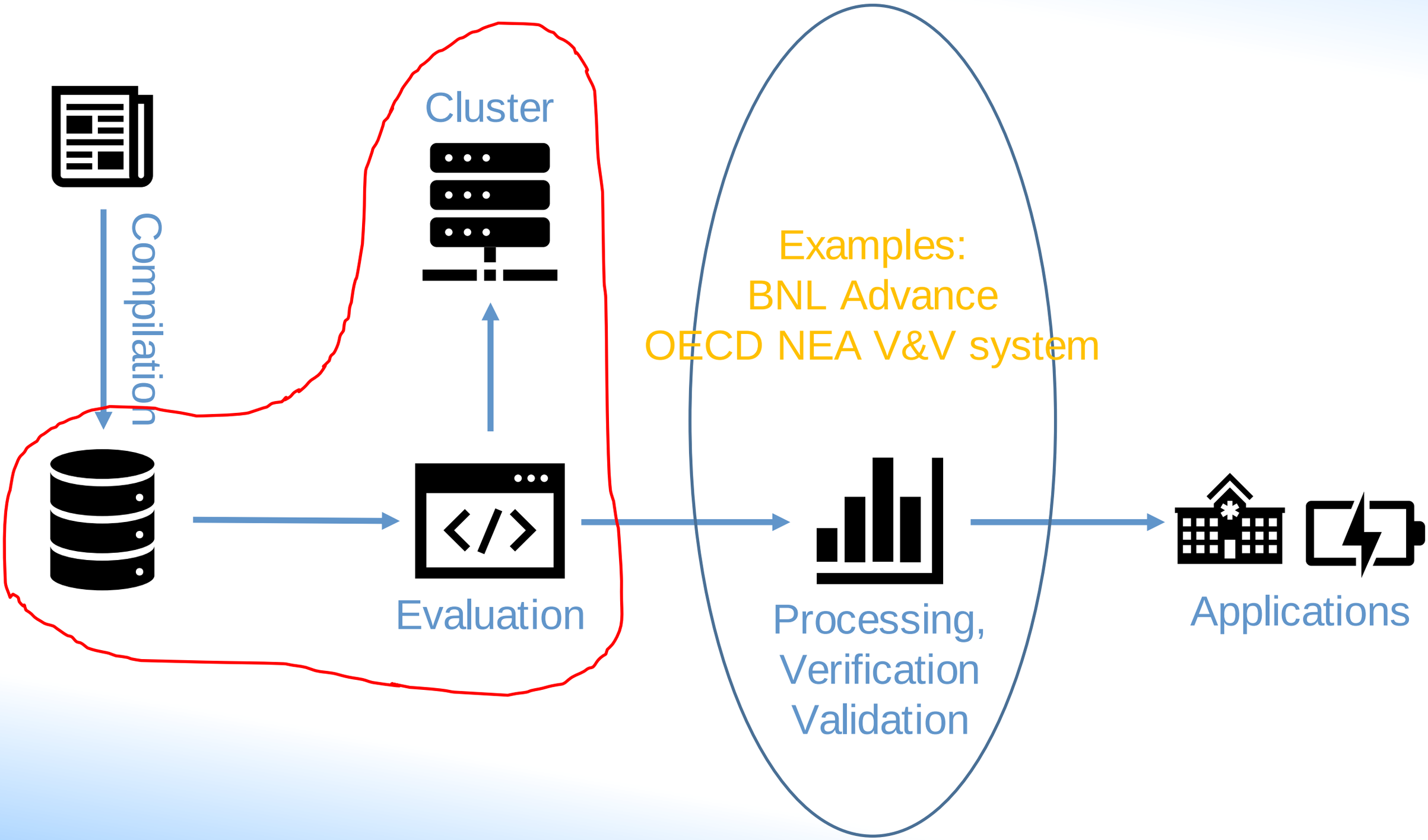
Advantages

- 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



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

UPPSALA UNIVERSITET

PAUL SCHERRER INSTITUT PSI

IAEA

Links & References

- Pipeline paper
  https://arxiv.org/abs/2009.00521
- 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

  https://nds1.gitlab.io/nds/files/jeffdocs/jefdoc-2020.pdf

  https://nds1.gitlab.io/nds/files/jeffdocs/jefdoc-2012.pdf
- BNL ADVANCE (V&V System)
- https://www.sciencedirect.com/science/article/pii/S0090375214001264