NARAD: Nuclear AI-Ready Accelerator Data and LinkML

EIC-BeamAI Collaboration Meeting

09/03/2026

Diana McSpadden

Data Scientist – Data Steward, JLAB

Jefferson Lab

U.S. Department of ENERGY

Agenda

1. The AI-ready accelerator data gap.
2. NARAD's approach and use cases
    - video of a reasoner model performing channel finding
3. The layered architecture + standards
4. How NARAD relates to PALS
5. Two real facility instances

The Problem:
An AI-Readiness Gap

What Makes Data AI-Ready

Many labs have developed AI-readiness eval tools. Example:
Hiniduma, K., Byna, S., Bez, J. L., & Madduri, R. (2024, July). Ai data readiness inspector (aidrin) for quantitative assessment of data readiness for ai. In *Proceedings of the 36th International Conference on Scientific and Statistical Database Management* (pp. 1-12). https://doi.org/10.1145/3676288.3676296

Model-centric

Can I train a model with this data?

Clean data

Consistent units

No manual preprocessing required

Task-contextual

Ready for what type of AI?

E.g., Classification, surrogate modeling, foundation models

There is no universal AI-ready standard, only AI-ready for X.

Provenance-aware

Can I reproduce this workflow?

Traceability, auditability, reproducibility

Lineage, transformations, and versions are documented

Path from raw to the intermediate steps is clear.

Knowledge-enriched

Data + meaning + structure

Data coupled with context/explicit knowledge

Ontologies, knowledge graphs, and semantic relationships

Can enable reasoning, not just prediction

FAIR + Machine actionable

Data is FAIR by design

Persistent IDs link people, projects, datasets, code

Shared vocabularies & ontologies align meaning across terms
Rich structured metadata is complete and consistent
Enables cross-lab AI use

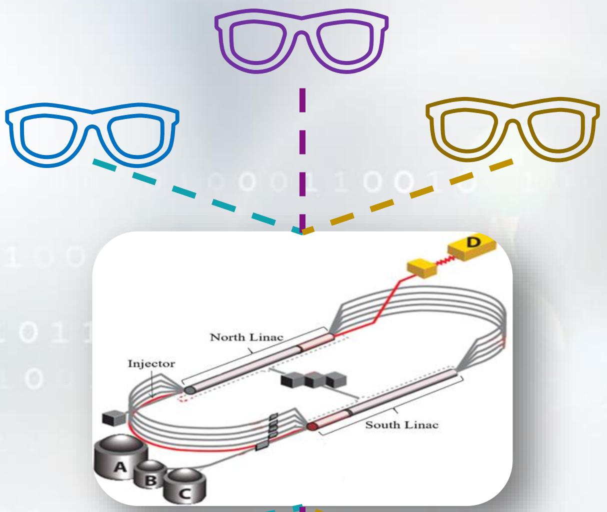
Not always ready to train without other views in place (e.g., **model-centric** and **provenance**)

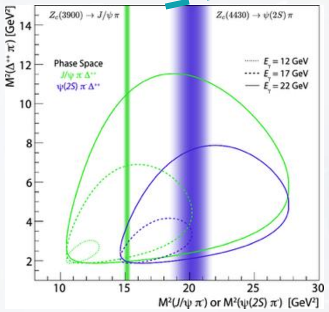
Hiniduma, Kaveen, et al. "Data Readiness for AI: A 360-Degree Survey," ACM Comput. Surv., vol. 57, no. 9, Apr. 2025. DOI:10.1145/3722214

Brewer, W., Widener, P., Anantharaj, V., Wang, F., Beck, T., Shankar, A., & Oral, S. (2025, September). Data Readiness for Scientific AI at Scale. In Workshop Proceedings of the 54th International Conference on Parallel Processing (pp. 18-24). DOI:10.1145/3750720.3757282

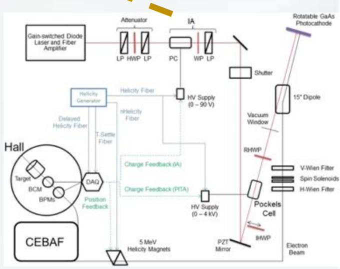
Motivation: the accelerator AI-readiness gap

- **AI promises**: autonomous tuning, anomaly detection, predictive maintenance, data discovery, co-design
- **Assumption**: data that is FAIR and machine-interpretable for *a* facility and *across* facilities
- **Reality**:
  - facilities are bespoke;
  - the same accelerator is described differently by:
    - **physicists**: phase space, lattice
    - **operators**: setpoints/readbacks
    - **engineers**: component design/limits
  - Mapping between a control-system PV and the physics quantity it represents can be implicit, undocumented, or "in someone's head"
  - Fine-*ish* for a human; a hard blocker for an autonomous agent

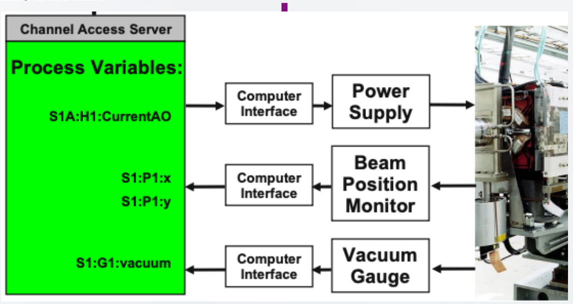




[1]



[2]



[3]

[1] Tennant, Chris & Carpenter, Adam & Powers, Tom & Solopova, Anna & Vidyaratne, Lasitha & Iftekharuddin, Khan. (2020). Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory. Physical Review Accelerators and Beams. 23. 10.1103/PhysRevAccelBeams.23.114601.
[2] Accardi, A., Achenbach, P., Adhikari, D. et al. Strong interaction physics at the luminosity frontier with 22 GeV electrons at Jefferson Lab. Eur. Phys. J. A 60, 173 (2024). https://doi.org/10.1140/epja/s10050-024-01282-x
[3] Johnson, Andrew. (2011). Introduction to EPICS. https://epics.anl.gov/docs/AES2011/01-Intro.pdf

NARAD: Nuclear AI-Ready Accelerator Data

- NARAD: an operational ontology + knowledge-graph framework
- Goal: give the operational description of an accelerator:
  - ordered elements,
  - parameters,
  - control points,
    - *with* typed relationships for containment, control, ordering, relative position, and connectivity
- Design:
  1. A layered operational architecture (stable concepts vs. facility-local channels)
  2. Ontology: LinkML from day one

NARAD and lattice standards (e.g., PALS) are peers, not a hierarchy

FAIR and machine actionable data:
3 use cases

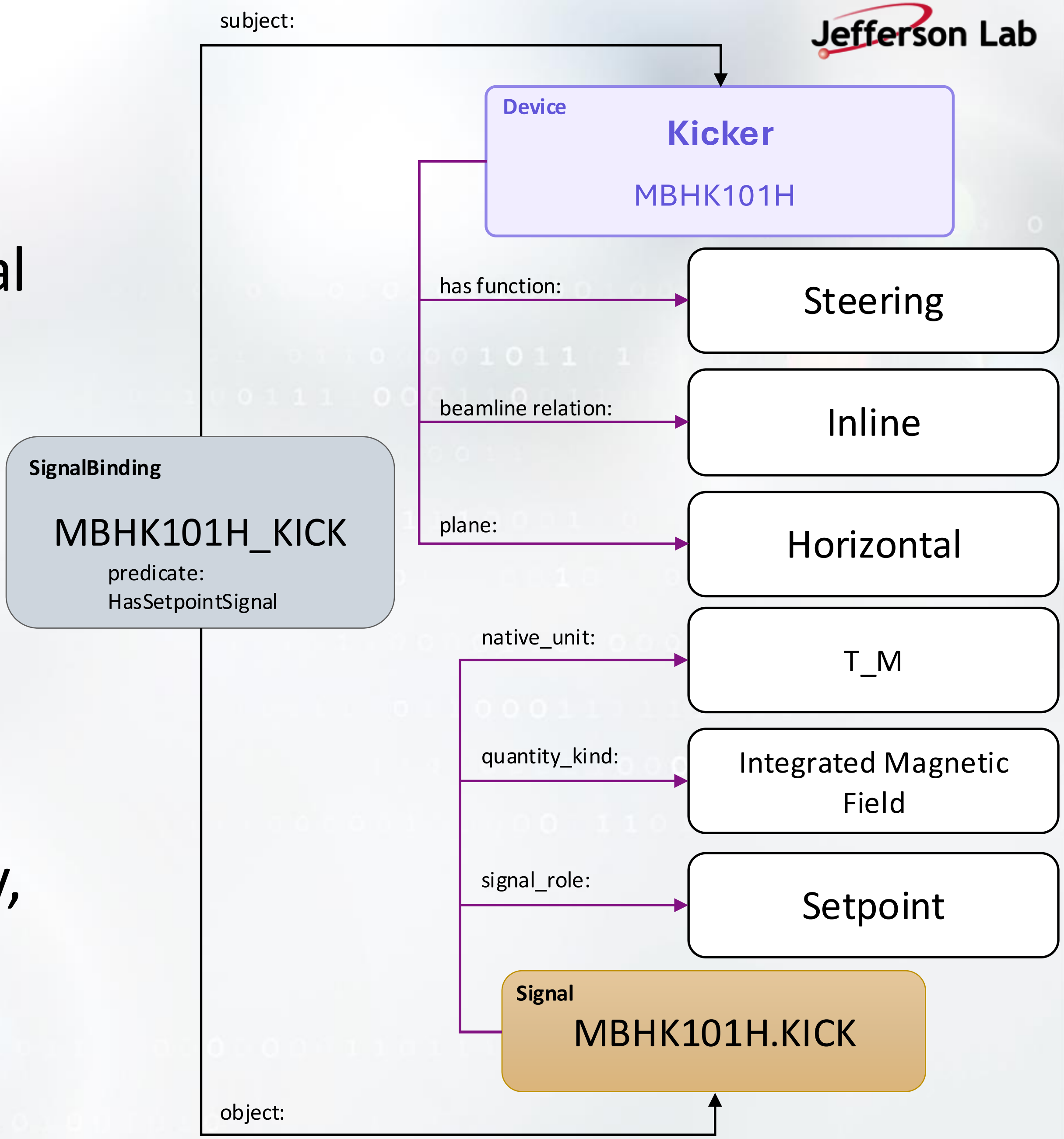
Three Use Cases for FAIR and machine actionable data

- Facility understanding & channel finding
  - an agent resolves "the horizontal position of the first BPM in the injector" to an exact PV
- Facility context for shared data
  - a dataset carries its own ontology/context so it's self-describing wherever it lands (e.g., AmSC / Genesis Mission)
- Model & digital-twin generation
  - walk a NARAD beamline (with live setpoints) to emit a lattice for madx / elegant / Xsuite / ImpactX

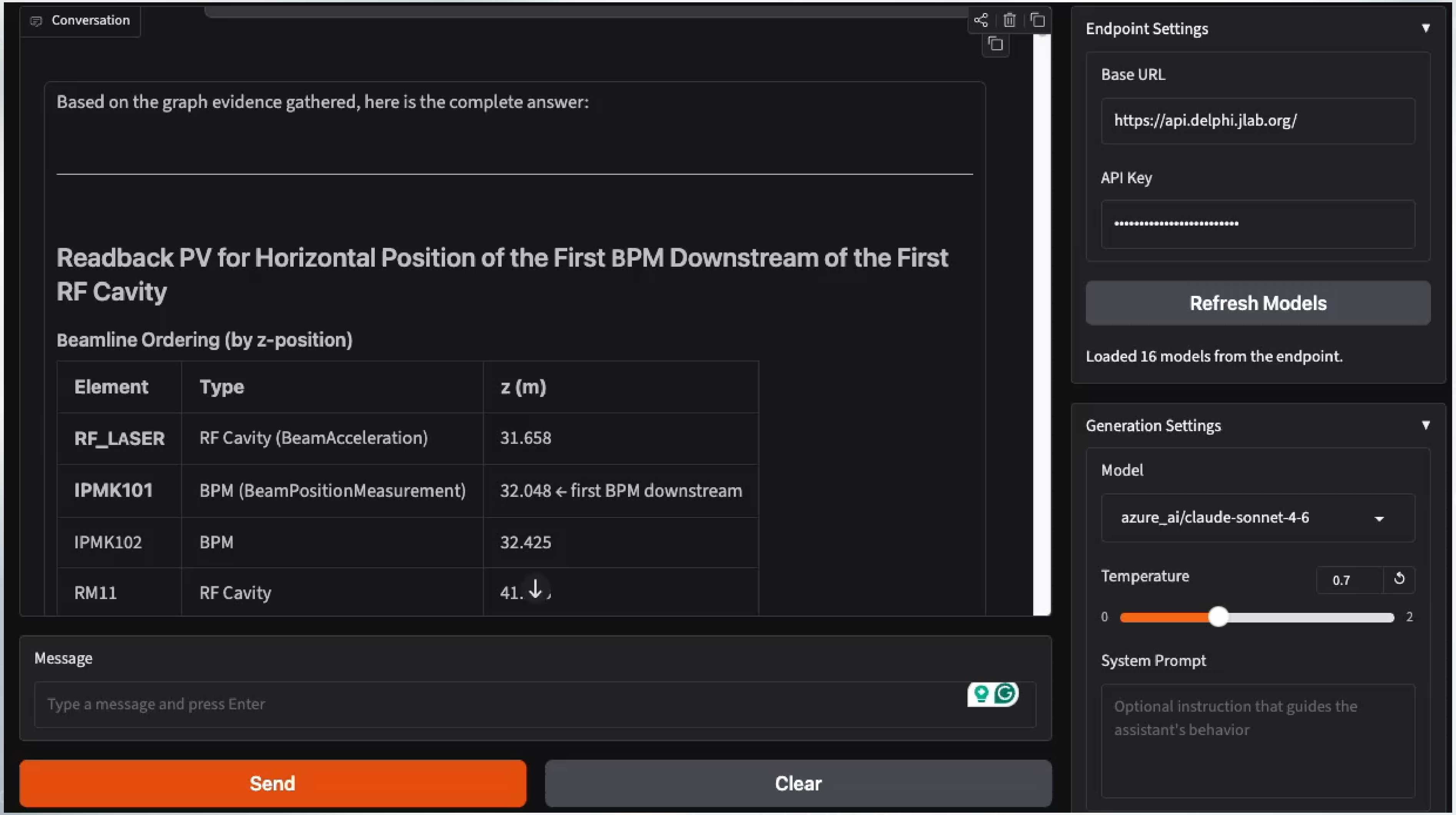
Jefferson Lab

Use case: channel finding

- Agent question: "What's the horizontal position of the first BPM in the injector?"
- Traversal: device → signal binding → signal → control-system name
- Every step is standards-grounded and auditable
    - Agents can reason on a new facility, given the ontology

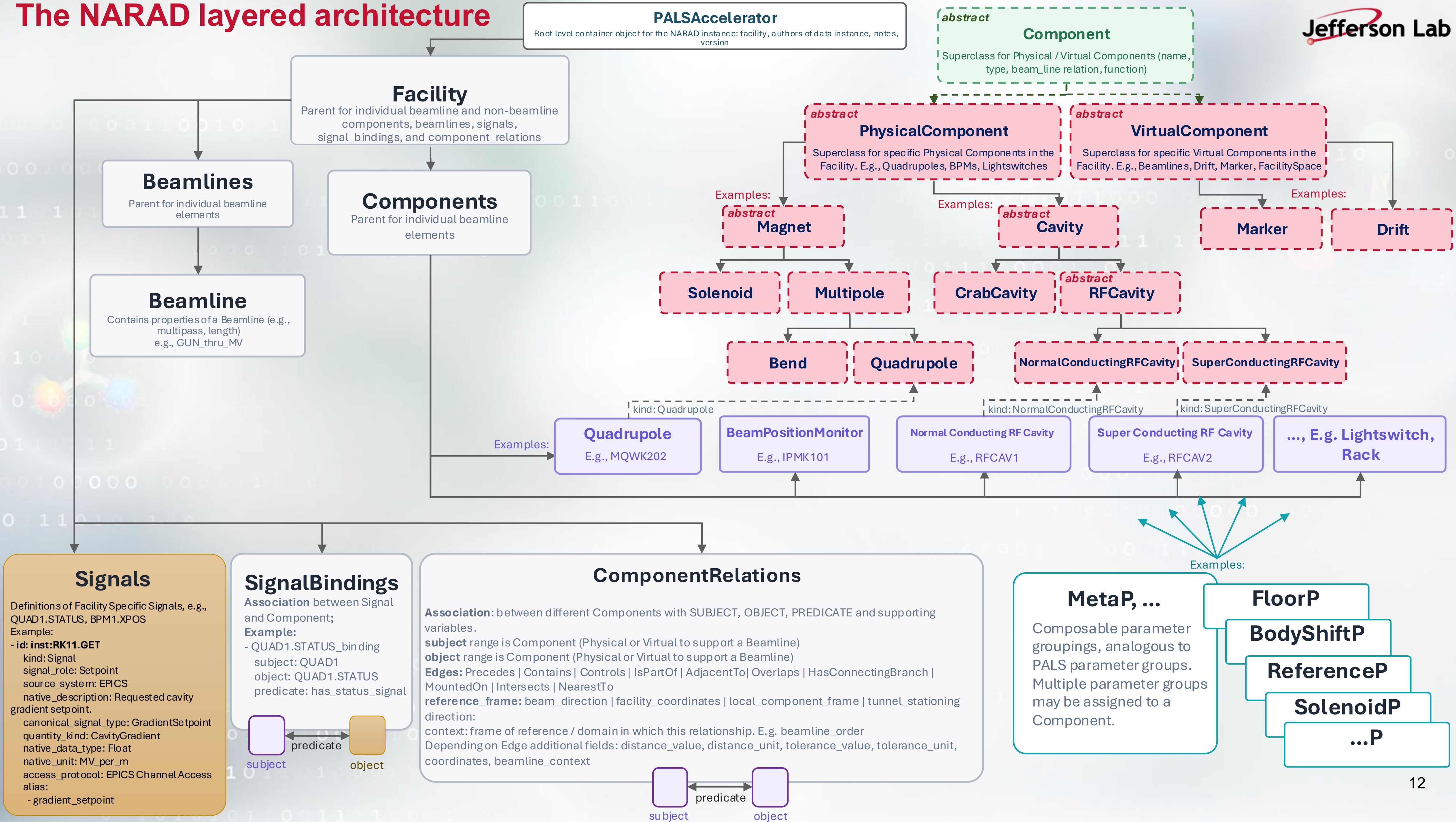


Use case: AI-ready channel finding



Layered Architecture and Standards:
Schema elements and ontology

The NARAD layered architecture



Standards grounding for FAIR and machine actionable data

Standard(s)	Concern
SOSA/SSN	Sensing/observation pattern
QUDT	Units & quantity kinds
RO, BFO, SAREF	Relations (precedes, part-of, adjacent-to…)
PALS	Lattice/beamline layout.

Ontologies?

Standards grounding for FAIR and machine actionable data

Standard(s)	Concern
SOSA/SSN	Sensing/observation pattern
QUDT	Units & quantity kinds
RO, BFO, SAREF	Relations (precedes, part-of, adjacent-to…)
PALS	Lattice/beamline layout.

```yaml
SignalRoleEnum:
  description: Roles for control-system signals.
  permissible_values:
    Setpoint:
      description: Requested or configured value.
      meaning: sosa:ActuatableProperty
    Readback:
      description: Measured or reported value.
      meaning: sosa:ObservableProperty
```

Jefferson Lab

Standards grounding for FAIR and machine actionable data

Standard(s)	Concern
SOSA/SSN	Sensing/observation pattern
QUDT	Units & quantity kinds
RO, BFO, SAREF	Relations (precedes, part-of, adjacent-to…)
PALS	Lattice/beamline layout.

exact mapping

NARAD schema Octupole → pals:Octupole

```
Octupole:
  description: >-
    A magnet primarily used to focus or defocus the beam. 8 poles.

    Only one of the electric or magnetic multipole components should be present, not both.
  is_a: Multipole
  exact_mappings:
    - pals:Octupole
```

NARAD schema includes ability to provide aliases for field names:

```
BodyShiftParamGroup:
  is_a: ParamGroup
  description: >-
    Mixin class for floor plan shift parameters that are aligned with PALS BodyShiftP (Floor Plan Shift Parameters)
    for interoperability with the Particle Accelerator Language Standard (PALS).
  attributes:
    x_offset:
      description: Shift in the X direction in the facility floor plan.
      range: float
      slot_uri: pals:x_offset
      aliases:
        - delta_x
        - x_shift
        - GlobalOffsetX
```

LinkML and semantic web standards

- **LinkML [1]**:
  - Linked data Modeling Language (linkml.io)
  - An object-oriented, schema-first data modeling framework (YAML-based classes, slots, mixins, enums, validation rules)
  - Built to solve problems not unique to accelerators [2]:
    - heterogeneous data,
    - ad hoc formats,
    - poor FAIR compliance
  - Developed by a team centered at LBNL, as part of the biomedical-ontology
  - Supports simple flat checklist-style schemas AND complex, interrelated, polymorphic data with inheritance

[1] Moxon, S. A., Solbrig, H., Harris, N. L., Kalita, P., Miller, M. A., Patil, S., ... & Mungall, C. J. (2026). LinkML: an open data modeling framework. GigaScience, 15, giaf152. https://doi.org/10.1093/gigascience/giaf152

[2] Shoup, A., Russell, A., & Nicol, G. (2025, November). LinkML for Collaborative Petrochemical Knowledge Graph Development. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (pp. 6014-6020). https://doi.org/10.1145/3746252.3761514

The NARAD project and the PALS standard

PALS & NARAD

- PALS describes the physics/lattice layer
- The NARAD project is focused on extending to the operational layer and providing FAIR data
- We envision the relationship between them is a defined interface between peers
- PALS itself is evolving
- That expansion is being coordinated with NARAD and MOAT; not competing with it
- A NARAD facility instance compatible with the"PALS-shape" *and* carries the operational metadata.
- Call out to Lucy Weijian and her PALS + signals → NARAD transformation

Two facility instances: UITF and BTA

Two facility instances, one schema

Example:	JLab UITF	BNL BTA
Scope	Example (17 devices) of the injector beamline	Booster-to-AGS transport line lattice (100+ elements)
Source	Ported from a prior JLab instance format	Generated from a PALS lattice export
Control system	EPICS	Local
What's populated	Components + signals + bindings + beamline precedes order	Components + beamline structure; signals layered in from local naming; some model-only constructs intentionally omitted for now
Contact	He Zhang (hezhang@jlab.org)	Lucy Weijian (wlin1@bnl.gov)

Open issues and roadmap

- Cross-facility vocabulary agreement (canonical signal types, quantity kinds): continuing community consensus
  - Unit alignment: e.g., reference momentum appears as MeV at UITF, GeV at CEBAF, while PALS expects eV — a known, unresolved mismatch
- Broader device coverage: some real elements (controllers, constants) have no schema representation *yet*
- **Next: quantitative query benchmarks demonstrating FAIR & Machine Actionable capabilities.**

Resources

PALS: https://pals-project.readthedocs.io/en/latest/

LinkML: https://linkml.io/linkml/

Moxon, S. A., Solbrig, H., Harris, N. L., Kalita, P., Miller, M. A., Patil, S., ... & Mungall, C. J. (2026). LinkML: an open data modeling framework. GigaScience, 15, giaf152. https://doi.org/10.1093/gigascience/giaf152

Shoup, A., Russell, A., & Nicol, G. (2025, November). LinkML for Collaborative Petrochemical Knowledge Graph Development. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (pp. 6014-6020). https://doi.org/10.1145/3746252.3761514

Semantic channel finding:

Hellert, T., Agladze, N., Giovannone, A., Jug, J., Mayet, F., Sherwin, M., ... & Tennant, C. (2026). From natural language to control signals: a conceptual framework for semantic channel finding in complex experimental infrastructure. Machine Learning: Science and Technology, 7(4), 045010. https://iopscience.iop.org/article/10.1088/2632-2153/ae8219/meta

Thank you

Diana McSpadden: dianam@jlab.org

Backups

Use case: FAIR (findable, accessible, interoperable, reusable) accelerator data

- **Problem**: Data across accelerator facilities is difficult for both machines and humans to understand and reuse consistently.
- **Approach**:
  - Persistent identifiers for ontology classes (signal types, component types)
  - Persistent identifiers for facility data instances (components, signals)
  - Pointers from NARAD types to data stores with corresponding measurements
- **Result**: data published on a shared platform carries the context required to interpret the data
  - What each channel measures
  - In what units
  - On what device
  - In what beamline location

Use case: digital twins & model generation

- A generator walks the ordered beamline; for each element reads: semantic class, physics parameters, modeled_as target, and (if present) live setpoints
- Same validated instance drives both channel finding and model generation → digital twin stays coherent with the operational description by construction

PALS & NARAD

- NARAD's beamline layer maps onto a PALS lattice (inst:GUN_thru_MV example)
- component_relations (pals:precedes) encode beamline order the same way a PALS lattice would
- Signal bindings attach live control-system state to a PALS-describable element
- A NARAD facility instance is simultaneously compatible with the"PALS-shape” or identical to PALS shape, *and* carries the operational metadata.

Standards grounding for FAIR and machine actionable data

Standard(s)	Concern
SOSA/SSN	Sensing/observation pattern
QUDT	Units & quantity kinds
RO, BFO, SAREF	Relations (precedes, part-of, adjacent-to…)
PALS	Lattice/beamline layout.

```yaml
UnitEnum:
  description: >-
    Engineering units using ASCII-safe enum names. Values map to QUDT unit
    IRIs with the LinkML meaning field when an exact QUDT unit is available.
  permissible_values:
    A:
      description: ampere
      meaning: qudt:A
    degree:
      description: degree
      meaning: qudt:DEG
    eV:
      description: electron volt
      meaning: qudt:EV
```

Standards grounding for FAIR and machine actionable data

Standard(s)	Concern
SOSA/SSN	Sensing/observation pattern
QUDT	Units & quantity kinds
RO, BFO, SAREF	Relations (precedes, part-of, adjacent-to…)
PALS	Lattice/beamline layout.

```yaml
Precedes:
  is_a: ComponentRelation
  description: >-
    A relationship indicating that the subject component precedes the object component in a specified context,
    such as along a beamline.
  class_uri: pals:PrecedesRelation
  exact_mappings:
    - BFO:0000063
```

LinkML and NARAD

- Chosen from the start, not retrofitted onto a finished ontology
- One LinkML schema → many generated artifacts:
  - Instance validation: validation for facility data instances
  - Property-graph artifacts (neo4j / neosemantics)
  - JSON Schema for interchange
  - Auto-generated documentation
  - Machine-readable mappings to external standards (exact_mappings / broad_mappings)

JLab Upgraded Injector Test Facility (UITF)


```yaml
- id: inst:IPMK101
  kind: BeamPositionMonitor
  functions: [BeamPositionMeasurement]
  MetaP:
    description: Digital receiver (stripline) beam position monitor.
    additional_metadata:
      - {key: Controlled_by, value: iocipmk101}
# ...signals...
- id: inst:IPMK101.XPOS
  signal_role: Readback
  canonical_signal_type: HorizontalBeamPosition
  quantity_kind: BeamPosition
  native_unit: mm
# ...binding...
- id: inst:IPMK101_XPOS
  subject: inst:IPMK101
  predicate: pals:has_readback_signal
  object: inst:IPMK101.XPOS
```

BNL's Booster to AGS (BtA)


```yaml
- id: inst:MW006
  kind: BeamProfileMonitor
  functions: [BeamProfileMeasurement]
  native_types: [Marker]
  MetaP:
    description: Multiwire beam-profile monitor.
    additional_metadata:
      - {key: nbr_wires_x, value: '32', datatype: Integer}
# ...signal (ADO)...
- id: inst:btaMW006Cntl:horInDataM
  signal_role: Readback
  source_system: Other
  canonical_signal_type: HorizontalBeamPosition
  quantity_kind: BeamPosition
  native_data_type: FloatArray
  native_unit: V
  access_protocol: ADO
# ...binding...
- id: inst:MW006.horizontal_profile_binding
  subject: inst:MW006
  predicate: pals:has_readback_signal
  object: inst:btaMW006Cntl:horInDataM
```