Model-Driven Integration of Domain Knowledge into Machine Learning Workflows: A Case for Multidisciplinary System Design

-An Envisioned Framework

Neeraj Katiyar, Jhelum Chakravorty

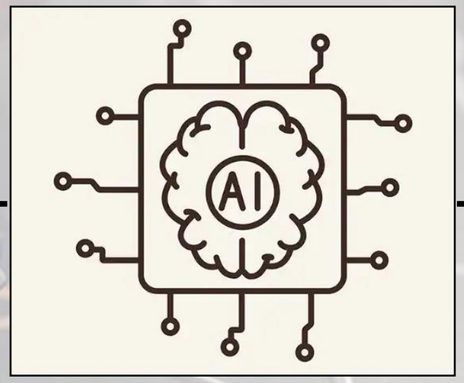
Hitachi Energy Research, Canada

Presentation flow

1

Motivation & Problem (O1)

2



Defining Research Scope (RQs)

3

Conceptualized Solution (O2-O4)

4

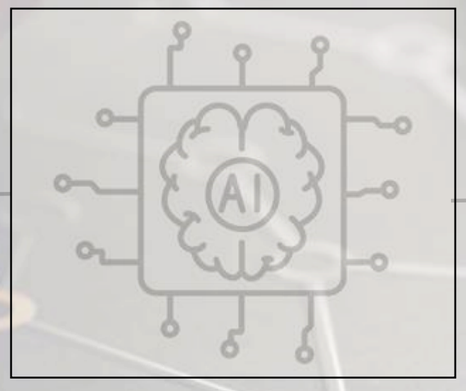
Conclusion

Presentation Flow

1

Motivation & Problem (O1)

2



Defining Research Scope (RQs)

3

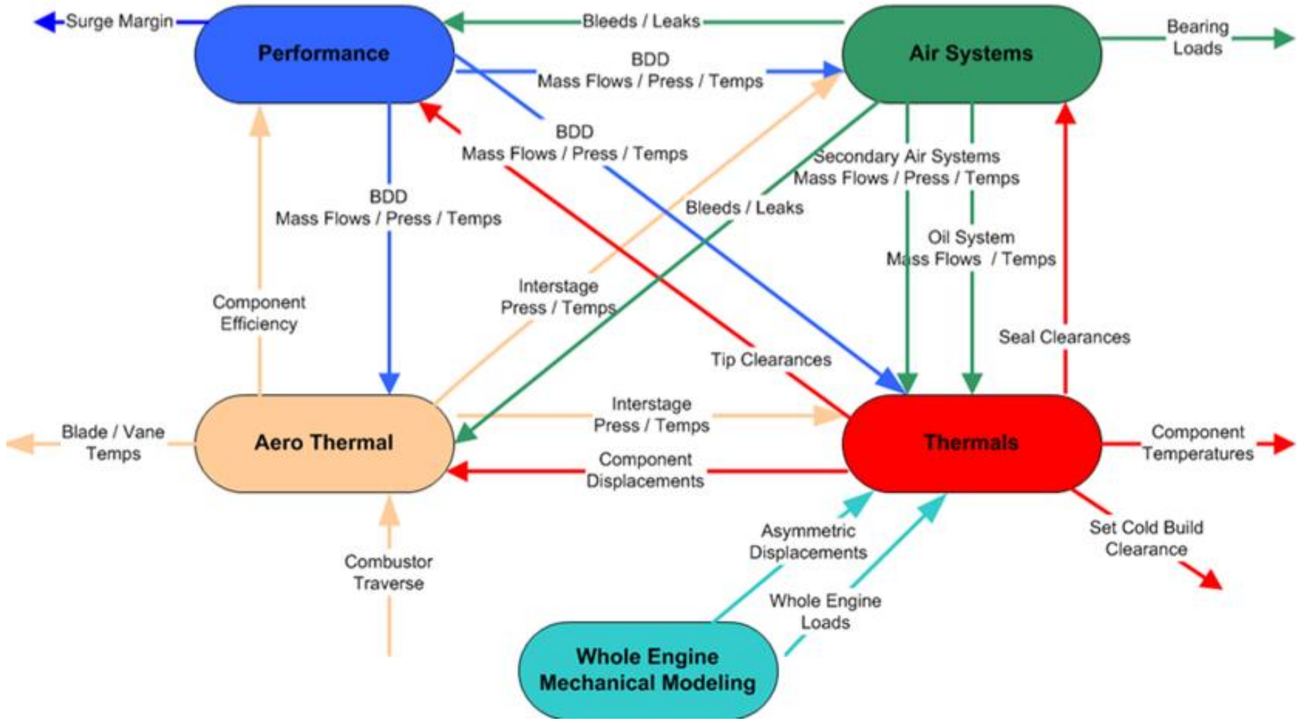
Conceptualized Solution (O2-O4)

4

Conclusion

Motivation

Complex System: Components of a Gas Turbine



- Multidisciplinary Systems are Complex
  – Gas turbines → multiple subsystems (combustion, turbine, compression, cooling, materials).

- Knowledge is Fragmented
  – Spread across documents, simulations, and informal expert discussions.

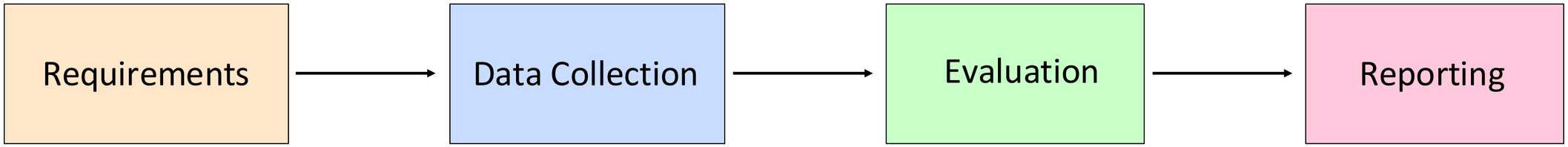
- ML Adoption Grows
  – Surrogate models & predictive tools increasingly used.

Key Challenge:

No standardized process for capturing and reusing domain knowledge → repetitive cycles & delays

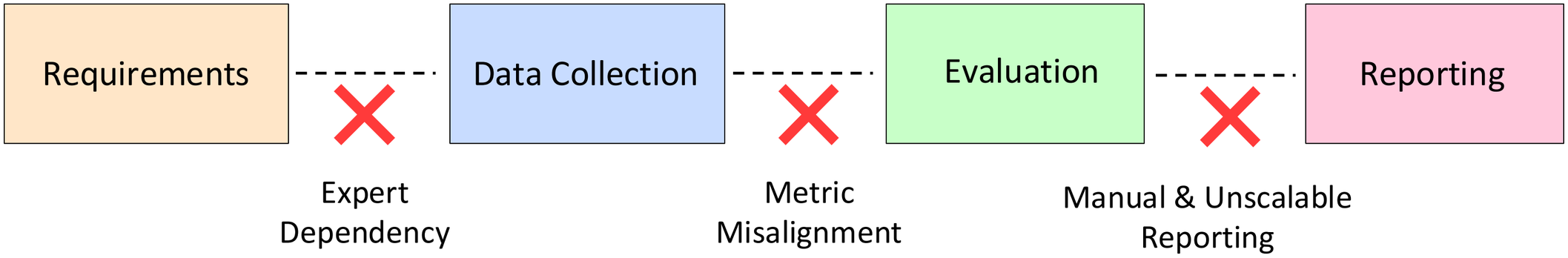
Problem: Ideal ML Pipeline

A standard ML Pipeline



Problem: Fragmented ML Pipeline

A standard ML Pipeline: Broken Links



Result

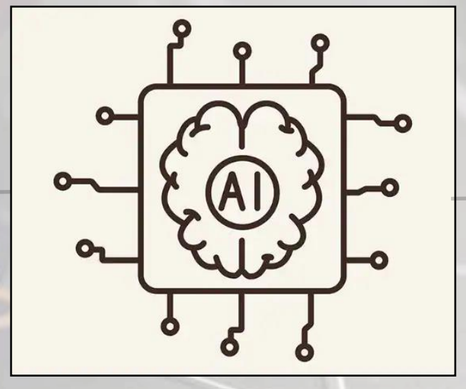
Slow iterations, inefficiency, and risk of wrong design decisions

Presentation Flow

1

Motivation & Problem (O1)

2



Defining Research Scope (RQs)

3

Conceptualized Solution (O2-O4)

4

Conclusion

Research Scope

How to systematically integrate domain knowledge into ML workflows for multidisciplinary systems?

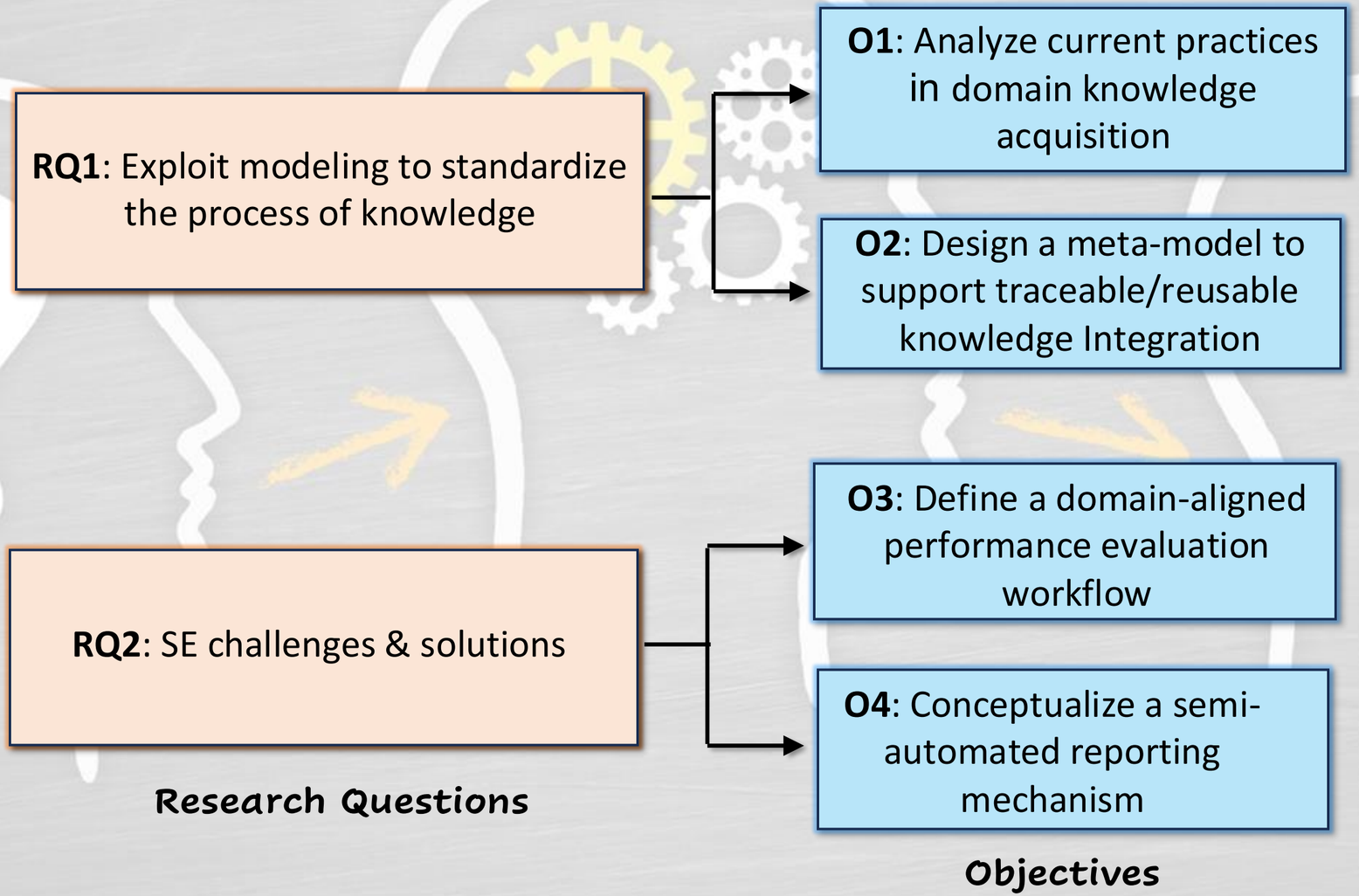
High-level research theme

RQ1: How can we exploit modeling to standardize the process of knowledge acquisition in a multi-disciplinary industrial setting for ML?

RQ2: How can we overcome identified SE challenges in ML evaluation and reporting?

Research Questions

Research Objectives

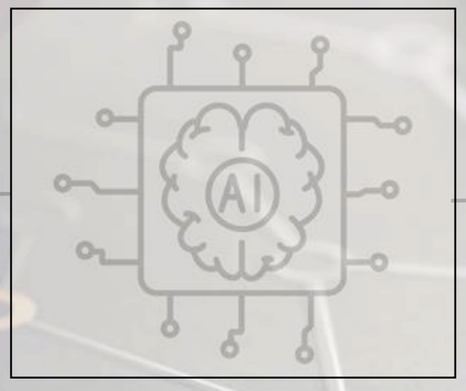


Presentation Flow

1

Motivation & Problem (O1)

2



Defining Research Scope (RQs)

3

Conceptualized Solution (O2-O4)

4

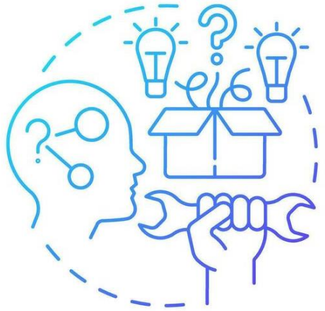
Conclusion

O1: Evidence from Current Practices (Recap)

Findings

Multiple Meetings

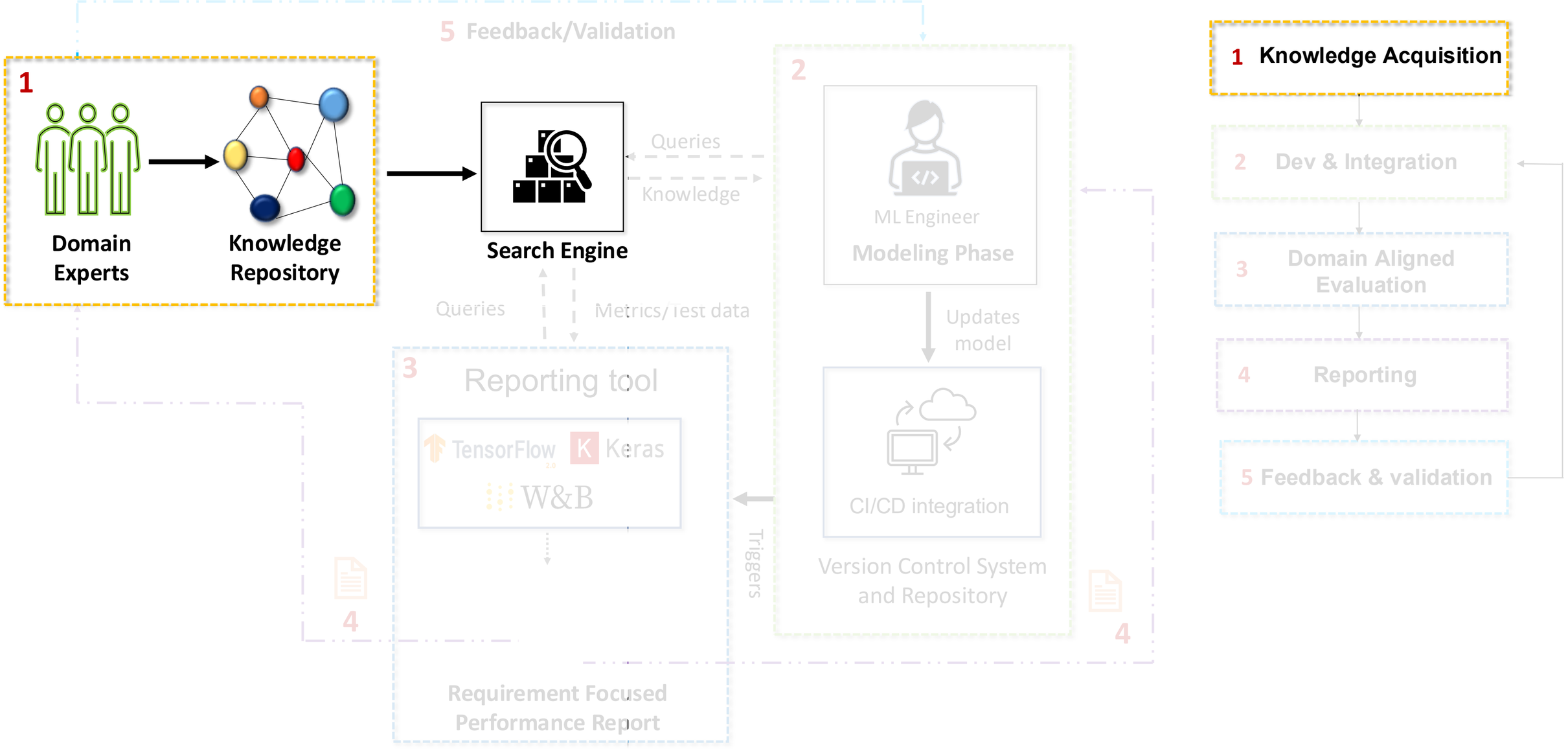
Scattered Knowledge



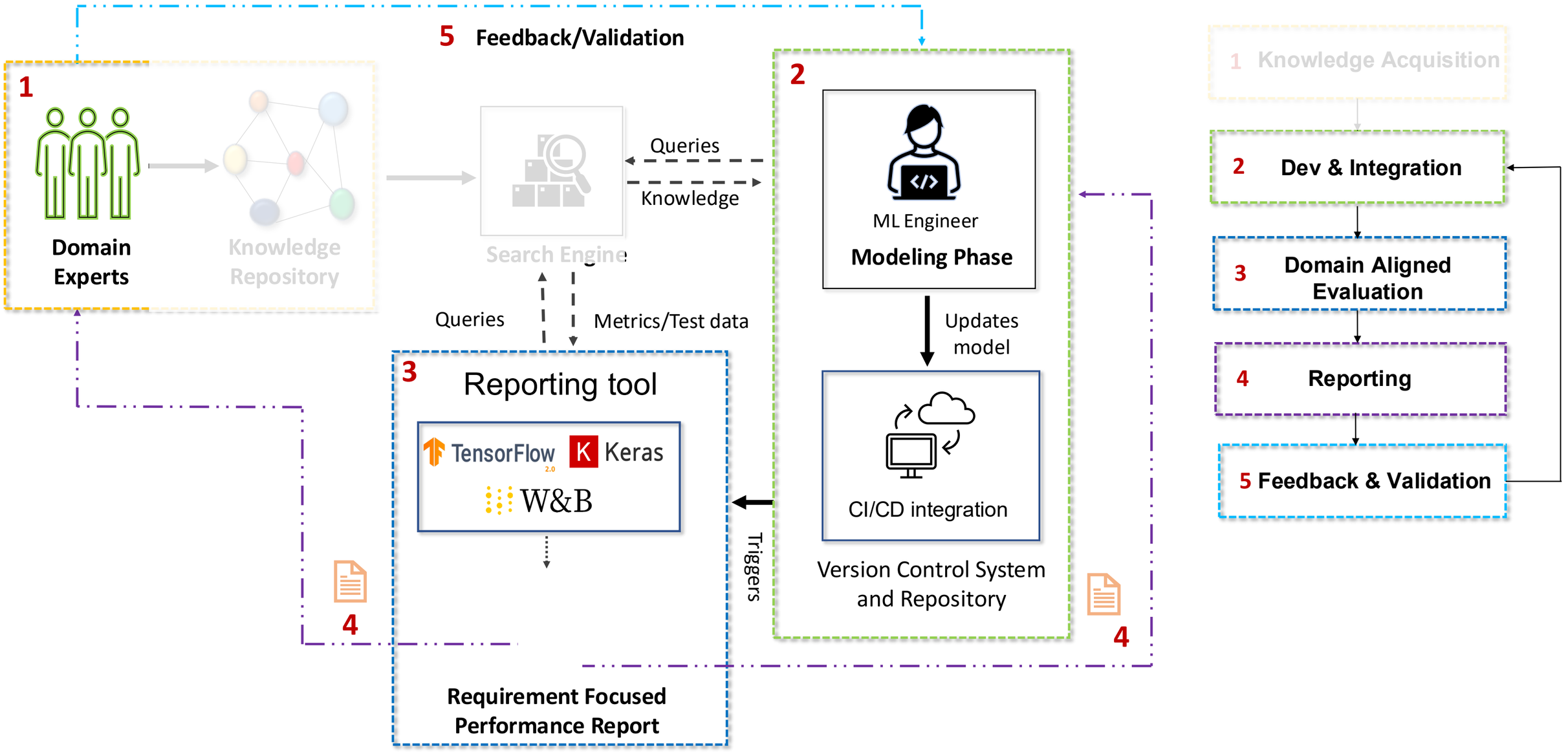
Tacit Notes

“Knowledge management is ad hoc, and fragmented”

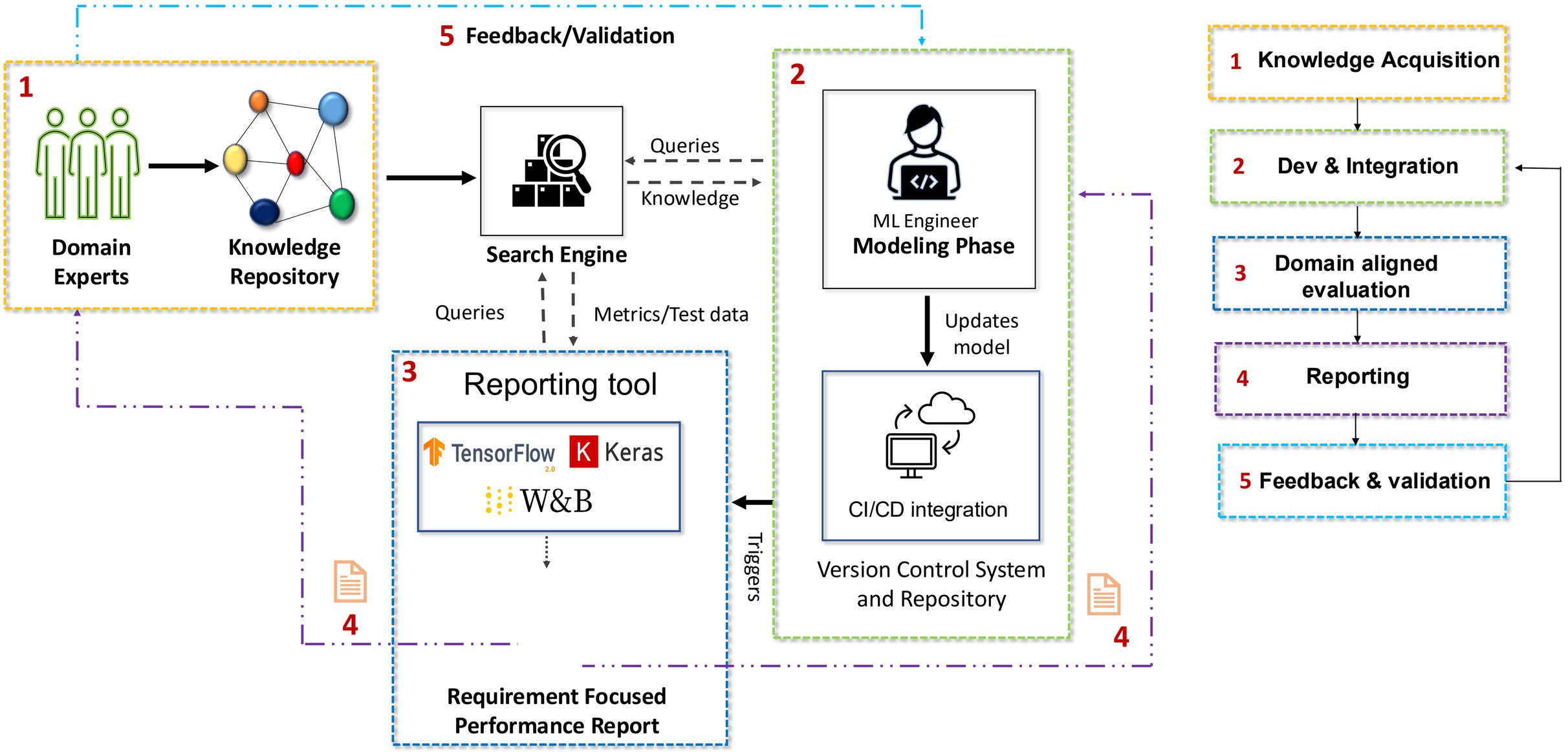
Part 1: Standardizing Knowledge Acquisition (RQ1)



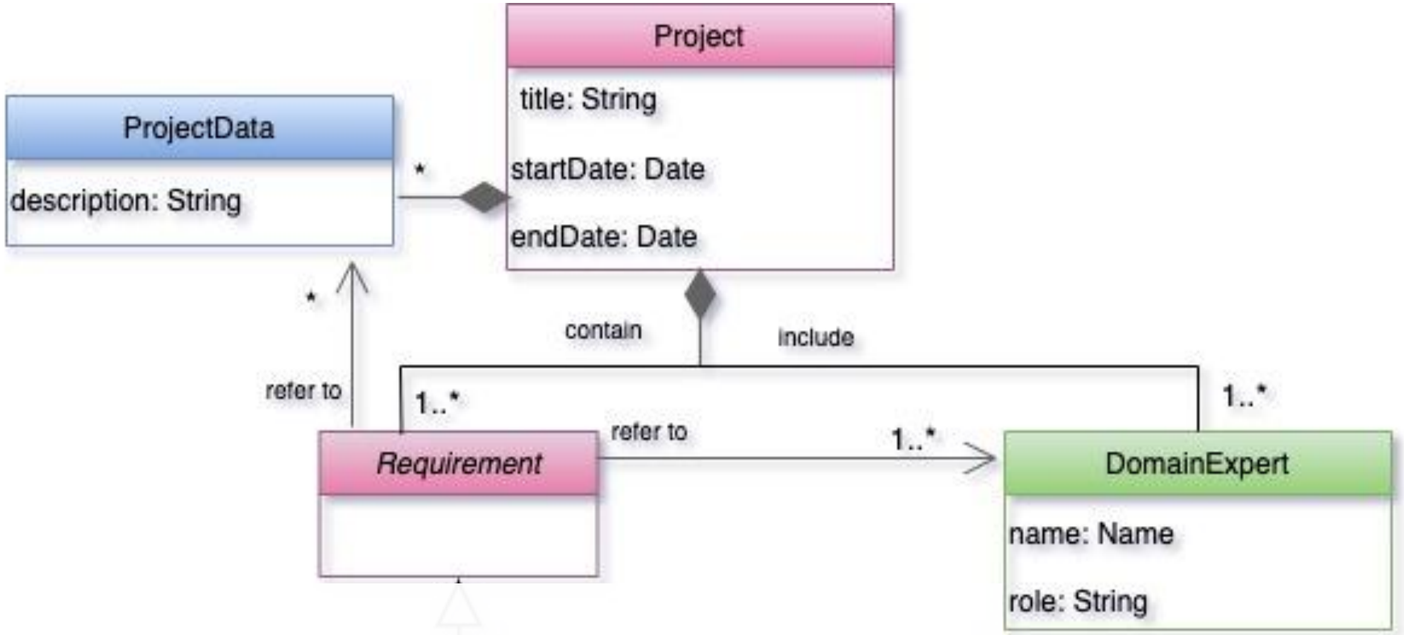
Part 2: Aligning Evaluation & Reporting with Domain Needs-RQ2



Big Picture: Proposed Our Envisioned Framework (RQ1 + RQ2)

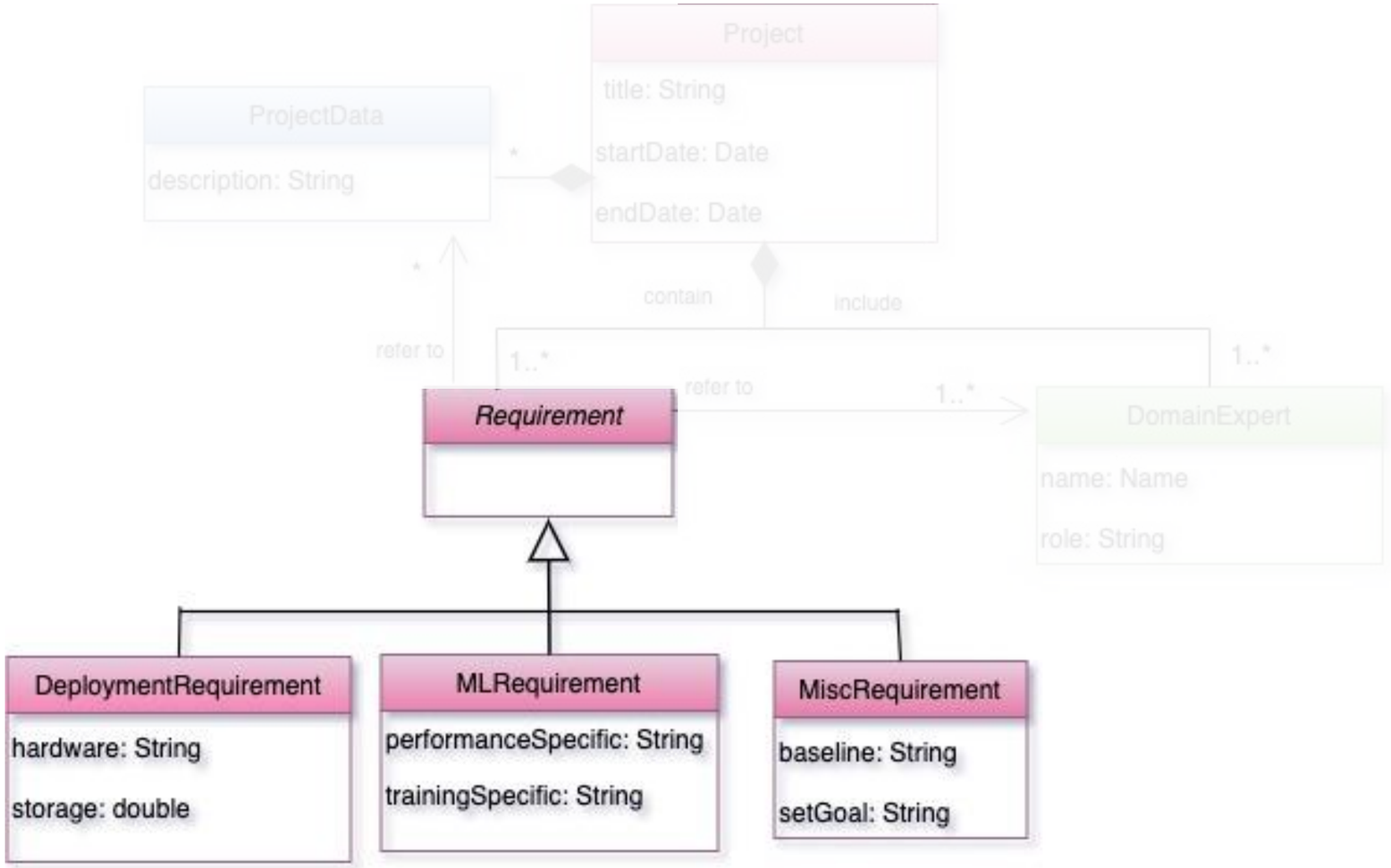


O2: Meta-Model for Knowledge Acquisition

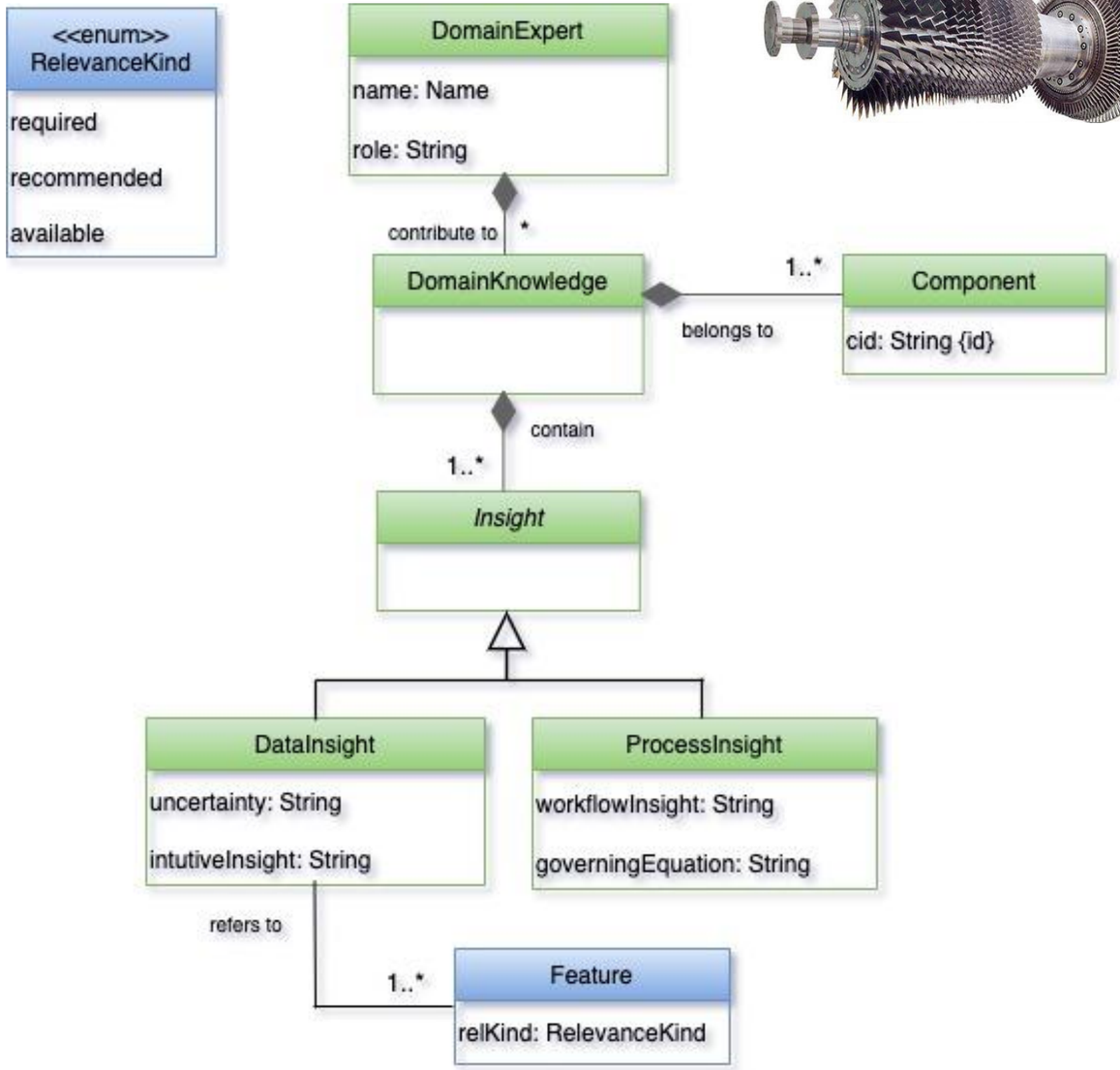


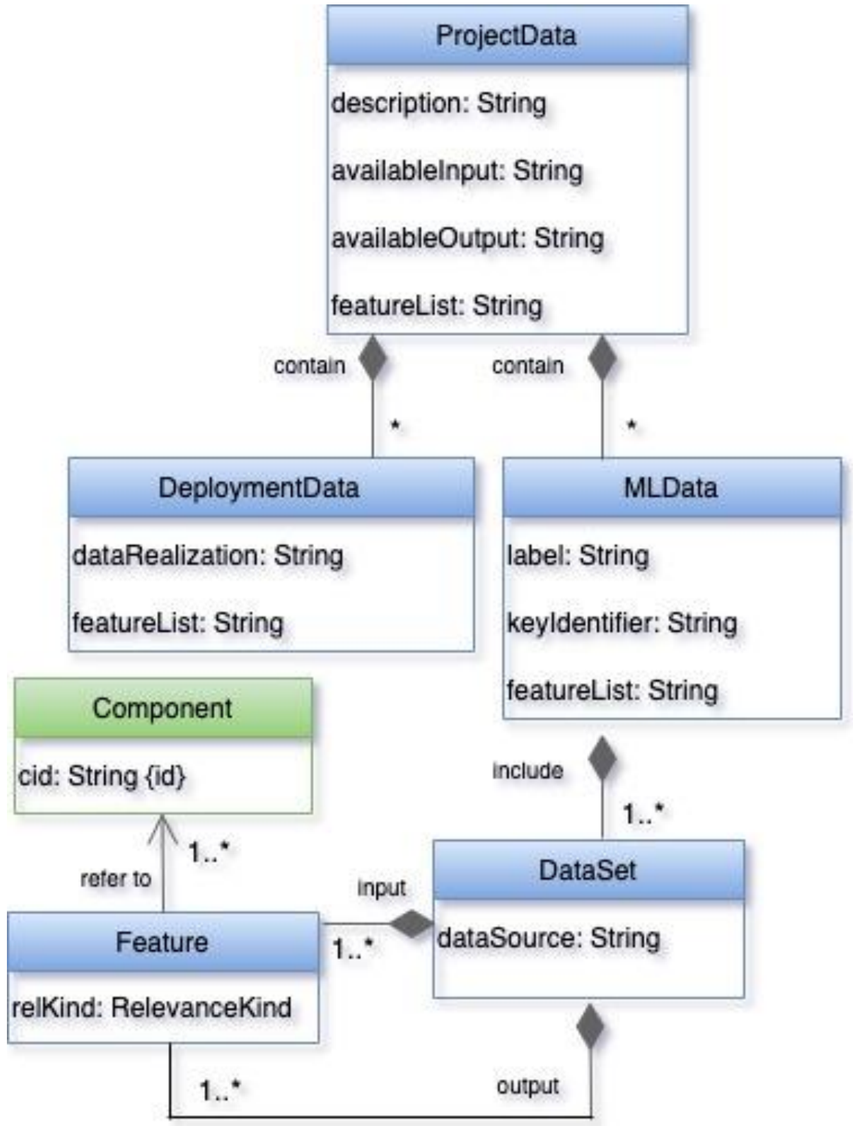
Formal representation of domain knowledge → reusable, extensible, traceable

O2: Requirement View

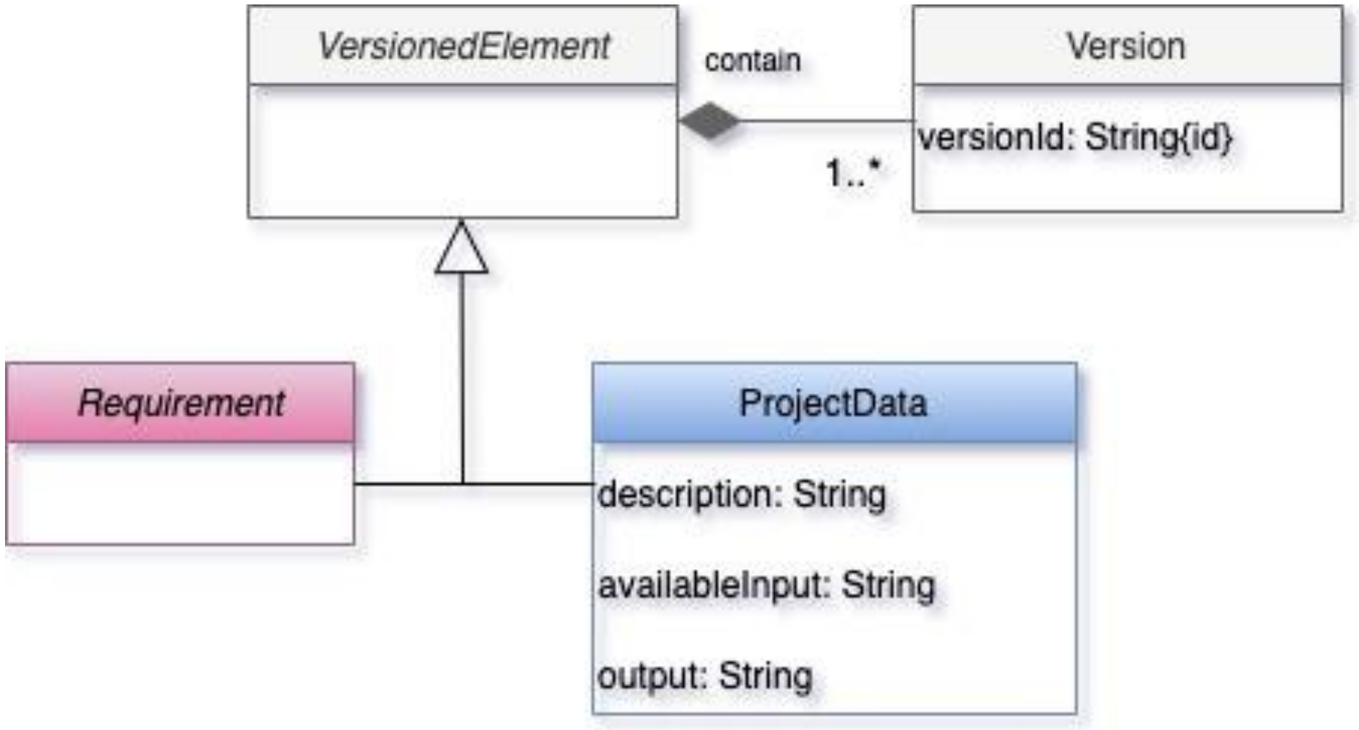


O2: Knowledge & Data View





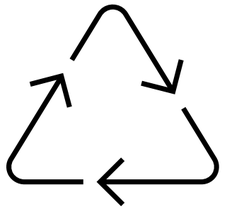
O2: Version View



O2 Wrap-Up: Meta-Model for Knowledge Acquisition

Requirement View → Captures constraints & goals

Knowledge & Data View → Links domain insights & datasets



Version view → Ensures traceability over time

Formal, reusable, extensible foundation for ML projects

O1 (evidence of gaps) → O2 (meta-model foundation) → leads to O3 (evaluation)

Test Slice: Partitioned Evaluation Data

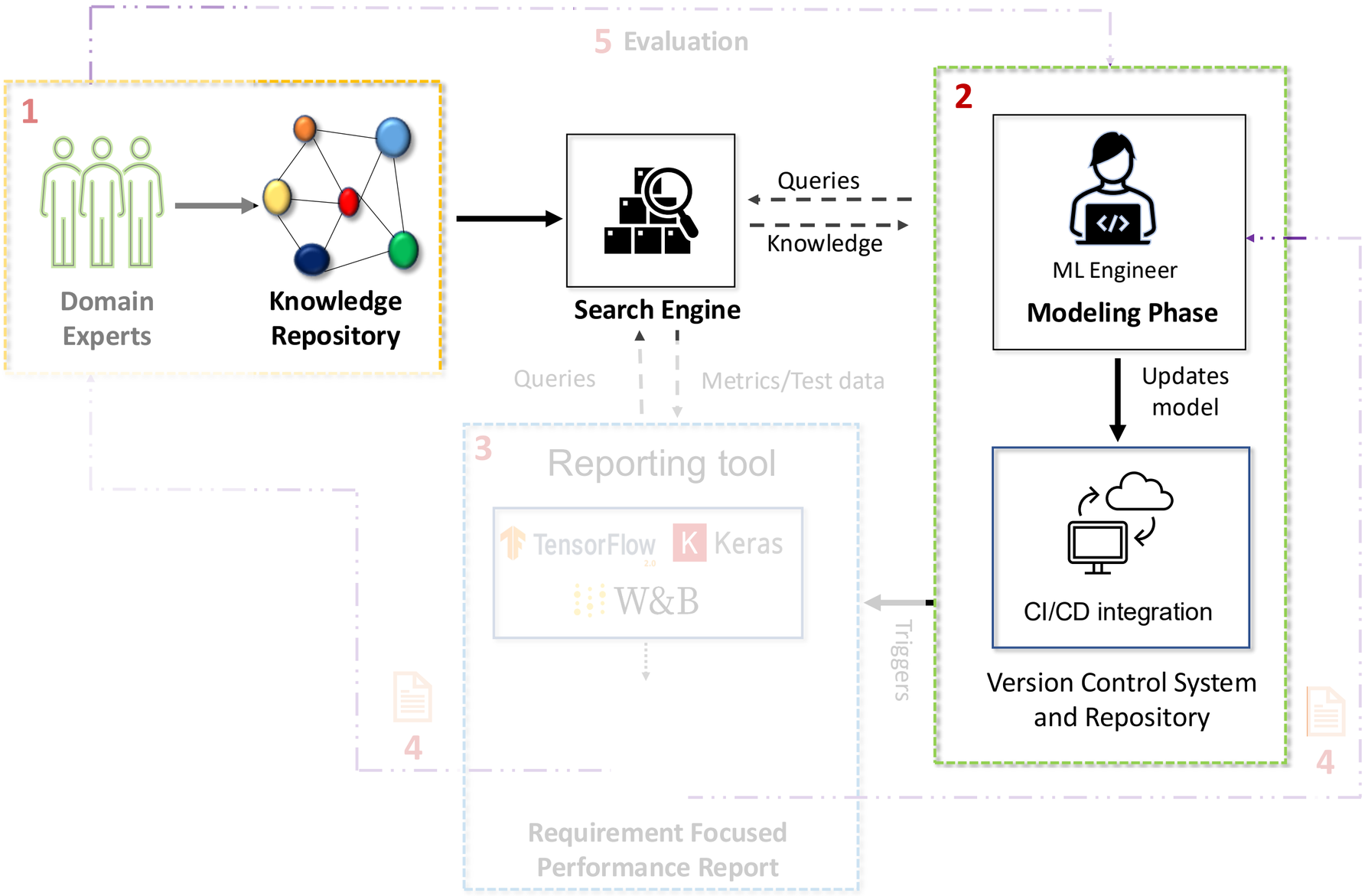
Why Test Slices Matter?

	Nominal	Degraded	Extreme
Load	✓	✓	✓
Ambient	✓	✓	✗
Transient	✓	✓	✓
Maintenance	✓	✓	✓

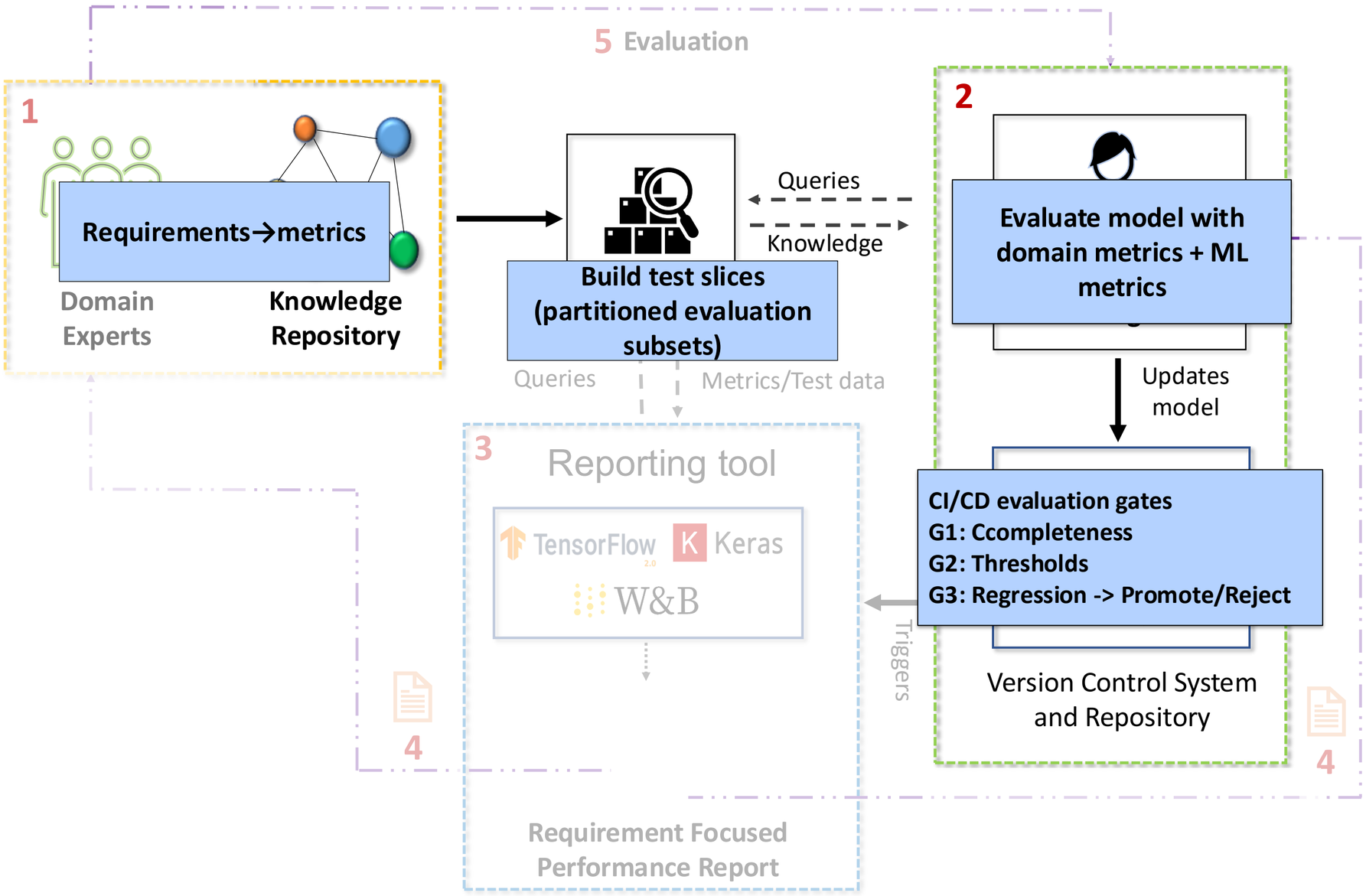
- Each cell = performance in a scenario × slice
- Most are ✓ (pass), but one ✗ (fail)
- Overall RMSE may still look 'good'

→ Test slices reveal hidden weaknesses

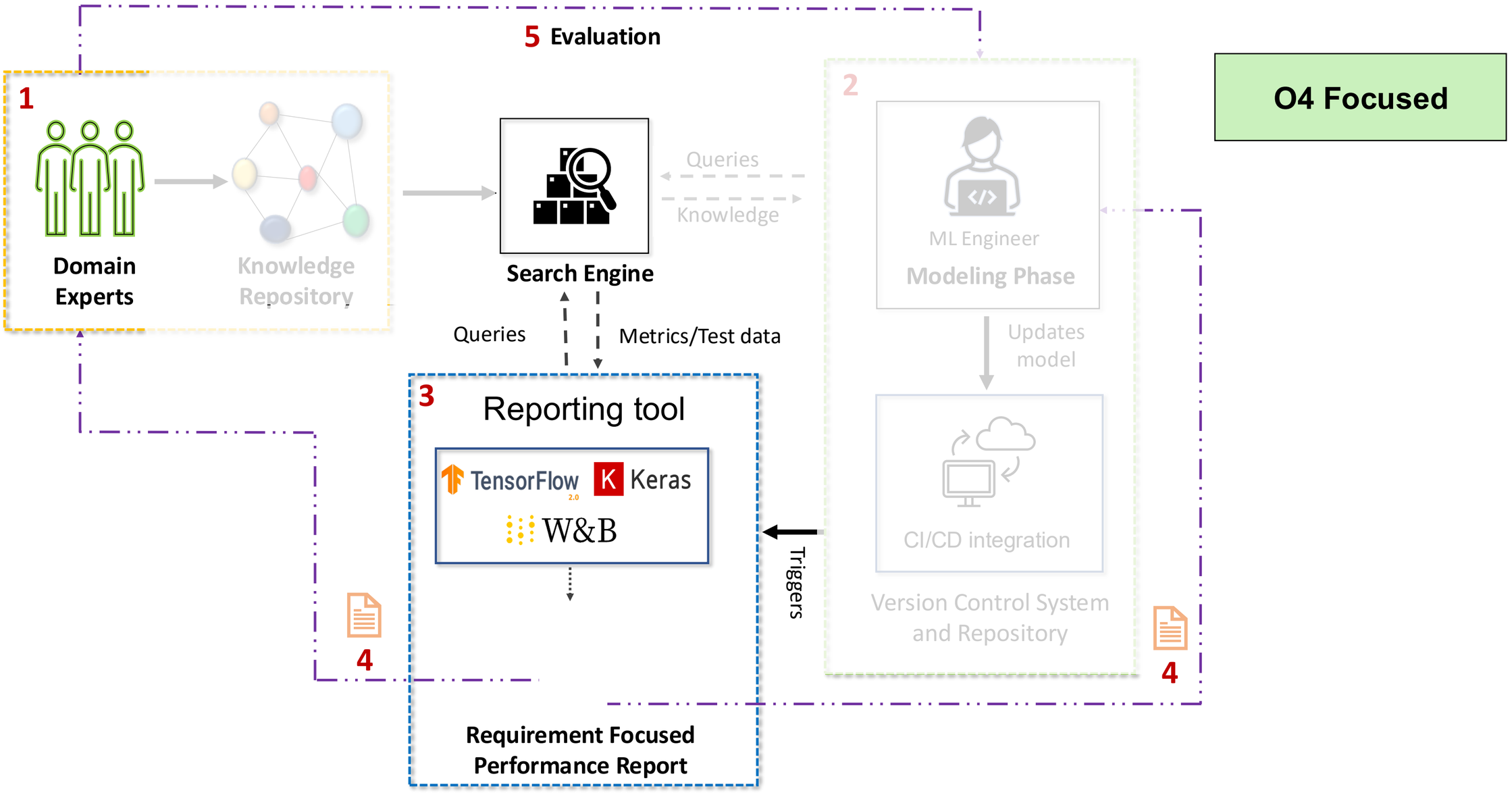
O3: Domain-Aligned Evaluation- Linking Metrics to Requirements



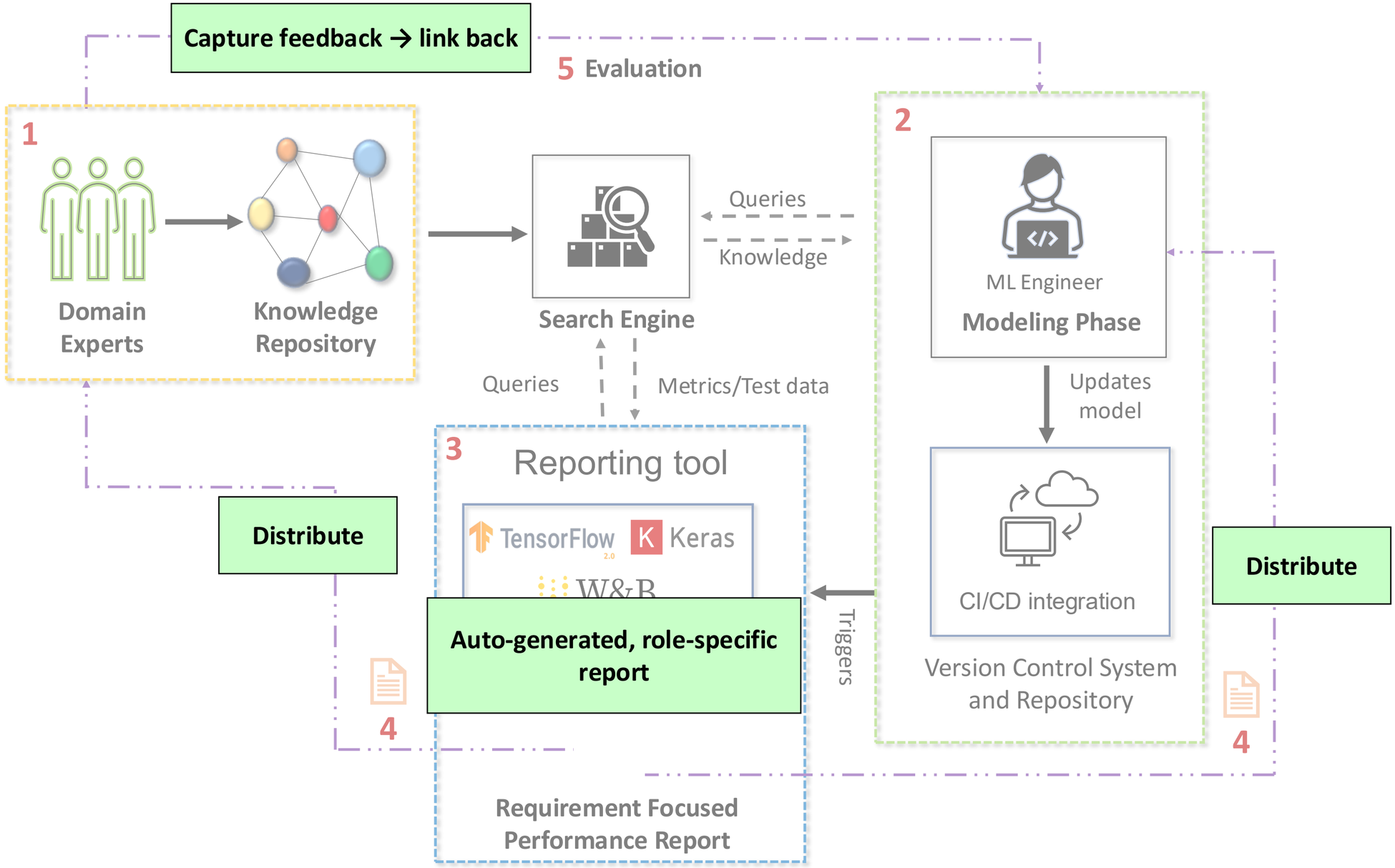
O3: Domain-Aligned Evaluation- Linking Metrics to Requirements



O4: Semi-Automated Reporting- Closing the Feedback Loop



O4: Semi-Automated Reporting- Closing the Feedback Loop

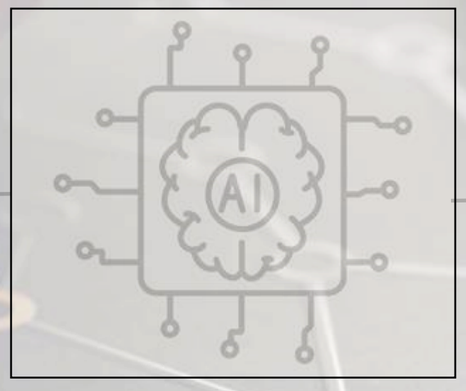


Presentation Flow

1

Motivation & Problem (O1)

2



Defining Research Scope (RQs)

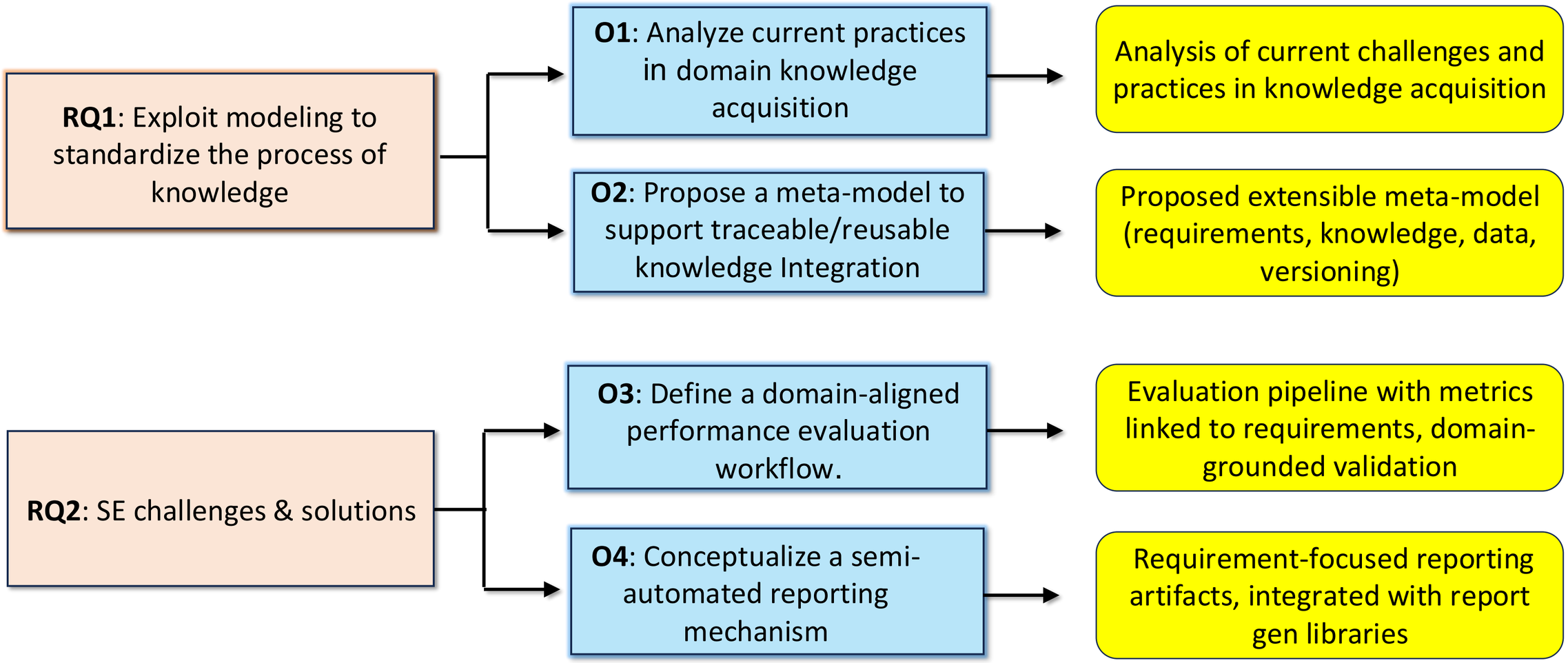
3

Conceptualized Solution (O2-O4)

4

Conclusion

Conclusion


RQ1: Exploit modeling to standardize the process of knowledge
O1: Analyze current practices in domain knowledge acquisition
Analysis of current challenges and practices in knowledge acquisition
O2: Propose a meta-model to support traceable/reusable knowledge Integration
Proposed extensible meta-model (requirements, knowledge, data, versioning)
RQ2: SE challenges & solutions
O3: Define a domain-aligned performance evaluation workflow.
Evaluation pipeline with metrics linked to requirements, domain-grounded validation
O4: Conceptualize a semi-automated reporting mechanism
Requirement-focused reporting artifacts, integrated with report gen libraries

Future Direction

Prototype the Framework
Build PoC for knowledge acquisition + evaluation flow

Integrate with ML Tooling
Link meta-model artifacts to ML platforms & CI/CD

Domain PoC Validation
Test with multidisciplinary case study (e.g., power system).

Extend Meta-Model
Add missing evaluation view & richer feedback

Thank you!