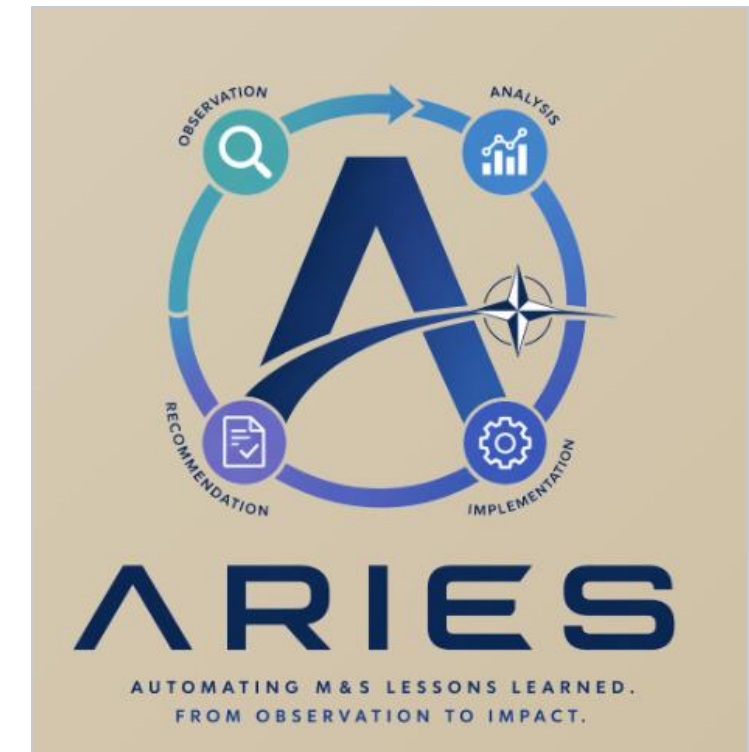




NATO M&S Lessons Learned process through a M&S-Specific Taxonomy and AI-Enabled Knowledge Management



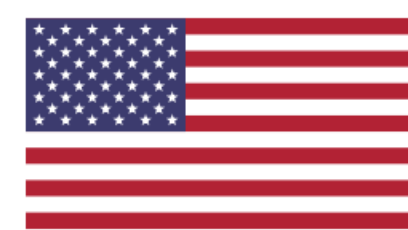
NATO MSG-242 ST: NATO M&S Services Lessons Learned Process

Tom Lasch – 4H4 Management GmbH, Germany

Jan Hodicky (Chair) – NATO Allied Command Transformation (ACT), Norfolk, VA, USA

Marco Gotelli – Simulation Team, Genoa, Italy

ARIES
*Autonomous Retrieval
and Intelligence Engine
for Synthesis*





Supporting the M&S LL Process: From Observation to Implementation

Applying AI to NATO process for Modelling & Simulation

1 Heterogeneous material

AARs, logs, exercise metrics, STO Technical Reports.
Manual analysis becomes increasingly difficult as the number and complexity of exercises grow

2 Defining M&S Taxonomy

Without a controlled Taxonomy and Vocabulary, lessons from one CAX are difficult to retrieve, compare and reuse across nations and federations.

3 LI → LL conversion

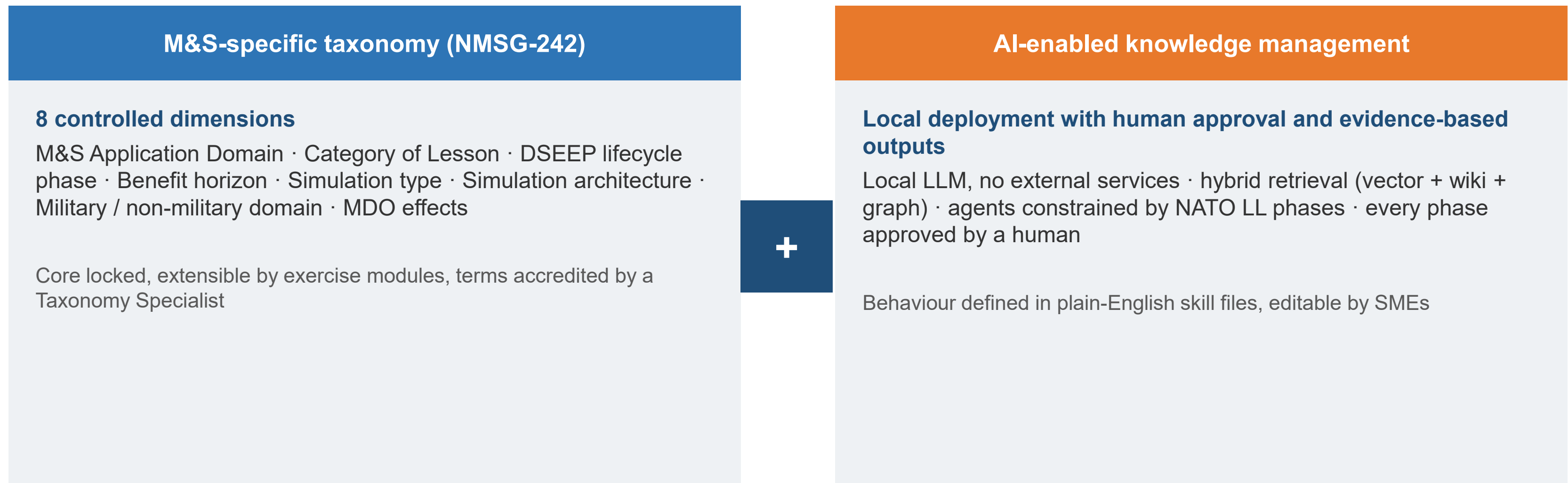
Observations are documented, but completing the Analysis, Remedial Action and Implementation phases remains challenging. As a result, many lessons remain at the 'Lesson Identified' stage

Reference process: NATO Lessons Learned Handbook (JALLC) · context: NMSG-242 activity on M&S Lessons Learned



Two Complementary Enablers

The paper's thesis

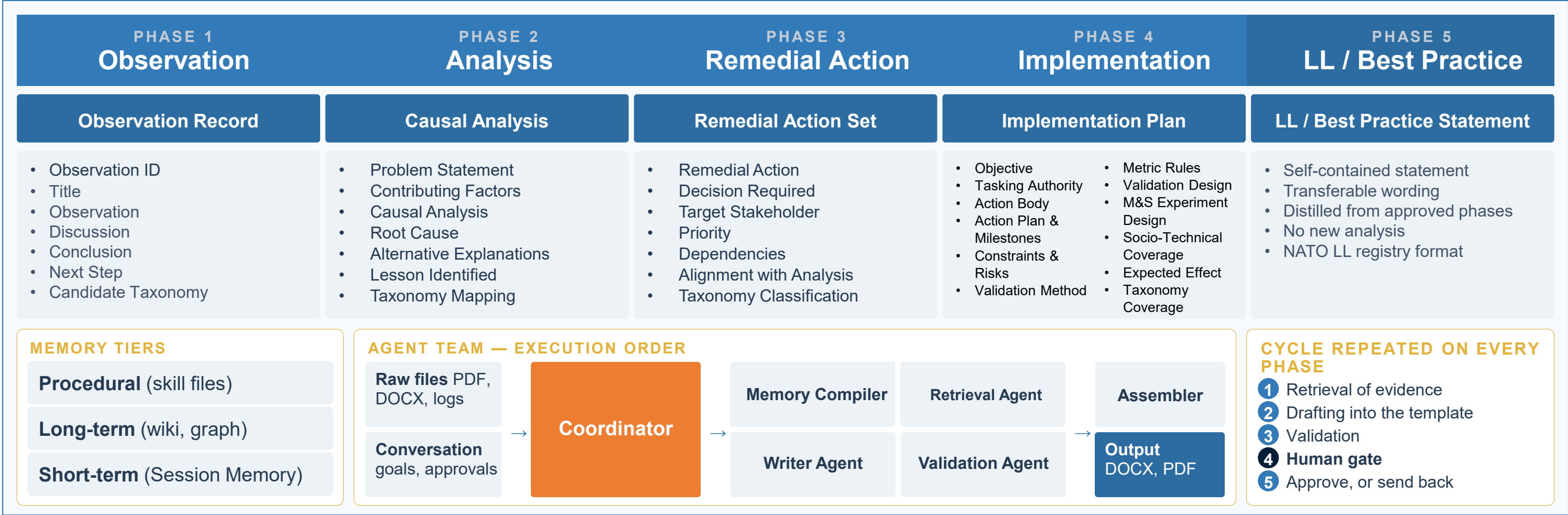


The taxonomy supports consistent classification and comparison across exercises, while AI facilitates its systematic application to larger volumes of exercise material.



INPUT DOSSIER
Exercise material, AARs, logs, metrics, etc.

ARIES AGENT TEAM
Local AI Based Solution to Support LI/LL process



ROLE OWNERSHIP AND ARIES USERS - Human Approval is Necessary at each Phase

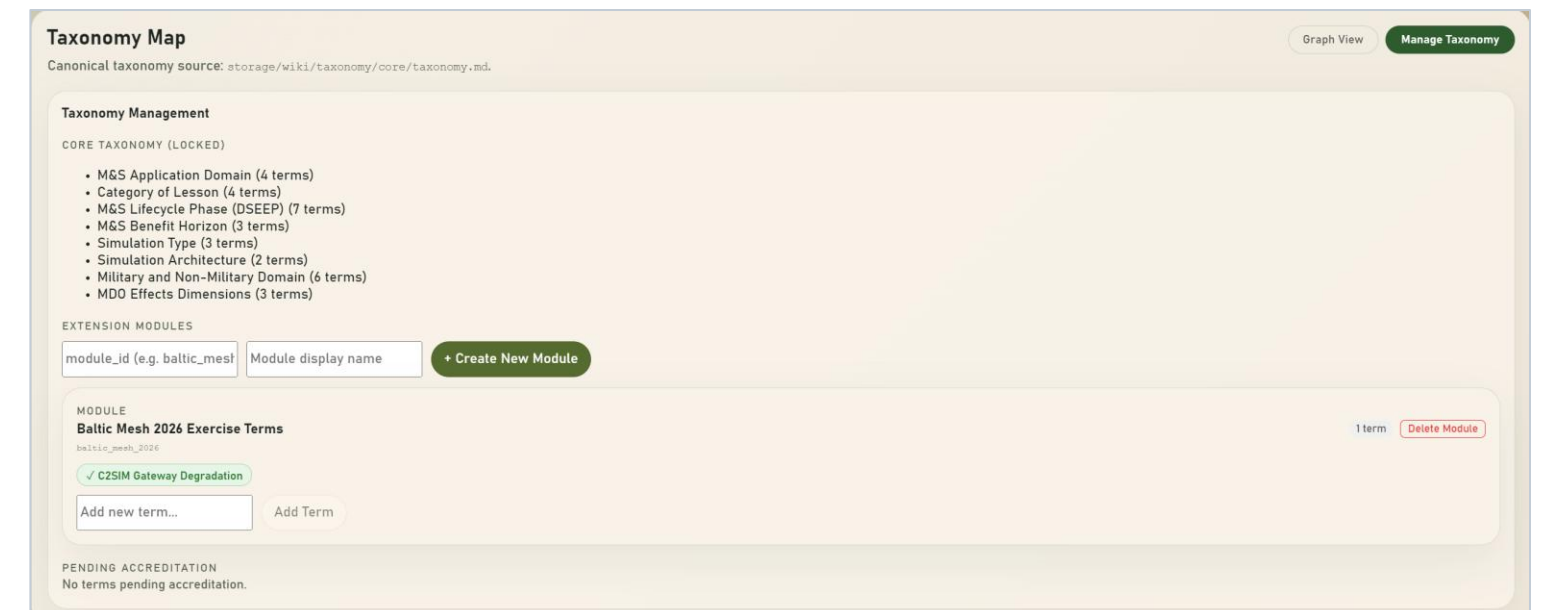
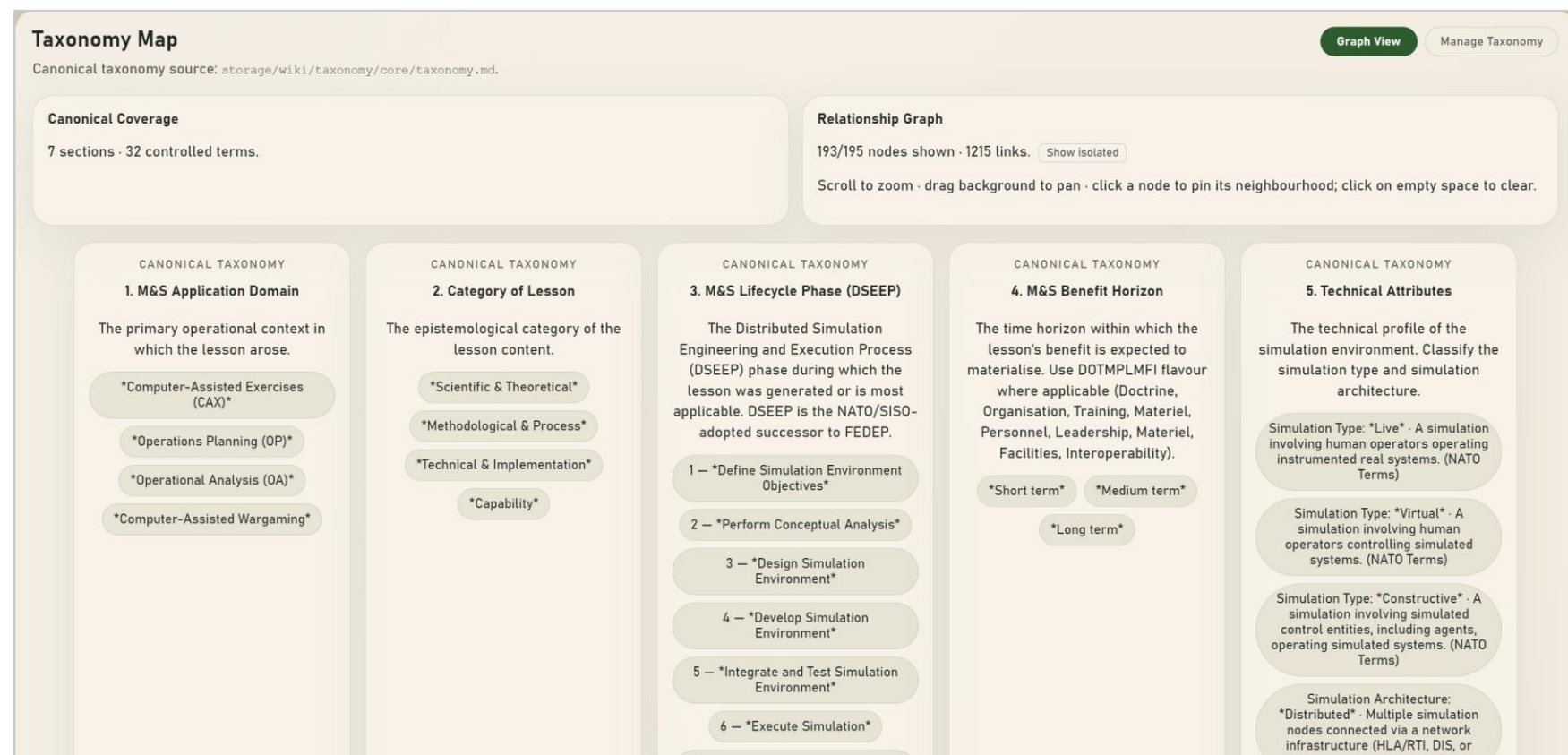
ROLES	PHASE 1	PHASE 2	PHASE 3	PHASE 4	PHASE 5
LLSO Lessons Learned Staff Officer	Coordinates the LL process end to end — runs the LL group and coordination team, tracks all LL tasks, supports the commander, contributes to Joint Analysis Requirements				
OA Originating Authority	Documents observation, analysis and conclusion — endorses the Lesson Identified, recommends Remedial Actions, names the Tasking Authority			Not Engaged	Uploads to the NLLP
TA Tasking Authority	Not Engaged		Decides on Remedial Actions, commits resources, appoints Action Bodies — tracks Lessons Identified through to full Lesson Learned status		
AB Action Body		Not Engaged		Implements the tasked Remedial Action	Not Engaged
M&S COMMUNITY ROLES					
M&S SME M&S Subject Matter Expert	Create and maintains the skill files that govern agent behaviour; defines, in plain English, what each phase must produce, the evidence rules and the validation criteria				
Taxonomy Specialist	Maintains NMSG-242 core taxonomy and accredits the terms added through extension modules				

Legenda: Potential ARIES Users



NMSG-242 Taxonomy in ARIES

A stable core taxonomy with exercise-specific extensions and an accreditation process



Canonical taxonomy

Defined sections, controlled terms and a relationship graph connecting taxonomy nodes, wiki articles, lessons and technical reports.

Core is immutable

The NMSG-242 core taxonomy cannot be modified at exercise level, ensuring consistent classification across nations and exercises.

Controlled extension process

Exercise-specific terms, such as 'C2SIM Gateway Degradation', are introduced through extension modules and reviewed by the Taxonomy Specialist before accreditation



Walkthrough 1: Ingestion and Classification

One exercise metrics file becomes a tagged, linked, auditable source

Validated Dossier

Detach from this session — drop this file from the dossier here; the global archive and other sessions keep their copy. Purge from long-term memory — also remove the stored raw file, raw summary, and manifest entry when you are the only user (otherwise the action falls back to a detach). Hover a button for details.

AVAILABLE

baltic_mesh_2026_event_metrics.xlsx

storage\raw\baltic_mesh_2026_event_metrics.xlsx

GENERATED

17/06/2026, 20:28:41

TYPE

xlsx

SIZE

5.9 KB

RAW SUMMARY

RS-sess_20260617_124414-2026-001

TAXONOMY

CATEGORY OF LESSON

Technical & Implementation

M&S APPLICATION DOMAIN

Operational Analysis (OA)

M&S BENEFIT HORIZON

Long term

M&S LIFECYCLE PHASE (DSEEP)

Step 5 - Integrate and Test Simulation Environment

MDO EFFECTS DIMENSIONS

Not classified

MILITARY AND NON-MILITARY DOMAIN

Not classified

TECHNICAL ATTRIBUTES / SIMULATION ARCHITECTURE

Distributed

TECHNICAL ATTRIBUTES / SIMULATION TYPE

Not classified

Auto-assigned taxonomy

Technical & Implementation · Operational Analysis · Long term · DSEEP step 5 · Distributed. When the available evidence does not support classification, the dimension remains ‘Not classified’

Link audit

6 direct and 158 second-level links into the domain wiki; 12 tags resolved, 0 unresolved. Accepted explicitly by the analyst.

LINK AUDIT

Accepted

6 direct158 second-level

RESOLVED TAGS

12

UNRESOLVED TAGS

0

Direct: OA (taxonomy), Technical & Implementation (taxonomy), Step 5 (taxonomy), Long term (taxonomy), Distributed (taxonomy), Baltic Mesh 2026 Event Metrics (knowledge)

Second level: M&S Lessons Learned Taxonomy (taxonomy), Conceptual Modelling (CM) (domain_knowledge), Cyber Defence Simulation (domain_knowledge), Data Farming (domain_knowledge), FEDEP / DSEEP — Federation Development and Execution Process (domain_knowledge), Federation Architecture and Design (DSEEP / FEDEP) (domain_knowledge), High Level Architecture (HLA) (domain_knowledge), Human Factors in Simulation (domain_knowledge), Simulation Interoperability (domain_knowledge), Modelling and Simulation as a Service (MSaaS) (domain_knowledge), NATO Modelling and Simulation Group (NMSG) (domain_knowledge), Simulation Fidelity (domain_knowledge), Synthetic Environment (SE) (domain_knowledge), Verification, Validation, and Accreditation (VV&A) (domain_knowledge), Wargaming Methodology (domain_knowledge), V&V Must Produce an Evidence-Based Acceptance Recommendation, Not a Checklist (lesson)

Accept audit

Refresh audit

Modify tags



Walkthrough 2: Inspectable Knowledge Base

Transparent and editable knowledge sources

Knowledge Build

0 source summaries | 0 knowledge articles

New Knowledge

Select documents

PDF, DOCX, TXT, MD, CSV, XLSX, and LOG files are compiled into the wiki.

[Choose files](#) No file chosen

[Ingest \(0\)](#) [Clear Queue](#)

No files queued for ingest.

Compiled Wiki

FIND

Title, tag, id, status

0

Raw summaries

0

Knowledge articles

104

Runtime pages

DOMAIN_KNOWLEDGE

shared KB

After-Action Review (AAR) and Lessons Learned Capture

draft

[Open markdown](#)

[Delete shared](#) [Delete total](#)

DOMAIN_KNOWLEDGE

shared KB

Conceptual Modelling (CM)

A Conceptual Model (CM) is a non-executable representation of the real-world system that a simulation is intended to model. It describes, in terms that both domain experts and simulation engineers can understand, what entities, relationships, states, events, and behaviours need to be captured — and with what scope and level of abstraction — to support the simulation's stated objectives.

draft

[Open markdown](#)

DOMAIN_KNOWLEDGE

shared KB

Cyber Defence Simulation

Cyber Defence Simulation is the application of M&S techniques to represent, analyse, and train for cyber operations — both defensive (protecting networks and systems) and offensive (understanding adversary cyber capabilities). It sits at the intersection of traditional constructive simulation and technical cyber range infrastructure.

draft

[Open markdown](#)

DOMAIN_KNOWLEDGE

shared KB

Data Farming

Data Farming is an analytical methodology that uses large-scale simulation experimentation — running many thousands of simulation replications across a structured design space — to produce data that supports decision-making. The term reflects the metaphor of "farming" insight from the fertile ground of simulation output, in contrast to "mining" insight from fixed historical data.

draft

DOMAIN_KNOWLEDGE

shared KB

Distributed Interactive Simulation (DIS)

Distributed Interactive Simulation (DIS) is the IEEE 1278.x family of standards defining a connectionless, datagram-based protocol for real-time exchange of entity state information between simulation systems. DIS predates HLA and remains in widespread use, particularly in vehicle, flight, and naval training simulators developed in the 1990s and 2000s.

draft

DOMAIN_KNOWLEDGE

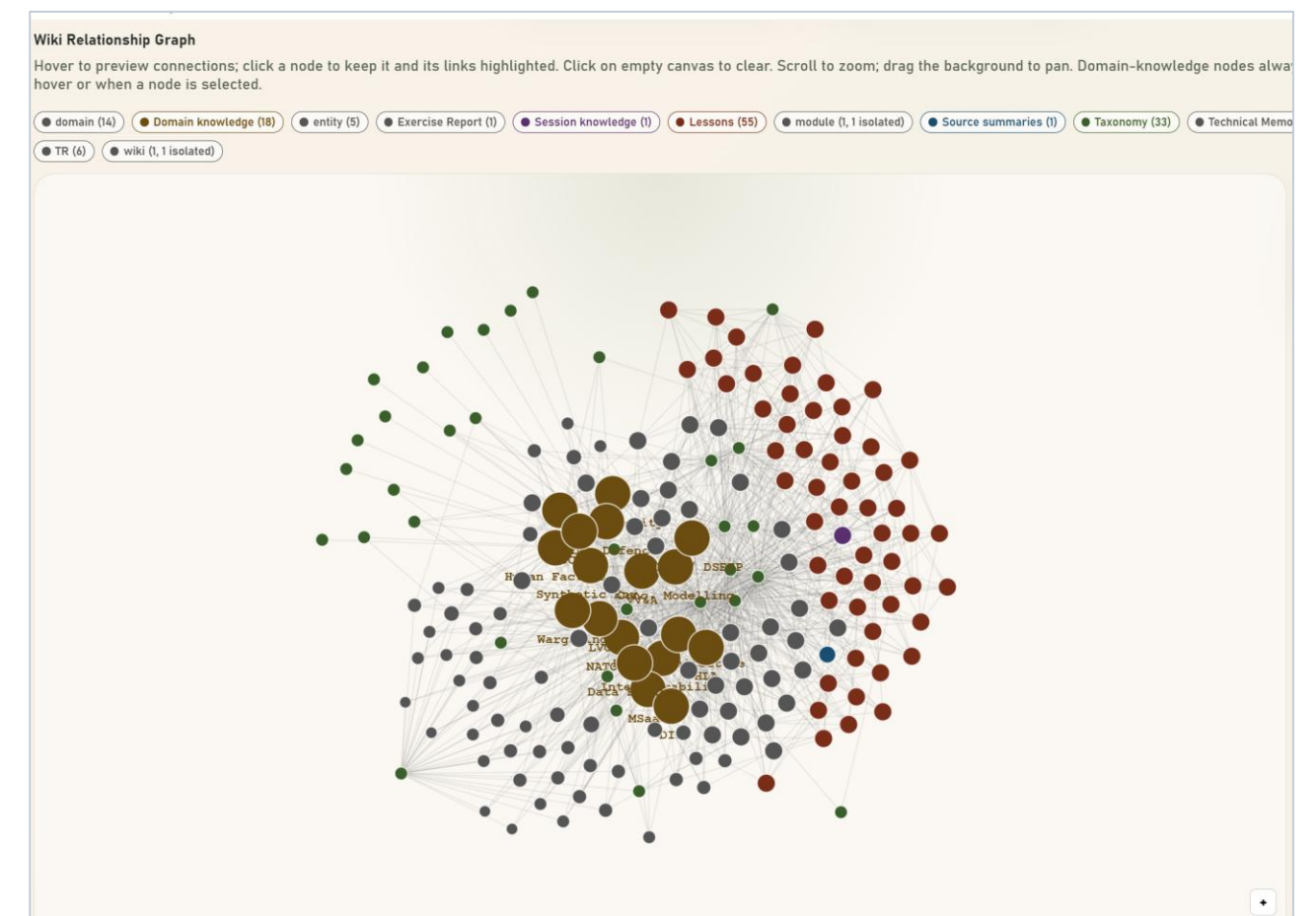
shared KB

FEDEP / DSEEP — Federation Development and Execution Process

FEDEP (Federation Development and Execution Process) was the original IEEE/DoD process model for the lifecycle of HLA-based distributed simulations. It defined six steps from requirements definition through federation execution and post-execution analysis.

draft

[Open markdown](#)



Seeded from NATO material

STO MSG Technical Reports (WP1 Appendix 1) + NATO LL Handbook + domain articles (CM, DIS, HLA, DSEEP, VV&A...). All content is stored in Markdown format and can be inspected, edited or removed by authorised users.

Wiki relationship graph

56 Technical Reports · 55 Lessons · 33 Taxonomy nodes · 18 Domain knowledge articles, linked by tags.



Walkthrough 3: Observation Drafting

Template fields filled by the Writer Agent, every sentence traceable

Observation ID

sess_20260617_124414

Edit

Observation

► During Exercise Baltic Mesh 2026, C2SIM/HLA gateway was affected when {fact: 'Primary Standards Used', 'value': 'HLA Evolved, C2SIM, FMN service profiles'}. Performance metrics indicated several breaches, including a high count of rejected logistics messages (peak 418 vs. Key events observed included the successful...

Edit

Conclusion

The available evidence is sufficient to treat this as a bounded lesson candidate for follow-on analysis, but it does not yet establish causal attribution.

Edit

Candidate Taxonomy

Operational Analysis (OA) / Technical & Implementation / Step 5 - Integrate and Test Simulation Environment / Distributed / [c2sim gateway degradation]

Edit

Title

C2SIM/HLA gateway disruption during Baltic Mesh 2026

Edit

Discussion

► Supporting evidence notes the Baltic Mesh 2026 was a federated interoperability test held on March 9, 2026. The exercise focused on Coalition C2 and logistics federation interoperability, utilizing a mission network laboratory to connect national logistics services, C2 viewers, and a synthetic movement model. The...

Edit

Next Step

Carry C2SIM/HLA gateway disruption during Baltic Mesh 2026 forward for causal analysis under Technical & Implementation and confirm scope, owner, and impact during Step 5 - Integrate and Test Simulation Environment.

Edit

Epistemic Separation

KIND	STATEMENT	EVIDENCE	GRADE
fact	► During Exercise Baltic Mesh 2026, C2SIM/HLA gateway was affected when {fact: 'Primary Standards Used', 'value': 'HLA Evolved, C2SIM, FMN service profiles'}. Performance metrics indicated several breaches, including a...	EV-OBS-001; EV-OBS-002; EV-OBS-003	observed

During Exercise Baltic Mesh 2026, C2SIM/HLA gateway was affected when {fact: 'Primary Standards Used', 'value': 'HLA Evolved, C2SIM, FMN service profiles'}. Performance metrics indicated several breaches, including a high count of rejected logistics messages (peak 418 vs. Key events observed included the successful joining of baseline services (07:50), the publication of convoy streams by the Logistics Service (08:27), and subsequent failures such as missing convoy tracks on Nation B Viewer (08:43) and confirmation of version mismatch via the Adapter Registry (09:31). [1] [2] [3]

[1] Baltic Mesh 2026 Event Metrics

session source summary

Summary

source: baltic_mesh_2026_event_metrics.xlsx

[2] Baltic Mesh 2026 Event Metrics

session source summary

Key Facts

source: baltic_mesh_2026_event_metrics.xlsx

[3] Exhaustive V&V Is Theoretically Impossible — Risk-Based Prioritisation Is the Only Viable Approach

global lesson

Observation

NATO Observation template
ID · Title · Observation · Discussion · Conclusion · Next Step · The system proposes a taxonomy classification for analyst review and approval

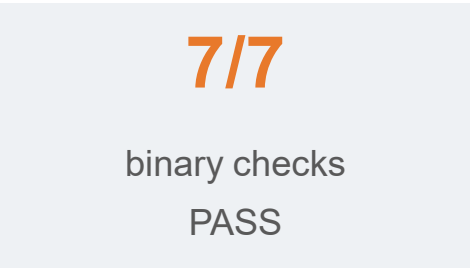
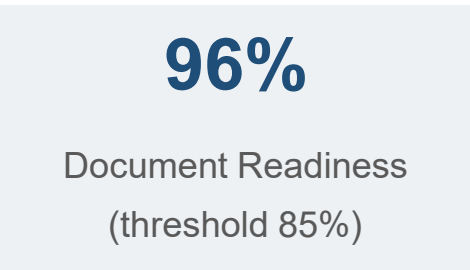
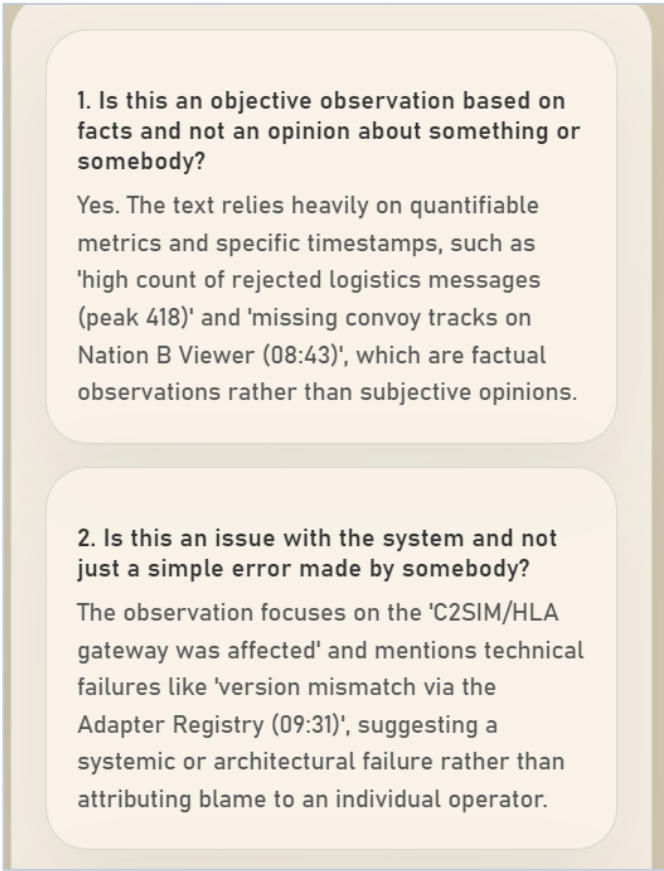
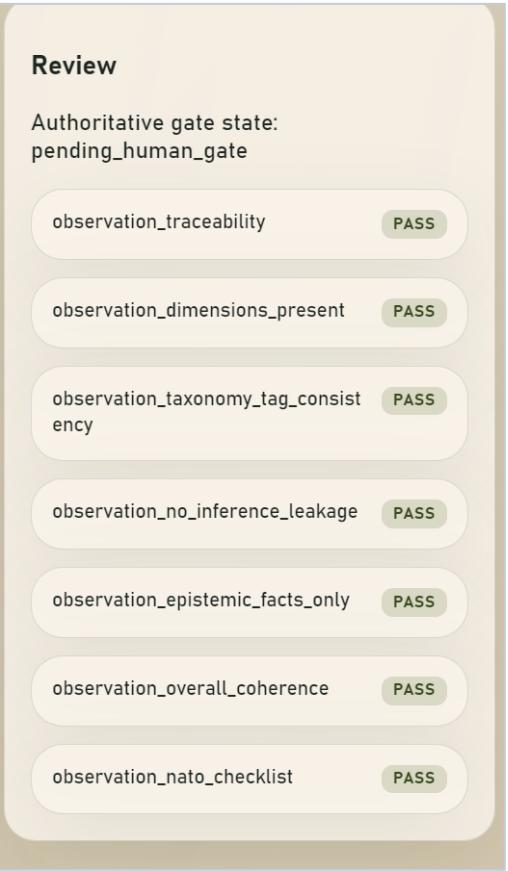
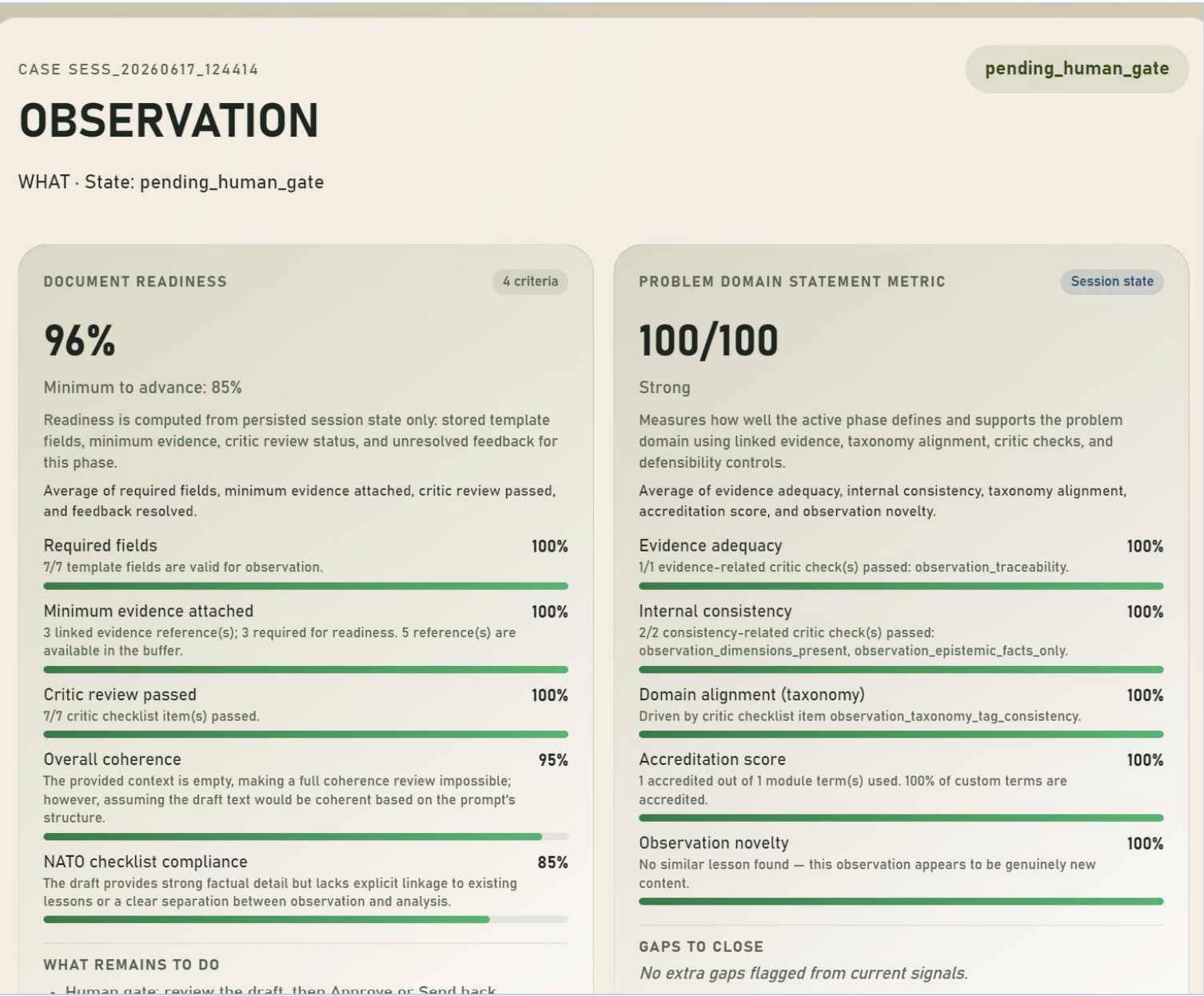
NATO STO - NATO MSG-242 ST: NATO M&S Services Lessons Learned Process

8



Walkthrough 4: Key Performance Indicators

Two KPI families, binary checks, 85% threshold



What is measured

Readiness = procedural compliance (fields, evidence, review, feedback).

PDS = content quality (evidence adequacy, consistency, taxonomy alignment, accreditation, novelty). Checks are binary and live in skill files. Override only with audit.



Walkthrough 5: Human Approval and Traceability

Each phase requires analyst approval, and every action is recorded.

GATE STATUSpending_human_gate

Run PhaseRe-check QualityApprove PhaseGenerate Final DOCX

KNOWLEDGE TOUCHPOINTS

18 document node(s)

Zoom inZoom outReset

Observation
ACTIVE PHASE

1

GM-VV Tailoring Is Mandatory — Ge...

global lesson
Observation

V&V Must Produce an Evi...
Exhaustive V&V Is Theor...
The V&V Composite Model...
Risk-Based Tailoring of...
Generic Methodology for...
Simulation Fidelity
Verification, Validatio...
FEDEP / DSEEP — Federat...
Federation Architecture...

1

Knowledge Network for Federation

global KB
Summary

FEDEP / DSEEP — Federat...
High Level Architecture...
Federation design knowl...
Wiki-based collaborativ...
Federation architecture...
PATHFINDER - Integratio...
Definition of a Framewo...

draft_passed | observation

The 'Baltic Mesh 2026' exercise provided critical operational metrics regarding federation interoperability (HLA Evolved, C2SIM). Key findings include a peak of 418 rejected messages and a Mean Time to Root Cause of 64 minutes. The primary remedial action identified is the creation of a signed adapter version register for the federation baseline. This observation significantly enhances the knowledge base by providing quantitative evidence of system weaknesses in complex, multi-domain environments.

Agent	critic		
Artifact	document_model.sections.observation.version=70		
Issues	None		
EV-OBS-001	Baltic Mesh 2026 Event Metrics	Summary	rrf via graph, pinned, rank 1
EV-OBS-002	Baltic Mesh 2026 Event Metrics	Key Facts	rrf via pinned, rank 2
EV-OBS-003	Exhaustive V&V Is Theoretically Impossible — Risk-Based Prioritisation Is the Only Viable Approach	Observation	rrf via qmd, rank 3

Revision history

draft_created by the Writer → draft_passed by the Validation Agent → observation version 70. Each evidence item carries its retrieval channel and rank (graph, pinned, wiki).

The gate

Run Phase · Re-check Quality · Approve Phase · Generate Final DOCX. Approval stores the phase text, creates a checkpoint and unlocks the next phase; send-back reopens the loop with the analyst's feedback.



Roles and Responsibilities

Human approval is required at each phase

LLSO	Lessons Learned Staff Officer	Coordinates the LL process end to end; runs the LL group; tracks all LL tasks
OA	Originating Authority	Documents observation, analysis and conclusion; endorses the Lesson Identified; names the Tasking Authority
TA	Tasking Authority	Decides on Remedial Actions, commits resources, appoints Action Bodies; tracks LI through to LL status
AB	Action Body	Implements the tasked Remedial Action
M&S SME	M&S Subject Matter Expert	Creates and maintains the skill files that govern agent behaviour: what each phase must produce, evidence rules, validation criteria
Taxonomy Specialist		Maintains the NMSG-242 core taxonomy and accredits terms added through extension modules

Blue: NATO LL Handbook roles · Gold: process owner · Orange: M&S community roles introduced by the approach



Current Limitations and Next Steps

Current scope of validation

Limits now	Next	Ask to this audience
• Local model size and RAM bound inference quality (design choice: no external services) • The phases from Analysis to Lesson Learned have been implemented but have not yet been tested with operational exercise data	• Multi-analyst calibration of KPI thresholds • Skill-file authoring by JALLC / NMSG SMEs	• Exercise datasets (AARs, logs, metrics) for a pilot • Reviewers for the Observation → Analysis hand-off



Takeaways

1

A shared M&S taxonomy is the precondition for lessons that travel across exercises and nations.

2

AI can support the Lessons Learned process when it operates on controlled evidence, remains subject to human approval and complies with information-handling requirements.

3

ARIES provides a working proof of concept that can support a pilot activity with NMSG-242



Backup — Technology stack

All components run on a single Windows laptop

Desktop app	Electron · React 18 · TypeScript · Vite
Backend	FastAPI · Python · Pydantic-validated session state (Global State Object)
Models	Ollama, locally hosted generation and embedding models
Knowledge store	LanceDB vector index · markdown wiki · NetworkX graph
Retrieval	Three channels fused by Reciprocal Rank Fusion
Export	Jinja2 templates · Pandoc → DOCX / PDF



Backup — Sessions and Case Summary

Persistent, resumable sessions; per-phase version baselines

PENDING_HUMAN_GATE
kptest2

Phase: observation | Updated: 05 Jul 2026, 21:04 CEST

Resume SessionDelete

BLOCKED
KPItest3

Phase: observation | Updated: 05 Jul 2026, 20:54 CEST

Resume SessionDelete

BLOCKED
sess_20260701_175614

Phase: observation | Updated: 05 Jul 2026, 19:01 CEST

Resume SessionDelete

IDLE
sess_20260619_133908

Phase: observation | Updated: 19 Jun 2026, 15:50 CEST

Resume SessionDelete

Case Summary

SESSION
kptest2

Phase: observation | Status: pending_human_gate

MISSION CONTEXT
Exercise not set

Domain: Unscoped | Classification: UNCLASSIFIED

WORKFLOW
0/4 phases approved

Active agent: coordinator | Evidence in scope: 5

PENDING FEEDBACK
None

No open send-back or critic payload.

SESSION SUMMARY
Analyst Context

The user initiated a session, which was acknowledged by the assistant.

PHASE BASELINES

observation	pending_human_gate	v70
analysis	not_started	v34
remedial_action	not_started	v36
implementation	not_started	v35
lesson_learned_best_practices	not_started	v43