ESnet Data and AI Workshop Report

June 2025

LBNL-2001684

Report Co-Authors

Chin Guok	ESnet
Ed Balas	ESnet
Sowmya Balasubramanian	ESnet
Justas Balcas	ESnet
Jaber Daneshamooz	UCSB
Sukhada Gholba	ESnet
Mike Haberman	ESnet
Shawn Kwang	ESnet
John MacAuley	ESnet
Sam Moats	ESnet
Matthew Nikahd	UCSB
Sam Oehlert	ESnet
Cody Rotermund	ESnet
Chris Robb	ESnet
Garrett Stewart	ESnet
Jiachuan Tian	ESnet
Chris Tracy	ESnet
Andrew Wiedlea	ESnet
John Wu	ESnet
Xi Yang	ESnet
Se-young Yu	ESnet

Table of Contents

1. Executive Summary

The United States Department of Energy (DOE) operates many cutting-edge experimental facilities that produce enormous amounts of data. These facilities include particle accelerators, high-intensity light sources, genomics and nanoscience centers, neutron-scattering facilities, and many others. The sheer volume and complexity of the data generated by these facilities necessitate a robust and high-performance network infrastructure to facilitate data transmission, analysis, and sharing. In this context, the Energy Sciences Network (ESnet), the DOE's high-performance network for science, plays a vital role as the critical infrastructure for high-volume data transmission, enabling researchers to collaborate, share data, and accelerate scientific discovery across the DOE ecosystem. To meet this need, ESnet is developing strategic initiatives to transform our operational framework and expand our service offerings, with a focus on harnessing emerging technologies to enhance our value to stakeholders.

The U.S. government has recognized the potential of artificial intelligence and machine learning (AI/ML) to drive scientific advancements and has launched initiatives such as Frontiers in Artificial Intelligence for Science, Security and Technology (FASST) to accelerate the integration of AI into scientific research. In line with these efforts, ESnet has embarked on a structured investigation to identify areas in which AI technologies and methodologies can best be leveraged. In February 2025, ESnet held a two-and-a-half day Data and AI Workshop in Berkeley, California. The objective of the workshop was to identify challenges within ESnet that could be addressed through data-driven methods, to help define ESnet's data-analysis requirements, and to shape our AI strategy, guiding data-stewardship efforts and the direction of AI research and AIOps exploration for ESnet7, the next iteration of ESnet's network. This report summarizes the multi-faceted discussions and findings and presents a set of recommendations for next steps.

The primary motivation for integrating AI into networking is to address the growing complexity of modern networks. Today's digital infrastructure is characterized by an enormous volume of telemetry data, highly dynamic user behavior, and rapidly evolving threat landscapes. Traditional methods based on static rules and manual interventions can no longer cope with the scale and sophistication required. Instead, AI-driven new approaches are essential for effective network planning, real-time operations, and proactive threat mitigation. In this rapidly changing environment, the workshop's focus on "understanding AI" and the role that data plays is not only timely but also critical for developing automation strategies that are both secure and efficient.

To turn data into information and information into insights, the quality of the data has a direct impact on the value of the insight. ESnet's extensive and varied operational data, which comes from a range of sources, can present some challenges in terms of accessibility and comprehension, largely due to the different methods, structures, and tools used across various sources [Section 3.1.1, DM.F1]. The underlying datasets can exhibit some inconsistencies in curation, documentation, and metadata that can make larger-scale programmatic analysis more complicated [Section 3.1.2, DM.F2]. Additionally, understanding the operational data across the entire end-to-end cyberinfrastructure paths used in

science pipelines is crucial. These paths often cross multiple administrative domains, and harvesting the operational data involves a manual and time-consuming process. [Section 3.1.3, DM.F4]. Apart from operational data, ESnet does possess a substantial amount of human-generated content, such as documentation and incident response tickets; with the help of AI, this data can be leveraged more effectively to provide valuable assistance in the future. [Section 3.1.2, DM.F3]

Currently, ESnet primarily employs traditional analysis techniques to derive insights that inform decision-making, including monitoring network performance, generating alerts, and predicting hardware failures [Section 3.2.3, TAM.F1]. However, traditional analysis methods have limitations [Section 3.2.3, TAM.F2], and data quality and accessibility can add to the challenge [Section 3.2.3, TAM.F3]. Many problems require integrating data from multiple systems and applying advanced multivariate techniques [Section 3.2.3, TAM.F4], with some necessitating hybrid approaches, combining classical statistical techniques with Natural Language Processing (NLP) or AI/ML [Section 3.2.3, TAM.F5]. With the enormous scope of possible analysis tasks, priority focus should be on efforts with high operational value and readily available high-quality data, to maximize the return on investment and ensure effective resource utilization [Section 3.2.3, TAM.F6].

ESnet is just beginning to explore the application of AI in its operations, and the potential for widespread adoption is significant. The organization is particularly interested in using AI for anomaly detection in three key areas: analyzing trends, identifying logical or rule-based anomalies, and examining semi-structured and unstructured data [Section 3.3.1, AIM.F1]. AI has already shown promise in root-cause analysis, data management, and quality assurance, and ESnet believes it can enhance early detection, causal understanding, and historical context [Section 3.3.5, AIM.F5; Section 3.3.6.4, AIM.F6]. To support AI-driven automation, ESnet is considering central workflow engines, modular AI agents, and action-oriented automation [Section 3.3.2, AIM.F2], as well as data preparation, decision support, and document generation to facilitate NLP queries and improve data accessibility [Section 3.3.3, AIM.F3]. Additionally, predictive modeling is a key area of focus for ESnet, with applications in operational intelligence, network performance, and administrative automation [Section 3.3.4, AIM.F4].

To ensure effective and trustworthy AI integration, ESnet believes it is crucial to prioritize transparency, accountability, and user trust [Section 3.4.2.8, UX.F1]. This can be achieved by providing contextual data explanations, aligning AI workflows with ESnet-specific data and use cases [Section 3.4.2.2, UX.F2], and investing in data hygiene and accessibility [Section 3.4.2.2, UX.F3]. Additionally, implementing Human-in-the-Loop (HITL) practices [Section 3.4.3.1, UX.F7], clear retraining and monitoring processes [Section 3.4.2.4, UX.F5], and version-controlled documentation [Section 3.4.2.5, UX.F6] can help maintain effective AI systems [Section 3.4.2.3, UX.F4]. It is also essential to manage trust, ethics, and bias [Section 3.4.3.4, UX.F10] through transparency, labeling, and user control, and to provide clear and contextual information around AI inputs (Section 3.4.3.2, UX.F8; Section 3.4.3.3, UX.F9). Furthermore, continuous user feedback [Section 3.4.3.6, UX.F12], thoughtful design choices [Section 3.4.3.7.3, UX.F13], and transparent data-usage disclosures [Section 3.4.3.5, UX.F11] are vital for ensuring AI systems are accurate, relevant, and aligned with dynamic operational contexts, while also protecting sensitive information and promoting societal acceptance [Section 3.4.3.8, UX.F14].

The 31 recommendations that follow comprise a mix of strategic and tactical guidance for ESnet.

Data Management Recommendations	
DM.R1	ESnet should improve data discovery, comprehension, and confidence using NLP and enterprise search and should explore new ways to gain insight from existing data by augmenting it with improved metadata and using AI- and ML-based analysis techniques that are aware of network and service topologies.
DM.R2	To support a zero-trust architecture and operational innovation, ESnet should establish a consistent information architecture, providing Application Programmable Interface (API)-driven access to well-structured data, and ensuring uniform access control across all data sources. This includes creating a comprehensive data catalog, facilitating incident response through consistent data access, and identifying and addressing gaps in existing data sources to enable increased automation and analysis.
DM.R3	ESnet should develop a comprehensive and integrated view of production services by providing seamless access to both technical and business data, and establishing a detailed operational service model that includes data retention and lifecycle management.
DM.R4	ESnet should develop a standardized data flow to unify and normalize data from various operational domains, including telemetry, system logs, and network configuration, to make data "ready for AI." This data flow should include automated checks for data consistency, enforcement of schema format, and metadata records to ensure high-quality data for AI analysis. It will require human-led control and data engineering investments to support effective AI use within ESnet.
DM.R5	To enhance operational efficiency, ESnet should focus on bridging key information gaps, including: tracking underutilized allocated resources, such as unused bandwidth reservations, gathering essential data to facilitate improved automation and analysis, and establishing a unified security analysis framework that integrates with all data sources and adheres to industry best practices.
DM.R6	To deliver optimal services, it is necessary to monitor the health and status of end-to-end cyberinfrastructure beyond ESnet's facility boundaries. ESnet should collaborate with the R&E networking community to develop end-to-end awareness through trusted and federated authentication, and to correlate identifiers across different domains.

Traditional Analytical Methods Recommendations

TAM.R1	To maximize impact, ESnet should prioritize work-packages that present a clear opportunity for statistical analysis, focusing on areas where data quality is high and operational impact is significant. Specifically, initial efforts should concentrate on network operations, capacity planning, and predictive maintenance, where the potential benefits of statistical analysis are likely to be most pronounced.
TAM.R2	Investing in data quality and integration is essential to unlocking the full potential of statistical analysis. Before applying advanced analytical methods, ESnet should improve data standardization, metadata enrichment, and automated collection processes. Statistical analysis should be built on a foundation of clean, reliable data, rather than attempting to apply analytical techniques to subpar data, which can lead to inaccurate or misleading results.
TAM.R3	To maximize analytical effectiveness, ESnet should leverage hybrid approaches that combine the strengths of statistical methods with those of NLP and AI/ML. By applying traditional statistical techniques to structured data and utilizing NLP for free-form or unstructured data, teams can unlock a more comprehensive understanding of their data and drive more informed decision-making.
TAM.R4	To ensure the reliability and accuracy of statistical analysis and lay the groundwork for future AI/ML initiatives, ESnet should promote data stewardship and ownership across the organization. This involves assigning clear responsibility for maintaining data integrity, establishing normalization protocols, managing data life cycles, and setting metadata standards for key systems. Additionally, reinforcing clarity around data custodianship will help to ensure that data is properly managed, maintained, and utilized, ultimately supporting informed decision-making and driving business value.
TAM.R5	Effective statistical analysis requires collaboration and iteration with stakeholders. To ensure that insights are relevant and useful, ESnet should validate approaches with operational teams, including engineers, operators, and business staff. This involves working closely with stakeholders to ensure that statistical insights are actionable, understandable, and align with business objectives, ultimately driving meaningful outcomes and informed decision-making.

AI Methods Recommendations

AIM.R1	For anomaly detection, ESnet should invest in time-series analysis of the various metrics, identify patterns in the data, and make these insights available across ESnet through dashboards as well as programmable interfaces. This analysis is fundamental to any automated intelligent monitoring or validation systems that ESnet wants to build and is essential for improving observability, detecting failures, creating AI troubleshooting assistants, checking for Service Level Agreement (SLA) compliance, and building predictive models.
AIM.R2	Although automation can significantly amplify efficiency and productivity, it is crucial to identify and prioritize areas where automation can yield the greatest return on investment. ESnet should examine three approaches for automating workflows and processes: (1) using LangGraph or another central workflow engine to orchestrate tasks such as alerting, (2) employing modular AI agents to expand the automation to distributed workflows, and (3) exploring action-oriented automation for HITL processes such as site deployment and contract generation, preserving human oversight for complex or risky decisions.
AIM.R3	ESnet should explore NLP technologies with domain-specific fine tuning for improving data integration, data access, decision support, and document generation. This effort could enhance network visibility by integrating data from various sources, including alerts and trouble tickets, with network telemetry. This could also provide assistance and recommendation for generating new documents such as ticket resolution, network configuration, and service contracts.
AIM.R4	Managing the ESnet Wide Area Network (WAN) is growing exponentially more complex, resulting in tasks that are increasingly time intensive and impractical to manage manually. To ensure operational efficiency, ESnet should adopt predictive modeling to enhance incident management, optimize resource allocation, and automate processes through data-driven intelligence.
AIM.R5	Reducing the time to resolution (TTR) is an important operational goal for ESnet. ESnet could leverage AI tools and methods to automate and enhance root-cause analysis, improving network resilience and stability by analyzing vast amounts of operational data, identifying correlations, and suggesting potential root causes.
AIM.R6	To make data AI-ready, ESnet should leverage AI methods to address several data-management requirements, including ensuring consistent data formatting, deploying efficient data-access methods, and enhancing data quality assurance.

User eXperience Recommendations

UX.R1	To ensure effective adoption and sustained trust in AI-assisted operations at ESnet, AI workflow management must prioritize intuitive, transparent, and task-aligned user experiences. ESnet should incorporate layered User eXperience (UX) elements that expose intermediate steps, model outputs, and decision rationales to users with varying levels of technical expertise. Workflow interfaces should offer clear visual indicators of progress, highlight key dependencies, and present actionable outputs in context. Parameter tuning and configuration options should be abstracted to user-friendly controls wherever possible, without sacrificing flexibility.
UX.R2	To unlock the full potential of AI at ESnet, data-management UX must be treated as a first-class design consideration, emphasizing data quality, context, and hygiene while facilitating seamless data access and collaboration among data owners and ML engineers. ESnet should architect a cohesive data management framework that enforces data hygiene and normalization, supports interoperable formats, and provides intuitive tooling for data exploration and preparation. Data access workflows should include clear metadata, provenance, and usage guidance, while respecting Controlled Unclassified Information (CUI) and Personally Identifiable Information (PII) restrictions. APIs must be well documented and enable frictionless collaboration by data owners and ML practitioners.
UX.R3	To ensure safe, reliable, and user-aligned AI behavior, especially in operational environments, it is necessary to prioritize robust validation, especially for edge cases, and implement protocols for retraining and continuous performance monitoring. ESnet should implement comprehensive edge-case testing frameworks that simulate rare or ambiguous scenarios, as well as adversarial testing techniques to stress-test model boundaries. Where feasible, training datasets should be augmented with noise and synthetic edge cases to improve robustness. Additionally, automated test harnesses should include validation checkpoints for high-risk actions, and all AI-driven recommendations that impact systems or users should be gated through Human-in-the-Loop approval mechanisms.
UX.R4	To ensure sustained effectiveness and reliability, AI systems must be supported by robust maintenance and monitoring frameworks that adapt to evolving data, user needs, and operational contexts. ESnet should implement retraining protocols with clear triggers, such as performance thresholds, scheduled intervals, or significant input distribution shifts, and ensure that retraining datasets are curated for quality and diversity.

UX.R5	To ensure successful AI implementation at ESnet, interdisciplinary collaboration and rigorous documentation must be treated as core components of system design and operation. ESnet should have structured collaboration workflows that facilitate continuous knowledge exchange, such as regular cross-team reviews, shared glossaries, and integrated feedback loops. All models, prompts, decision logic, and configuration changes should be version-controlled in a central repository to ensure reproducibility and auditability.
UX.R6	To ensure AI systems deployed at ESnet remain transparent, accountable, and aligned with user intent, HITL design must be embedded as a core architectural principle. ESnet should design AI interfaces that clearly distinguish between suggestions and actions, provide intuitive editing and approval workflows, and highlight high-risk or ambiguous outputs for manual review. Systems should offer unobtrusive but accessible override and rollback options, and integrate correction inputs directly into retraining or feedback loops to improve model performance over time.
UX.R7	To maximize the effectiveness of AI systems within ESnet, it is essential to embed context-aware guidance into user interactions. Suggested context, such as sample queries, autocomplete options, and dynamic prompt scaffolding, helps users formulate clearer, more precise inputs, leading to better AI responses and a more intuitive overall experience. ESnet should integrate context-sensitive help features that adapt to the user's task, role, and system state, along with onboarding tools that introduce users to effective prompt strategies. Additionally, natural language interfaces should proactively offer clarifying suggestions or corrections when ambiguous or incomplete inputs are detected.
UX.R8	To ensure AI outputs are actionable, trustworthy, and properly understood within operational environments like ESnet, it is essential to embed rich, interpretable context directly into system responses. ESnet should explore mechanisms and processes to ensure that all AI outputs include clear explanations of the underlying data, the reasoning behind recommendations, and, where applicable, confidence levels and risk indicators. The tone and assertiveness of outputs should be adjusted based on uncertainty or impact, signaling whether a result is a strong recommendation or a tentative suggestion. Where decisions carry potential operational consequences, outputs must clearly communicate fallback options and the scope of impact in the event of error.
UX.R9	To ensure the responsible and ethical deployment of AI systems across ESnet, it is imperative to implement mechanisms that promote transparency, user agency, and clarity around AI-generated content. ESnet should use visual indicators, disclaimers, and metadata tags to differentiate AI-generated outputs from human-authored content. Interfaces should maintain an auditable history that distinguishes between AI-driven and manual actions, supporting traceability and accountability. Where feasible, provide users with the ability to opt out of specific AI features or adjust the level of automation based

	on their role or context. Additionally, communicate the system's capabilities, limitations, and known biases clearly to set accurate expectations and avoid misuse.
UX.R10	To uphold ethical standards and protect sensitive information within AI-driven systems at ESnet, transparent data usage disclosure must be integrated into all user-facing interfaces and workflows. ESnet should incorporate clear, persistent disclaimers or visual indicators when data could be stored, analyzed, or influence future AI behavior. Interfaces should include context-aware warnings, particularly when inputs are entered into free-form fields that may inadvertently capture sensitive content, and provide guidance on safe data-entry practices. Additionally, administrative controls must allow data owners to configure data-collection policies, with granular options for opting in or out of training pipelines.
UX.R11	Integrating user feedback into AI systems is a foundational practice for ensuring ongoing model accuracy, adaptability, and domain relevance within ESnet's dynamic operational environment. ESnet should embed lightweight, intuitive feedback tools, such as approval buttons, correction prompts, or rating scales, directly into the user interface to encourage participation without interrupting workflow. Where feasible, feedback should be captured in a structured format and integrated into retraining pipelines, allowing for supervised fine tuning that reflects real-world performance. Systems should also prioritize transparency by indicating how user feedback is used and offering visibility into the impact of cumulative input over time.
UX.R12	Selecting the appropriate interface granularity is essential for delivering AI features that align with user needs, operational contexts, and the intended depth of interaction. Based on work-package analysis, ESnet should tailor AI integration using one of three established UX design frameworks, depending on task complexity and user engagement levels: immersive, assistive, or embedded.
UX.R13	ESnet should evaluate each AI-enabled work-package to determine the optimal interface granularity and carefully map that decision to user roles, task criticality, and environment constraints.
UX.R14	To ensure clarity, foster appropriate trust, and prevent misinterpretation of AI capabilities, user interfaces should be explicitly designed to avoid anthropomorphizing AI systems. ESnet should use neutral, technical language in system prompts and responses, avoiding terms that imply emotion, intention, or personality. Visual elements, such as avatars or icons, should reinforce that users are engaging with a system, not a person. Additionally, disclaimers or contextual indicators should clarify the deterministic or probabilistic nature of AI outputs.

As the DOE's primary scientific data network, ESnet is uniquely positioned to leverage AI not only to optimize our own operational efficiency, but also to contribute to the broader goals of the FASST program by providing a cutting-edge network infrastructure that supports AI-driven scientific discovery. ESnet's focus on AI directly addresses key challenges identified by the DOE, including the need to enhance data quality, establish robust metadata standards, and ensure the trustworthiness of data and AI-driven insights (Bertino et al., 2021; Brown et al., 2023; Carter et al., 2023; Dart et al., 2023; Stevens et al., 2020).

2. Workshop Details

ESnet's Feb. 2025 Data and AI workshop was attended by roughly 50 staff members, representing a broad cross-section of ESnet's 144-person organization at the time of the workshop, and 6 invited experts from academia and industry. The workshop consisted of 5 main sessions, each building on the output of the preceding session. The sequence of the sessions is presented in Figure 2.a and described below:

- **Session 1: Understanding the questions / problems** — This session focused on identifying the key questions and problems that ESnet must address as we prepare for ESnet7.
- **Session 2: Understanding our data** — This session provided a high-level understanding of relevant data sources available today within ESnet, which are expected to be available soon, or likely will be needed to support the ESnet Facility's complete life cycle.
- **Session 3: Understanding AI** — This session brought together a curated group of external researchers and developers to explore the possibilities of advanced analytics and AI technologies. Through a series of concise presentations and an interactive Q&A discussion, participants acquired a fundamental understanding of AI's strengths, limitations, and applications. This primer aimed to equip participants with the knowledge to discern when AI is the optimal solution for specific challenges and when traditional analytical methods are more effective.
- **Session 4: Bringing it together, building work-packages** — This session analyzed the outputs of the prior sessions and defined work-packages that included the specific questions/problems identified (Session 1), the data required to address the questions/problems (Session 2), relevant AI or traditional methods (Session 3), and the expected user experience in interacting with the solution.
- **Session 5: Where do we go from here?** — This session performed an initial cross-cut analysis across the various work-packages. It focused on four areas: Data Management, Traditional Analysis, AI Methods, and User eXperience.

Figure 2.a. Workshop session sequence.

For each session, the workshop attendees were divided into groups of 6 to 12 to work on session-specific tasks. Groups were intentionally composed of attendees with a mix of technical backgrounds, organizational roles, and years of experience.

A detailed agenda of the workshop can be found in Appendix A1 and the list of attendees in Appendix A2.

2.1. Session 1: Understanding the Questions / Problems

Session 1 was designed as a foundational step in defining ESnet's data analysis requirements and shaping the organization's data & AI strategy for ESnet7 and beyond. The workshop recognized the importance of capturing a holistic view of operational challenges by tapping into the diverse experience and expertise of staff across the organization. The session was structured around a set of "charge questions" designed to elicit not just surface-level frustrations but the fundamental, systemic barriers to effective operations and decision-making:

1. What are your most significant challenges, operational or otherwise?
2. What decisions are hardest to make with current information?
3. What are the biggest bottlenecks in your current workflows?
4. What capabilities would most improve your productivity?

Groups were encouraged to focus on clearly articulating problems—using real-world examples and "looping" ideas back to the core issue—rather than proposing solutions at this stage. Participants were reminded to ask, "What is the problem we are solving?" and to clarify the distinction between symptoms and root causes.

At the end of the session, each group delivered a 6- to 7-minute readout in a plenary. The readouts provided an opportunity for cross-group questioning and clarification, fostering a shared understanding of the challenges and reinforcing the importance of "admiring the problem" before proposing solutions. Common themes and high-priority issues were captured in real time, ensuring that both nuance and consensus were documented.

2.1.1. Key Themes and High-Impact Problems Identified

The breakout discussions and readouts revealed a set of cross-cutting themes that spanned technical, organizational, and cultural dimensions. These themes represent the most impactful and persistent challenges facing ESnet as we seek to modernize our operations and leverage data and AI-driven approaches.

2.1.1.1. Data Integrity, Consistency, and Accessibility

Fragmentation and Siloes: Essential operational data—such as circuit inventories, service statuses, contracts, and configuration details—are distributed across multiple, often disconnected systems (e.g., internal tools such as the ESnet database, or ESDB; IP Address Management, or IPAM, systems; Business Office databases). This fragmentation leads to inconsistencies, duplication, and ambiguity about which system constitutes the "source of truth."

Quality, Curation, and Lifecycle Management: Data is frequently incomplete, outdated, or unstructured. There is no standardized process for regular data audits, validation, or lifecycle management. Documentation, when it exists, is scattered across wikis, Google Docs, and ticketing systems and is often stale or sometimes contradictory.

Information Architecture Gaps: The absence of a dynamic, queryable information model means that relationships between low-level network data and high-level business or service constructs are difficult to establish or maintain. This impedes impact analysis, troubleshooting, and the ability to adapt to evolving requirements.

Accessibility and Searchability: Staff struggle to locate, access, and correlate relevant information quickly. There is a strong desire for universal search capabilities, including natural language interfaces, that can traverse structured and unstructured data repositories.

2.1.1.2. Workflow Inefficiencies and Process Bottlenecks

Manual and Repetitive Tasks: Many routine activities—such as updating documentation, triaging tickets, managing email communications, and handling administrative logistics—are performed manually, consuming valuable staff time and increasing the risk of errors.

Configuration Management Challenges: Generating, validating, and deploying network configurations is labor intensive and prone to mistakes, especially when relying on a mix of vendor tools and homegrown scripts. The lack of robust staging, rigorous validation, and automation tools slows service fulfillment and increases the risk of outages or misconfigurations.

Inconsistent and Non-Standard Workflows: Teams often develop their own processes and tools in isolation, leading to inconsistencies, duplicated effort, and barriers to collaboration. This lack of standardization complicates onboarding, knowledge transfer, and cross-team initiatives.

2.1.1.3. Operational Visibility, Correlation, and Predictive Capability

Alert Fatigue and Event Correlation: The network itself generates large volumes of alerts, logs, and fault data, often with a poor signal-to-noise ratio. This makes it difficult to distinguish actionable events from background noise, correlate incidents and alarms across systems, or understand the true impact on services and users.

Service Dependency Mapping: There is a limited ability to model, visualize, and analyze dependencies between services, policies, and infrastructure. This hampers root cause analysis, proactive maintenance, and the ability to anticipate cascading impacts of changes or failures.

Intent and Context Capture: The rationale ("intent") behind operational actions, configuration changes, or service deployments is only occasionally documented. Without capturing the "why," it becomes difficult to automate processes, audit decisions, or learn from past actions. There is interest in developing tools that can infer intent from existing workflows/configurations or prompt staff to document it at the point of action.

2.1.1.4. Knowledge Transfer and Organizational Memory

Siloed Expertise: ESnet's operational knowledge is largely concentrated in the expertise of its staff members, which can create challenges for knowledge retention and transfer. As staff members leave or new ones join, there is a risk that valuable knowledge and experience may be lost. Additionally, the absence of structured mechanisms for capturing and disseminating this expertise may limit the organization's capacity for growth and innovation.

Ticket and Incident Summarization: Ticket histories may be either extremely terse or long and convoluted, with duplicative updates and unclear resolution paths. This makes it challenging to extract lessons learned, identify recurring issues, or automate triage using prior solutions.

Barriers to Learning and Collaboration: The lack of searchable, well-curated documentation and knowledge repositories hinders both individual productivity and organizational learning.

2.1.1.5. Security, Access Control, and Responsible Data Stewardship

Sensitive Data Management: As data sharing and AI/ML adoption increase, so do concerns about protecting sensitive information (e.g., personally identifiable information (PII), internal operational data, contract and pricing details). There is a need for robust data tagging, access controls, and policy frameworks that balance openness and security.

Federated and Granular Access: Current access models are often binary (internal-only or public-only), lacking the granularity needed for nuanced data sharing with internal and external collaborators. This limits both research productivity and the ability to respond to partner needs.

AI/ML Safety and Governance: The use of AI/ML tools with sensitive or operational data raises questions about explainability, auditability, and organizational standards. There is a need for clear guidelines on model selection, customization, deployment, and ongoing support.

2.1.1.6. Resource and Capacity Planning

Beyond Network Capacity: Effective planning now requires integrating data on compute, storage, power, and human resources—not just network bandwidth. Capacity bottlenecks can emerge in any of these domains, and current planning tools are insufficiently integrated.

Evolving Requirements: User and operational requirements change rapidly; systems/processes often lag behind. There is a need for mechanisms to capture, track, and respond to changing needs across time domains (immediate, short term, and long term).

2.1.1.7. Tooling and Automation Needs

Advanced Search and Query Tools: Participants expressed a strong need for universal, user-friendly search and query tools—ideally leveraging NLP—that can access both structured and unstructured data across all internal repositories.

Automation of Repetitive Tasks: Opportunities exist to reduce administrative burden through robotic process automation, AI-driven configuration management, and automated generation of documentation, RFPs, and other routine outputs.

AI/ML Integration: There is enthusiasm for applying AI/ML to tasks such as configuration validation, ticket summarization, documentation organization, and intent extraction. However, concerns remain about data quality, model transparency, and the risks of automating critical operations.

2.1.2. Representative Example Problem Statements

To further illustrate the session findings, Table 2.1.2.a summarizes representative example problem statements for each of the major challenge themes identified by participants. These statements are synthesized from the breakout group discussions and verbal readouts. They are not direct quotes, but rather are concise expressions of the recurring issues and barriers raised throughout the session. The examples are intended to capture the essence of the problems in a way that is actionable and easily referenced as ESnet moves forward with its data and AI strategy.

Theme	Example Problem Statement
Data Integrity & Accessibility	Critical data exists in multiple systems with no easy way to correlate or verify accuracy.
Workflow Inefficiency	Configuration processes are manual, slow, and error-prone,

	delaying service delivery.
Operational Visibility	We cannot easily correlate alerts, logs, and service dependencies to understand root cause or predict impacts.
Knowledge Transfer	Key operational knowledge is undocumented and siloed within individuals.
Security & Access Control	We lack granular access controls and policy frameworks for sensitive data and AI usage.
Resource Planning	Capacity planning is hampered by lack of integrated data on compute, storage, and human resources.
Tooling & Automation	There is no universal search or natural language interface for our internal data and documentation"

Table 2.1.2.a. Example problem statements.

2.1.3. Additional Insights from Group Readouts

Throughout the readouts from Session 1, several additional nuances and priorities emerged:

Desire for a "Self-Driving" Network: Participants envision a future where ESnet leverages data, AI/ML, and automation not just for engineering tasks, but for end-to-end operational management. Achieving this vision requires robust, accurate sources of truth and the ability to augment human decision-making with Data/AI-driven insights.

Data Lifecycle and Retention: Questions were raised about how long to retain high-resolution operational data, how to compress or summarize historical data without losing critical context, and how to ensure data remains actionable over time.

Risk Management and Data Sharing: As ESnet collaborates more closely with external partners and the broader DOE community, the risks associated with data sharing, federated access, and derived data products grow in complexity and importance.

User Experience and Cultural Change: There is recognition that technological solutions alone are insufficient; cultural factors—such as resistance to change, unconscious competence, and the need for better user experience—must also be addressed to realize the full benefits of data and AI initiatives.

2.1.4. Session 1 Summary

Session 1 of the workshop surfaced and clarified the most significant operational problems facing ESnet as we seek to advance our data and AI strategy. The identified themes—spanning data quality, workflow, operational insight, knowledge transfer, security, resource planning, and tooling—reflect

both longstanding challenges and new complexities introduced by emerging technologies and evolving user needs.

These insights would directly inform subsequent workshop sessions and the development of actionable work-packages. In particular, Session 2 would focus on a deeper exploration of ESnet's existing data collections and management practices, aiming to identify data gaps and requirements necessary to address the challenges identified in Session 1.

2.2. Session 2: Understanding Our Data

ESnet is a production network facility producing an immense quantity of data of varying types, quality, and constraints, which is collected and used by dozens of groups. The focus of Session 2 was to expand participants' knowledge of the current data and how it could be used to address issues and problems identified in Session 1. This session had the following secondary goals:

1. Develop a common understanding of the available types of data, how they can be used, and what security or privacy issues constrain their use.
2. Review the importance of data curation, normalization, and metadata in the act of converting raw data into insights.
3. Review strengths, weaknesses, opportunities, and threats of our current datasets and curation practices.
4. Discuss and identify what changes we should make in our data curation practices.

The facility data is used in technical and business operations, planning, engineering, and community engagement. In all use cases, the diverse data sources must use common metadata to allow us to look at multiple data types together in the same context and control access to ensure appropriate use. For instance, in order to combine contract details with performance metrics, we need identifiers such as a customer identifier that are used in all related documents so that related information can be retrieved easily.

Prior efforts within ESnet have worked to drive consistency and quality of data, such as an internal Measurement Working Group convened in 2021 and an internal Monitoring Working Group convened in 2025. Building upon these efforts, we conducted a participant survey of gaps and challenges within the realm of institutional data .The findings of the pre-workshop survey are as follows:

2.2.1. Workshop Data Survey

Prior to the workshop, a survey of participants was conducted. The survey was designed to solicit ideas and feedback about the current state of data and analysis within ESnet. Four questions were posed, each with free-form responses. The responses revealed competing ambitions that need to be managed:

1. There is always more data that can be collected to improve situational awareness, but often the breadth of available data combined with current discovery and access methods makes it difficult to derive optimal value without a shift in our analytic approach.
2. ESnet staff know the data we collect is valuable to external users, but it's hard to share the data due to external data sharing being a fairly manual process
3. ESnet has a wealth of fantastic data, but the relationship between types of data is not clearly defined and some data needs additional curation to make it usable.

Question 1: What types of data are not currently available for ESnet but would be helpful in our mission and execution if they did?

ESnet maintains an extensive set of network operations data; however, there are a few types of data that were identified as desired but not readily available today. They include:

1. Routing Table / Network Control Plane activity logs, such as details of BGP routing table updates.
2. A well-structured repository of Router, Firewall and Host-Based Firewall Access Control Lists, to assist with debugging.
3. Ubiquitous Availability and mean time between failures (MTBF) metrics for all production services.

Respondents also noted the need to collect data from beyond the ESnet network infrastructure in support of end-to-end science. Specifically:

1. A comprehensive list of the services we offer and the instances of each.
2. End-user application logs to see how the science pipeline is performing.
3. Science project metadata to correlate with network metadata.
4. DTN usage logs.

Question 2: What analysis use cases (questions you would like answers to) do we wish we could answer, where we have the data but not the analytical solution in place?

Respondents' interests were focused on three areas: Business Analytics, Operations, and Capacity Planning. In all of these a common thread was the desire to combine multiple sources of data to perform analysis, and in particular to use the network topology and service dependencies as a common frame of reference to provide Service Aware Analysis.

Business Analytics: The central driver in this area was the desire to maintain better situational awareness of the ongoing business aspects of running the network; in particular, that we have effective ways of understanding resource use, contracts, and hardware lifecycle. This can be difficult in

a WAN, as there are many contracts for the physical circuits on the ground and 356 locations where we have equipment.

Operations: ESnet is a facility that operates a tiered portfolio of services for its users. Users may access physical links, Layer2 virtual circuits, Internet net access, or Layer 3 Virtual Private Networks (L3VPNs). These different services may or may not have common components, making understanding the impact of outages difficult. Responses in this area highlighted the need for all services to have programmatically accessible operational state or health status. Additionally, when combining individual service status with a structured description of service-to-service dependency, the relationships between services could be used to assess cross-service impact of outages, improve user situational awareness through better communications, and help predict future failure scenarios.

Capacity Planning: A primary challenge in any WAN is the cost and time required to acquire new capacity between major metropolitan areas. Respondents noted that while we do routine capacity planning for our optical and routed network layers, it is desired to do the same for our Layer2 VPN services, where users can reserve bandwidth that is not in use. Additionally, respondents felt we could improve upon our current techniques by augmenting capacity planning with more extensive failure scenario planning (based on service-to-service dependency and topology), and to include the integration of compute and storage into the capacity-planning calculus.

Question 3: Are you aware of any use cases where external stakeholders would like access to either raw data or derived analysis that would be useful to our mission and its execution?

ESnet is different from commodity Internet providers in that it partners with stakeholders to collaboratively develop network services along with the science pipelines that depend on those services. This collaboration requires us to support external stakeholders by providing controlled and appropriate access to relevant network data. A key implication for these use cases is that we need to ensure we have the correct metadata included in our network data to enforce appropriate access, and we need well-defined ways to authenticate and authorize users in a federated environment. Respondents identified three classes of external stakeholders.

Sites: Sites use one or more network services provided, and they would like access to all data relevant to site activity, including traffic statistics, power consumption of equipment located at the site, and detailed packet traces for debugging.

Science Projects: Desire similar access as sites, but sometimes with users from different institutions.

Peers: Want access to network data to create an end-to-end federated understanding of cyberinfrastructure.

Question 4: Are there any impediments to effectively using current or desired data that you wish to call out?

ESnet strives to continually improve the services it provides to DOE and associated stakeholders. In this context, respondents noted the following challenges to using data.

Zero Trust: We need to continue to prioritize Zero Trust initiatives that provide federated authentication, fine-grained access control, and ubiquitous microservice API-based access to data.

Data Curation: We should invest more heavily in data curation to ensure all data is machine readable, has adequate quality control, and adequate metadata to support access control and cross-data-source correlation.

Usability: We need to improve user training and user experience generally for available data.

2.2.2. Prework: Data Inventory

Additionally, the ESnet workshop organizers also undertook an effort to crowdsource a comprehensive inventory of all known and desired data sources required for the facility's operation. This initiative yielded a total of 160 identified data sources, which were categorized into 32 distinct groups.

2.2.3. Session 2 Summary

Ultimately, without reliable, complete, and accessible data there can be no useful analysis and no data-driven decision making. Session 2 helped draw out precursor requirements in the data space that would enable or impede advanced analytical techniques such as AI/ML along with the creation of the equivalent of a network digital twin.

2.3. Session 3: Understanding AI

In the past few years, ESnet has been developing a vision of self-driving networks: systems that can autonomously monitor, analyze, and optimize network performance, all while maintaining a high degree of explainability and trust. A central challenge is bridging the gap between academic research—with controlled environments in which many AI models achieve high accuracy—and the messy realities of production networks, where data quality issues, unexpected anomalies, and the need for real-time decision making can undermine even the most promising models.

Session 3 featured a panel discussion with leading AI experts from academia and industry, who shared their perspectives on the applications and limitations of AI. The conversation covered areas where AI shows promise, as well as domains where traditional approaches have proven to be more effective.

The panel included:

- **Arpit Gupta**, Professor, University of California, Santa Barbara, and Berkeley Lab Faculty Scientist working with ESnet, who outlined a technical roadmap for transitioning from rule-based automation to AI-powered network operations.

- **Claudionor Coelho, Jr.**, Chief AI Officer, Zscaler, who focused on the role of large language models and generative AI in AIOps.
- **Sangeetha Abdu Jyothi**, Assistant Professor, University of California, Irvine, who addressed the urgent need for explainability in deep learning–based systems.
- **Taghrid Samak**, Engineering Manager, Meta, who presented on how machine learning is transforming network planning and optimization.
- **Vyas Sekar**, Professor, Carnegie Mellon University, who provided a high-level architectural view of enabling AIOps for next-generation networks.
- **Walter Willinger**, Chief Scientist, NIKSUN, who offered a critical perspective on the current state of AI/ML in networking.

Throughout the session, each panelist provided not only technical insights but also reflective commentary on the broader challenges facing the field, including issues of reproducibility, data quality, human–machine collaboration, and the practical aspects of deploying AI in high-stakes environments.

2.3.1. Panel Presentations Summaries

2.3.1.1 Arpit Gupta (UCSB/ESnet): Making the Self-Driving "Net" Work

Arpit Gupta's presentation, "Making the Self-Driving 'Net' Work: Realizing Production-Ready AIOps for R&E Networks," focused on the technical and operational roadmap for transitioning from rule-based network automation to AI-powered network operations. The presentation covered:

1. **The need for transition:** Gupta contrasted current rule-based network automation with the vision for AI-powered operations, which can autonomously detect anomalies, diagnose issues, and remediate and optimize network performance.
2. **Core technical requirements:** Gupta identified key technical requirements for production-ready AIOps, including:
    a. LLM-powered query interface for natural language queries
    b. Meaningful latent space representations to capture complex network relationships
    c. Hybrid retrieval engine for accurate and actionable insights
3. **Closed-loop ML and exogenous data integration:** Gupta emphasized the importance of closed-loop ML systems, where models are refined based on real-world performance, and integrating exogenous data sources to build comprehensive models.
4. **Roadmap for prototyping and iterative development:** Gupta proposed a roadmap that includes:

a. Identifying key use cases and defining clear KPIs
b. Developing modular prototypes and testing them in controlled environments
c. Scaling and integrating prototypes into live network environments

Gupta's presentation provided a detailed and pragmatic vision for achieving production-ready AIOps, highlighting the technical and operational requirements for transitioning from rule-based to AI-powered network operations.

2.3.1.2 Claudionor Coelho, Jr (Zscaler): LLMs for AIOps

Claudionor Coelho, Jr's presentation, "LLMs for AIOps," explored the role of large language models (LLMs) in AI for IT operations. The presentation covered:

1. **Evolution of Generative AI:** Coelho outlined the evolution of generative AI, from:
    a. Gen AI 1.0: Current LLMs, which excel at tasks like summarization and language generation, but have limited reasoning capabilities.
    b. Gen AI 1.5 and Beyond: Future AI agents that combine LLMs with additional data sources and analytical tools, enabling more advanced capabilities.
2. **Limitations and the Need for Hybrid Systems:** Coelho highlighted the limitations of LLMs, including:
    a. Reasoning limitations and inability to handle numerical calculations.
    b. Risk of "hallucination," where LLMs generate inaccurate or unreliable information.
    c. Need for hybrid systems that combine LLMs with dedicated analytical tools and data processing modules to compensate for their weaknesses.
3. **Security and Trust in LLM-Driven Systems:** Coelho emphasized the importance of security in LLM-driven systems, including:
    a. Preventing vulnerabilities and ensuring that LLMs do not leak sensitive information or produce misleading outputs.
    b. Embedding LLMs in robust frameworks that address both performance and security concerns.

Coelho's presentation provided a balanced view of the potential of LLMs in AIOps, highlighting their strengths and weaknesses, and emphasizing the need for hybrid systems and robust security frameworks to ensure their effective and secure deployment.

2.3.1.3. Sangeetha Abdu Jyothi (UCR): Opening the Black Box—Explainability for Learning-Enabled Systems

Sangeetha Abdu Jyothi's presentation, "Opening the Black Box: Explainability for Learning-Enabled Systems," emphasized the importance of explainability in AI deployment, particularly in network operations. The presentation covered:

1. **The Imperative of Explainability:** Jyothi argued that explainability is a necessity, not a luxury, for AI systems to be trusted and effectively used in production networks. Without transparent models, operators cannot diagnose failures, fine-tune performance, or gain confidence in automated systems.
2. **Existing Techniques for Explainability:** Jyothi reviewed existing techniques, such as LIME, SHAP, and decision tree-based methods, which provide insights into feature importance but have limitations:
    a. Limited scope: focus on input features alone.
    b. Low-level explanations: outputs are often too technical and granular for operators to interpret.
3. **Advancing Explainability:** Jyothi presented advanced methods for explainability:
    a. Future-Based Explainability: techniques like CrystalBox, which forecast future network performance to provide insight into the potential impact of an action.
    b. Concept-Based Explainability: approaches that map complex internal representations to high-level, human-understandable concepts, enabling natural language interaction between operators and the system.

Jyothi's presentation highlighted the need for clear, actionable, and intuitive explanations in AI systems, particularly as they become more integral to network operations. She argued that investing in advanced explainability techniques is essential to ensure that AI can become a trusted partner, rather than an inscrutable black box.

2.3.1.4 Taghrid Samak (Meta): ML for Network Planning and Optimization

Taghrid Samak's presentation discussed the evolution of network planning at Meta, shifting from traditional static forecasting and manual configuration to dynamic, data-driven approaches. The presentation covered:

1. **Challenges in Network Planning:** Modern networks face challenges such as:
    a. Dynamic traffic patterns: unpredictable user behavior and fluctuations in demand.
    b. Multi-tiered planning: planning across multiple time horizons, from short-term anomaly detection to long-term capacity forecasting.
    c. Complex data requirements: integrating various data sources, including historical traffic patterns, real-time telemetry, and synthetic data.
2. **Predictive Modeling and Optimization Techniques:** Samak detailed the predictive models employed at Meta, including:

    a. Mid- to long-term predictive models: forecasting network demand and detecting potential bottlenecks.
    b. Short-term models for anomaly detection: monitoring current network conditions and flagging anomalies.
    c. Optimization and simulation: determining the optimal configuration of network resources using graph theory and optimization algorithms.
3. **Challenges and the MLOps Cycle:** Samak discussed the challenges in ML-driven network planning, including:
    a. Data quality and observability: ensuring high data quality is crucial for reliable predictions.
    b. Alignment across time horizons and network layers: planning models must reconcile short-term and long-term needs.
    c. Risk mitigation: balancing the risks associated with underforecasting and overforecasting.

    To manage these challenges, Meta has implemented a closed-loop MLOps cycle, which includes data quality assurance, model reliability checks, and release management.
4. **Operational Impact and Future Directions:** Samak shared case studies illustrating the benefits of ML-driven planning, including improved network performance and reduced congestion. Future research directions include:
    a. Integration of reinforcement learning: enabling models that adapt in real-time.
    b. Enhanced model explainability: providing clearer insights into the decision-making process.
    c. Bridging the simulation-reality gap: refining synthetic data generation techniques to better mimic live network conditions.

Samak emphasized that ML for network planning requires an integrated approach that combines technical innovation with robust operational processes. She also stressed that the reliability of any predictive model is only as good as the underlying data, and Meta invests heavily in data observability and quality assurance, ensuring that models are trained on accurate, representative datasets.

2.3.1.5. Vyas Sekar (CMU): Enabling AIOps for Next Generation Networks

Vyas Sekar's talk, "Enabling AIOps for Next Generation Networks," provided a high-level architectural perspective on integrating data analytics, modeling, and AI workflows to build resilient and intelligent network operations systems. The presentation covered:

1. **A 20,000-Foot View of AIOps:** Sekar argued that successful AIOps systems must operate across multiple time scales:

    a. Real-time operations: immediate detection, troubleshooting, and control actions.
    b. Longitudinal analysis: trend analysis, capacity planning, and predictive maintenance.
2. **Overcoming Data Bottlenecks with Synthetic Data:** Sekar discussed the use of synthetic data, such as the Rockfish system, to generate high-fidelity, privacy-preserving synthetic session data that mirrors real network conditions. This helps overcome data bottlenecks, reducing development costs, and enabling faster testing and validation.
3. **The Importance of Stateful Analytics:** Sekar emphasized the need for stateful analytics methods that capture the sequence, timing, and context of events. This is essential for:
    a. Root cause analysis: pinpointing the sequence of events leading to an anomaly.
    b. Predictive maintenance: identifying long-term trends that may indicate future network degradation.
    c. Dynamic response: enabling more nuanced and effective control actions.
4. **Data Processing and AI Workflows:** Sekar highlighted the need for sophisticated AI workflows that integrate diverse data sources, providing a comprehensive view of network health. By developing abstractions that bridge the gap between low-level data and high-level insights, network operators can better diagnose issues and optimize performance.

Sekar's presentation provided a strategic view of AIOps, emphasizing the need for a synthesis of real-time and longitudinal perspectives, stateful analytics, and sophisticated AI workflows to build resilient and intelligent network operations systems.

2.3.1.6. Walter Willinger (NIKSUN): AI/ML for Networking

Walter Willinger's presentation, "AI/ML for Networking," critically examined the evolution of AI/ML in networking, providing a historical overview and a forward-looking critique. The key points were:

1. **History of AI/ML in Networking:** AI/ML has been applied to networking challenges for over a decade, but despite thousands of publications, real-world impact remains limited.
2. **The Standard ML Pipeline and Its Limitations**: the conventional ML pipeline focuses on demonstrating "effect" (high accuracy metrics) but neglects the "cause" behind model performance, leading to limitations in real-world applications.
3. **Limitations of Existing Approaches:** an informal survey of 100 published papers revealed that only about 10 produced reproducible results, while the rest suffered from issues such as underspecification and reliance on spurious correlations.
4. **The Need for Next-Generation ML Pipelines**: Willinger called for a paradigm shift towards developing next-generation ML pipelines that emphasize a deep understanding of "cause" and incorporate real-world feedback to continuously refine training data and improve model robustness.

Willinger's message was clear: to truly harness AI/ML for networking, researchers must be humble about the limitations of existing approaches and commit to developing models that can explain themselves, which is the only path to achieving the vision of self-driving networks.

2.3.2. Analysis of the Q&A Session

Following the presentations, the panel opened the floor for a 90-minute Q&A session. This session provided an opportunity for both the panelists and the audience to delve deeper into technical nuances, share experiences from the field, and discuss the future trajectory of AI in networking. The following is a summary of the panel Q&A session.

2.3.2.1. Defining AIOps: Observability and Agentic AI

A recurring theme during the Q&A was the definition of AIOps. Several panelists—Willinger, Sekar, and Samak—spoke about the need to distinguish between two key components:

- **Responsible AI for Observability:** Systems that are designed to detect anomalies, monitor network performance, and provide insights into what is happening in real time.
- **Agentic AI for Mitigation:** Systems that not only detect problems but also autonomously decide on and implement corrective actions.

A common consensus emerged that both aspects are essential. One audience question focused on what it means to truly “trust” an AI system. The panelists agreed that trust comes from transparency: systems must provide clear explanations for their decisions, and their outputs must be verifiable against known benchmarks. This directly ties back to the earlier presentations on explainability and the need for closed-loop ML systems.

2.3.2.2. Balancing Automation with Human Oversight

Another major topic was the extent to which AI should be allowed to operate autonomously versus the need for human oversight. Several panelists, including Coelho and Gupta, emphasized that while AI can greatly enhance operational efficiency, there are many scenarios where human judgment remains indispensable. The discussion centered on:

- **Determining Boundaries for Automation:** Panelists discussed criteria for when AI can be trusted to take full control—for example, in tasks with clear, objectively verifiable outcomes. In contrast, ambiguous or high-risk decisions should still involve human operators.
- **Iterative Automation:** A gradual approach was advocated, where AI tools are first introduced as assistants (or co-pilots) before transitioning to more autonomous roles. This allows

operators to build trust in the system incrementally while retaining control over critical decisions.

The consensus was that a hybrid approach—where human expertise is complemented by AI's computational power—offers the best of both worlds. This perspective was supported by examples from the presentations, such as Gupta's roadmap for closed-loop ML systems and Jyothi's emphasis on explainability.

2.3.2.3. Data Management and Quality Assurance

Data management was another hot topic during the Q&A. Many questions were directed at understanding how organizations can ensure the quality and relevance of the data used to train AI models:

- **Challenges of Data Quality:** Samak and Sekar both underscored that data quality is the "crown jewel" of successful AI deployments. Inconsistent, noisy, or incomplete data can lead to unreliable models, which in turn can undermine operational confidence.
- **Observability and Metrics:** The discussion also covered the need for robust observability systems to continuously monitor data quality and model performance. Metrics that go beyond conventional accuracy scores—such as the reduction in downtime, improvements in service quality, or cost savings—were recommended as better indicators of real-world impact.
- **Integration of Diverse Data Sources:** Panelists highlighted the importance of integrating data from various sources—telemetry, logs, synthetic data, and even external datasets—to build a more comprehensive understanding of network performance. This integration is essential for building models that can generalize well to different operational scenarios.

2.3.2.4. Evaluation and Success Metrics

Evaluating the success of AI systems in networking was a recurring subject. The panelists agreed that traditional evaluation metrics (e.g., accuracy, F1-score) are insufficient for high-stakes network operations. Instead, success should be measured in terms of:

- **Real-World Impact:** Metrics such as reduced SLA violation minutes, lower downtime, improved user experience, and operator efficiency gains were emphasized as more meaningful indicators.
- **Continuous Improvement:** The need for iterative feedback loops was discussed extensively. By continuously monitoring system performance and incorporating operator feedback, AI models can be refined over time to better meet operational needs.

- **Transparency in Evaluation:** There was also a call for more transparent evaluation methodologies. Operators need to understand not only that an AI system is performing well but also why it is performing well. This transparency is critical for building trust and ensuring that the models can be relied upon in critical situations.

2.3.2.5. Use Cases and Collaborative Opportunities

Several audience members inquired about specific use cases and opportunities for collaboration, particularly with organizations not already connected with ESnet:

- **Automated Incident Response:** One popular topic was the potential for AI-driven incident response. Panelists discussed how closed-loop ML systems could eventually enable fully automated responses to common network disruptions, while still allowing human intervention in more complex scenarios.

- **Root Cause Analysis and Capacity Planning:** The value of AI in identifying the underlying causes of network failures and in planning for future capacity was also highlighted. Both Gupta and Samak provided insights into how predictive models can be used to forecast demand and adjust network configurations proactively.

- **Security Applications:** Security was another area of keen interest. While some panelists noted that AI systems have the potential to enhance cybersecurity (by detecting anomalies or automating threat responses), they also cautioned that integrating AI into security frameworks introduces its own set of vulnerabilities that must be addressed.

- **Cross-Disciplinary Collaboration:** Finally, there was broad agreement that solving these challenges requires collaboration across disciplines—combining the expertise of network engineers, data scientists, AI researchers, and cybersecurity professionals. This collaborative approach was seen as essential for developing robust, production-ready solutions.

2.3.2.6. Balancing Innovation with Operational Realities

The Q&A session also featured a spirited debate about the balance between cutting-edge innovation and the practical realities of operating large-scale networks:

- **Realism vs. Ambition:** Some panelists cautioned against over-engineering AI solutions without a clear understanding of the operational environment. For instance, while the promise of self-driving networks is exciting, many agreed that the journey toward fully autonomous systems will be incremental and fraught with challenges.

- **Prioritizing Use Cases:** There was a consensus that organizations should start by targeting the most pressing and well-defined problems—those where AI can offer a clear advantage over

existing methods. Once these foundational use cases are established, the scope can gradually be expanded to more complex challenges.

- **Long-Term Vision:** Despite the immediate challenges, panelists were optimistic about the future. They envisioned a gradual evolution where AI systems become increasingly sophisticated and reliable, eventually enabling a level of autonomy that could transform network operations.

In summary, the Q&A session served as a microcosm of the broader debates in the field—highlighting both the immense potential of AI in networking and the numerous practical challenges that must be overcome.

2.3.3. Synthesis: Key Themes and Future Directions

Drawing together the insights from the presentations and the Q&A session, several overarching themes emerge that encapsulate the current state and future potential of AI for networking.

2.3.3.1. Moving Beyond the Standard ML Pipeline

A recurring critique across multiple presentations—especially in Willinger's talk—is that the conventional ML pipeline is inadequate for addressing the complexities of real-world network operations. The standard approach, focused on achieving high accuracy on IID data, fails to account for the nuanced "cause" behind model predictions. Moving forward, the development of closed-loop ML pipelines is essential. These systems must:

- Continuously refine training data based on real-world feedback.
- Offer transparent, explainable outputs that build operator trust.
- Provide robustness across varying operational conditions.

By addressing the "garbage in, garbage out" problem, next-generation ML pipelines have the potential to transform AI/ML research in networking, making it more reproducible and practically relevant.

2.3.3.2. The Imperative of Explainability

As emphasized by Jyothi, explainability is critical for the adoption of AI in mission-critical network environments. Operators must be able to understand and validate the decisions made by AI systems. Advanced techniques—such as future-based explainability (exemplified by CrystalBox) and concept-based approaches—represent promising directions for making AI models more interpretable. In a high-stakes context where even minor errors can have significant consequences, explainability is not merely a technical nicety but a fundamental requirement for operational safety and trust.

2.3.3.3. Integrating Diverse Data Streams and Embracing Statefulness

Sekar's emphasis on synthetic data generation and stateful analytics points to a broader trend in modern network management: the need to integrate and analyze diverse data streams. Successful AI systems must combine real-time telemetry, historical logs, and even synthetic data to create a holistic picture of network performance. Capturing the temporal dynamics of events—beyond mere "count events"—is essential for both anomaly detection and root cause analysis. In this regard, innovations like Rockfish and Time-State Analytics pave the way for more nuanced and effective network monitoring systems.

2.3.3.4. Hybrid Systems and the Role of LLMs

Coelho's discussion of LLMs in the context of AIOps highlights an important point: while large language models are powerful, they are not a standalone solution. The future lies in hybrid systems where LLMs serve as the interface for natural language processing while specialized agents handle data extraction, analysis, and execution. This multi-agent approach—envisioned as Generative AI 1.5 and Generative AI 2.0—has the potential to overcome the inherent limitations of LLMs (such as their reasoning gaps and tendency to hallucinate) by embedding them in larger, purpose-built systems.

2.3.3.5. Roadmap for a Self-Driving Network

Gupta's detailed roadmap for transforming network operations offers a concrete path forward. The key elements include:

- **Identifying high-impact use cases:** Prioritizing problems where AI can deliver immediate value (e.g., automated incident response, capacity planning, and root cause analysis).
- **Defining robust KPIs:** Moving beyond conventional accuracy metrics to include real-world performance indicators such as reliability, efficiency, and cost savings.
- **Developing modular prototypes:** Starting with small-scale implementations that can be iteratively refined based on operator feedback.
- **Integrating diverse data sources:** Leveraging both endogenous and exogenous data to build comprehensive, high-fidelity models.
- **Building closed-loop systems:** Ensuring that models are continuously updated and improved in response to operational feedback.

This roadmap underscores that the journey to self-driving networks is incremental—requiring sustained innovation, rigorous evaluation, and close collaboration between researchers and operators.

2.3.3.6. Network Planning: A Case Study in ML-Driven Optimization

Samak provided a detailed look at how ML is being applied to the complex task of network planning and optimization at Meta. Her insights reveal that:

- **Predictive modeling and optimization are key:** By combining long-term forecasting with real-time anomaly detection, it is possible to optimize resource allocation and prevent congestion.
- **Data quality is paramount:** High-quality, observable data is essential for building reliable models. Continuous MLOps cycles that ensure data quality and model performance are critical.
- **Organizational integration is as important as technical innovation:** Successful AI-driven network planning requires close collaboration between data scientists, domain experts, and operations teams. This collaborative approach ensures that technical innovations translate into tangible operational benefits.

2.3.3.7. Human-AI Collaboration and the Future of Network Operations

Throughout the session, a clear theme emerged: despite the promise of advanced AI, human expertise remains indispensable. Whether it is through:

- **Hybrid systems that blend automated decision-making with human oversight,**
- **Explainability techniques that empower operators to understand and trust AI outputs,** or
- **Iterative feedback loops that enable continuous refinement of models…**

… the future of network operations will be defined by effective human-AI collaboration. As AI tools become more sophisticated, they will serve as powerful assistants rather than complete replacements for human judgment.

2.3.4. Session 3 Summary

The discussions in Session 3 served as both a reflection on the current state of AI in networking and a roadmap for future innovation. As the industry continues to evolve, several areas of focus will be critical:

- **Enhanced Data Quality and Observability:** Continued investment in data management systems will be essential to ensure that models are trained on accurate and representative

datasets. Innovations in synthetic data generation and advanced stateful analytics will play a key role.

- **Integration of Closed-Loop Learning:** The next wave of innovation may well involve the integration of closed-loop learning techniques, which can enable systems to learn from their operational outcomes and adapt in real time.
- **Security and Robustness:** As AI systems become more integrated into network operations, ensuring their security will be paramount. This includes not only protecting AI models from external threats but also ensuring that they do not inadvertently introduce vulnerabilities into the network.
- **Operational Scalability:** The challenges of scaling AI solutions across heterogeneous and geographically distributed networks will require innovative approaches to model deployment, monitoring, and continuous improvement.
- **Fostering Cross-Disciplinary Collaboration:** Finally, the future of AI in networking will depend on breaking down traditional silos between academia, industry, and government. Collaborative efforts that bring together diverse expertise will be essential for addressing the multifaceted challenges of network management.

While many of the current approaches remain in the early stages of development, the collective vision is clear: By embracing next-generation ML pipelines, enhancing explainability, integrating diverse data streams, and fostering robust human-AI collaboration, the goal of self-driving networks is within reach.

As network environments become ever more complex and critical to our digital infrastructure, the lessons from this session can serve as a roadmap for the industry. By aligning technical innovation with practical operational needs, the vision of fully autonomous, resilient, and intelligent network management systems can be realized. The journey will undoubtedly be challenging, but the potential rewards—in terms of efficiency, security, and user experience—are immense.

2.4. Session 4: Bringing It Together—Building Work-Packages

Session 4 of the ESnet Data and AI Workshop focused on transforming conceptual problems or opportunities into actionable work-packages. Attendees were first guided through a structured methodology to: (1) articulate the opportunity statements, (2) determine the required datasets, (3) identify the requirements and constraints, (4) explore opportunities and potential solutions, and (5) perform a gap analysis. The goal was to produce clearly scoped, implementation-ready documents to guide future development and collaborations within ESnet.

2.4.1. Opportunity Statement

Defining a clear problem or opportunity statement is a critical first step when considering the use of AI to address organizational challenges. It helps focus efforts on solving specific, meaningful issues rather than pursuing AI for its own sake. A well-articulated statement narrows the scope, making it easier to determine whether AI is the appropriate solution and what form that solution should take.

This clarity also aligns stakeholders across technical, business, and leadership teams. Everyone can work from a shared understanding of the goal, making it easier to prioritize resources, gain support, and define success. It also allows for a realistic assessment of feasibility by surfacing the data, skills, and infrastructure needed to implement a solution.

Moreover, a good problem statement supports better planning and evaluation. It enables the team to estimate the potential return on investment and identify key performance indicators. By starting with the problem—not the technology—you ensure that AI is applied where it can have the greatest impact, in a way that is both strategic and measurable.

For this session, the organizers asked participants to clearly define the core issue or opportunity the group is addressing. This should be articulated from the user's perspective and reflect an identifiable need or desired improvement. The workshop organizers asked for the opportunity statements, if possible, to be written in user story format:

> "As a **[user role]**, I want to **[achieve some goal]**, so that **[reason/benefit]**."

The resulting set of user stories could then be used to drive the data collection process for the remainder of the session. Here is an example of an AI-oriented user story to demonstrate the outcome of the statement definition:

> "As a **scientific computing team** at a national lab, we want to **enhance our software development velocity and security posture** by leveraging **locally hosted generative AI tooling**, so that we can **accelerate science** while **ensuring strict data governance compliance**."

2.4.2. Identification of Datasets

As part of the opportunity definition, the participants were asked to identify the datasets needed to solve the problem or realize the opportunity. This question can be a crucial part of defining an AI initiative because data is the foundation of any AI solution. Without the right data (clean, relevant, and sufficiently comprehensive), AI models cannot be trained, validated, or deployed effectively. By requiring participants to think about the data early, we ensured that proposed solutions were not only conceptually sound but also practically feasible.

This step also helps reveal important realities, such as whether the necessary data exists, where it resides, and whether it is accessible or needs to be collected. It encourages critical thinking about data quality, coverage, biases, and privacy concerns, all of which can significantly influence the outcome of an AI project.

Additionally, identifying data requirements early facilitates collaboration between business teams and data owners, aligns expectations, and reduces the risk of stalled projects due to missing or inadequate data. It moves the conversation from “can we build this?” to “do we have what we need to build this well?”, which is key to selecting high-impact, achievable opportunities.

Continuing with our “enhance software development velocity” story, below is an example that demonstrates the outcome of the datasets definition:

- Define data classification for code repositories (e.g., public, internal, sensitive).
- Normalize metadata across codebases to support search and tagging.
- Implement data curation process for model training (scrubbing sensitive tokens, proprietary IP).
- Establish audit logs for all AI-generated suggestions.
- Ensure all datasets used for fine-tuning are documented with provenance and versioning.

2.4.3. Known Constraints and Requirements

Enhancing user stories and dataset information with known constraints and identifiable requirements is important because it grounds the AI opportunity in real-world operational and technical context. Constraints, such as legal restrictions, data privacy requirements, system limitations, or resource availability, help shape what is realistically achievable and prevent costly missteps later in the development process.

By identifying requirements, such as accuracy thresholds, response times, integration needs, or user expectations, you ensure that the solution will be fit for purpose and aligned with stakeholder goals. These requirements guide model selection, data preparation, and evaluation criteria, helping teams focus on delivering value within defined boundaries.

Incorporating constraints and requirements early also improves cross-functional alignment, enabling developers, domain experts, and decision-makers to make informed trade-offs. It leads to better-scoped projects, reduces the risk of rework, and increases the likelihood that the final solution will be usable, compliant, and impactful.

As output for this session, we asked participants to capture any known constraints and requirements they could identify that would need to be considered when implementing a solution to their defined user stories. These results should include any of the following:

- Legal, technical, security, or operational constraints.
- Policies or sensitive data issues.
- Technical/hardware limitations.
- Required tool support (e.g., languages, environments, toolchain compatibility).

As an example, the “enhance software development velocity” user story could have the following session output for this section:

Legal/Policy Constraints:

- All code and data must remain confidential within the ESnet security boundary.
- Compliance with DOE cybersecurity and data handling policies is mandatory.

Technical Constraints:

- Tools must integrate with existing Python, Goland, and Java development environments.
- Support for secure CI/CD workflows is required.
- AI models must be hosted on-premises; no cloud-hosted inference allowed.

Functional Requirements:

- Solutions must support code generation, refactoring, bug detection, and test case creation.
- Preference for tools that support plugin-based integration with JetBrains or Visual Studio.

2.4.4. Opportunities and Potential Solutions

In this phase, the goal was to explore and evaluate a range of ways to address the defined problem or opportunity. This exploration was informed by the previously identified user stories, datasets, constraints, and requirements, and provided an opportunity to creatively and pragmatically assess how AI, and possibly other technologies, could be applied.

Potential solutions might include leveraging existing tools or platforms already in use within the organization, reducing development time and integration complexity. Alternatively, they might involve new AI model development, particularly if the problem is novel or highly specific. Solutions may also take the form of hybrid approaches, where AI components are integrated into existing systems or workflows to enhance functionality or automate tasks.

As part of identifying potential solutions, it’s important to examine key considerations that influence the success and sustainability of any proposed approach. These include how users will interact with the solution, what data is required to support it, and whether the solution is feasible given current constraints.

Critically, this phase should also consider non-AI alternatives, such as rule-based systems, improved user interfaces, or process redesign, to ensure that AI is used only where it adds real value. Evaluating solution paths with respect to feasibility, scalability, maintainability, and alignment with organizational goals helps ensure that the chosen approach is both effective and sustainable.

The expected session output for this section is a list of proposed solutions or opportunity areas for investigation. The “enhance software development velocity” user story could, as an example, have the following session output for this section:

Solution 1: Deploy a locally hosted LLM-based coding assistant (e.g., Mistral or Code Llama) accessible via a web interface and IDE plugin.

Solution 2: Evaluate the JetBrains AI Service for ESnet internal deployment with containerized model hosting.

Solution 3: Investigate using Visual Studio Intellicode with ESnet-specific training data for domain-adapted assistance.

Solution 4: Explore fine-tuning an open-source model using anonymized internal code repositories.

2.4.5. Gap Analysis

Gap analysis is the process of identifying what is missing or insufficient in the current state relative to the desired future state. This step is essential for understanding the obstacles that must be addressed to move from concept to implementation. It provides a clear picture of the readiness level of the organization and helps prioritize actions needed to close the gaps.

The following questions were addressed in this session:

- **Are all required datasets available and complete?**

  Evaluating data availability, completeness, and quality helps determine whether the inputs needed to support AI development are accessible or require additional collection, cleansing, or enrichment.

- **Are metadata and governance frameworks adequate?**

  This assesses whether data is well-documented, traceable, and compliant with internal and external policies. Strong metadata and governance are essential for responsible AI development, reproducibility, and auditability.

- **What processes, technologies, or information are missing?**

  Identifying missing components, such as integration mechanisms, model deployment pipelines, domain knowledge, or end-user feedback loops, ensures that the solution can be operationalized and maintained effectively within the existing environment.

By conducting a thorough gap analysis, teams can develop realistic implementation plans, mitigate risks early, and ensure that AI solutions are both technically viable and aligned with organizational readiness.

As an example gap analysis for the “enhance software development velocity” user story we defined the following session output for this section:

Current State:

- No approved and secure AI code assistance tools are deployed at ESnet today.

- Development teams use traditional IDEs with manual testing and refactoring.

Gaps Identified:

- Lack of secure, vetted LLM infrastructure within ESnet's environment.
- No internal policy for AI tool use or output auditing.
- Incomplete understanding of developer pain points and workflow bottlenecks.
- Need for budget allocation to pilot tool deployments.

Data Gaps:

- Internal codebases are not fully labeled for model training.
- Metadata on existing repositories is inconsistent or missing.

2.4.7. Work-Package Summary

Twenty-nine initial work-packages were generated by participants during the workshop. After a post-examination, WP18 (Automating Configuration) was assimilated into WP17 (Automating Site Deployment) resulting in 28 final work-packages. Each of these final 28 provides a concise overview of the proposed initiatives that emerged from the opportunity identification and analysis process. Each represents a well-defined effort aimed at addressing a specific problem or opportunity using AI, data analytics, or related technologies. These summaries are intended to capture the scope, objectives, and strategic relevance of each initiative, along with key considerations such as required datasets, expected outcomes, and implementation constraints.

Table 2.4.7.a shows a functional clustering of the work-packages, with many of them overlapping in multiple categories. For example, WP11 (Predict Hardware Failures) falls both in "Network Operations and Automation" and "AI and Advanced Analytics", and WP22 (Automated ServiceNow Ticket Summarization) is in both "AI and Advanced Analytics" and "Knowledge Management and User Experience." This reflects the interconnected nature of ESnet's challenges.

Category	Description	Work-Packages
Data Management and Infrastructure	Focuses on establishing a robust data foundation, including data quality control, lifecycle management, cataloging, consistent data management practices, unified querying capabilities, information architecture, and unified dataset access.	WP03 (Data Quality) WP04 (Lifecycle) WP05 (Data Catalog) WP15 (Consistent Data Management) WP16 (Query All Data) WP26 (Information Architecture)

	These WPs ensure data is discoverable, reliable, and readily available.	WP29 (Dataset Unified Query)
Network Operations and Automation	Addresses the automation and optimization of network operations, including alerting, rules correlation, managing network services, parsing outage notifications, automating ticket resolution, correlating alarms, predicting hardware failures, detecting configuration anomalies, and automating site deployment. These WPs aim to make network operations more proactive and efficient.	WP01 (Alerting) WP02 (Rules Correlation) WP06 (Network Services) WP08 (Outage Notification Parsing) WP09 (Ticket Resolution) WP10 (Correlate Alarms) WP11 (Predict Hardware Failures) WP12 (Detect External Configuration Anomalies) WP17 (Automating Site Deployment)
Business and Administrative Support	Covers work-packages related to business operations and administrative tasks, such as supporting business operations, enabling fast contract lookup, building RFPs and contracts, managing requirements, and supporting mission management. These WPs focus on streamlining administrative processes.	WP07 (Business Ops) WP14 (Fast Contract Lookup) WP20 (RFP/Contract Builder) WP27 (Requirements Management) WP28 (Mission Support Management)
AI and Advanced Analytics	Centers on leveraging AI and advanced analytics techniques, including predicting hardware failures, creating an AI sandbox for experimentation, automating ticket summarization using NLP, and building NLP interfaces to systems. These WPs explore and apply cutting-edge technologies.	WP11 (Predict Hardware Failures) WP19 (AI Sandbox) WP22 (Automated ServiceNow Ticket Summarization) WP25 (NLP Interfaces to Systems)
Knowledge Management and User Experience	Focuses on managing knowledge, improving user experience, including building a unified document search, using AI to summarize tickets, creating NLP interfaces for easier system interaction, and designing the information architecture for better knowledge accessibility and usability.	WP21 (Unified Document Search) WP22 (Automated ServiceNow Ticket Summarization) WP25 (NLP Interfaces to Systems) WP26 (Information Architecture)

Technical Debt and Legacy Systems	Targets technical debt and managing legacy systems, specifically capturing the intent behind configurations and addressing challenges related to legacy code.	WP13 (Capture Configuration Intent) WP24 (Legacy Code)
Security and Access Control	Addresses security-related work packages, including detecting external configuration anomalies and implementing federated authentication.	WP12 (Detect External Configuration Anomalies) WP23 (Federated Authentication)

Table 2.4.7.a. Work-package functional groupings.

The following sections contain a summary of each work-package. The complete, unabridged work-packages are included in Appendix B.

2.4.7.1. WP01 (Alerting)

This work-package outlines issues with current monitoring and alerting practices. It highlights a lack of complete monitoring, inconsistent methodologies, and difficulties in assessing service reliability and root causes.

WP01 (Alerting)	
Opportunity	Not all services have complete monitoring or a consistent alerting and availability tracking methodology.
Datasets	The document identifies six key datasets useful for input: network connectivity and performance data, host- and system-level data, application-specific performance metrics, service instance specific dependency graphs, configuration changes, and a data source defining hosts, nodes, and applications.
Constraints	Include data sensitivity, high uptime requirements, and privacy law restrictions.
Potential solutions	Developing a shared monitoring pattern, generating configurations declaratively, using analysis middleware to catch anomalies, building dependencies within a

	middleware framework, and creating an intelligent system for incident prioritization.
Gap analysis	Reveals deficiencies such as an incomplete set of services and service instances, lack of a comprehensive approach to measurement, missing or incomplete datasets, lack of policy-based monitoring definitions, no system to convert monitoring data into actionable alerts, and absence of a well-structured representation of topology and interdependency. Additionally, a dataset of all configuration changes is missing, as are effective techniques for detecting multi-variate anomalies.

Table 2.4.7.1.a. WP01 (Alerting) summary.

Development of solutions identified in this work-package would improve service reliability, efficiency, and enable data-driven decision-making, while addressing gaps in data availability, policy definitions, and automatic anomaly detection.

2.4.7.2. WP02 (Rules Correlation)

This work-package outlines the challenges and potential solutions for correlating network data to improve connectivity troubleshooting at ESnet.

WP02 (Rules Correlation)	
Opportunity	Currently, correlating data from various sources (firewall rules, route tables, traffic logs, etc.) is difficult due to fragmented data, lack of a unified view, and insufficient insights. This leads to wasted engineer time on manual correlation, errors, and prolonged downtime.
Datasets	The required datasets can be broadly categorized into four high-level classes. Network configuration data includes firewall rules, routing tables, and host-specific settings that define how traffic is managed and directed. Network observability data covers traffic logs, packet captures, and blackhole routes, providing visibility into real-time and historical network activity. Topology and infrastructure data represents the structural layout of the network, including device interconnections

	and subnet definitions. Finally, historical and incident data is essential for training AI models and understanding past network behaviors and anomalies.
Constraints	The solution must consider data sensitivity, maintain high uptime, and comply with privacy regulations like GDPR and CCPA
Potential solutions	Three potential solutions are proposed: a Correlation Engine (graph-based modeling), an Automated Rule Validator (predictive modeling), and NLP Reachability Analysis (natural language interface). These solutions aim to provide a unified platform for real-time visualization and analysis, reducing troubleshooting time.
Gap analysis	Current challenges include a lack of an integrated data repository, inconsistent metadata, limited analytics tools, and incomplete data availability. Addressing these gaps requires collecting and normalizing data, developing systems for actionable alerts, and implementing anomaly detection techniques.

Table 2.4.7.2.a. WP02 (Rules Correlation) summary.

The development of a network configuration correlation tool would be complex but could significantly improve network reliability, reduce downtime, and free up engineer time for more important tasks and innovation.

2.4.7.3. WP03 (Data Quality)

The work-package discusses data quality issues hindering security operations, specifically the lack of structured, normalized, and consistently available data across critical infrastructure. This data fragmentation makes incident response slow and inaccurate.

WP03 (Data Quality)	
Opportunity	Key problems include inconsistencies, outdated records, and missing entries in data sources like DNS, LDAP, SN, and ESDB. This impedes automated detection, slows investigations, and creates security blind spots.

Datasets	The document identifies essential datasets needed for improved incident response: network telemetry, system logs, host/identity data, and virtual infrastructure data.
Constraints	Known constraints include the resources needed to solve the problem, compliance with government mandates (M-21-31 and M-22-09), legal obligations like GDPR and CCPA, poor metadata tracking, and operational risks of integrating with legacy systems.
Potential solutions	Potential solutions focus on auditing, normalizing, and centralizing existing security data sources in Splunk. Specific solutions include a Log Coverage Auditor, Normalization Pipeline, Centralized Data Access Strategy, and Flexible Interaction Modes.
Gap analysis	Reveals that while most data exists in Splunk, critical gaps remain, such as incomplete MAC/ARP/ND/IP/VLAN/Port data and poor PCAP integration. Efforts are needed to audit missing logs, validate data completeness, and implement normalization to improve data quality, incident response, and enable AI/ML in security analytics.

Table 2.4.7.3.a. WP03 (Data Quality) summary.

Implementing the solutions in this document, primarily focused on data quality improvements for security operations, would lead to more effective and efficient incident response. By auditing, normalizing, and centralizing existing security data sources, particularly those within Splunk, inconsistencies and gaps would be addressed. This would result in better data fidelity, reduced time-to-insight during investigations, and improved accuracy in answering fundamental operational questions. Furthermore, these improvements would create a solid foundation for leveraging AI/ML in security analytics for enhanced correlation and anomaly detection, as well as ensuring compliance with government and legal mandates regarding data handling and visibility.

2.4.7.4. WP04 (Lifecycle)

This work-package outlines a solution for managing the lifecycle of business documents and contracts.

WP04 (Lifecycle)	
Opportunity	Organizations struggle to manage and track the lifecycle of business documents and contracts. The central problem is that current document management systems lack sufficient metadata and centralized tracking, leading to outdated decision-making, compliance risks, and operational inefficiencies. The desired outcome is to establish a centralized mechanism that tracks lifecycle metrics, flags outdated documents, and enhances document relevance scoring.
Datasets	Relevant data sources include document metadata such as author, version, and last updated information, user interaction logs, legal/compliance policy references, business repositories like Google Docs and Confluence, and potentially expanding to Jira tickets, Git logs, email, and Slack content. The data also encompasses unstructured formats such as PDFs and Excel files.
Constraints	Legal and privacy compliance with regulations like GDPR/CCPA due to sensitive data, technical limitations such as inadequate metadata tracking in existing DMS platforms, organizational barriers like resistance to workflow changes and inconsistent document cleanup, and security considerations regarding access permissions for AI/ML models.
Potential solutions	Metadata tagging system to standardize information, an automated reminder engine to notify stakeholders about stale documents, AI relevance scoring to predict staleness, statistical summary techniques, document clustering using LLMs, a data governance framework, and a central repository with federated search capabilities across various platforms.
Gap analysis	Current state is incomplete metadata, lack of centralized tracking, missing links between documents and context, limited accessibility, and unstructured content. The desired state is standardized metadata, integrated lifecycle monitoring, contextual awareness, federated search, and enriched content. Gaps identified include missing metadata standards, lack of accountability, and inconsistent data cleanup.

Table 2.4.7.4.a. WP04 (Lifecycle) summary

Implementing the solutions described in this document would result in a robust, centralized document lifecycle management system. This system would improve document tracking, enhance compliance, ensure better access and organization of information, and leverage AI to maintain document relevance, ultimately leading to more efficient and informed decision-making across the organization.

2.4.7.5. WP05 (Data Catalog)

This work-package defines the need for an up-to-date inventory of ESnet's datasets. It highlights why staff to know what data is available, its type, schema, access methods, and responsible parties.

WP05 (Data Catalog)	
Opportunity	There is a need for an up-to-date data catalog within ESnet. Staff members currently lack a comprehensive view of available datasets, their types, schemas, update times, access methods, and responsible parties. This lack of visibility hinders efficiency and the potential use of existing data for machine learning and other tools. The primary goal is to enable staff to quickly discover and access relevant data, thus avoiding duplication of effort and maximizing data utilization.
Datasets	To address this issue, a variety of datasets are required. These include data from structured databases, documentation, ServiceNow, Google Docs, and other business systems. The document also acknowledges an existing data sources spreadsheet but notes its limitations in terms of accuracy and completeness. The desired datasets would need to provide information on data types, schemas, last updated timestamps, responsible parties, and access methods such as APIs.
Constraints	Some datasets will contain sensitive information, and the visibility or even existence of these datasets may need to be controlled. This means that access policies and restrictions must be implemented. Additionally, the mere existence of some datasets might be considered sensitive information, adding another layer of complexity to the cataloging process. These sensitivity issues necessitate careful consideration of access controls and data security.
Potential solutions	Leveraging commercially available machine learning tools to crawl various data storage locations and produce an inventory of datasets. These tools would periodically and on-demand create the inventory. The process would require a starting point, similar to the data sources spreadsheet already collected. This approach aims to automate data discovery and cataloging, ensuring that the information is current and accessible.
Gap analysis	The existing data sources spreadsheet has limited accuracy and lacks critical information about data access and usage. There is a need for more metadata about datasets, including access policies, schemas, and APIs. While metadata for datasets in database systems might be readily available, metadata for informal datasets like wikis and spreadsheets will need to be created.

Table 2.4.7.5.a. WP05 (Data Catalog) summary.

Implementing the suggested solutions would result in an up-to-date and easily accessible data catalog for ESnet staff. This catalog would provide a comprehensive overview of available datasets, their metadata, and access methods. By using machine learning tools to automate the cataloging process, ESnet could ensure the information is current and accurate. This would improve efficiency, reduce duplication of effort, and enable greater use of existing data for machine learning and other purposes, ultimately optimizing data management and utilization within the organization.

2.4.7.6. WP06 (Network Services)

This work-package outlines the needs and potential solutions for managing and understanding ESnet's Guaranteed Bandwidth Service.

WP06 (Network Services)	
Opportunity	There is a lack of understanding of ESnet's Guaranteed Bandwidth Service, including the management of these services. The problem statement highlights the need for network engineers, engagement staff, and senior leadership to access information about service commitments, utilization, and potential violations. This includes the number of customers, bandwidth amounts, usage percentages, and any instances where ESnet failed to meet bandwidth commitments.
Datasets	Several types of data were identified as required to address this problem. These include a list of business agreements with guaranteed bandwidth commitments (SLAs/SLEs), data from OSCARS (Guok et al., 2006) to monitor usage and identify entities, circuit counts, and allocation percentages, and telemetry data, such as queue drops, to detect service violations. The use of Stardust to correlate OSCARS data with other telemetry is also suggested.
Constraints	A key known constraint is the potential sensitivity of IP addresses or other data, necessitating a method to filter information or implement role-based access. However, no other known constraints are identified.
Potential solutions	Several opportunities and potential solutions were identified, including establishing an intake form for bandwidth guarantee requests, exporting statistics from OSCARS to a dashboard using tools like Grafana, providing API access to data for integration with other applications, and performing statistical analysis to understand bandwidth consumption trends.

Gap analysis	Analysis revealed a significant challenge: the lack of a formal service intake process and a common repository for existing Service Level Agreements (SLAs). Additionally, it notes that usage and telemetry data are distributed across various sources, necessitating an integration method or data import strategy. Implementing role-based access control is also identified as a requirement.

Table 2.4.7.6.a. WP06 (Network Services) summary.

Implementing the suggested solutions would result in a centralized, accessible system for managing and understanding ESnet's Guaranteed Bandwidth Service. This would include a formal intake process for bandwidth requests, a dashboard visualizing key metrics, API access for integration with other applications, and comprehensive analysis of usage trends. Ultimately, this would improve transparency, efficiency, and the ability to monitor and manage service commitments effectively.

2.4.7.7. WP07 (Business Ops)

This work-package is focused on improving the Business Office's data management and access for invoice validation and procurement renewals at ESnet.

WP07 (Business Ops)	
Opportunity	The challenge faced by business operations staff is mapping services/inventory to sub-contracts/POs for invoice validation and procurement renewals. The central problem is the lack of a defined database or centralized dashboard for the Business Office to cross-reference POs and invoices, hindering efficient operation and validation.
Datasets	Datasets necessary to address the problem include information currently spread across various databases, namely ServiceNow, FMS, BAR, ESDB, and Google Drive. These sources hold critical data related to services and inventory, but there is currently no clear direction on which should be referenced for the most up-to-date and relevant information.
Constraints	There are no data sensitivity issues of concern within ESnet, and the target access policy for the potential solutions is "Internal Only". The work-package is primarily an integration challenge, not an AI problem. The suggested interface should be presented to humans.

Potential solutions	Create a dashboard that links data from multiple existing sources like ESDB and ServiceNow. This dashboard should link POs to circuit IDs, cross-connects, and colocation via information available on ESDB, such as vendor details, contact information, end dates, and services. It is important that all information on ESDB is validated.
Gap analysis	Revealed a lack of data mapping and verification between the various systems currently in place. While systems exist, there is no clarity on how information from one source relates to another. Furthermore, processes to ensure data is kept up to date are missing, along with clear identification of the responsible parties for data maintenance. There is also a large amount of manual labor in data entry currently.

Table 2.4.7.7.a. WP07 (Business Ops) summary.

Implementing the suggested solutions would result in a centralized dashboard that provides the Business Office with a unified and validated view of service/inventory data, enabling efficient mapping to sub-contracts/POs for accurate invoice validation and procurement renewals. This would streamline operations, reduce manual labor, and ensure data accuracy, ultimately improving overall efficiency and data management.

2.4.7.8. WP08 (Outage Notification Parsing)

This work-package focuses on automating the parsing of outage notifications received via email from network providers and entering them into systems like ServiceNow and ESDB.

WP08 (Outage Notification Parsing)	
Opportunity	Automate the current procedures of manually processing outage notifications received via email from network providers. The manual process of entering this information into systems like ServiceNow is time-consuming, error-prone, and struggles to keep up with updates. Automating this process is identified as a key opportunity to reduce errors and improve efficiency for Network and NOC Engineers.
Datasets	There are two primary Datasets outage notifications from network providers (Lumen, Internet2, GEANT, etc.), typically received as emails, and inventory data from existing systems such as ServiceNow and ESDB. This data would form the basis for an automated parsing and updating system.

Constraints	There is a potential for some providers to consider outage details or circuit IDs sensitive or confidential, possibly requiring adherence to NDAs. Therefore, any automation system must ensure secure processing and storage of this data. Additionally, the new tool must integrate seamlessly with existing ticketing and inventory systems.
Potential solutions	Use an AI-based or automated parsing tool to harvest emails from designated sources and convert their content into a structured format. This parsed data would then trigger the creation of change records via the ServiceNow API, ideally linking specific items mentioned in the notifications to their corresponding entries in ServiceNow.
Gap analysis	The analysis indicates that available data sources appear sufficient for initial automation stages, but consistent naming conventions and robust lookups are crucial for accurate matching. Validation of changes, extensions, or cancellations from providers across all systems is critical. Incomplete data in ServiceNow or ESDB might require additional updates or new data flows.

Table 2.4.7.8.a. WP08 (Outage Notification Parsing) summary.

Implementing the suggested solutions would result in significantly reduced manual work for Network Engineers and NOC staff, as outage notifications would be automatically parsed and entered into ServiceNow and other systems. This automation would decrease human error, speed up operations, and improve workflow efficiency, while also allowing other automation tools to leverage this information for proactive system management.

2.4.7.9. WP09 (Ticket Resolution)

This work-package outlines a plan to improve ticket resolution at ESnet by leveraging past troubleshooting experiences. The goal is to create a system that allows staff, particularly NOC and Network Engineers, to quickly find solutions to new, similar problems by analyzing prior ticket resolutions.

WP09 (Ticket Resolution)	
Opportunity	Improve ESnet's ticket resolution process by leveraging past troubleshooting experiences. The problem statement revolves around the need for NOC and Network Engineers to quickly determine solutions for new but similar issues by analyzing prior ticket resolutions. Instead of starting from scratch, they aim to utilize a knowledge base of past solutions to expedite the troubleshooting process.

Datasets	To achieve this, several datasets are identified, including ServiceNow trouble ticket data, Jira tickets, Slack channel discussions, email conversations, and Zoom meeting transcripts. These sources contain valuable information about past incidents and their resolutions, which can be analyzed to provide insights for resolving current issues.
Constraints	Potential sensitivity of ticket data, which may contain PII, IP addresses, or site access information that shouldn't be shared externally. There's also a risk of providing poor guidance to junior staff if the system is not well-developed. Additionally, limitations include the potential for ticket export data to overwhelm the AI model and concerns about data completeness and consistency. Customer confidentiality and special handling of sensitive site data are crucial considerations.
Potential solutions	Create an interactive command prompt or web interface focused on troubleshooting, formalizing ticket structures, filtering data to focus on ESnet-made solutions, introducing audit steps to validate information, feeding ticket data and related discussions into the solution, implementing Model Context Protocol (MCP) technology for dynamic data queries, and exploring vendor-provided AI solutions. These solutions aim to provide engineers with a comprehensive knowledge base for efficient problem resolution.
Gap analysis	Need to define troubleshooting workflows, develop a proof of concept, establish a common ticket format, introduce procedural steps for documentation, determine what information to feed into the solution, build MCP integrations, and address the challenge of ticket quality assurance at scale. While the needed datasets are available, they are often incomplete or inconsistent, requiring normalization.

Table 2.4.7.9.a. WP09 (Ticket Resolution) summary.

Implementing these solutions would result in a system that enables ESnet engineers to efficiently resolve network and system issues by leveraging past troubleshooting experiences. This would lead to faster resolution times, reduced downtime, and improved overall network reliability. The system would provide a centralized knowledge base, facilitate dynamic data queries, and ensure consistency in ticket data and resolution processes.

2.4.7.10. WP10 (Correlate Alarms)

This work-package focuses on correlating alarms and maintenance notifications to real-world impacts for NOC engineers.

WP10 (Correlate Alarms)	
Opportunity	Improve alarm correlation and maintenance notification analysis for NOC engineers, aiming to quickly link these alerts to real-world service impacts. The core problem is the difficulty in determining affected services during outages due to issues such as multiple alarm feeds with correlation problems, decentralized network topology data, and inconsistent data structures.
Datasets	The datasets required to solve this problem include alarm data, infrastructure documentation that links back to service impact, and a defined availability model for outage categorization (e.g., Up, Impaired, Down). These datasets are essential for understanding the scope and severity of incidents and their effects on services.
Constraints	The creation of a correlation tool is hindered by the inability to programmatically interpret unstructured information from vendor/site maintenance notifications, data inconsistency or missing information in network documentation, and the absence of a service instance database, which makes comprehensive correlation challenging. Additionally, sensitive data in troubleshooting activities raises concerns.
Potential solutions	Leveraging existing alarm inputs like Spectrum, TNMS, Syslog, and email notifications, and considering the use of AI for alarm correlation logic as built into existing auto-ticketing scripts. Ideally, solutions would integrate with existing workflow tools like ServiceNow and allow users to query and refine generated responses.
Gap analysis	Reveals that most needed data exists but is often not structured, complete, or accurate enough for reliable use without human intervention. Network model data may be missing or incorrect, a complete service catalog is lacking, and notifications from sites/peers may not contain sufficient information for automated analysis.

Table 2.4.7.10.a. WP10 (Correlate Alarms) summary.

Implementing the suggested solutions would result in a system capable of rapidly correlating alarms and maintenance notifications to real-world service impacts, enhancing the efficiency of NOC engineers. This would involve improved data structures, AI-powered parsing of natural language notifications, and seamless integration with existing workflow tools, ultimately leading to quicker identification and resolution of service-affecting incidents.

2.4.7.11. WP11 (Predict Hardware Failures)

This work-package focuses on predicting hardware failures in networking equipment. The goal is to identify early warning signs from logging and telemetry data to proactively replace at-risk hardware, minimizing unplanned outages.

WP11 (Predict Hardware Failures)	
Opportunity	Reduce the disruptive nature of networking equipment hardware failures by attempting to predict these failures using log and telemetry data. The goal is to proactively replace at-risk hardware before failure occurs, thus reducing unplanned outages and improving network stability.
Datasets	Several key datasets are needed to predict hardware failures, including optical performance metrics (voltage levels), transponder and transceiver bit error rate (BER) counts, Forward Error Correction (FEC) state changes, and hardware SNMP and Syslog data reporting equipment state and changes over time, such as temperature, voltage, and error counts. Specifically, Stardust, LibreNMS, Netlog/Syslog data, and DNA for Open Line System (OLS) metrics are targeted for analysis.
Constraints	Potential reluctance of vendors to authorize proactive Return Material Authorizations (RMAs) before an actual failure. Additionally, while optical metrics are not considered sensitive, Syslog data may contain proprietary information. The primary requirement is the ability to accurately predict hardware failure likelihood using data analysis.
Potential solutions	Analyze syslog event facility and severity logs for patterns that precede hardware failures, along with utilizing Optical Performance Monitoring (PM) to detect deterioration based on increasing voltage levels, elevated BER, and changes in FEC compensation. Algorithms would need to be developed for both syslog and optical data streams. User interaction would be facilitated through a report or dashboard indicating the most likely hardware to fail, with links to the data used for analysis to verify prediction accuracy.
Gap analysis	Need further investigation into existing datasets to identify key failure indicators and the development of algorithms for predictive purposes. It also highlights the potential need for a system to capture, display, and link alarms to source time-series data. Further investigation and development may be required to retrieve and store OLS PM Data to a centralized location. The collection of

	performance measurements for new hardware types is also necessary, as well as historical forensic data for model training.

Table 2.4.7.11.a. WP11 (Predict Hardware Failures) summary.

If implemented, the solutions outlined in this work-package would provide network and NOC engineers with a report of at-risk hardware based on log and telemetry data analysis. This would enable proactive scheduling of maintenance events to replace potentially failing hardware before a failure occurs, monitor equipment health over time, and make informed decisions to prioritize hardware upgrades based on estimated lifespans, ultimately improving network stability and reducing unplanned outages.

2.4.7.12. WP12 (Detect External Configuration Anomalies)

This work-package focuses on detecting external configuration anomalies that can disrupt network traffic. It aims to provide engineers with a tool that analyzes received network configuration data from peers, highlights routing path changes, and notifies them of detected anomalies.

WP12 (Detect External Configuration Anomalies)	
Opportunity	There is a desire to reduce the number of network disruptions and degradations caused by misconfigurations in data shared between autonomous networks, or "peers." The opportunity lies in developing a tool that analyzes received network configuration data, identifies routing path changes, and notifies engineers of anomalies and their impact on the network.
Datasets	Received routing updates (like BMP and BGP data), external third-party tools such as Routeviews and PeeringDB for a global network view, flow data to detect traffic changes, SNMP/Netlog data for hardware-level traffic changes, geo-locating IP addresses, and configuration validation via orchestration tools. These datasets are crucial for monitoring and analyzing network behavior and detecting anomalies.
Constraints	Potential data sensitivity issues, particularly with flow data, which is considered private. Legal constraints and government mandates are not specifically mentioned but may apply. There's also the challenge of differing naming and data conventions between networks, and while specific data sources may be network-proprietary, sharing summarized data with a lower TLP classification could facilitate collaboration.

Potential solutions	Focus on individual data sources initially, with cross-correlation in future iterations. Specifically, solutions include leveraging the existing Stardust ingest pipeline for flow data, monitoring ingress BGP attributes, and monitoring ingress interface counters via Stardust for near-zero counts. These solutions aim to detect external network changes and generate events upon detection of anomalies.
Gap analysis	Need to implement application logic for monitoring flow data, BGP properties, and SNMP thresholds. There's also the challenge of managing high-volume, low signal-to-noise "firehose logs" and the need to ingest control plane updates into a long-term structured format. While most data exists and is accessible, improvements in data processing and analysis are required.

Table 2.4.7.12.a. WP12 (Detect External Configuration Anomalies) summary.

Implementing the suggested solutions would result in a network analysis tool that provides real-time monitoring of received network configuration data, identifies routing path changes, highlights configuration anomalies, and notifies network engineers of potential impacts on the current network state, ultimately improving network stability and efficiency.

2.4.7.13. WP13 (Capture Configuration Intent)

This work-package focuses on capturing configuration intent for system configuration changes to speed up the process and eliminate blind spots.

WP13 (Capture Configuration Intent)	
Opportunity	Capture "configuration intent" for engineers making system configuration changes. The goal is to record the context associated with code or metadata changes to expedite the configuration process and avoid potential errors. Target users are system administrators and network engineers who need to correlate issues with previous changes and understand the reasoning behind them.
Datasets	Git repositories for code and Ansible scripts, LDAP updates for user and host databases, Jira tickets, wiki/Confluence documentation, Google Drive files, and ESDB (likely a specialized database). The document references a list of data sources located in an external Google Sheet.
Constraints	Need for access control for Git repositories and other data sources due to data sensitivity. Additionally, there is currently no clear method to link a ticket to the

	resulting configuration changes, indicating a need to add fields for tracking this association.
Potential solutions	Examine the hierarchy of configuration files to develop a tool for understanding intent. The work-package suggests referencing research papers on specification mining and related projects like Pybatfish for inspiration. It also lists several research papers and related works on network configuration and validation.
Gap analysis	Need to better define "configuration intent" and determine which existing tools are suitable for ESnet's needs. Developing an ontology of intent for network engineering is proposed, along with leveraging Large Language Models (LLMs) due to intent being expressed in natural language. This would likely require developing an application with a query interface and historical data viewing capabilities, with humans acting as users to review and validate results.

Table 2.4.7.13.a. WP13 (Capture Configuration Intent) summary.

Implementing the suggested solutions would result in a tool or application that allows engineers to easily understand the intent behind configuration changes by querying various data sources and historical records. This system would provide context, potentially linking changes to trouble tickets and documentation, leading to faster configuration processes, reduced errors, and improved overall network management.

2.4.7.14. WP14 (Fast Contract Lookup)

This work-package outlines the need for a system to quickly access contract information and validate vendor billing for ESnet. Currently, these tasks are time-consuming.

WP14 (Fast Contract Lookup)	
Opportunity	There is a need for a more efficient way to look up contract information and validate billing at ESnet. Currently, Network Engineers and business staff spend significant time locating applicable contracts for specific equipment and verifying vendor charges against invoices, which include various services like equipment, circuits, and licenses.
Datasets	Contract details/Invoices (found in Google Sheets and purchasing systems) and ESDB (for equipment and circuit information). These data sources are essential for providing comprehensive and accurate information for contract lookups and billing validations.

Constraints	Presence of sensitive data within the data sources, which necessitates the implementation of role-based access controls. Furthermore, the requirement for normalized and searchable contract information to support keyword searches and retrieval of relevant data related to services or equipment is crucial.
Potential solutions	Ingest contract and invoice data into a common database and make it searchable, leveraging AI technologies like Gemini for processing documents, and considering ServiceNow's contracts module for associating contracts with network elements. Additionally, building an application or AI agent that can retrieve and combine information from service contracts and the inventory database is suggested.
Gap analysis	Contracts are currently scattered across emails, Google Sheets, and purchasing systems, requiring integration and normalization. There are also issues with the accuracy and ownership of data in the inventory database, as well as the imperfect association of purchase orders to network elements.

Table 2.4.7.14.a. WP14 (Fast Contract Lookup) summary.

Implementing the suggested solutions would result in a more streamlined process for contract lookups and billing validation, reducing the time spent by staff and improving accuracy. This would be achieved by centralizing and normalizing contract information, leveraging AI for data retrieval, and enhancing the inventory database's accuracy and data ownership.

2.4.7.15. WP15 (Consistent Data Management)

This work-package outlines the problem of inconsistent data management and analysis across teams within the organization. Different teams use varying methods and formats, leading to fragmentation and hindering collaboration.

WP15 (Consistent Data Management)	
Opportunity	ESnet suffers from inconsistent data management and analysis across teams within an organization. Currently, teams employ various methods and formats, leading to fragmentation, hindering cross-team collaboration, and complicating the development of operational tooling and unified infrastructure views. The core problem is the lack of a shared, normalized data format and a common model for describing resources and their components.
Datasets	Realizing a unified, composable model of the system will require access to all available organizational datasets, including infrastructure metrics, resource inventories, monitoring telemetry, service dependencies, and domain-specific

	information. This broad data access is crucial for building a holistic understanding of the infrastructure and services.
Constraints	Potential sensitive information within data sources requiring review, the possibility that not all systems can adopt a shared format due to legacy or technical limitations, and the potential for organization-wide debates regarding a unified schema due to differing team needs. Variations in access control models and automation challenges are also recognized as complexities to be addressed.
Potential solutions	Establish a consistent set of naming conventions, ideally mandated across domains, and utilizing an extensible ontology framework like OWL to define resource types and relationships. Namespacing strategies could be introduced to accommodate team-specific extensions, while Domain Owners would be identified for each area to ensure alignment and long-term sustainability. An iterative approach to ontology coverage is recommended.
Gap analysis	There are organizational challenges, such as conflicting opinions across teams and the absence of a clearly defined domain owner. The document also highlights the lack of a framework for representing access control within the proposed model and the absence of a method for prioritizing areas of misalignment. These gaps need to be addressed to prevent standardization efforts from stalling or leading to incomplete solutions.

Table 2.4.7.15.a. WP15 (Consistent Data Management) summary.

Implementing the suggested solutions would result in an organization with a unified, consistent data model, enabling seamless cross-team collaboration, improved operational efficiency, and accelerated innovation. The organization would benefit from a holistic understanding of its infrastructure and services, enhanced reusability of data and tools, and a more agile development process.

2.4.7.16. WP16 (Query All Data)

This work-package outlines the need for a unified search system for ESnet engineers to access information across various platforms. Currently, searching for data is fragmented and time-consuming due to disparate interfaces and syntax.

WP16 (Query All Data)	
Opportunity	ESnet engineers currently face problems when searching for information across various systems and documentation sources. The current process is fragmented and inefficient, requiring manual searches across disparate platforms and leading to duplicated effort, slower decision-making, and reduced situational awareness. The opportunity is to create a unified, intelligent search capability that allows

	engineers to query all available data through a single interface, improving operational efficiency and decision quality.
Datasets	Documentation repositories like Google Docs, Confluence Wiki, and Lucidcharts; operational records such as Jira tickets, ServiceNow entries, and ESDB entries; network telemetry and logs from Stardust, syslog, and NetFlow; and historical incident summaries and engineering analyses. These diverse data sources hold valuable information that engineers need to access quickly and effectively.
Constraints	Strict adherence to access and compliance boundaries, particularly with Controlled Unclassified Information (CUI), Personally Identifiable Information (PII), and other sensitive content. The system must support fine-grained access controls, robust user authentication and authorization, and consistent formatting standards for unstructured documentation to ensure data security, privacy, and reliable processing.
Potential solutions	Build an intelligent query interface powered by Retrieval-Augmented Generation (RAG) and relevance-based ranking, such as PageRank. This system would enable engineers to ask natural-language questions and receive ranked, contextualized answers along with direct links to relevant documents and records. A programmatic API would also support automation workflows and embedding the search capability within existing tools and dashboards.
Gap analysis	There are several key challenges between the current state and the proposed solution, including the lack of a centralized index across all documentation silos, non-vectorized and inconsistently structured documents, the absence of a unified access control framework, and no existing infrastructure for real-time semantic ranking and retrieval. These gaps need to be addressed to enable an effective search experience.

Table 2.4.7.16.a. WP16 (Query All Data) summary.

Implementing the suggested solutions would result in a unified query system that streamlines access to institutional knowledge, reduces cognitive overhead, increases operational agility, and preserves institutional knowledge for both current and future teams. Engineers would be able to quickly find relevant information across all data sources, validate hypotheses, find authoritative references, and act with greater confidence, ultimately improving overall operational efficiency and decision-making.

2.4.7.17. WP17 (Automating Site Deployment)

This work-package discusses automating site deployment for network engineers to streamline pre-orchestration processes, reducing errors and increasing efficiency.

WP17 (Automating Site Deployment)	
Opportunity	Automate site deployment for network engineers, who currently wish to streamline the process by automating pre-orchestration tasks like base configuration and bootstrapping. This will reduce manual errors and improve deployment efficiency.
Datasets	Inventory management data (Bill of Materials), staging and shipping coordination details, communication protocols between tools and teams, credential access information, and data currently stored across multiple spreadsheets, all necessary for supporting site deployment automation.
Constraints	Complexity due to numerous steps, potential issues with asynchronous execution, and constraints identified during user interviews. Specific constraints noted are accurate inventory management, staging and shipping coordination, inter-tool communication, secure credential access, consolidating spreadsheet information, limited resources, managing configurable credentials, and the sensitivity of some IP information.
Potential solutions	Leverage hybrid AI to automate parts of the pre-orchestration process, using an AI-assisted deployment agent to provide real-time guidance, and establishing a clear framework that documents all steps, requirements, dependencies, and constraints. A proposed solution includes an AI agent that enumerates and manages steps from documented information, identifies asynchronous steps, manages blocked steps, reserves resources, and performs safe actions.
Gap analysis	Analysis is based on limited feedback and further user interviews are needed for a clearer picture of improvements. The issues appear to be more process-related than AI automation issues, and standardization of elements like IP address block assignment and site naming schemas needs to be addressed.

Table 2.4.7.17.a. WP17 (Automating Site Deployment) summary.

Implementing the suggested solutions would result in a streamlined site deployment process using an AI agent, which provides real-time guidance and automation, reduces errors, increases productivity, and allows network engineers to focus on more strategic activities. It also emphasizes the need for standardized processes and better data management for successful automation.

2.4.7.18. WP18 (Automating Configuration)

This work-package was assimilated into WP17 (Automating Site Deployment).

2.4.7.19. WP19 (AI Sandbox)

This work-package describes the need for an AI Sandbox environment for ESnet staff to experiment with AI models and ESnet data.

WP19 (AI Sandbox)	
Opportunity	ESnet personnel want to analyze ESnet data with AI models but lack a dedicated space to test and evaluate different models against their specific data, hindering proof-of-concept development.
Datasets	Necessary datasets exist within ESnet but need to be collected and collated for this development purpose. Subject matter experts will need to incorporate the data into the sandbox, particularly for sanitization or redaction purposes due to potential sensitivity issues.
Constraints	Potential sensitivity of the data, requiring sanitization or other measures to keep data within ESnet. There may be challenges to provisioning enough storage and a programming environment for running the models, noting sizing the sandbox space might be difficult.
Potential solutions	Deploy a storage system containing datasets in a structured directory, a set of AI models, and wiki instructions on using them. Possibility of using LBL's CBorg and ESnet's NERSC allocation for training models. Finally, it suggests PIPE talks to disseminate knowledge gained from using the sandbox.
Gap analysis	Datasets must be provided by domain experts, AI models need to be incorporated, systems and storage resources must be allocated, and documentation/instructions need to be written for the sandbox environment.

Table 2.4.7.19.a. WP19 (AI Sandbox) summary.

Implementing these solutions would create a working AI Sandbox for ESnet staff to experiment with various AI models and techniques on ESnet data. This would enable proof-of-concept development, knowledge sharing within ESnet, and the potential exploration of advanced AI applications using ESnet resources.

2.4.7.20. WP20 (RFP/Contract Builder)

This work-package describes the need for a more efficient way to generate RFPs and contracts at ESnet. Currently, the process is time-consuming, requiring manual input and coordination from various sources, which can lead to delays and inconsistencies.

WP20 (RFP/Contract Builder)	
Opportunity	Automate/assist in the creation of Requests for Proposals (RFPs) and contracts, which is currently a time-consuming and inefficient process involving gathering fragmented information from various sources. This leads to delays, inconsistencies, and risks of misaligned contracts. There is an opportunity to streamline this process by developing a structured and minimal-input method for generating RFPs and contracts using standard templates, reusable components, and auto-populated fields.
Datasets	Will require datasets for training an AI model, primarily focusing on existing RFPs and contracts sourced from ESnet and Network Services (NS). Additionally, it suggests incorporating high-quality RFPs and contracts from across LBNL and, optionally, anonymized RFP responses for further refinement. The document also highlights the importance of metadata, such as creation dates and outcome-based confidence ratings, to improve the model's accuracy.
Constraints	Sensitive nature of the data, which is often classified as TLP:Green or TLP:Red and protected by NDAs. Any solution must ensure appropriate handling, storage, and access controls to protect this information. The system also needs to adhere to LBNL-required phrasing and terminology, such as the correct use of "must," "should," and "shall."
Potential solutions	Develop an AI-assisted framework to streamline the creation of RFPs and contracts using a language model trained on curated datasets. This model would incorporate rule-based constraints, and a natural language interface would allow users to generate new RFPs by specifying asset types and requirements. The system would assemble tailored documents from historical RFPs and provide a draft with inline scoring for review.
Gap analysis	Although there is sufficient internal data, the utility of incorporating RFP responses into the training set raises privacy concerns. It also notes that relying solely on publicly available documents might not align with LBNL's legal language. Metadata and access to internal documentation outlining legal phrasing changes are needed to increase the model's accuracy and adaptability.

Table 2.4.7.20.a. WP20 (RFP/Contract Builder) summary.

Implementing the suggested solutions would result in a more efficient and streamlined process for generating high-quality RFPs and contracts, reducing manual input, delays, and inconsistencies. The

AI-assisted system would provide a tailored document with a structured summary, enabling better clarity and reducing the risk of misinterpretation during the bidding process.

2.4.7.21. WP21 (Unified Document Search)

This work-package discusses the need for a unified document search system for ESnet employees. Currently, information is scattered, making it difficult to find. The goal is to create a global search interface to improve collaboration and decision-making.

WP21 (Unified Document Search)	
Opportunity	ESnet employees face challenges searching for information scattered across various documents and knowledge bases. There is an opportunity to develop a unified global search interface to simplify this process, improve collaboration, enhance decision-making, and foster innovation. This search platform aims to consolidate disparate sources into a single accessible point.
Datasets	ESnet's internal documents and knowledge bases, encompassing a wide range of data formats and quality levels. The search solution must be able to handle this variety while also accommodating varying levels of access control and security requirements. Integration with existing rules and restrictions from multiple data sources is a key consideration.
Constraints	Need for data-level and query-level access control to protect sensitive information and limit query types. There is also a preference for a self-hosted solution due to the internal nature of the data, which must align with lab policy and legal advice. Existing rules and restrictions must be integrated, considering RBAC/ACL considerations across systems.
Potential solutions	Leveraging AI or hybrid approaches, including known AI/ML solutions like Google Gemini, docq.ai, meilisearch, NotebookLM, and RAG. Possibly generate a custom AI solution or utilizing off-the-shelf products, along with the need for a projected timeline for development and deployment. Technical challenges, such as data quality and computational resources, as well as organizational challenges like staffing, must be addressed.
Gap analysis	There is concern on how RBAC/ACLs and general security will be handled, particularly regarding users with different access levels on different systems. Multiple approaches were proposed, including running everything behind a proxy for user validation and access control, or having each underlying system provide its

	own query interface and enforce access control, with a higher-level search tool propagating the user's identity.

Table 2.4.7.21.a. WP21 (Unified Document Search) summary.

Implementing the suggested solutions would result in a unified search interface spanning ESnet's entire knowledge landscape. This would streamline workflows, unlock new insights, drive progress in time-sensitive work, and improve the overall efficiency and effectiveness of information retrieval for ESnet employees.

2.4.7.22. WP22 (Automated ServiceNow Ticket Summarization)

This work-package focuses on automating ServiceNow ticket summarization. ServiceNow tickets contain large amounts of unstructured data, making it difficult to understand incident timelines and status.

WP22 (Automated ServiceNow Ticket Summarization)	
Opportunity	ServiceNow tickets have inconsistent data, including redundant content and poorly maintained fields. This complicates understanding incident status and timelines. The goal is to create concise AI-generated summaries for ServiceNow operators and staff to aid in incident response, management, and documentation.
Datasets	The primary data input for this project is ServiceNow ticket logs, with potential additional data from the ESDB.
Constraints	Several constraints and requirements are identified, including data sensitivity (PII), data quality issues with ServiceNow tickets (poorly maintained fields like resolution notes), UX and integration limits, privacy and compliance with ESnet's standards, and limited in-house expertise in prompt engineering for LLMs.
Potential solutions	Two implementation paths are proposed: 1) ServiceNow Paid Add-On, offering easy deployment but with unknown cost and quality; and 2) Custom Implementation, leveraging ServiceNow APIs and LLMs for better control and customization.
Gap analysis	Crafting effective prompts for LLMs is the largest challenge, as this directly impacts the quality of generated summaries. Inconsistent population of fields like resolution notes and missing links to After Action Reports also pose difficulties. Additionally, the desired level of UI integration is unclear, and any third-party tools must undergo security compliance reviews.

Table 2.4.7.22.a. WP22 (Automated ServiceNow Ticket Summarization) summary.

Implementing the suggested solutions would result in a tool that generates concise, organized summaries of ServiceNow tickets directly within the platform. This would improve incident response, documentation, and managerial oversight, potentially paving the way for broader AI/ML use within the organization.

2.4.7.23. WP23 (Federated Authentication)

This work-package focuses on developing a federated authentication system for ESnet to allow external collaborators secure access to resources while maintaining security and compliance.

WP23 (Federated Authentication)	
Opportunity	Challenge of providing secure access to ESnet's sensitive scientific data and HPC resources for external collaborators while meeting strict security, compliance, and usability requirements. The goal is to balance these often conflicting needs, enabling streamlined collaboration without compromising data protection. The desired outcome is a system allowing secure data sharing, streamlining collaboration via Single Sign-On (SSO) and Just-In-Time (JIT) provisioning, ensuring compliance with government regulations, allowing for scalable onboarding of new collaborators, and enabling rapid breach detection and mitigation.
Datasets	User data (affiliation, role, clearance), Access Control Lists (ACLs), Federated Identity Metadata, dataset metadata (classification, ownership), logs for real-time monitoring, certificates, and agreements such as Memoranda of Understanding (MOUs) and regulatory compliance documents. These diverse datasets are crucial for the system to accurately manage access and maintain security.
Constraints	Legal compliance with DOE regulations and federal law, hardware and software limitations including legacy systems, potential incompatibility between Federated Identity Providers (IdPs) and ESnet's Service Provider (SP), and data sensitivity concerns covering classified information, personally identifiable information (PII), and export-controlled data. Addressing these constraints is vital to creating a secure and functional system.
Potential solutions	Enhance InCommon Federation Integration by leveraging the existing InCommon Federation and OIDC. This would streamline authentication for DOE labs already part of the federation while addressing gaps for non-InCommon labs. This solution involves utilizing InCommon Metadata, user attributes (eduPersonPrincipalName, eduPersonAffiliation, roles), and access logs. This approach aims to simplify authentication and improve security for diverse users.

Gap analysis	Inefficiencies like manual authentication for non-InCommon labs, inconsistent attribute release causing access control mismatches, insufficient logging and analytics capabilities, and missing data elements such as lab-specific authentication details, unified role definitions, dataset licensing terms, user feedback channels, and real-time security metrics. Overcoming these gaps requires collecting and normalizing data, developing a system for actionable alerts, and implementing techniques to detect anomalies.

Table 2.4.7.23.a. WP23 (Federated Authentication) summary.

Implementing the suggested solutions would result in a robust and secure platform for data sharing and collaboration, improving security, productivity, and compliance. By addressing the current inefficiencies and gaps, the system would enable researchers and scientists to focus on their work while ensuring the data's security and integrity.

2.4.7.24. WP24 (Legacy Code)

This work-package focuses on replacing legacy Perl scripts and programs at ESnet, to convert them to Python or another readable language. The process involves using an LLM with coding capabilities, creating unit tests for both the original and converted software, and validating the new software.

WP24 (Legacy Code)	
Opportunity	There is a desire to replace legacy Perl scripts and programs at ESnet, particularly those used in Network Services, with more readable software like Python. The problem statement identifies the need for engineers to remove and rewrite these legacy Perl scripts, which are often not well-understood or documented.
Datasets	Collection of all existing Perl software currently in use at ESnet.
Constraints	Both the input Perl scripts/programs and the resulting replacement software should not be publicly exposed unless decided otherwise after the work is completed.
Potential solutions	Use an LLM with coding capabilities to rewrite the Perl scripts into Python or another readable language. Alternatively, create more readable Perl that clarifies "magic variables" to aid in understanding the original software's function. Another proposed solution is the creation of unit tests for the original Perl scripts to validate their functionality and subsequently developing unit tests for the converted Python code to ensure accurate conversion.

Gap analysis	The primary issue is missing institutional knowledge due to the departure of the original authors of the legacy code. This knowledge is now fragmented, making it difficult to fully understand the original software's functionality. Additionally, there is a lack of a testing harness to validate any software conversion, and a need to establish ownership and maintenance procedures for the new software to prevent future recurrence of this problem.

Table 2.4.7.24.a. WP24 (Legacy Code) summary.

Implementing the suggested solutions would result in the replacement of legacy Perl programs at ESnet with equivalent software written in Python or another more readable language. This would include creating unit tests to validate both the original and converted software, ensuring accurate functionality, and establishing clear ownership and maintenance plans for the new software.

2.4.7.25. WP25 (NLP Interfaces to Systems)

This work-package focuses on integration of NLP interfaces to systems within ESnet.

WP25 (NLP Interfaces to Systems)	
Opportunity	The goal is to provide an NLP interface to ESnet engineers and customers, allowing them to interact with systems without needing coding, API, or CLI knowledge. This aims to improve accessibility and understand user needs better.
Datasets	Various datasets are required, including system configurations, API manuals, ESDB data, wiki pages, configuration histories, Ansible playbooks, example user intents, ServiceNow tickets, and external configuration examples.
Constraints	Risks include users misinterpreting NLP interpretations, integrating AAA policies, potential leaks of sensitive information from training data, and ambiguity in natural language. Software limitations include dealing with diverse systems and the large amount of work for training data preparation. Data sensitivity is a concern, particularly regarding system configuration data leaks.
Potential solutions	NLP interfaces can save training time for engineers and clarify user intent. For customers, it can serve as an additional interface, reduce training time, and act as a single point of touch. Use case scenarios include creating VLAN interfaces, summarizing network states, creating Ansible scripts, and reserving network paths using natural language.

Gap analysis	Gaps include determining where to start, which LLM product to use, how to fine-tune, how to anonymize data, missing or unstructured documents, and integration of user admission control. Recommendations include implementing a gatekeeping mechanism to review and validate NLP-generated system configuration changes and detecting "risky" actions.

Table 2.4.7.25.a. WP25 (NLP Interfaces to Systems) summary.

Implementing the suggested solutions would result in the creation of an NLP interface that allows users to interact with ESnet systems using natural language, simplifying complex technical procedures. This would reduce training time for engineers, enhance accessibility for customers, improve service order processing, and provide a more intuitive way to manage network configurations and tasks.

2.4.7.26. WP26 (Information Architecture)

This work-package outlines the need for better understanding and documentation of data systems at ESnet. Engineers require knowledge of data production, consumption, and flow to improve accuracy, avoid breaking dependencies, and efficiently update data.

WP26 (Information Architecture)	
Opportunity	There is a need for a clear understanding of data production and consumption within ESnet. The problem statement focuses on engineers needing to know which systems produce and consume data, understand existing designs to avoid breaking dependencies, identify data sources for updates, know who to work with for schema changes, and comprehend system taxonomies for effective communication.
Datasets	List of existing datasets produced by different systems at ESnet, including the taxonomy of each system. Detailed information about each system or collection of systems, such as the services provided, authorization for changes, inputs and outputs, data updating dynamics, and freshness requirements of workflows, along with understanding how data flows through each system to others.
Constraints	There are currently no identified constraints for this project. Requirements include the need for comprehensive data documentation with metadata.
Potential solutions	Develop a directed graph illustrating data flow between systems, maintaining a detailed spreadsheet listing systems of record and their respective data sources and fields, implementing a change control process for managing system and dataset

	changes, and providing the document to large language models for understanding data locations and order. Initially, low-tech solutions like an embedded wiki diagram can be used before developing more complex user experiences.
Gap analysis	Previous attempts to document datasets have been incomplete, there is no designated owner for this process, no management sponsor, and the need to ensure all necessary datasets and metadata are available and complete. Existing spreadsheets need consolidation and formal ownership.

Table 2.4.7.26.a. WP26 (Information Architecture) summary.

Implementing the suggested solutions would result in a comprehensive and well-maintained information architecture for ESnet, enabling engineers to efficiently manage data, understand system dependencies, and effectively communicate about data structures and flows, leading to improved data accuracy, synchronization, and overall system reliability.

2.4.7.27. WP27 (Requirements Management)

This work-package outlines a problem with ESnet's current requirements management. The current processes are not integrated, lack a system-level view, and communication across groups is inconsistent.

WP27 (Requirements Management)	
Opportunity	The current processes for handling project requirements are fragmented and lack a system-level view. This leads to difficulties in communicating requirements across teams and efficiently fulfilling customer needs. The desired outcome is a common schema, standardized evaluation processes, and a queryable repository for requirements that would facilitate dashboards, natural language queries, and resource forecasting.
Datasets	Measurements of the resource utilization of our existing system, timelines, target and baseline metrics, security requirements, and a central repository for requirements with a standardized intake process. These datasets are essential for understanding the current state and developing effective solutions for future management of requirements.
Constraints	There are no known legal constraints, government mandates, risks, hardware, or software limitations. Similarly, there are no data sensitivity issues of concern. This

	simplifies the solution space as there are no external factors limiting the scope or execution of the proposed requirements management system.
Potential solutions	Standardize how and when requirements are collected and provided, establishing a transparent process for prioritizing work, and integrating into a common requirements management system with verification and validation capabilities. Designate a responsible party for maintaining a requirements register and providing forward-looking roadmaps for resource planning and prioritization. End-user interaction would focus on accessing and utilizing data within this system for planning and fulfillment.
Gap analysis	Requirements are currently gathered and used differently by various groups and that the process for prioritizing work is not transparent. The necessary data exists but is scattered and lacks a common repository or process for generation and sharing. This indicates a need for consolidation and standardization of requirements management practices.

Table 2.4.7.27.a. WP27 (Requirements Management) summary.

Implementing the suggested solutions would result in a definitive source for clear, complete, and accurate requirements, supporting more efficient customer need fulfillment. It would also improve engineer work planning, consistency of requirements across the organization, and enhanced communication and sharing of these requirements within ESnet. Ultimately, it aims to create a more organized, transparent, and effective system for managing and utilizing requirements across the organization.

2.4.7.28. WP28 (Mission Support Management)

This work-package outlines the need for ESnet to develop a system to track and understand who they support and what their needs are.

WP28 (Mission Support Management)	
Opportunity	Unlike other DOE user facilities, ESnet lacks a mechanism to track the users it supports and their specific needs. The opportunity lies in developing a system to gather information about ESnet users, their projects, and their requirements, enabling ESnet to better serve science program data mobility needs.
Datasets	A census of scientific end-users and their activities, a method to link flow-IDs to science programs, and a system to extract and categorize information from the

	existing ticketing system by science program. These datasets would help provide a comprehensive understanding of user activities and needs.
Constraints	Potential data sensitivity issues, particularly regarding Personally Identifiable Information (PII) and the reluctance of some research programs to share detailed information. The solution must be flexible enough to accommodate incomplete or missing data, as complete information availability is unlikely.
Potential solutions	Leverage data from existing external transfer schedules, building an external user registry, implementing a CRM system to organize engagement data, and potentially utilizing an AI model to summarize information and answer queries. End-users would interact with the data through natural language processing, graphical displays, and API or search queries.
Gap analysis	There is a significant challenge for obtaining programmatic data associated with flows from external User Facilities, which will require collaboration with DOE Program Managers and other facility leadership. The remaining capabilities are generally considered off-the-shelf and primarily involve integrating data from internal ESnet sources.

Table 2.4.7.28.a. WP28 (Mission Support Management) summary.

Implementing the suggested solutions would result in a system that allows ESnet to track and understand its user base, their scientific programs, and their data mobility needs. This would enable Science Engagement to map ESnet's contribution to science, Network Engineering to better respond to requests and forecast needs, and ultimately, improve support for scientific endeavors.

2.4.7.29. WP29 (Dataset Unified Query)

This work-package outlines the need for a unified query interface to simplify access to various data sources within ESnet.

WP29 (Dataset Unified Query)	
Opportunity	There is a desire for a unified query interface to access various data sources within ESnet. The problem statement highlights the difficulty users face in retrieving information from multiple systems, and the opportunity lies in simplifying this process. A centralized system would allow network engineers, LLMs, and software engineers to access information efficiently, correlate data, and ensure data synchronization across systems.

Datasets	The required datasets already exist within ESnet. The solution focuses on building client libraries and a universal search index that can ingest high-level documents from all systems. This would provide a more accessible and streamlined method for retrieving data. The project also emphasizes the need for sufficient metadata to enable LLMs to understand and utilize the data sources effectively.
Constraints	Access control issues and data sensitivity. Any new system must have robust access control mechanisms to ensure users can only view information they are authorized to see. Additionally, the universal search index has the potential for broader access than individual end-users might typically have, necessitating careful consideration of data sensitivity. These constraints highlight the need for security and privacy measures.
Potential solutions	Develop client libraries for all ESnet APIs and then use these libraries to populate a single index with documents from all systems. MCP tools will be developed for both the client libraries and the search index to facilitate LLM interaction. These solutions also necessitate comprehensive metadata for each data source, including docstrings for MCP tools, to enable LLMs to infer the types of information available.
Gap analysis	Lack of sufficient examples for using ESnet APIs and the tools necessary for LLMs to access existing data. There is also a need for better API documentation and examples to inform the development of client libraries. While the data exists, it is not currently accessible to LLMs. Bridging this gap involves creating clear, consistent documentation and tools that are readily usable by both humans and LLMs.

Table 2.4.7.29.a. WP29 (Dataset Unified Query) summary.

If the suggested solutions are implemented, ESnet would have a unified query interface with accessible client libraries, a universal search index, and robust metadata. This would significantly simplify data access for network engineers, LLMs, and software engineers, enabling more efficient data retrieval, correlation, and synchronization. The resulting system would support improved decision-making, reduced errors, and greater understanding of data dependencies within ESnet.

2.4.8. Session 4 Summary

The 28 final work-packages primarily focus on improving information access, management, and search capabilities within ESnet. Key themes include document lifecycle management, with a focus on tracking, maintaining, and updating business documents to avoid outdated decision-making and compliance risks. Another major theme is the development of a unified document search system, addressing the current fragmentation of information across various platforms. This includes exploring

AI/ML solutions to create a global search interface capable of handling different data formats and access controls, enhancing collaboration and decision-making. Finally, there's a significant emphasis on the need for a system to query all data, enabling ESnet engineers to efficiently retrieve and correlate information from diverse sources through a single interface, utilizing Retrieval-Augmented Generation (RAG) and addressing security and access control constraints.

2.5. Session 5: Where Do We Go from Here? A Crosscut Analysis

Session 5 served as a capstone of the ESnet Data and AI Workshop, distilling the main themes and outcomes from the earlier sessions and providing a comprehensive overview of the event. This session involved a sweeping cross-cut analysis of the work-packages produced in Session 4, with the objective of pinpointing the most influential bodies of work that cut across multiple work-packages and providing actionable insights to inform ESnet's prioritization and resource allocation strategies. The cross-cut analysis were done on the four areas identified below:

1. **Data Management**: This refers to the end-to-end management of data, covering data creation, collection, processing, storage, retrieval, sharing, and preservation
2. **Traditional analysis**: This refers to widely accepted and established methods of data analysis, including statistical analysis, data modeling, and data interpretation
3. **AI methods**; This includes advanced methods such as NLP, ML, and LLMs to analyze complex data, generate predictive models, and to drive insights and decision-making.
4. **User eXperience**: This focuses on creating an intuitive, efficient, and engaging interaction between users and systems, encompassing aspects like interactivity, visual appeal, and overall usability.

The results of the cross-cut analysis are detailed in Section 3.

3. Crosscut analysis

3.1. Data Management

This section examines the data-related findings and recommendations that emerged as either opportunities or impediments to applying advanced analytics techniques in network engineering and operations.

The success of AI/ML-driven initiatives within ESnet is dependent on the availability of structured, harmonized, and trustworthy data (Afzal et al., 2021; Kidwai-Khan et al., 2024; Liang et al., 2022). Data unification and normalization are a requirement for a foundational layer, without which the rest of the AI/ML stack cannot operate reliably. This underscores the critical importance of integrating and harmonizing data from various sources across ESnet's infrastructure, which includes data from monitoring systems, logging platforms, email notifications, service agreements, and procurement

contracts. AI plays a crucial role in this unification process by providing tools for data cleaning, data transformation, and schema mapping. Furthermore, AI/ML techniques could enhance querying capabilities, enabling efficient access and analysis of unified data. A well-integrated and trustworthy data framework is a fundamental prerequisite for the successful deployment and effectiveness of all other AI applications, as accurate and consistent data is essential for training robust AI models and generating reliable insights.

To fully utilize AI capabilities within ESnet, data unification and normalization are a strategic need and requirement for an AI component. This includes the need to build a standardized data flow (ingest → clean → synthesize → index), enforce metadata format and standard, expand log coverage, and create a unified registry. Without these elements, AI efforts within ESnet are vulnerable to falling into the classic trap of "garbage in, garbage out." Additionally, AI can be used to improve data readiness, such as identifying anomalies, metadata information gaps, or automatically suggesting normalisation rules, as identified in WP19 and WP25 (see Section 3.3.6).

AI can be also leveraged to assist in preparing data for AI. For example, WP19 (AI Sandbox) would provide a safe environment for experimenting with data cleaning, synthesis, reduction and model testing with ESnet datasets. As also highlighted in WP25 (NLP Interfaces), it could help to abstract complex configuration systems information and guide to a consistent and normalized data representation. All of this needs to be done with security and policy constraints in mind and respecting data access control, permissions, and data privacy.

3.1.1. Data Access

ESnet has an immense and diverse set of data that is, at times, hard to find and comprehend. As identified in Section 2.2, ESnet has more than 160 data sources in over 30 categories; in some cases, the repositories have billions of records.

Accessing and understanding individual data sources can be a significant challenge for humans, even with highly accurate, well-structured, and well-documented data. Each data source may have its own unique access methods, data structure, and analysis tools that can change over time. To overcome this challenge, leveraging AI-assisted enterprise search capabilities could be beneficial for initiatives such as WP16 (Query All Data), WP21 (Unified Document Search), and WP29 (Dataset Unified Query). Furthermore, utilizing Natural Language Processing (NLP) for forming requests and summarizing data (Wilkinson et al.), could enhance capabilities in areas like WP10 (Correlate Alarms), WP22 (Automated ServiceNow Ticket Summarization), and WP25 (MLP Interfaces to Systems).

To address the searchability issue more fundamentally, WP15 (Consistent Data Management) and WP26 (Information Architecture) emphasize the need for a comprehensive information architecture and a data catalog. These elements are crucial for ensuring data consistency, discoverability, and accessibility, ultimately facilitating more efficient and effective data management.

Certain work-packages, including WP07 (Business Ops), WP27 (Requirements Management), and WP28 (Mission Support Management), require the integration of diverse datasets, such as technical and

business data, from multiple sources. This necessity highlights the importance of ensuring that data is accurate, consistently formatted, well-documented, and governed by suitable access controls. The importance of consistent access control across the facility is specifically called out in WP23 (Federated Authentication)

To address facility-wide access control, ESnet is advancing our Zero Trust architecture, implementing fine-grained, non-location-based access control for all data services. Furthermore, to maximize data accessibility and usability across the organization, ESnet should continue maturing our data systems, enabling API-driven access to well-structured data that spans all services.

Findings:

DM.F1	ESnet has an immense and diverse set of data, comprising over 160 sources across 30 categories, with some repositories containing billions of records. However, even high-quality data can be challenging for humans to access and understand due to the diversity of access methods, data structures, and analysis tools used across different sources, which can change over time.

Recommendations:

DM.R1	ESnet should improve data discovery, comprehension, and confidence using Natural Language Processing (NLP) and enterprise search, and explore new ways to gain insight from existing data by augmenting it with improved metadata and using Artificial Intelligence (AI) and Machine Learning (ML)-based analysis techniques that are aware of network and service topologies.
DM.R2	To support a zero-trust architecture and operational innovation, ESnet should establish a consistent information architecture, providing Application Programmable Interface (API)-driven access to well-structured data, and ensuring uniform access control across all data sources. This includes creating a comprehensive data catalog, facilitating incident response through consistent data access, and identifying and addressing gaps in existing data sources to enable increased automation and analysis.

3.1.2. Data Curation

ESnet's underlying datasets exhibit varying levels of curation, documentation, normalization, metadata structure, and access control, which creates challenges for large-scale programmatic

analysis, particularly when working with multiple data sources. Several use cases necessitate the combination of two or more data sources for analysis, highlighting the need for coordinated metadata across data sources. To enable data fusion and cross-data source analysis, it is crucial that data sources have identical metadata structures, allowing for seamless programmatic integration of related data elements. Currently, many techniques are designed for single data sources and do not fully leverage the benefits of fused data, presenting an opportunity for significant improvement in areas such as automated incident response and failure modeling.

ESnet has a wealth of human-generated content in the form of documentation and incident response tickets. However, this data tends to be semi-structured, making programmatic use difficult without using AI. In cases like WP10 (Correlate Alarms), some incident response data is expressed as free-form notes in a ticketing system. This type of data is often written by the engineer as notes for themselves without the expectation of later data mining. To address this, an effort needs to be undertaken to assess its quality and gaps. A cultural shift may be necessary to encourage individuals to invest effort in data entry, with the understanding that this investment will ultimately enhance the capabilities of automated agents to provide effective assistance in the future.

Findings:

DM.F2 Underlying datasets have varying levels of curation, documentation, metadata structure, and access control methodology. These differences need to be addressed to facilitate larger-scale programmatic analysis, especially those employing multiple data sources. In some cases, existing data may also need to be enhanced to facilitate automated analysis. Existing initiatives such as the push towards zero-trust architectures within ESnet, are working to address these differences.

DM.F3 ESnet possesses a significant amount of human-generated content, such as documentation and incident response tickets, but its semi-structured nature makes it challenging to utilize programmatically without AI. Automating the analysis of this data creates a positive feedback loop, where the more accurate and thorough the human-entered data, the more effective automated agents will be in providing assistance in the future.

Recommendations:

DM.R3 ESnet should develop a comprehensive and integrated view of production services by providing seamless access to both technical and business data, and establishing a detailed operational service model that includes data retention and lifecycle management.

DM.R4	ESnet should develop a standardized data flow to unify and normalize data from various operational domains, including telemetry, system logs, and network configuration, to make data "ready for AI". This data flow should include automated checks for data consistency, schema format enforcement, and metadata records to ensure high-quality data for AI analysis, and will require human-led control and data engineering investments to support effective AI use within ESnet.

3.1.3. Data Completeness

Challenging operations and engineering issues often require understanding the end-to-end path across all cyberinfrastructure used within a science pipeline. Today, this is a semi-manual process that impedes progress. In the most basic scenario, an end-to-end path would involve three organizations: Site A, ESnet, and Site Z. Commonly, the end-to-end path involves five or more organizations. As the number of external organizations increases, the amount of manual coordination increases proportionally. A key finding is that in order to provide optimal services, we need to be aware of the health and performance of the end-to-end cyber-infrastructure. Such awareness requires coordination and data sharing with each of the other organizations that make up the end-to-end paths. Such coordination is why ESnet has long championed end-to-end visibility with its role in the perfSONAR and MetrANOVA projects. In order to move from semi-automated to fully automated end-to-end awareness, federated techniques to programmatically access operational state within the entire community are needed.

Findings:

DM.F4	It is essential to have visibility into the health and performance of the entire end-to-end cyberinfrastructure. Understanding the end-to-end path of cyberinfrastructure used in science pipelines is crucial for resolving complex operations and engineering issues, but currently involves time-consuming semi-manual processes.

Recommendations:

DM.R5	To enhance operational efficiency, ESnet should focus on bridging key information gaps, including: tracking underutilized allocated resources, such as unused bandwidth reservations; gathering essential data to facilitate improved automation and analysis; and establishing a unified security analysis framework that integrates with all data sources and adheres to industry best practices.
DM.R6	To deliver optimal services, it is necessary to monitor the health and status of end-to-end cyberinfrastructure beyond ESnet's facility boundaries. ESnet should collaborate with the

R&E networking community to develop end-to-end awareness through trusted and federated authentication, and correlation of identifiers across different domains.

The provided Data Management Findings and Recommendations can be thought of as a two-part concept: how to make data ready for AI, and how to use AI to make the most use of the available data. AI alone cannot resolve this without human-led control and data engineering investments.

3.1.4. Data Management Conclusions

ESnet faces challenges in accessing and utilizing its vast dataset due to varying levels of curation and access control. To overcome these challenges, ESnet should establish a consistent information architecture, develop a comprehensive data catalog, and create a standardized data flow to unify and normalize data. Leveraging AI/ML-based analysis techniques can also be used to improve data discovery, comprehension, and confidence, ultimately supporting a zero-trust architecture and operational innovation.

3.2. Traditional Analytical Methods

The goal of the statistical analysis cross-cut during Session 5 was to review all defined work-packages and evaluate where, how, and to what extent traditional analysis methods apply. These methods include, in the broad sense, traditional statistical modeling, time-series analysis, correlation, trend detection, optimization, rule-based methods, formal validation techniques, and related approaches.

3.2.1. Methodology

Each work-package was reviewed using a consistent template, addressing:

- **Applicability**: Is statistical analysis relevant or not for the specific work-package?
- **Nature of Data**: Types of data involved and their statistical properties.
- **Data Structure**: Time-series, cross-sectional, or other forms and implications for analysis.
- **Forecasting**: Is prediction or forecasting appropriate, and if so, what approaches might be best?
- **Interdependencies**: Do multivariate or complex relationships exist (requiring correlation, principal component analysis, etc.)?
- **Anticipated Challenges**: Pitfalls such as missing data, high dimensionality, or data integration issues.

Table 3.2.1.a provides both a high-level synthesis of our findings and actionable recommendations for next steps The cross-cut analysis revealed that traditional analysis methods were only applicable to about 9 of the 28 final work-packages.

	Nature of Data	Complexity, Dependencies	Challenges
WP01 (Alerting)	Multiple sources (alerts, logs, perf. metrics)	High; time, place, cross-system	Missing data; diverse requirements
WP02 (Rules Correlation)	Firewall rules, routing configs, topology	Very high; need normalization	High dimensionality; missing data
WP04 (Lifecycle)	Cross-sectional business/contract data	Low; univariate summary sufficient	Missing/inaccurate metadata
WP06 (Bandwidth Guarantees)	Time-series, bandwidth usage, utilization	High; multi-dataset correlation	Documentation/ definition gaps
WP09 (Ticket resolution)	Freeform text, multi-system	Cross-system, unstructured	NLP needed; messy data
WP10 (Correlate alarms)	Alarm data, topology, inventory	Cross-alarm dependency	Data completeness/ modeling challenges
WP11 (Predict hardware failures)	Performance metrics, logs, telemetry	Predictive modeling; time-series	Data for rare events
WP12 (Detect external configuration anomalies)	Routing, flow, SNMP, orchestration	Possible; type unclear	Data integration; need for clarity
WP26 (Information Architecture)	Dependency graphs, data flow metadata	Highly interconnected; forecasting value	Requires clearly defined metadata

Table 3.2.1.a. Summary of applicability of traditional analytical methods to the work-packages.

3.2.2. Representative Methods and Approaches

The traditional analysis methods discussed the most often in the work-packages were as follows:

- **Time-Series Analysis**: For operational monitoring and prediction (e.g., ARIMA, trend decomposition, anomaly detection, thresholding).
- **Regression/Correlation**: For uncovering relationships between service, usage, and outcome metrics (e.g., capacity/usage analysis, reliability compliance, and understanding cross-metric dependencies)

- **Univariate/Multivariate Summary Statistics**: For alert scoring, root cause correlation, and identifying system bottlenecks.
- **Graph Analysis**: For modeling interdependencies, data architecture, and service dependency mapping.
- **Clustering/Dimensionality Reduction**: Where appropriate for massive or high-dimensional datasets.

3.2.3. Cross-Cutting Themes and Insights

3.2.3.1. Areas Well-Suited to Traditional Analysis

Below is a summary of cross-cut traditional analysis themes organized by recurring use case:

Operational Monitoring and Anomaly Detection (WP01, WP06, WP11):

- Time-series methods (univariate and multivariate) are a natural fit for analyzing network performance metrics, host health, bandwidth utilization, and optical telemetry. Techniques such as ARIMA, anomaly detection, and historical trend analysis surfaced repeatedly as key tools.

Support for Root Cause and Correlation Analysis (WP02, WP09, WP10):

- Many operational issues (e.g., service incidents, alarm/alert correlation, firewall debugging) require analyzing data dependencies. Cross-correlation, regression analysis, and clustering were seen as necessary for tying together network, service, and configuration data.
- Temporal and network correlation, often in conjunction with NLP, can help identify recurring problem/solution patterns.

Graph/Dependency Modeling (WP26):

- Graph analysis (centrality, cycle detection, dependency flows) can map and optimize operational/data dependencies across systems.

Service Compliance and Bandwidth Tracking (WP06):

- Descriptive statistics and correlation methods enable understanding of bandwidth use, service reliability and capacity planning.

3.2.3.2. Problems Requiring Mixed or Non-Traditional Approaches

A substantial set of work-packages are currently best addressed through data management, normalization, integration, API development, or the application of NLP/LLM tools—statistical analysis is either not directly applicable or premature (e.g., until data normalization and access are improved). Examples include:

Data Catalogs and Unified Search (WP05, WP16, WP21, WP29)

- These work-packages are less about analyzing data and more about making data findable, understandable, and actionable (Wilkinson et al., 2016). They require a mix of data engineering, information science, natural language processing, and organizational change—often preceding and enabling more traditional forms of statistical analysis.

Business Data Lifecycle Management (WP07, WP14, WP20)

- These work-packages require a mix of text analytics, workflow automation, human-centered review processes, and robust information management. Traditional statistical methods play a limited or supporting role. The focus is on organizing, tracking, validating, and automating document handling—making these areas inherently suited for mixed or non-traditional analytic approaches, such as NLP, RPA (Robotic Process Automation), and human-machine collaboration.

Configuration and Change Intent Capture (WP13, WP17, WP18)

- These work-packages require a combination of software engineering, workflow/process automation, metadata/schema design, and natural language processing. The goal is to make implicit human reasoning explicit and machinable, not to quantitatively analyze large datasets. As such, these problems require mixed or non-traditional analytics—bringing together data engineering, NLP, and DevOps—not classical statistical approaches.

In addition, the line between traditional statistics and machine learning is blurry (e.g., statistical learning, feature extraction for AI/ML pipelines, graph analysis for service dependencies). Some tasks, such as ticket summarization or root cause parsing, combine statistical and NLP needs, especially for handling unstructured data. Many of these tasks would require extensive validation that might require formal methods even though not explicitly mentioned (Bolton et al., 2013; Sinha et al., 2019; ter Beek et al., 2018) Additionally, several work-packages mention configuration building and document generation that might make use of template-based methods or rule-based methods, which are among well-known traditional analytics approaches.

3.2.3.3. Anticipated Challenges and Barriers

Data Gaps & Readiness: Many promising analytical approaches are impeded by poor or fragmented underlying data—missing fields, inconsistent labeling, or lack of meta-data.

Scale & High-Dimensionality: Where analysis is appropriate, dimensionality reduction, prioritization of features, and scalable infrastructure are required.

Prediction for Rare Events: For use cases like hardware failure, statistical prediction requires either large historical datasets or creative approaches to very sparse events.

Interdependency and Complexity: Correlation of multiple, cross-platform data streams (alerts, logs, traffic, configs) often requires unifying formats and careful cross-referencing.

Human Factors: Ensuring output is actionable and usable for operations, and reflecting the limits of statistical confidence or uncertainty.

Organizational Silos and Access: Data and knowledge remain siloed, hence standardization and cross-team cooperation are prerequisites for meaningful statistical analysis. Clear documentation and ownership of datasets is necessary, especially for newer or more complex metrics.

3.2.3. Traditional Analytical Methods Conclusions

The traditional analysis cross-cut group found that robust, meaningful analysis is feasible and high-value for a subset of ESnet's work-packages—chiefly those concerned with monitoring, operational automation, and proactive management. However, much foundational work on data normalization and integration remains to be done. Strategic investment in these areas will pay dividends both for statistical and AI-driven approaches in coming years, making ESnet's operations more efficient, resilient, and data-informed.

Findings:

TAM.F1	Statistical analysis plays a vital role in certain work-packages, particularly those that involve network performance monitoring, alerting, bandwidth and SLA compliance, and predictive hardware failure. These areas can greatly benefit from the application of statistical methods, including time-series analysis, anomaly detection, and regression, to extract insights and inform decision-making.
TAM.F2	Not all work-packages lend themselves to statistical analysis. Those primarily focused on data management, normalization, or integration, such as data cataloging, documentation search, and configuration intent capture, typically require that data be cleanly structured, accessible, and consistently curated before statistical methods can be effectively applied.
TAM.F3	A recurring challenge across nearly all work-packages that require statistical analysis is the issue of data quality and accessibility. Incomplete, inconsistent, or poorly documented data is a pervasive problem, often compounded by siloed data sources, lack of normalization, and missing metadata, which can significantly hinder the ability to perform advanced analysis.
TAM.F4	Many problems are characterized by complexity and a multivariate nature, requiring the integration of data from multiple systems. Use cases such as correlating alarms, root cause analysis, and policy conformance necessitate the application of advanced multivariate techniques, including correlation analysis, clustering, and dimensionality reduction methods like principal component analysis, to uncover meaningful insights and relationships.
TAM.F5	Certain work-packages require hybrid analysis approaches, combining classical statistical techniques with alternative methods such as NLP or AI/ML to effectively extract insights.

	This is particularly true for work-packages involving unstructured or free-form text data, such as tickets, emails, and changelogs, where traditional statistical methods may not be sufficient.
TAM.F6	The readiness and potential impact of applying statistical analysis vary significantly across different work-packages. As a result, prioritization is necessary to focus efforts on those work-packages that offer high operational value and have readily available, high-quality data, maximizing the potential return on investment and ensuring the most effective use of resources.

Recommendations:

TAM.R1	To maximize impact, ESnet should prioritize work-packages that present a clear opportunity for statistical analysis, focusing on areas where data quality is high and operational impact is significant. Specifically, initial efforts should be concentrated on network operations, capacity planning, and predictive maintenance, where the potential benefits of statistical analysis are likely to be most pronounced.
TAM.R2	Investing in data quality and integration is essential to unlocking the full potential of statistical analysis. Before applying advanced analytical methods, ESnet should improve data standardization, metadata enrichment, and automated collection processes. Statistical analysis should be built on a foundation of clean, reliable data, rather than attempting to apply analytical techniques to subpar data, which can lead to inaccurate or misleading results.
TAM.R3	To maximize analytical effectiveness, ESnet should leverage hybrid approaches that combine the strengths of statistical methods with those of NLP and AI/ML. By applying traditional statistical techniques to structured data and utilizing NLP for free-form or unstructured data, teams can unlock a more comprehensive understanding of their data and drive more informed decision-making.
TAM.R4	To ensure the reliability and accuracy of statistical analysis and lay the groundwork for future AI/ML initiatives, ESnet should promote data stewardship and ownership across the organization. This involves assigning clear responsibility for maintaining data integrity, establishing normalization protocols, managing data life cycles, and setting metadata standards for key systems. Additionally, reinforcing clarity around data custodianship will help to ensure that data is properly managed, maintained, and utilized, ultimately supporting informed decision-making and driving business value.
TAM.R5	Effective statistical analysis requires collaboration and iteration with stakeholders. To ensure that insights are relevant and useful, ESnet should validate approaches with

operational teams, including engineers, operators, and business staff. This involves working closely with stakeholders to ensure that statistical insights are actionable, understandable, and align with business objectives, ultimately driving meaningful outcomes and informed decision-making.

3.3. AI Methods

A preliminary examination of the 28 final work-packages revealed that only 12 of them mentioned AI-related components. The subsequent cross-cut analysis of the 12 work-packages identified 6 key thematic elements pertinent to AI methods and techniques that would benefit ESnet operations; (1) Anomaly Detection, (2) Automating Workflows and Processes, (3) NLP, (4) Predictive Modeling, (5) Root Cause Analysis, and (6) Data Unification (see Table 3.3.a). These themes are not isolated but rather interconnected (see also Table 2.4.7.a), creating many opportunities for AI-driven enhancements across various work-packages.

	Anomaly Detection	Automating Workflows & Processes	Natural Language Processing	Predictive Modeling	Root Cause Analysis	Data Unification
WP01 (Alerting)	X	X		X	X	X
WP02 (Rules Correlation)	X		X	X		X
WP03 (Data quality)	X	X	X		X	X
WP06 (Bandwidth Guarantees)	X	X		X		X
WP08 (Outage Notification Parsing)	X	X	X		X	
WP09 (Ticket resolution)	X	X	X		X	
WP10 (Correlate alarms)				X	X	X
WP11 (Predict hardware failures)	X	X	X	X		X
WP12 (Detect external configuration anomalies)	X	X	X	X		X
WP17 (Automating Site Deployment)		X				
WP20 (RFP Contract Builder)	X		X			
WP22 (Automated ServiceNow Ticket Summarization)	X	X	X	X	X	

Table 3.3.a. Common AI Method themes across the work-packages.

Looking deeper into each of these six AI topics provides a more granular examination of their potential applications within the context of ESnet's specific operational challenges. Sections 3.3.1 through 3.3.5 identify concrete ESnet operations that stand to gain significant benefits from the application of established AI techniques. These sections are produced by a combination of ESnet engineers with the help of several academic researchers. This combined set of expertise ensures that the identified AI topics are not only technically feasible but also directly address critical operational necessities, reflecting a pragmatic and results-oriented approach to AI adoption.

The sixth topic in Table 3.3.a describes the need to prepare data for AI and analytical applications, a core theme for Section 3.1. Section 3.3.6 explores popular AI-driven approaches for data management, including data integration, annotation, querying, and quality assurance. Developing high-quality, well-documented, and easily accessible analysis-ready data is critical to improving network observability, enhancing resilience, and optimizing network services. Beyond ESnet's internal operational benefits, establishing a validated and reliable data infrastructure would position ESnet to expand our service portfolio. This capability could directly support broader research initiatives, such as the DOE's FASST program, by providing a robust foundation for advanced analytics and decision-making.

The following assessment of potential AI method applications reflects the workshop participants' current understanding of their technical capabilities and potential utility. These methods are undergoing rapid evolution, and many may necessitate substantial engineering efforts to adapt them for operational deployment within ESnet's environment. Additionally, several of these techniques are being incorporated into commercial products and platforms, which may become viable options by the time implementation is planned.

To ensure alignment with ESnet's requirements, the workshop organizers intend to develop a formal implementation strategy over the next few months. This strategy will prioritize the evaluation of tool and algorithm maturity, including factors such as scalability, reliability, and compatibility with ESnet's infrastructure, as well as applicable rules and regulations about security, privacy, and confidentiality. Rigorous due diligence will be conducted to balance innovation with operational feasibility before proceeding with any deployment.

3.3.1. Anomaly Detection

Data has become an increasingly valuable resource, often likened to "the new oil" for its potential to drive insights and innovation. ESnet's substantial collection of diverse data types includes network telemetry data, giving insights into network performance and behavior; system metrics that provide insights about the health of hardware and software infrastructure; detailed logging from applications and devices that captures context of operational events; incident tickets for specific network events or failures; and data related to security from firewalls, ACLs, and more. ESnet also has a wide array of unstructured data related to business, such as contracts and Service Level Agreements (SLAs), as well as technical and project documentation in wikis, Google Suite documents, Jira tickets, etc.

Anomaly Detection plays a crucial role in ensuring the reliability and stability of ESnet's infrastructure (Fernandes et al., 2019; S. Wang et al., 2021). By leveraging AI algorithms, ESnet could proactively identify deviations from normal network behavior, enabling early intervention and preventing potential outages or performance degradation. This capability is explicitly mentioned in work-packages such as WP01 (Alerting), which could be significantly enhanced by intelligent alerting systems that filter noise and prioritize critical anomalies. Similarly, WP12 (Detecting External Configuration Anomalies) could benefit from AI-powered tools that continuously monitor and validate external configurations, ensuring adherence to security policies and operational best practices. The effectiveness of anomaly detection is intrinsically linked to data quality, as highlighted by WP03 (Data Quality), underscoring the need for accurate and reliable data as a foundation for robust AI applications.

However, due to the extensive amount of data, manual visualization and analysis by humans is impossible, necessitating automated intelligent systems. Finding trends and anomalies in the data naturally lends itself to the application of machine learning and artificial intelligence.

Findings:

AIM.F1	The three types of anomaly detection tasks pertinent to ESnet are: (1) trend analysis for outlier identification, (2) anomalies in logic or rules, and (3) analyzing semi/unstructured data to detect anomalies.

3.3.1.1. Trend Analysis and Outlier Detection

For the telemetry, system metrics and other timeseries data, ESnet can analyze patterns and establish baseline trends using classical statistical methods like moving averages, exponential smoothing, Autoregressive Integrated Moving Average Models (ARIMA) or Seasonal Autoregressive Integrated Moving Average Models (SARIMA) (Alimohammadi & Nancy Chen, 2022; Blázquez-García et al., 2022). These statistical methods can be combined with supervised machine learning like SVR or neural networks to build hybrid models. We could also analyze the time series data using Long Term-Short Term Autoencoders and use them to detect anomalies (Abdallah et al., 2021; Lindemann et al., 2021).

If the results of these analyses are made available in a common repository, it can be integrated in various applications or can be used to build more complex intelligent systems. Establishing baseline trends and detecting anomalies are needed in WP01 (Alerting), WP06 (Bandwidth Guarantees), and WP11 (Predict hardware failures), and for metric analysis in WP12 (Detect external configuration anomalies) to establish base patterns for the various metrics.

Trend analysis and outlier detection solutions are the foundational blocks that will enable ESnet to build intelligent monitoring and alerting solutions, check for policy or SLA compliance or serve as input for prediction models.

3.3.1.2. Rule Inference and Exception Identification

Detecting anomalies in configurations or rules will help avoid connectivity outage due to human error as well as save manual labor involved in troubleshooting such issues.

This will require correlating data across firewall rules (e.g., iptables, ACLs, Panorama), route tables (e.g., IS-IS, BGP, static routes), host configurations (e.g., IP addresses, interfaces, host-based routing tables for management), traffic logs (e.g., Flowdata, packet captures, Zeek conn logs, high touch), network topology (e.g., device interconnections, subnets), and blackhole routes (e.g., SCRAM) along with a well-structured representation of hierarchical topology and service interdependency. ESnet will also need labeled incident data that can be correlated with the configuration changes. While some of the data exists today, there is significant effort needed to create a normalized well-structured topology and service interdependency representation.

This type of correlation and configuration validator would be useful in WP02 (Rules correlation) and WP12 (Detect external configuration anomalies).

3.3.1.3. Analyzing semi/unstructured data to detect anomalies

Analyzing semi/unstructured data like logs and incident tickets to detect errors, failures or other events is another class of anomaly detection. This involves text processing which is covered extensively in Section 3.3.3 on Natural Language Processing, and correlating it with time series data and if applicable, with configuration data. This may require hybrid techniques like using NLP techniques or multi-modal models to analyze the unstructured or semistructured data like logs along with time-series analysis on relevant metrics that was discussed previously.

Recommendations:

AIM.R1	For anomaly detection, ESnet should invest in time series analysis of the various metrics, identify patterns in the data, and make these insights available across ESnet through dashboards as well as programmable interfaces. This analysis is fundamental to any automated intelligent monitoring or validation systems that we want to build, and is essential for improving observability, detecting failures, creating AI troubleshooting assistants, checking for SLA compliance, and building some predictive models.

3.3.2. Automating Workflows and Processes

Automating Workflows and Processes represents a significant opportunity to improve operational efficiency and reduce manual effort within ESnet (Adekunle et al., 2021); (Rafique & Velasco, 2018; M. Wang et al., March-April 2018). AI-powered automation could streamline repetitive tasks, optimize resource allocation, and accelerate response times. Examples of this include WP17 (Automating Site Deployment), where AI could orchestrate the complex steps involved in deploying new network sites,

minimizing human error and accelerating deployment timelines. Furthermore, WP20 (Generating RFP contracts) demonstrates the potential of AI to automate document generation, freeing up valuable human resources for more strategic activities. The intersection of automation with other AI themes, such as anomaly detection and predictive modeling, could lead to self-healing network capabilities and proactive issue resolution. Based on the requirements identified in these work-packages and the current available technology, three categories of approaches appear suitable for ESnet operations:

Findings:

AIM.F2	The three common automation workflow approaches relevant to ESnet are: (1) central workflow engine, (2) modular AI agents, and (3) action-oriented automation.

3.3.2.1. Central Workflow Engine

A centralized workflow engine (e.g., LangGraph (J. Wang & Duan, 2024)) paired with specialized AI agents could revolutionize ESnet's engineering operations by automating complex workflows across work-packages like WP01 (Alerting), WP06 (Bandwidth Guarantees), WP08 (Outage Notification Parsing), WP09 (Ticket Resolution), and WP12 (External Configuration Anomalies). This architecture could enable end-to-end automation: AI agents handle task-specific analyses (e.g., anomaly detection for alerts, NLP-based outage parsing, or configuration validation), while the engine orchestrates data flow, coordinates actions, and ensures consistency. A centralized approach is likely to be able to concentrate computing resources to provide (near-)real-time responsiveness for rapid detection of issues (e.g., hardware failures, SLA breaches) and immediate action (e.g., auto-remediation scripts, ticket updates). It would also be more future-proof by easily incorporating new AI tools or workflows as they emerge, while maintaining alignment with ESnet's policies and infrastructure. This approach is best-suited for ESnet's high-velocity, complex data environment.

Network equipment makers are also actively considering adding automation features to their management systems (Kalpage, n.d.; LangChain, 2025).

3.3.2.2. Modular AI Agents

The second approach, Modular AI Agents, involves an automated workflow that may consist of several smaller AI agents or workflows that do one specific task and excel at that task (Quarantiello et al., 2024; Sapkota et al., 2025). Then these agents can be strung together to create a workflow that accomplishes a major task (Zhang et al., 2024). Empowering each group within ESnet with lightweight AI agents, to which the group can assign tasks and string multiple agents, may allow larger goals to be accomplished more easily. For example, ESnet might have an Anomaly Detection Agent to monitor systems for anomalies (e.g., hardware failures, network outages) and a Predictive Analysis Agent to forecast hardware failures or bandwidth demand using historical data.

3.3.2.3. Action-Oriented Automation

Action-oriented automation combines software-driven workflows with human judgment or external system interactions to manage complex tasks requiring non-automated inputs, such as validating security alerts (e.g., WP01 (Alerting)) or coordinating physical actions like hardware shipping (e.g., WP17 (Automating Site Deployment)). A simple example of this approach may include a centralized orchestration engine (e.g., LangGraph) that manages automated steps (e.g., anomaly detection, logistics tracking) alongside human-in-the-loop (HITL) decision points, where experts review alerts, approve actions, or correct AI suggestions (Kumar et al., 2024; Mosqueira-Rey et al., 2023; Wu et al., 2022; Zhang et al., 2024). These systems integrate external tools (e.g., ticketing platforms, logistics APIs) and provide decision support (e.g., context summaries, recommendations) to balance efficiency and accuracy. Challenges include managing latency in critical paths, ensuring audit trails, and scaling HITL steps, while benefits include reduced manual effort, improved decision-making through AI insights, and adaptability to evolving tools and policies. This hybrid approach enables ESnet to handle high-stakes workflows with precision, leveraging automation for routine tasks while preserving human oversight for complex or risky decisions.

Recommendations:

AIM.R2	While automation can significantly amplify efficiency and productivity, it is crucial to identify and prioritize areas where automation can yield the greatest return on investment. ESnet should examine three approaches for automating workflows and processes: (1) using central workflow engine like LangGraph to orchestrate tasks such as alerting, (2) employing modular AI agents to expand the automation to distributed workflows, and (3) exploring action-oriented automation for human-in-the-loop processes such as site deployment and contract generation, preserving human oversight for complex or risky decisions.

3.3.3. Natural Language Processing

NLP offers powerful tools for extracting valuable insights from unstructured textual data, which is abundant in network operations. WP08 (Parsing Outage Notifications) exemplifies how NLP could automatically analyze outage reports, identify key information, and trigger appropriate remediation workflows. Similarly, WP22 (Summarizing Tickets) could leverage NLP to condense lengthy support tickets into concise summaries, enabling faster understanding and resolution of issues. The ability of NLP to process and understand human language opens up new avenues for intelligent automation and improved human-machine interaction within ESnet's operational environment.

ESnet’s needs for NLP encompass three main functional activities:

- Upstream data preparation as part of supporting other analytics, or the use of NLP to extract meaning, enhance data tagging, or improve subsequent data query and analytics [WP02 (Rules Correlation), WP03 (Data Quality)]
- Human decision support, or the use of NLP to streamline, summarize, or automatically translate from different data sources as part of supporting human decisionmaking [WP08 (Outage Notification Parsing), WP11 (Predict Hardware Failures), WP22 (Automated ServiceNow Ticket Summarization)]
- Document generation support, or the use of NLP to facilitate human readable documents more quickly and easily [WP09 (Tick Resolution), WP20 (RFP/Contract Builder)]

Findings:

AIM.F3	The three activities necessary to support ESnet's need for Natural Language Processing are: (1) upstream data preparation as part of supporting other analytics, (2) human decision support, and (3) document generation support.

3.3.3.1. Upstream NLP Needs

ESnet operates via a wide set of systems and databases; currently the organization relies on human skill to synthesize and understand operational states and events. A variety of work-packages propose to automate parts of this situational understanding through the use of anomaly detection, automated root cause analysis, or improved query, etc. These work-packages call for improved NLP capabilities to support cross-domain data normalization and/or data verification and auditing. This NLP use is primarily to support other kinds of AI, which will occur after disparate data sources are automatically interpreted and adapted for use as part of a domain-common schema.

NLP research in the areas of generative deep learning and multi-modal learning seem most applicable to support this set of ESnet NLP needs, and an extensive amount of existing open source and commercially available tools exist in these areas.

3.3.3.2. Human Decision Support

The operation of ESnet depends on being able to measure, monitor and interpret near-realtime data from a wide variety of sources including network flows, equipment status, architecture changes, customer tickets, etc. NLP would be called upon to support some degree of data extraction and synthesis from disparate sources, however these work-packages also call for use of predictive AI and NLP as part of supporting decision making by human operators in support of maintaining network operational performance (Z. Yang et al., 2019; Yin et al., 2024; Zhao et al., 2022).

NLP focus areas for this application include Human Centered NLP, including enhanced support for intelligent decision making, and improved abilities to understand AI NLP weights and neural network results. Advances in transfer learning, meta-learning, and data augmentation will also be useful since

ESnet's training corpus to support NLP pattern recognition accuracy on our complex system, will be limited. One model to study, particularly as a surrogate for the analytic functionality envisioned in portions of WP11 (Predict hardware failures) and WP12 (Detection external configuration anomalies) may be ORNL's ChatHPC conversational AI Assistant (Yin et al., 2024).

3.3.3.3. Document Generation Support

The use of NLP and generative AI to support document generation could also increase operational efficiency. This functional use of NLP would be intended to improve our ability to support interactions with human stakeholders, and better support documentation (Singh, 2024; Tateishi et al., 2019).

The use of NLP to support automated text or code generation using natural language expression of requirements is becoming increasingly common, and ESnet may be able to obtain these capabilities, using our corpus of tickets and contract documents, with relatively low need to support development or take on technology risk (Hassan et al., 2021; Priyadarshni, 2024). These work-packages may also lend themselves to cooperation with other User Facilities or LBNL-IT/Ops, since similar efforts (particularly on automating contracting) are underway.

Implementing NLP in ESnet requires careful consideration of domain-specific challenges and opportunities. Given the potential limitations of ESnet's training corpus, techniques like transfer learning and data augmentation are critical to build robust models, which can then be fine-tuned on pre-trained large language models (LLMs) to address data scarcity. NLP models must be deeply integrated with ESnet's unique lexicon, troubleshooting logic, and operational context (e.g., for work-packages like WP8 (Outage Notification Parsing), WP11 (Predict Hardware Failures), and WP22 (Automated ServiceNow Ticket Summarization)), ensuring they align with network-specific terminology and workflows. For critical decisions—such as validating outage fixes or hardware warnings—human oversight is essential to verify NLP outputs and reduce false positives, necessitating feedback loops where operator input refines models iteratively. Leveraging tools like ChatHPC-inspired assistants (e.g., ORNL's framework) and open-source/commercial solutions can accelerate implementation, while workflows must be tailored to ESnet's security requirements and operational patterns. Seamless integration with existing systems (e.g., ServiceNow for ticketing, Grafana for monitoring) is vital for workflow automation, particularly for standardizing language in tickets (WP9 (Ticket Resolution)) and contracts (WP20 (RFP/Contract Builder)), reducing manual effort. Collaboration with DOE facilities or LBNL-IT/Ops on shared needs—such as contract automation—can further reduce development costs and risks. By combining domain-specific customization, human-in-the-loop validation, and strategic partnerships, ESnet can deploy NLP solutions that enhance efficiency while maintaining reliability and compliance.

Recommendations:

AIM.R3	ESnet should explore NLP technologies with domain-specific fine-tuning for improving data integration, data accesses, decision support, and document generation. This effort could enhance network visibility by integrating data from various sources, including

alerts and trouble tickets, with network telemetry. This could also provide assistance and recommendation for generating new documents such as ticket resolution, network configuration, and service contracts.

3.3.4. Predictive Modeling

Predictive Modeling utilizes historical data and statistical techniques to forecast future events, enabling proactive decision-making and resource management. WP11 (Predicting Hardware Failures) highlights the critical role of predictive modeling in anticipating potential hardware failures, allowing for timely maintenance and minimizing network downtime. By analyzing patterns in hardware performance data, AI algorithms could identify components that are likely to fail, enabling proactive replacement and ensuring the continued availability of critical infrastructure. The insights derived from predictive modeling could also inform capacity planning and resource allocation strategies, optimizing network performance and efficiency.

This section outlines how predictive modeling can enhance ESnet's operations, ranging from proactive incident detection to bandwidth forecasting and automation of administrative processes. These efforts apply a variety of machine learning approaches including time-series forecasting, anomaly detection, supervised classification, and NLP, and support broader goals of increasing reliability, responsiveness, and efficiency.

A review of the work-packages have identified the following predictive modeling tasks needed to enhance ESnet's operations:

- **Operational Intelligence & Incident Management:** Predicting service degradation, correlating alarms, forecasting hardware failures, detecting external configuration anomalies, and summarizing incident tickets.
- **Network Performance & Resource Planning:** Ensuring SLA compliance by forecasting bandwidth usage trends and identifying future risks.
- **Administrative & Planning Automation:** Generating contract drafts and automating parts of the planning lifecycle using NLP.

Findings:

AIM.F4 The three areas in predictive modeling that are needed to support ESnet operations are: (1) operational intelligence & incident management, (2) network performance & resource planning, and (3) administrative & planning automation.

3.3.4.1. Operational Intelligence & Incident Management

These work-packages support proactive management of network issues by leveraging predictive analytics to reduce downtime, improve incident response, and anticipate faults (Gonzalez et al., 2017; Papageorgiou et al., 2022).

Together, they provide operational intelligence for ESnet. More specific tasks to provide this operational intelligence includes the following extracted from work-packages WP01 (Alerting), WP10 (Correlate Alarms), WP11 (Predict Hardware Failures), WP12 (Detect External Configuration Anomalies), and WP22 (Automated ServiceNow Ticket Summarization). Predictive modeling (e.g., LSTM, Prophet) forecasts service degradation using multivariate sensor data, enabling preemptive action. Anomaly detection (Isolation Forest (Al Farizi et al., 2021; H. Xu et al., 2023), autoencoders (Frehner et al., 2024)(Chalapathy & Chawla, 2019)) minimizes false positives by correlating alerts with dependency graphs, while NLP enriches alarm context from unstructured data. Graph-based models trace root causes in alarm floods, prioritizing critical incidents. Hardware failure predictions leverage telemetry data to schedule maintenance, reducing outages. External configuration anomalies are detected via LSTM analysis of BGP updates, isolating disruptions. Ticket summarization tools distill lengthy logs into concise summaries, streamlining incident resolution. This holistic approach balances automation with human oversight, ensuring efficient, proactive network operations.

3.3.4.2. Network Performance & Resource Planning

ESnet offers guaranteed bandwidth services, and predictive modeling can help ensure SLA compliance (WP06 (Guaranteed Bandwidth)) (Silva et al., 2021; Zhu et al., 2021). Better planning and compliance with performance guarantees can leverage predictive models to track trends and identify future risks. Time-series models like ARIMA or LSTM can be used to forecast usage trends across circuits, based on telemetry from OSCARS and other sources (Guok et al., 2006). For example, if traffic analysis shows consistent growth over weeks, the model can project a future violation of the bandwidth guarantee—triggering alerts or adjustment recommendations. Classification models can also assess the risk of SLA breaches based on usage patterns and routing behaviors. Regression models support performance evaluation and assist in planning future commitments.

3.3.4.3. Administrative & Planning Automation

Developing RFPs and contracts manually is time-consuming and error-prone [WP20 (RFP/Contract Builder)]. Predictive modeling using NLP can generate drafts based on prior RFPs, templated clauses, and historical contract outcomes. This can reduce cognitive load and improve consistency by automating parts of the contract and planning lifecycle (M. Wang et al., March-April 2018).

A language model trained on ESnet's corpus of RFPs and vendor responses can classify contract components, detect omissions, and suggest standard phrasing. For example, a user entering a query like "multi-site 400G optical transport system" can receive a pre-filled draft including technical requirements, compliance language, and evaluation criteria. Predictive scoring can highlight weak areas in the draft likely to reduce response quality or vendor clarity.

Recommendations:

AIM.R4	Managing the ESnet WAN is growing exponentially more complex, resulting in tasks that are increasingly time-intensive and impractical to manage manually. To ensure operational efficiency, ESnet should adopt predictive modeling to enhance incident management, optimize resource allocation, and automate processes through data-driven intelligence.

3.3.5. Root Cause Analysis

In a large-scale, mission-critical scientific network like ESnet, root cause analysis (RCA) is essential for maintaining service reliability, shortening incident response time, and institutionalizing operational knowledge (Gonzalez et al., 2017; Roy et al., 2024; Wankvar, 2024). However, traditional approaches to RCA—manual correlation, log inspection, and operator experience—are increasingly strained by the scale, complexity, and heterogeneity of ESnet's infrastructure. AI-powered tools could accelerate this process by analyzing vast amounts of operational data, identifying correlations, and suggesting potential root causes. By leveraging machine learning algorithms, ESnet could develop more effective strategies for identifying and resolving the fundamental issues that lead to network disruptions, improving overall network resilience and stability.

AI techniques such as Isolation Forest (D. Xu et al., 2017), LSTM Autoencoders (Lindemann et al., 2021), and graph-based neural networks (Yen et al., 2022) offer transformative potential for enhancing RCA across ESnet's operational workflows. These models enable proactive identification of "soft failures" (e.g., emerging anomalies in WP01 (Alerting) or WP03 (Data Quality) before they escalate, while graph-based methods clarify complex system dependencies (critical for WP10 Correlate Alarms), reducing guesswork in diagnosing issues. Large-language models (LLMs) further accelerate RCA by synthesizing historical knowledge to streamline ticket resolution (WP09 (Ticket Resolution)) and automate summarization of ServiceNow tickets (WP22 (Automated ServiceNow Ticket Summarization)), cutting time-to-resolution. However, realizing this potential requires aligning AI solutions with ESnet's specific needs: proactive failure detection for WP08 (Outage Notification Parsing), dependency mapping for WP10 (Alerting), and knowledge-driven automation for WP22 (Automated ServiceNow Ticket Summarization). By addressing these use cases, AI can bridge gaps in RCA workflows, but success hinges on tailoring tools to ESnet's operational challenges, such as data quality (WP03 (Data Quality)), alarm correlation complexity, and the need for human-augmented decision-making in high-stakes scenarios. This alignment ensures AI complements—not replaces—human expertise while advancing RCA efficiency and accuracy.

A number of AI tools for RCA are already available or under active development, for example, ZDX from Zscaler, Doctor Droid, BigPanda, and Logz. In academic research, works like Deep Network Analyzer (DNA) have also generated wide interest (K. Yang et al., 2017)(Banerjee et al., 2009). Others have

adopted multivariate anomaly detection techniques such as Isolation Forest and Prophet to monitor critical metrics (CPU, latency) and flag deviations early. Additional researchers have implemented variations of Graph Neural Networks for Topology Analysis, e.g., GraphSAGE to map relationships between alarms and devices, enabling faster identification of cascading failures.

Findings:

AIM.F5	AI tools have been effectively used for root cause analysis, particularly for improving early detection, increasing causal clarity, and integrating historical context.

Recommendations:

AIM.R5	Reducing the time to resolution (TTR) is an important operation goal for ESnet. ESnet could leverage AI tools and methods to automate and enhance root cause analysis, improving network resilience and stability by analyzing vast amounts of operational data, identifying correlations, and suggesting potential root causes.

3.3.6. Data Unification: Techniques to Improve Data Management

Since many of the work-packages touch on the issue of data management, this topic naturally appears in Table 3.3.a as well. Section 3.1 already covered the majority of the requirements and opportunities in data management, next we describe a handful of possible uses of AI methods for data integration, annotation, querying, and quality assurance.

3.3.6.1. Automating Data Integration and Preprocessing

AI/ML tools have been shown to automate the consolidation and preparation of raw data from diverse sources—such as network logs, sensor data, and control plane metrics (Althati et al., 2024; Tadi, 2021). Such automated processing is ensuring data is uniformly structured, cleaned, and enriched to support advanced analysis. This process begins with data standardization, where AI-driven systems align formats, units, and schemas across disparate datasets (Kozina, 2024; Sharma et al., 2023). For example, tools can automatically convert inconsistent timestamps into a unified time format or normalize numerical ranges (e.g., scaling metrics to a common scale) to eliminate discrepancies that arise from heterogeneous data sources. Data cleaning follows, leveraging ML algorithms to systematically remove duplicates, impute missing values using techniques like regression or transformer-based models (e.g., Chronos (Ansari et al., 2024)), and flag outliers that could skew analysis results. These steps ensure data integrity while reducing manual effort. Time synchronization is another important task, particularly for multimodal data streams, where AI models align temporally

disjointed datasets—such as correlating BGP updates with traffic logs—to create a cohesive timeline of events.

3.3.6.2. Automating Data Curation and Annotation

Implementing AI tools to automate data curation and metadata enrichment is essential for addressing manual maintenance challenges, such as inconsistent tagging, fragmented provenance tracking, and outdated entries. Key tasks include metadata tagging, where natural language processing (NLP) automatically generates descriptive tags for datasets (e.g., labeling files as "BGP update logs from 2023-Q4" to improve discoverability). Data lineage tracking ensures transparency by documenting the origin and transformation history of datasets (e.g., source systems, processing steps), while quality assurance leverages AI to flag inconsistencies (e.g., mismatched timestamps) and enforce validation rules, maintaining data reliability. Schema alignment uses graph-based models like knowledge graphs to map relationships between datasets and resolve semantic conflicts, ensuring compatibility across heterogeneous sources. For ESnet, this approach directly addresses the workshop's highlighted challenges of manual curation inefficiencies and abandoned efforts to maintain data accuracy (e.g., in Salesforce), enabling consistent, up-to-date datasets that support advanced analytics and decision-making.

3.3.6.3. Query Optimization

To improve querying efficiency across ESnet's diverse and high-velocity data sources (e.g., BGP updates, interface error logs, and unstructured text records), AI-driven systems can optimize storage and retrieval through smart indexing, dynamic resource allocation, and semantic search. Smart indexing automates the creation of indexes based on frequent query patterns (e.g., common time ranges or metrics), accelerating data access. Resource allocation employs reinforcement learning to dynamically adjust storage tiers (e.g., hot vs. cold storage) and computational resources, ensuring optimal performance while minimizing costs. Semantic search leverages large language models (LLMs) like Gemini or Glean to interpret natural language queries, enabling users to extract insights from unstructured data (e.g., free-text logs or incident reports) without requiring technical query syntax. Additionally, predictive caching and lossless compression prioritize frequently accessed data while reducing redundancy, further enhancing retrieval speed. For ESnet, these capabilities are critical for managing large-scale, high-velocity network data streams, ensuring efficient access to time-sensitive information like BGP updates or error logs. This streamlines troubleshooting, supports real-time decision-making, and aligns with ESnet's need to unify fragmented data sources into a cohesive, query-friendly ecosystem.

3.3.6.4. AI-Driven Data Quality Assurance

AI/ML tools could play a critical role in addressing data quality challenges (e.g., missing values, inconsistencies, noise, or inaccuracies) to ensure ESnet datasets are reliable and fit for analysis. Key applications include missing data imputation, where transformer-based models (e.g., Chronos (Ansari et al., 2024)) or generative adversarial networks (GANs) predict and fill gaps in time series or tabular

data (e.g., network performance metrics), while simpler datasets leverage statistical methods like k-NN imputation. Outlier and anomaly detection uses techniques such as autoencoders or isolation forests to identify anomalies in sensor readings, network metrics, or logs, flagging issues like unexpected BGP update patterns or interface errors. Data consistency checks employ graph-based models (e.g., knowledge graphs) to map relationships between datasets and resolve conflicts, ensuring alignment in records like BGP logs or error reports. For unstructured text data (e.g., ticket descriptions or free-text logs), NLP tools like large language models (LLMs, e.g., Gemini) standardize entries by correcting typos, normalizing terminology, or extracting structured data from unstructured sources. These capabilities directly address ESnet's challenges, such as filling gaps in network performance metrics to prevent skewed analyses, ensuring consistency in critical logs, and managing high-dimensional or unstructured data (e.g., inconsistent Salesforce ticket updates).

Findings:

AIM.F6	AI tools have shown great potential for data management, such as data integration, annotation, querying, and quality assurance.

Recommendations:

AIM.R6	To make data AI-ready, ESnet should leverage AI methods to address several data management requirements, including ensuring consistent data formatting, deploying efficient data access methods, and enhancing data quality assurance.

3.3.7. AI Methods Conclusions

ESnet can benefit from various AI applications, including anomaly detection, automation workflows, NLP, predictive modeling, root-cause analysis, and data management. For anomaly detection, ESnet should invest in time-series analysis and make insights available through dashboards and programmable interfaces. Central workflow engines, modular AI agents, and action-oriented automation can be leveraged for automation while preserving human oversight for complex or risky decisions. ESnet can also leverage NLP technologies with domain-specific fine-tuning to improve data integration, decision support, and document generation. The use of predictive modeling can augment incident management, optimize resource allocation, and automate processes, while AI tools can automate and enhance root-cause analysis and data management tasks, such as data integration, annotation, and quality assurance to improve network resilience, stability, and operational efficiency.

3.4. User eXperience

This section details the UX cross-cut analysis of various ESnet AI/ML work-packages, identifies best practices for data and workflow management, and offers UX guidelines for ensuring a positive end-user experience within the ESnet ecosystem. Its primary goal is to guide the development of AI/ML-powered tools and interfaces that are transparent, trustworthy, efficient, and human-centered. It addresses both backend AI workflow design and frontend end-user interactions, with a focus on ESnet's unique operational context.

This section provides key recommendations for designing user experiences (UX) in AI and machine learning (ML) workflows within the ESnet ecosystem, focusing on human-in-the-loop (HITL) practices, context-driven inputs and outputs, trust, ethics, and appropriate interface granularity.

3.4.1. UX Cross-Cut Analysis of Work-Packages

In Session 5 of the workshop, participants conducted a cross-cutting user experience (UX) analysis across all submitted work-packages to ensure consistent alignment with user needs and interaction models. This session focused on evaluating each work-package through four key UX-focused lenses identified in the work-package template. The goal was to capture how AI-driven functionality is intended to be delivered to users, the expected access and interaction patterns, and the role of human oversight. This workshop output was then used in follow-up sessions to drive the selection of key UX best practices for design of AI and machine learning (ML) workflows within the ESnet ecosystem (Tehsin, 2023).

3.4.1.1. Desired Target User(s)

For each work-package, we identified the primary users who will interact with or benefit from the proposed functionality. This ranged from specific technical roles, such as network engineers and data scientists, to broader archetypal groups like research facility operators or workflow developers. Documenting these user personas helps ground the design of AI interfaces in real-world needs and clarifies whose problems the solution is aiming to solve.

3.4.1.2. Access

ESnet classified each work-package based on its intended access policy. Categories included:

- **Public Facing:** Accessible to external users or the broader community.
- **Internal Only:** Restricted to ESnet personnel or authorized collaborators.
- **Mixed or Role-Based:** Offering differentiated access depending on a user’s role or affiliation.

This distinction helps determine appropriate levels of security, documentation, and user interface design depending on the scope of exposure and intended audience.

3.4.1.3. Granularity of Interface

Proposed user interfaces were assessed in terms of their granularity and integration within existing systems. Options included:

- **Immersive/Entire Application:** Full applications where AI is central to the experience.
- **Assistive/Complementary:** AI features that augment an existing tool or workflow.
- **Embedded/Single-Entity:** Lightweight or in-place features embedded in a specific UI element or context.
- **Command Line Interface:** For power users or developers needing low-level control.
- **Not Applicable:** For work-packages that are backend-focused, such as datasets or model training artifacts without direct user interaction.

Participants were asked to justify their interface choices based on intended user roles and the complexity of interaction required.

3.4.1.3. Human-in-the-Loop (HITL)

Finally, workshop participants examined whether each work-package incorporates HITL best practices, particularly important for AI systems where transparency, trust, and oversight are critical. This included identifying where human judgment is necessary for validation, override, or learning loop feedback. Work-packages were encouraged to highlight points in their workflows where humans provide value beyond automation, ensuring responsible and interpretable use of AI technologies.

3.4.1.4. Summary

This structured UX cross-cut session provided valuable clarity on how each work-package connects to real users and operational contexts, ultimately guiding design choices that support both usability and trust in AI-driven systems at ESnet.

3.4.2. Best UX Practices for Large Datasets and AI Model Workflows

As ESnet explores the integration of AI into its infrastructure and operations, thoughtful design of both the workflows and the supporting data ecosystem is paramount. Many of the themes defined in the work-packages introduced new workflow driven tasks, as well as tasks involving large (inconsistent) datasets. The UX working group focused on these themes when selecting the UX best practices within the following subsections.

This section outlines key considerations and best practices for building AI-driven systems that are not only technically robust but also accessible, understandable, and operationally relevant. It emphasizes the importance of aligning AI workflows with ESnet-specific use cases and datasets, maintaining high standards of data quality, and ensuring that workflows are designed with user experience (UX) in mind. From managing and visualizing AI workflows to addressing the complex realities of data

preparation and access, each component plays a critical role in delivering trustworthy, high-impact outcomes.

Equally important are the long-term processes of validation, monitoring, and retraining, which ensure models remain effective over time, as well as the need for strong collaboration across domains to make sense of complex system behavior. The following subsections explore these principles in detail, offering a comprehensive approach to operationalizing AI at ESnet.

3.4.2.1. AI Workflow Management UX

AI Workflow Management UX is important because it directly impacts the usability, effectiveness, and adoption of AI tools within complex operational environments like ESnet.

Investigations found that eight of the defined work-packages request user-driven domain-specific workflow capabilities using AI assistance, or fully-automated system-managed workflows for tasks such as anomaly detection. WP03 (Data Quality) discusses the importance of using AI in automated detection and response workflows, while WP17 (Automating Site Deployment) describes a user-driven AI-assisted site deployment workflow that can automate setting up the device's base configuration, bootstrapping, and performing other essential configuration tasks. For all work-packages defining AI-assisted workflows it is important to create a UX experience that not only speeds up the specific task, but invokes confidence in the solution.

Findings:

UX.F1	AI workflow tools were most effective, and promoted trust, when users had access to contextual data explanations associated with AI generated output.

Here are the key workflow management UX best practices the working group has identified as applicable to use case:

- AI workflows must be user-friendly and intuitive, designed to minimize complexity while empowering users.
- Expose underlying steps or calculations to enhance transparency and user comprehension of the workflow’s control flow wherever possible and appropriate.
- Simplify configuration and parameter tuning without requiring deep technical expertise wherever possible or appropriate.
- Provide clear visualizations of workflow progress, dependencies, and outputs for non-technical stakeholders.

AI workflows, such as those defined in WP03 (Data Quality) and WP17 (Automating Site Deployment), should be designed to guide users seamlessly through complex tasks while maintaining visibility into the logic and data underpinning AI-driven decisions. Workflow interfaces should offer clear visual

indicators of progress, highlight key dependencies, and present actionable outputs in context. Parameter tuning and configuration options should be abstracted to user-friendly controls wherever possible, without sacrificing flexibility. Ultimately, successful AI workflow UX ensures that users can confidently understand, manage, and collaborate with AI systems, transforming automation into a trusted operational partner rather than a black-box tool.

In summary, AI Workflow Management UX isn't just a cosmetic layer, it is a critical enabler for successful, sustainable, and impactful AI deployment, especially in mission-driven organizations like ESnet where interdisciplinary collaboration and operational reliability are essential.

Recommendations:

UX.R1	To ensure effective adoption and sustained trust in AI-assisted operations at ESnet, AI workflow management must prioritize intuitive, transparent, and task-aligned user experiences. ESnet should incorporate layered User eXperience (UX) elements that expose intermediate steps, model outputs, and decision rationales to users with varying levels of technical expertise. Workflow interfaces should offer clear visual indicators of progress, highlight key dependencies, and present actionable outputs in context. Parameter tuning and configuration options should be abstracted to user-friendly controls wherever possible, without sacrificing flexibility.

3.4.2.2. Data Management UX

Data Management UX is essential because the success of AI workflows depends not just on having data, but also on understanding, accessing, and preparing the right data in the right context. Poor user experience around data management can introduce friction, delay insights, and ultimately compromise the quality of AI outcomes.

The cross-cut activity uncovered that many work-packages have identified data management as a major issue to the success of AI-based solutions. WP03 (Data Quality) contains a superset of the issues, identifying government mandated compliance for data visibility, resistance to changes in existing data governance rules, and lastly, the most identified issues of lack of data consistency, interoperability, or completeness among ESnet datasets. WP15 (Consistent data management) identifies that teams across ESnet use different methods and formats for data management and analysis that contributes to the consistency issue. WP16 (Query All Data) also identifies issues relating to operating within strict access and compliance boundaries. It must not access or expose Controlled Unclassified Information (CUI), Personally Identifiable Information (PII), or any other sensitive content that is restricted under federal or organizational policies as identified by WP03 (Data Quality). General conclusions were that ESnet data will require improved normalization, correlation, and completeness across the diverse data sources.

Findings:	
UX.F2	Aligning AI workflows with ESnet-specific data and use cases is critical to ensure context and relevance.
UX.F3	Investing in data hygiene, accessibility, and maintenance is crucial to ensuring high-quality data that drives successful outcomes.

Here are the key data management UX best practices the working group has identified for this use case:

- **For AI workflows to succeed, the system must deeply understand your specific context and data. Hence, data is meaningful only if you know the context—without this understanding, even high-quality data loses relevance.**
- Critical to this process is the burden of data preparation: cleaning, processing, and structuring data for training ML models falls on ML engineers or those responsible for model training.
- To enable this, API or data owners must ensure seamless access to required datasets for engineers, fostering collaboration and transparency.
- Enforce strict data hygiene where required: Poor-quality inputs (e.g., incomplete or inconsistent data) leads to unreliable outputs (garbage in, garbage out)
- **Prioritize specialized, operation-specific AI use cases: These drive the most value. Identify high-impact areas and clearly map their associated data requirements to align development efforts with ESnet's needs.**

Successful AI workflows rely not only on data availability but also on the user's ability to access, interpret, and prepare that data efficiently and within policy constraints. Work-packages such as WP03 (Data Quality), WP15 (Consistent Data Management), and WP16 (Query All Data) have highlighted widespread challenges including inconsistent formats, data silos, and strict compliance boundaries around sensitive information. By embedding contextual understanding and operational relevance into data workflows, ESnet can ensure that its AI initiatives are both scalable and grounded in trustworthy, actionable data.

Effective data management UX is foundational to the success of AI workflows at ESnet. It ensures that users can not only access data, but also understand its context, prepare it efficiently, and apply it to high-value, domain-specific problems. The working group identified key design patterns to reduce friction in data preparation, enforce data hygiene, improve access, and align datasets with operational priorities—ultimately enabling more accurate, trustworthy, and impactful AI outcomes.

Recommendations:

UX.R2	To unlock the full potential of AI at ESnet, data management UX must be treated as a first-class design consideration, where we emphasize data quality, context, and hygiene while facilitating seamless data access and collaboration between data owners and Machine Learning (ML) engineers. ESnet should architect a cohesive data management framework that enforces data hygiene and normalization, supports interoperable formats, and provides intuitive tooling for data exploration and preparation. Data access workflows should include clear metadata, provenance, and usage guidance, while respecting Controlled Unclassified Information (CUI) and Personally Identifiable Information (PII) restrictions. Application Programmable Interfaces (APIs) must be well-documented and enable frictionless collaboration between data owners and ML practitioners.

3.4.2.3. Validation & Testing

In the context of AI UX, validation and testing are critical to ensuring that models behave reliably and transparently under real-world conditions, especially in scenarios that fall outside the norm. Hallucinations revolve around the AI systems misinterpreting data, creating false relationships, providing inaccurate summaries, or generating non-existent information. These potential issues highlight the need for careful validation and oversight of any AI solutions implemented.

Based on discussions during the cross-cut sessions, the working group was extremely concerned about edge case scenarios, and hallucinations that could cause invalid results. There are a number of work-packages that could be negatively impacted by hallucinations, and therefore, need special consideration around testing. For example:

- **Language Intent (WP25 (NLP Interfaces to Systems)):** In the context of "NLP Interfaces to Systems," if an NLP interface is used, it could misinterpret a user's natural language request to configure a network device. For example, a user might ask to "increase bandwidth on port 1," but the system might hallucinate a different action.
- **Generating Non-Existent Connections Between Data (WP16 (Query ALL Data), WP29 (Dataset Unified Query)):** When querying all data or providing a unified query, an AI system could create false connections between documents or datasets. For example, it might link a document to a ticket that is completely unrelated, or hallucinate that a specific configuration is associated with a particular host when it's not.
- **Providing Incorrect or Inaccurate Code Conversions (WP24 (Legacy Code)):** In "Legacy Code," when asking an LLM to rewrite Perl scripts into Python, the LLM might hallucinate code that doesn't function as intended or omits critical parts of the original script. This could lead to significant issues when replacing the legacy code.

Findings:

UX.F4	AI hallucinations can negatively impact both the UX experience, as well as actions taken by AI on behalf of the user. AI actions that have consequences should be vetted and approved by a human before being performed.

Here are the key validation and testing UX best practices the working group has identified for this use case:

- **Edge Case Testing:** Develop tools wherever value is justified to identify and test model behavior in edge cases and rare events (e.g., outliers, low-frequency scenarios) that may be underrepresented in training data. This ensures robustness in real-world applications.
    - It may be wise to add noise or rare events to your training data if not already present to detect edge cases when deployed to production.

Robust AI UX requires proactive validation strategies, particularly edge case testing, to uncover model weaknesses in rare or unexpected conditions. Introducing synthetic noise or rare event data can enhance model resilience and support trustworthy performance in production environments.

Recommendations:

UX.R3	To ensure safe, reliable, and user-aligned AI behavior, especially in operational environments, it is necessary to prioritize robust validation, especially for edge cases, and implement protocols for retraining and continuous performance monitoring. ESnet should implement comprehensive edge case testing frameworks that simulate rare or ambiguous scenarios, as well as adversarial testing techniques to stress-test model boundaries. Where feasible, augment training datasets with noise and synthetic edge cases to improve robustness. Additionally, automated test harnesses should include validation checkpoints for high-risk actions, and all AI-driven recommendations that impact systems or users should be gated through Human-In-The-Loop (HITL) approval mechanisms.

3.4.2.4. Maintaining & Monitoring

Maintaining and monitoring AI systems is essential to sustaining long-term performance, ensuring that models remain accurate, relevant, and aligned with changing data and operational conditions.

The working group identified cross-cutting issues within the work-packages relating to identifying incorrect AI generated results, as well as giving feedback to retrain the model. As an example, WP17 (Automating Site Deployment) describes AI-assisted automation of the site deployment process,

where the user can provide feedback to this AI agent at the end of each site deployment for further improvements.

Continuous monitoring should be implemented to track key indicators such as model accuracy, data drift, and anomaly rates in real-time, with alerting mechanisms that notify responsible teams of degradation or unexpected behavior. For example, in WP17 (Automating Site Deployment), integrating a structured feedback loop allows users to flag issues or suggest improvements after each deployment, which should feed directly into retraining pipelines. These practices not only safeguard model relevance but also reinforce user confidence by demonstrating a commitment to transparency and continuous improvement. Over time, this approach helps embed AI more deeply and safely into operational workflows, aligning its behavior with real-world performance expectations.

Findings:

UX.F5	Clear retraining and monitoring processes enable timely detection of performance issues and help maintain effective, trustworthy AI systems over time.

Based on these discussion the working group has recommended the following best practices for this use case:

- **Retraining Protocols:** Establish schedules and criteria for retraining models with fresh data to adapt to evolving environments and new datasets.
- **Continuous Monitoring:** Track model performance, data drift, and degradation over time. Deploy real-time alerts for anomalies or declining accuracy.

By implementing clear retraining protocols and continuous monitoring, teams can detect performance drift, respond to anomalies, and keep AI systems effective and trustworthy throughout their lifecycle.

Recommendations:

UX.R4	To ensure sustained effectiveness and reliability, AI systems must be supported by robust maintenance and monitoring frameworks that adapt to evolving data, user needs, and operational contexts. ESnet should implement retraining protocols with clear triggers, such as performance thresholds, scheduled intervals, or significant input distribution shifts, and ensuring that retraining datasets are curated for quality and diversity.

3.4.2.5. Collaboration and Documentation

An effective AI implementation at ESnet relies on strong collaboration across teams and thorough documentation practices to ensure transparency, traceability, and shared understanding.

The working group identified cross-cutting issues within the work-packages WP01 (Alerting) and WP03 (Data Quality) relating to collaboration within ESnet in the context of improving data refinement through access to domain specific knowledge. Although not explicitly called out in the work-package, WP16 (Query All Data), WP25 (NLP Interfaces to Systems), and WP29 (Dataset Unified Query) would also benefit from this capability.

Findings:

UX.F6	Version-controlled documentation combined with close interdisciplinary collaboration leads to faster troubleshooting, clearer insights, and more resilient AI systems.

The working group suggests the following best practices for the use case:

- **Documentation Updates:** Maintain version control for models and workflows to ensure traceability and accountability.
- **Collaboration:** To derive meaningful insights from ESnet data we will need solid collaboration between software and network engineering teams, if software engineering teams come up with anomalies in data - network engineering teams will need to provide expertise on correlating the anomaly to hardware failure or real-world network failure. This technical foundation supports faster issue resolution, accelerates development, and ensures that AI outputs are grounded in operational reality, thereby enhancing system trust and long-term maintainability.

Maintaining version-controlled documentation and fostering close collaboration between software and network engineering teams enables faster troubleshooting, clearer insights, and more accountable, resilient AI systems.

Recommendations:

UX.R5	To ensure successful AI implementation at ESnet, interdisciplinary collaboration and rigorous documentation must be treated as core components of system design and operation. ESnet should have structured collaboration workflows that facilitate continuous knowledge exchange—such as regular cross-team reviews, shared glossaries, and integrated feedback loops. All models, prompts, decision logic, and configuration changes should be version-controlled in a central repository to ensure reproducibility and auditability.

3.4.3. Best Practices for End-User UX

AI applications can benefit from well-defined UX best practices which are essential for ensuring that the applications are accessible, understandable, and trustworthy. They help bridge the gap between complex AI functionality and real-world user needs by guiding how AI is presented, interacted with, and controlled. By following UX best practices, developers can:

- Increase adoption by designing intuitive and approachable interfaces,
- Build trust through transparency, feedback, and human-in-the-loop controls,
- Improve effectiveness by aligning AI outputs with user goals and contexts,
- Reduce errors by clarifying system behavior and limitations,
- And support collaboration across roles through shared, interpretable interfaces. (Pai, 2024) (Tehsin, 2023)(Hsiao & Tang, 2024; Weisz et al., 2024)

In short, strong UX practices turn AI from a black box into a usable, reliable tool in decision-making and operations. The next section summarizes the work-package end user UX cross-cut output, with working group findings and suggested recommendations for designing UX in AI/ML workflows within the ESnet ecosystem, focusing on HITL practices, context-driven inputs and outputs, trust, ethics, and appropriate interface granularity.

3.4.3.1. Human-in-the-Loop (HITL) UX Mechanisms

HITL UX mechanisms are essential for AI-driven applications because they provide critical oversight, improve decision quality, and help build trust in the system.

Twenty-five of the 28 final work-packages capture some degree of HITL practices. Of the four that were not, most were work-packages focusing on organization policies or workflows such as defining a common vocabulary for more consistency across our data management solutions.

WP08 (Outage Notification Parsing), is a great example of HITL where an end user is presented with an AI/LLM generated outage notification, along with links and sources connected to that outage. No automated action to resolve the outage is taken, it simply prompts the user to investigate and confirm. Automated responses would have to be approved by the end user before AI/ML can take action.

WP25 (NLP Interfaces to Systems), specifies a single natural language interface to provide network configuration operations and end user facing network characteristics. An HITL design will need to be leveraged as a “gatekeeping” mechanism to review and validate the NLP generated system configuration changes. With humans in the loop, any suggested “risky” or “invalid” actions should be detected in a manual review.

Findings:

UX.F7	Human-in-the-Loop in AI ensures oversight, improves decision quality, and builds user trust. HITL UX practices ensure AI remains a tool for empowerment, balancing

automation with accountability and human oversight.

Based on analysis for the work-packages, the working group suggests the following best practices for HITL:

- **Assist, not replace:** Design AI systems to assist rather than replace users, positioning AI as a supportive copilot that enhances human decision-making.
- **Ease of retrainability:** Make AI-generated content easy to edit or correct directly in the interface. Provide obvious, low-friction entry points for humans to intervene or modify AI suggestions.
- **Allow for overrides:** Ensure users can easily challenge or bypass AI recommendations.
- **Action correctability:** Let users easily undo or revert AI actions.

In short, HITL UX design practices ensure that AI remains a tool for empowerment, not automation without accountability. They strike the right balance between efficiency and oversight, allowing AI systems to scale while staying aligned with human intent and judgment. This approach transforms AI into a trusted copilot, augmenting user decision-making without displacing human judgment, and ensures that the benefits of automation are realized without compromising operational integrity or user confidence.

Recommendations:

UX.R6	To ensure AI systems deployed at ESnet remain transparent, accountable, and aligned with user intent, HITL design must be embedded as a core architectural principle. ESnet should design AI interfaces that clearly distinguish between suggestions and actions, provide intuitive editing and approval workflows, and highlight high-risk or ambiguous outputs for manual review. Systems should offer unobtrusive but accessible override and rollback options, and integrate correction inputs directly into retraining or feedback loops to improve model performance over time.

3.4.3.2. Provide Suggested Context for Inputs and Prompts

Providing suggested context for inputs and prompts is important for an AI application because it helps users interact more effectively with the system, leading to clearer intent, better outputs, and a smoother user experience (Liu, 2024).

A number of work-packages leverage these practices but WP16 (Query All Data) in particular is a great example. Given the potential magnitude of the dataset, its UX describes listing capabilities, and helps guide the user through the most effective way to use this tooling.

Findings:

UX.F8	Supplying appropriate context empowers users, improves communication with AI, and enhances usability and outcomes.

The working group discussions focused on the topic of improved context, especially for AI applications used for daily operations. The following best practices are proposed for this use case:

- **Provide context:** Whenever possible, provide context for the writing of effective prompts for your ML or AL powered feature.
- **Guide interactions:** When working with a natural language input, provide users with contextually relevant prompt suggestions to guide interactions and leverage AI capabilities effectively.
- **Onboarding:** Provide onboarding features to educate users on how to collaborate with the AI system effectively.

Helping users understand how to interact with AI systems is just as important as the system's underlying intelligence. By offering suggested context, prompt guidance, and onboarding support, AI applications become more approachable, reduce user error, and deliver more meaningful results. These practices are especially critical in operational environments, where clarity and effectiveness in human-AI collaboration directly impact efficiency and trust.

Recommendations:

UX.R7	To maximize the effectiveness of AI systems within ESnet, it is essential to embed context-aware guidance into user interactions. Suggested context, such as sample queries, autocomplete options, and dynamic prompt scaffolding, helps users formulate clearer, more precise inputs, leading to better AI responses and a more intuitive overall experience. ESnet should integrate context-sensitive help features that adapt to the user's task, role, and system state, along with onboarding tools that introduce users to effective prompt strategies. Additionally, natural language interfaces should proactively offer clarifying suggestions or corrections when ambiguous or incomplete inputs are detected.

3.4.3.3. Provide Context in Output

Providing context for output from an AI application is critical because it helps users interpret results accurately, make informed decisions, and build trust in the system. Without context, even correct answers can be misunderstood, misused, or dismissed.

Output context was identified as important in 10 of the work-packages. WP01 (Alerting) and WP08 (Outage Notification Parsing) touch on these practices but make sure to provide the end user with references for what triggered the outage notification and sources for a human to go investigate.

Findings:

UX.F9	Providing clear, contextual information around AI outputs enhances user understanding, trust, and the responsible use of AI-driven decisions.

This topic was well discussed by the working group during cross-cut activities, and identified the following best practices for the use case:

- **Dataset context:** Provide insights into how AI outputs are generated.
- **Decision making:** Offer simple, contextual explanations of why the AI made a decision. For example, “We suggest using a Nokia router because those are already being used at the same target location”
- **Confidence level:** To indicate an accurate level of confidence in the output when possible. Dynamically adjust how assertively the AI presents itself based on confidence levels and risk (Glaros, 2024).
- **Communicate risk:** When allowing AI/ML to execute actions based on human input, indicate the expected impact radius and fallback options for when the AI fails or produces unexpected results.

Providing clear, contextual explanations for AI outputs is essential to ensuring that users understand not just what the system is telling them, but why and how it reached its conclusions. By embedding context, confidence levels, and risk indicators into AI-driven results—as identified in numerous work-packages including WP01 and WP08—teams can support more informed decision-making, reduce misuse, and build lasting trust in AI-powered tools across ESnet's operational landscape.

Recommendations:

UX.R8	To ensure AI outputs are actionable, trustworthy, and properly understood within operational environments like ESnet, it is essential to embed rich, interpretable context directly into system responses. ESnet should explore mechanisms and processes to ensure that all AI outputs include clear explanations of the underlying data, the reasoning behind recommendations, and, where applicable, confidence levels and risk indicators. The tone and assertiveness of outputs should be adjusted based on uncertainty or impact, signaling whether a result is a strong recommendation or a tentative suggestion.

Where decisions carry potential operational consequences, outputs must clearly communicate fallback options and the scope of impact in the event of error.

3.4.3.4. Trust, Ethics, and Bias Management

Trust, ethics, and bias management are crucial in AI applications because they directly impact the system's reliability, fairness, and societal acceptance. As AI becomes more integrated into decision-making processes, overlooking these areas can lead to harmful consequences, both technical and human.

A number of work-packages generate output that needs to be explicitly labeled as AI output. WP12 (Detect External Configuration Anomalies), WP22 (Automated ServiceNow Ticket Summarization), and WP28 (Mission Support Management) explicitly call out having appropriate labeling in order to satisfy the HITL requirements. It is expected that a responsible human end user will do additional investigation before taking AI labeled content at face value.

Findings:

UX.F10	Trust, ethics and bias management are critical for ensuring AI system reliability, fairness and societal acceptance. Managing trust, ethics, and bias through transparency, labeling, and user control is essential for safe and responsible AI integration.

The work group identified the following best practices for the use case:

- **Visual cues:** Clearly indicate AI/ML generated content through visual cues or textual disclaimers
- **Context history:** When displaying a history of changes or decisions made indicate which were made by AI/automation and which ones were made by humans.
- **Tailorable features:** When an AI feature is not required, consider providing ways for a user to turn off that feature via customization.
- **Set expectations:** Be transparent about the AI's capabilities and potential shortcomings to set realistic user expectations.

Upholding trust, ethics, and bias management in AI applications is not just a design choice, it is a responsibility. By clearly labeling AI-generated content, distinguishing human and machine decisions, and setting transparent expectations, we ensure that AI remains a supportive tool rather than a source of confusion or risk. These practices are essential to maintaining accountability, enabling informed human oversight, and fostering the responsible use of AI across ESnet's mission-critical environments.

Recommendations:

UX.R9	To ensure the responsible and ethical deployment of AI systems across ESnet, it is imperative to implement mechanisms that promote transparency, user agency, and clarity around AI-generated content. ESnet should use visual indicators, disclaimers, and metadata tags to differentiate AI-generated outputs from human-authored content. Interfaces should maintain an auditable history that distinguishes between AI-driven and manual actions, supporting traceability and accountability. Where feasible, provide users with the ability to opt out of specific AI features or adjust the level of automation based on their role or context. Additionally, communicate the system's capabilities, limitations, and known biases clearly to set accurate expectations and avoid misuse.

3.4.3.5. Data Usage Transparency/Disclosure

Data usage transparency and disclosure is essential in AI applications to maintain user trust, protect sensitive information, and comply with ethical and legal standards. Providing upfront, understandable disclosure not only supports informed consent but also promotes responsible data stewardship and helps organizations avoid unintended misuse of sensitive inputs. One of the most concerning discussions during the workshop was around whether existing AI tools like ChatGPT or internally deployed LLM tools used data as input for the user to train their models, and would this possible sensitive data be leaked to other users.

Any solution including but not limited to WP08 (Outage Notification Parsing), WP09 (Ticket Resolution), and WP29 (Dataset Unified Query) that involves the training or refinement of our own models/datasets needs to include very specific warnings regarding the safety of our data usage. At ESnet we handle sensitive data at varying levels and severities.

Findings:

UX.F11	Transparent data usage disclosures are vital to ensure user awareness, protect sensitive information, and align with ethical AI practices.

Based on this information, the work group identified the following best practices for the use case:

- **Indicate training data:** Clearly indicate to users with visual cues or textual disclaimers when data input or feedback could be collected and used for AI/ML training data, especially when working with sensitive access controlled data. For example, "a user adds notes to a location field containing a passcode for entry into a server closet."

As AI capabilities continue to evolve, maintaining transparency around data usage is not optional, it is foundational to ethical and secure deployment. By clearly communicating how user data may be used,

especially in sensitive environments like ESnet, we not only safeguard information but also reinforce user trust and uphold our commitment to responsible AI practices.

Recommendations:

UX.R10	To uphold ethical standards and protect sensitive information within AI-driven systems at ESnet, transparent data usage disclosure must be integrated into all user-facing interfaces and workflows. ESnet should incorporate clear, persistent disclaimers or visual indicators when data could be stored, analyzed, or influence future AI behavior. Interfaces should include context-aware warnings, particularly when inputs are entered into free-form fields that may inadvertently capture sensitive content, and provide guidance on safe data entry practices. Additionally, administrative controls must allow data owners to configure data collection policies, with granular options for opting in or out of training pipelines.

3.4.3.6. User Feedback and Training

Providing user feedback to AI results is important because it creates a feedback loop that directly improves model training, accuracy, and relevance over time. User feedback, such as flagging incorrect results, approving useful outputs, or suggesting better alternatives, gives the model real-world data about how it's performing. This helps fine-tune the model to reduce errors and better align with actual user expectations and domain-specific needs. AI models often struggle to keep up with changing environments, policies, or user goals. Continuous user feedback enables models to evolve alongside the context in which they operate, improving long-term usefulness and reducing drift.

Similarly to section 3.4.2.4 (Data Usage Transparency/Disclosure), WP08 (Outage Notification Parsing), WP09 (Ticket Resolution), and WP29 (Dataset Unified Query) involve the training or refinement of our own models/data. Collecting user feedback can help us improve or understand the effectiveness of our AI/ML end user work-packages.

Findings:

UX.F12	Continuous user feedback is essential for keeping AI systems accurate, relevant, and aligned with dynamic operational contexts.

After discussion on this topic, the work group identified the following best practices for the use case:

- **Model feedback:** Provide functionality for, and encourage, users to correct errors and provide feedback on AI quality or accuracy. When possible, create workflows for capturing these corrections back into your training workflows.

- **HITL feedback:** Offer simplified UX tools to label data or approve AI predictions. For example, a “this was helpful” button.

Providing user feedback on AI results is a critical best practice for improving model accuracy and relevance over time. By enabling users to correct errors, approve predictions, or flag issues, feedback creates a valuable loop that helps AI systems adapt to changing contexts and better meet real-world needs, especially in work-packages like outage parsing, ticket resolution, and unified query systems.

Recommendations:

UX.R11	Integrating user feedback into AI systems is a foundational practice for ensuring ongoing model accuracy, adaptability, and domain relevance within ESnet’s dynamic operational environment. ESnet should embed lightweight, intuitive feedback tools, such as approval buttons, correction prompts, or rating scales, directly into the user interface to encourage participation without interrupting workflow. Where feasible, feedback should be captured in a structured format and integrated into retraining pipelines, allowing for supervised fine-tuning that reflects real-world performance. Systems should also prioritize transparency by indicating how user feedback is used and offering visibility into the impact of cumulative input over time.

3.4.3.7. Choose an Appropriate Granularity of Interface for GUI Features

UX design patterns for AI are critical when considering interface granularity because they help ensure that AI capabilities are delivered in a way that matches user needs, technical skill levels, and the context of use. Granularity defines how deeply integrated AI features are within a user’s workflow, ranging from fully immersive applications to lightweight, embedded tools or command-line utilities. Choosing the right level of integration isn’t just a design choice, it directly affects usability, adoption, and the effectiveness of the AI itself.

3.4.3.7.1. Immersive Framework

Use full-screen, focused interfaces for tasks requiring deep engagement with AI-generated content. Work-package analysis shows 19 proposals recommending use of immersive frameworks. An example of this is the text-based conversational user interface used by ChatGPT.

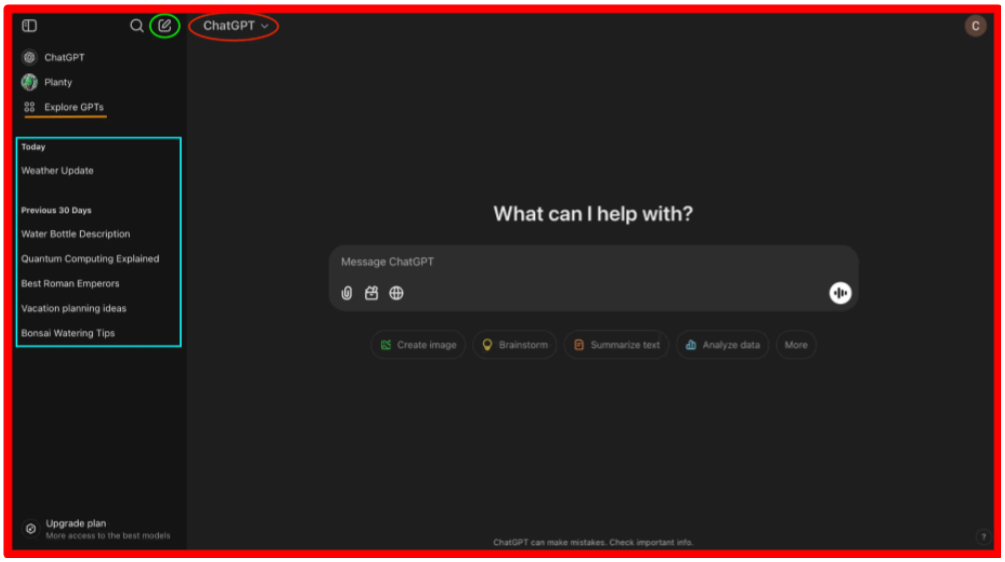

Figure 3.4.3.7.2.a. ChatGPT interface.

3.4.3.7.2. Assistive Framework

Integrate AI assistance within existing applications to provide contextual support without disrupting user workflows. Work-package analysis shows 13 proposals recommending use of assistive frameworks. An example of this is the Gemini chat interface which has a “side bar assistant” as seen in Figure 3.4.2.7.2.a.

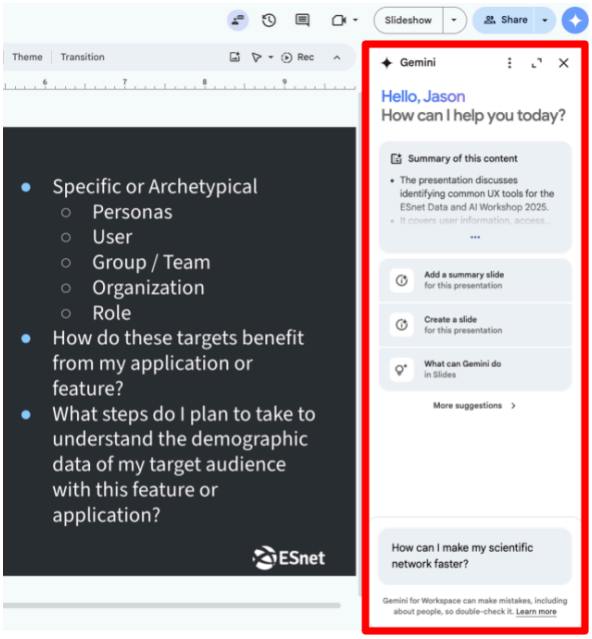

Figure 3.4.3.7.2.a. Gemini chat interface “side bar assistant” in Google Drive.

3.4.3.7.3. Embedded / Single Entity Framework

Embed AI functionalities within specific components or features, offering targeted assistance for particular tasks. Work-package analysis shows 7 proposals recommending use of assistive frameworks. An example of this is the Grammarly pop-up suggestion window see in Figure 3.4.3.7.3.a.

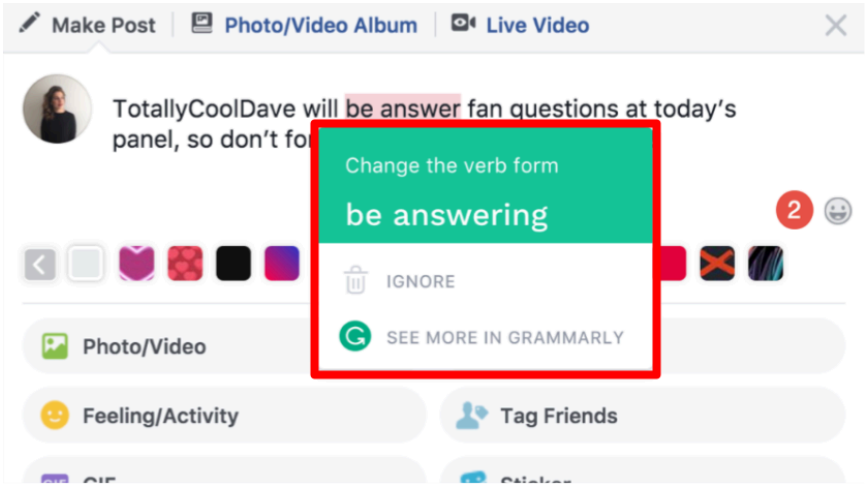

Figure 3.4.3.7.3.a. A Grammarly pop-up suggestion.

In summary, interface granularity is a key lens for designing AI systems that are usable, adaptable, and fit for purpose. Thoughtful UX design patterns ensure that AI is integrated at the right level of the stack, maximizing its value without disrupting existing workflows.

Findings:

UX.F13 Thoughtful design choices help integrate AI at the appropriate level in the user experience, maximizing utility without interfering with workflows.

Work-package	Immersive	Assistive	Embedded	CMD line	N/A
WP01 (Alerting)	X	X	X		
WP02 (Rules Correlation)	X				
WP03 (Data Quality)	X	X	X		
WP04 (Lifecycle)	X	X			
WP05 (Data Catalog)	X				

WP06 (Network Services)	X	X			
WP07 (Business Ops)	X	X	X		
WP08 (Outage Notification Parsing)			X		
WP09 (Ticket Resolution)	X	X			
WP10 (Correlate Alarms)	X	X	X		
WP11 (Predict Hardware Failures)			X		
WP12 (Detect External Configuration Anomalies)	X				
WP13 (Capture Configuration Intent)	X				
WP14 (Fast Contract Lookup)		X			
WP15 (Consistent Data Management)					X
WP16 (Query All Data)	X				
WP17 (Automating Site Deployment)		X			
WP19 (AI Sandbox)	X				
WP20 (RFP Contract Builder)	X				
WP21 (Unified Document Search)		X			
WP22 (Ticket Summarization)		X	X		
WP23 (Federated Authentication)					X
WP24 (Legacy Code)	X				
WP25 (NLP Interfaces to Systems)	X				
WP26 (Information Architecture)	X				
WP27 (Requirements Management)					X
WP28 (Mission Support Management)	X	X			
WP29 (Dataset Unified Query)	X	X			
Totals	19	13	6	0	4

Table 3.4.3.7.3.a. Summary of findings for Granularity of Interface.

Recommendations:

UX.R12	Selecting the appropriate interface granularity is essential for delivering AI features that align with user needs, operational contexts, and the intended depth of interaction. Based on work-package analysis, ESnet should tailor AI integration using one of three established UX design frameworks; immersive, assistive, or embedded, depending on task complexity and user engagement levels.
UX.R13	ESnet should evaluate each AI-enabled work-package to determine the optimal interface granularity and explicitly mapping that decision to user roles, task criticality, and environment constraints.

3.4.3.8. Avoid Anthropomorphizing AI

Anthropomorphizing AI is the act of attributing human traits, emotions, intentions, or consciousness to artificial intelligence systems. Humans are naturally inclined to interpret behavior in human terms, especially when interacting with systems that use natural language, have conversational interfaces, or mimic social cues (like tone or facial expressions in avatars). In some cases, anthropomorphism can make AI more approachable and user-friendly (e.g., virtual assistants like Siri or Alexa), but in other cases it can lead to misunderstanding AI's true capabilities and limitations, overestimating its intelligence, trustworthiness, or autonomy, which may result in poor decision-making or misplaced trust.

During cross-cut analysis, it became clear that users wanted not only a clear context of any AI generated output, but also wanted it to be clear they're interacting with a tool, not a sentient being.

Findings:

UX.F14	Users should know they are interacting with a tool, not a sentient being. This can help frame trust context around nondeterministic outputs from an AI powered feature.

Based on this feedback, the working group requests the following UX design practices be followed:

- **No anthropomorphizing:** Use language and visuals that reflect the AI's capabilities honestly — don't pretend it's human.
- **Set expectations:** Avoid over-promising what the AI can do.

While anthropomorphizing AI can make systems feel more approachable, it often leads to misunderstandings about their capabilities. To support clarity and trust, the working group

recommends avoiding human-like language or visuals and ensuring users clearly understand they are interacting with a tool and not a sentient agent.

Recommendations:

UX.R14	To ensure clarity, foster appropriate trust, and prevent misinterpretation of AI capabilities, user interfaces should be explicitly designed to avoid anthropomorphizing AI systems. ESnet should use neutral, technical language in system prompts and responses, avoiding terms that imply emotion, intention, or personality. Visual elements, such as avatars or icons, should reinforce that users are engaging with a system, not a person. Additionally, disclaimers or contextual indicators should clarify the deterministic or probabilistic nature of AI outputs.

3.4.4. UX Conclusions

These recommendations provide a comprehensive framework for designing effective and trustworthy user experiences within ESnet's evolving AI/ML ecosystem. Key areas of focus include streamlined workflow management, robust data practices, rigorous validation and monitoring, and user-centered design that emphasizes transparency, ethical considerations, and appropriate interface granularity. By implementing these best practices, ESnet can ensure that our AI-driven tools are not only technically sound but also highly usable, reliable, and aligned with the organization's mission-critical operations. Ultimately, these efforts will foster greater confidence in AI systems and maximize their potential to enhance ESnet's capabilities.

3.5. Workshop Conclusions

The February 2025 Data and AI Workshop offered ESnet a valuable opportunity to identify key challenges hindering progress as well as potential opportunities for managing infrastructure beyond human-scale limitations. The workshop's resulting work-packages provide a concrete foundation for discussions, enabling practical and actionable conversations about addressing gaps and establishing achievable expectations. This report captures ESnet's current perspective on integrating AI into its operational ecosystem, recognizing that the rapidly evolving AI landscape will likely lead to changes in this perspective over time.

References

Abdallah, M., An Le Khac, N., Jahromi, H., & Delia Jurcut, A. (2021, August 17). A hybrid CNN-LSTM based approach for anomaly detection systems in SDNs. *The 16th International Conference on Availability, Reliability and Security*. ARES 2021: The 16th International Conference on Availability, Reliability and Security, Vienna Austria. https://doi.org/10.1145/3465481.3469190

Adekunle, B. I., Chukwuma-Eke, E. C., Balogun, E. D., & Ogunsola, K. O. (2021). Machine learning for automation: Developing data-driven solutions for process optimization and accuracy improvement. *International Journal of Multidisciplinary Research and Growth Evaluation*, *3*(1), 800–808. https://doi.org/10.54660/.ijmrge.2021.2.1.800-808

Afzal, S., Rajmohan, C., Kesarwani, M., Mehta, S., & Patel, H. (2021, September). Data Readiness Report. *2021 IEEE International Conference on Smart Data Services (SMDS)*. 2021 IEEE International Conference on Smart Data Services (SMDS), Chicago, IL, USA. https://doi.org/10.1109/smds53860.2021.00016

Al Farizi, W. S., Hidayah, I., & Rizal, M. N. (2021, September 23). Isolation forest based anomaly detection: A systematic literature review. *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*. 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia. https://doi.org/10.1109/icitacee53184.2021.9617498

Alimohammadi, H., & Nancy Chen, S. (2022). Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis. *Expert Systems with Applications*, *191*(116371), 116371. https://doi.org/10.1016/j.eswa.2021.116371

Althati, C., Tomar, M., & Shanmugam, L. (2024). Enhancing Data Integration and Management: The Role of AI and Machine Learning in Modern Data Platforms. *Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023*, *2*(1), 220–232. https://doi.org/10.60087/jaigs.v2i1.154

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Arango, S. P., Kapoor, S., Zschiegner, J., Maddix, D. C., Wang, H., Mahoney, M. W., Torkkola, K., Wilson, A. G., Bohlke-Schneider, M., & Wang, Y. (2024). Chronos: Learning the language of time series. In *arXiv [cs.LG]*. https://doi.org/10.48550/ARXIV.2403.07815

Banerjee, D., Madduri, V., & Srivatsa, M. (2009, September). A framework for distributed monitoring and root cause analysis for large IP networks. *2009 28th IEEE International Symposium on Reliable Distributed Systems*. 2009 28th IEEE International Symposium on Reliable Distributed Systems (SRDS), Niagara Falls, New York, USA. https://doi.org/10.1109/srds.2009.22

Bertino, E., Kantarcioglu, M., Akcora, C. G., Samtani, S., Mittal, S., & Gupta, M. (2021, April 26). AI for Security and Security for AI. *Proceedings of the Eleventh ACM Conference on Data and Application*

Security and Privacy. CODASPY '21: Eleventh ACM Conference on Data and Application Security and Privacy, Virtual Event USA. https://doi.org/10.1145/3422337.3450357

Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2022). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys*, *54*(3), 1–33. https://doi.org/10.1145/3444690

Bolton, M. L., Bass, E. J., & Siminiceanu, R. I. (2013). Using formal verification to evaluate human-automation interaction: A review. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, *43*(3), 488–503. https://doi.org/10.1109/tsmca.2012.2210406

Brown, B., Miller, W., Bard, D., Boehnlein, A., Fagnan, K., Guok, C., Lançon, E., Ramprakash, S. (jini), Shankar, M., & Schwarz, N. (2023). *Integrated research infrastructure architecture blueprint activity (final report 2023)*. US Department of Energy (USDOE), Washington, DC (United States). Office of Science; Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States). https://doi.org/10.2172/1984466

Carter, J., Feddema, J., Kothe, D., Neely, R., Pruet, J., Stevens, R., Balaprakash, P., Beckman, P., Foster, I., Iskra, K., & Others. (2023). *Advanced Research Directions on AI for Science, Energy, and Security: Report on Summer 2022 Workshops*. Argonne National Laboratory (ANL), Argonne, IL (United States). https://doi.org/10.2172/1986455

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. In *arXiv [cs.LG]*. https://doi.org/10.48550/ARXIV.1901.03407

Dart, E., Zurawski, J., Hawk, C., Brown, B., & Monga, I. (2023). *ESnet requirements review program through the IRI lens: A meta-analysis of workflow patterns across DOE office of science programs (final report)*. Office of Scientific and Technical Information (OSTI). https://doi.org/10.2172/2008205

Fernandes, G., Jr, Rodrigues, J. J. P. C., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L., Jr. (2019). A comprehensive survey on network anomaly detection. *Telecommunication Systems*, *70*(3), 447–489. https://doi.org/10.1007/s11235-018-0475-8

Frehner, R., Wu, K., Sim, A., Kim, J., & Stockinger, K. (2024). Detecting Anomalies in Time Series Using Kernel Density Approaches. *IEEE Access*, *12*, 33420–33439. https://doi.org/10.1109/ACCESS.2024.3371891

Glaros, M. (2024, September 20). *Creating a dynamic UX: guidance for generative AI applications*. Microsoft Learning. https://learn.microsoft.com/en-us/microsoft-cloud/dev/copilot/isv/ux-guidance

Gonzalez, J. M. N., Jimenez, J. A., Lopez, J. C. D., & Parada G, H. A. (2017). Root cause analysis of network failures using machine learning and summarization techniques. *IEEE Communications Magazine*, *55*(9), 126–131. https://doi.org/10.1109/mcom.2017.1700066

Guok, C., Robertson, D., Thompson, M., Lee, J., Tierney, B., & Johnston, W. (2006). Intra and Interdomain Circuit Provisioning Using the OSCARS Reservation System. *2006 3rd International*

Conference on Broadband Communications, Networks and Systems, 1–8. https://doi.org/10.1109/BROADNETS.2006.4374316

Hassan, F. ul, Le, T., & Lv, X. (2021). Addressing legal and contractual matters in construction using natural language processing: A critical review. *Journal of Construction Engineering and Management*, *147*(9), 03121004. https://doi.org/10.1061/(asce)co.1943-7862.0002122

Hsiao, H.-L., & Tang, H.-H. (2024). A study on the application of generative AI tools in assisting the user experience design process. In *Artificial Intelligence in HCI* (pp. 175–189). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60611-3_13

Kalpage, H. (n.d.). *Outshift*. Outshift by Cisco. Retrieved May 23, 2025, from https://outshift.com/blog/jarvis-agentic-platform-engineering-outshift

Kidwai-Khan, F., Wang, R., Skanderson, M., Brandt, C. A., Fodeh, S., & Womack, J. A. (2024). A roadmap to artificial intelligence (AI): Methods for designing and building AI ready data to promote fairness. *Journal of Biomedical Informatics*, *154*, 104654. https://doi.org/10.1016/j.jbi.2024.104654

Kozina, A. (2024). Data transformation review in deep learning. *CEUR Workshop Proceedings*, *3716*. https://www.wir.ue.wroc.pl/info/article/UEWR05a965c7711849849f790bcbb943c8c4

Kumar, S., Datta, S., Singh, V., Datta, D., Kumar Singh, S., & Sharma, R. (2024). Applications, challenges, and future directions of human-in-the-loop learning. *IEEE Access: Practical Innovations, Open Solutions*, *12*, 75735–75760. https://doi.org/10.1109/access.2024.3401547

LangChain. (2025, May 4). *How Outshift by Cisco achieved a 10x productivity boost with their Agentic AI Platform Engineer*. LangChain Blog. https://blog.langchain.dev/cisco-outshift/

Liang, W., Tadesse, G. A., Ho, D., Li, F.-F., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence*, 1–9. https://doi.org/10.1038/s42256-022-00516-1

Lindemann, B., Maschler, B., Sahlab, N., & Weyrich, M. (2021). A survey on anomaly detection for technical systems using LSTM networks. *Computers in Industry*, *131*(103498), 103498. https://doi.org/10.1016/j.compind.2021.103498

Liu, F. (2024, August 2). *Prompt Controls in GenAI Chatbots: 4 Main Uses and Best Practices*. Nielsen Norman Group. https://www.nngroup.com/articles/prompt-controls-genai/

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, *56*(4), 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

Pai, S. (2024, January 26). UX considerations for generative AI apps and agents. *Google Cloud Blog*. https://cloud.google.com/blog/products/ai-machine-learning/how-to-build-a-genai-application

Papageorgiou, K., Theodosiou, T., Rapti, A., Papageorgiou, E. I., Dimitriou, N., Tzovaras, D., & Margetis, G. (2022). A systematic review on machine learning methods for root cause analysis towards zero-defect manufacturing. *Frontiers in Manufacturing Technology*, *2*.

https://doi.org/10.3389/fmtec.2022.972712

Priyadarshni, S. (2024). AI-driven document automation and compliance in contract lifecycle management. *2024 International Conference on Communication, Control, and Intelligent Systems (CCIS)*, 1–6. https://doi.org/10.1109/ccis63231.2024.10931892

Quarantiello, L., Marzeddu, S., Guzzi, A., & Lomonaco, V. (2024). LuckyMera: a modular AI framework for building hybrid NetHack agents. *Intelligenza Artificiale*, *18*(2), 191–203. https://doi.org/10.3233/ia-230034

Rafique, D., & Velasco, L. (2018). Machine learning for network automation: Overview, architecture, and applications [invited tutorial]. *Journal of Optical Communications and Networking*, *10*(10), D126. https://doi.org/10.1364/jocn.10.00d126

Roy, D., Zhang, X., Bhave, R., Bansal, C., Las-Casas, P., Fonseca, R., & Rajmohan, S. (2024, July 10). Exploring LLM-based agents for root cause analysis. *Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering*. FSE '24: 32nd ACM International Conference on the Foundations of Software Engineering, Porto de Galinhas Brazil. https://doi.org/10.1145/3663529.3663841

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. In *arXiv [cs.AI]*. https://doi.org/10.48550/ARXIV.2505.10468

Sharma, A., Li, X., Guan, H., Sun, G., Zhang, L., Wang, L., Wu, K., Cao, L., Zhu, E., Sim, A., Wu, T., & Zou, J. (2023). Automatic Data Transformation Using Large Language Model - An Experimental Study on Building Energy Data. *2023 IEEE International Conference on Big Data (BigData)*, 1824–1834. https://doi.org/10.1109/BigData59044.2023.10386931

Silva, F. S. D., Neto, E. P., Oliveira, H., Rosario, D., Cerqueira, E., Both, C., Zeadally, S., & Neto, A. V. (2021). A survey on long-range wide-area network technology optimizations. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 106079–106106. https://doi.org/10.1109/access.2021.3079095

Singh, D. (2024). Legal documents Text Analysis using Natural Language Processing (NLP). *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 1302–1307. https://doi.org/10.1109/icssas64001.2024.10760929

Sinha, R., Patil, S., Gomes, L., & Vyatkin, V. (2019). A survey of static formal methods for building dependable industrial automation systems. *IEEE Transactions on Industrial Informatics*, *15*(7), 3772–3783. https://doi.org/10.1109/tii.2019.2908665

Stevens, R., Taylor, V., Nichols, J., Maccabe, A., Yelick, K., & Brown, D. (2020). *AI for science: Report on the department of energy (DOE) town halls on artificial intelligence (AI) for science*. Argonne National Laboratory (ANL). https://doi.org/10.2172/1604756

Tadi, V. (2021). Revolutionizing data integration: The impact of AI and real-time technologies on modern data engineering efficiency and effectiveness. *International Journal of Science and Research (Raipur, India)*, *10*(8), 1278–1289. https://doi.org/10.21275/sr24709210525

Tateishi, T., Yoshihama, S., Sato, N., & Saito, S. (2019). Automatic smart contract generation using controlled natural language and template. *IBM Journal of Research and Development*, *63*(2/3), 6:1–6:12. https://doi.org/10.1147/jrd.2019.2900643

Tehsin, A. (2023, October 10). *Best practices for building collaborative UX with Human-AI partnership*. Microsoft Learning. https://learn.microsoft.com/en-us/community/content/best-practices-ai-ux

ter Beek, M. H., Gnesi, S., & Knapp, A. (2018). Formal methods and automated verification of critical systems. *International Journal on Software Tools for Technology Transfer: STTT*, *20*(4), 355–358. https://doi.org/10.1007/s10009-018-0494-5

Wang, J., & Duan, Z. (2024). Agent AI with LangGraph: A modular framework for enhancing machine translation using large language models. In *arXiv [cs.CL]*. https://doi.org/10.48550/ARXIV.2412.03801

Wang, M., Cui, Y., Wang, X., Xiao, S., & Jiang, J. (March-April 2018). Machine Learning for Networking: Workflow, Advances and Opportunities. *IEEE Network*, *32*(2), 92–99. https://doi.org/10.1109/MNET.2017.1700200

Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B. (2021). Machine learning in network anomaly detection: A survey. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 152379–152396. https://doi.org/10.1109/access.2021.3126834

Wankvar, T. (2024). *Automatic root cause analysis of network failure on IP/MPLS network using machine learning and case-based reasoning* [Dataset]. Thammasat University. https://doi.org/10.14457/TU.THE.2024.102

Weisz, J. D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024). Design Principles for Generative AI Applications. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, *10*, 1–22. https://doi.org/10.1145/3613904.3642466

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generations Computer Systems: FGCS*, *135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014

Xu, D., Wang, Y., Meng, Y., & Zhang, Z. (2017, December). An improved data anomaly detection method based on isolation forest. *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*. 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou. https://doi.org/10.1109/iscid.2017.202

Xu, H., Pang, G., Wang, Y., & Wang, Y. (2023). Deep isolation forest for anomaly detection. *IEEE*

Transactions on Knowledge and Data Engineering, *35*(12), 12591–12604. https://doi.org/10.1109/tkde.2023.3270293

Yang, K., Liu, R., Sun, Y., Yang, J., & Chen, X. (2017). Deep network analyzer (DNA): A big data analytics platform for cellular networks. *IEEE Internet of Things Journal*, *4*(6), 2019–2027. https://doi.org/10.1109/jiot.2016.2624761

Yang, Z., Cui, Y., Li, B., Liu, Y., & Xu, Y. (2019, July). Software-defined wide area network (SD-WAN): Architecture, advances and opportunities. *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. 2019 28th International Conference on Computer Communication and Networks (ICCCN), Valencia, Spain. https://doi.org/10.1109/icccn.2019.8847124

Yen, C.-C., Sun, W., Purmehdi, H., Park, W., Deshmukh, K. R., Thakrar, N., Nassef, O., & Jacobs, A. (2022, April 25). Graph neural network based root cause analysis using multivariate time-series KPIs for wireless networks. *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary. https://doi.org/10.1109/noms54207.2022.9789858

Yin, J., Hines, J., Herron, E., Ghosal, T., Liu, H., Prentice, S., Lama, V., & Wang, F. (2024). chatHPC: Empowering HPC users with large language models. *The Journal of Supercomputing*, *81*(1), 1–27. https://doi.org/10.1007/s11227-024-06637-1

Zhang, H., Wang, Z., Lyu, Q., Zhang, Z., Chen, S., Shu, T., Dariush, B., Lee, K., Du, Y., & Gan, C. (2024). COMBO: Compositional world models for embodied multi-agent cooperation. In *arXiv [cs.CV]*. https://doi.org/10.48550/ARXIV.2404.10775

Zhao, L., Alhoshan, W., Ferrari, A., Letsholo, K. J., Ajagbe, M. A., Chioasca, E.-V., & Batista-Navarro, R. T. (2022). Natural Language Processing for Requirements Engineering. *ACM Computing Surveys*, *54*(3), 1–41. https://doi.org/10.1145/3444689

Zhu, H., Gupta, V., Ahuja, S. S., Tian, Y., Zhang, Y., & Jin, X. (2021, August 9). Network planning with deep reinforcement learning. *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. SIGCOMM '21: ACM SIGCOMM 2021 Conference, Virtual Event USA. https://doi.org/10.1145/3452296.3472902

Acronym Glossary

AAA	Authentication, Authorization, and Accounting
ACL	Access Control List
AI/ML	Artificial Intelligence / Machine Learning
AIOps	AI Operations
API	Application Programmable Interface
ARIMA	Autoregressive Integrated Moving Average
ARP	Address Resolution Protocol
BER	Bit Error Rate
BGP	Border Gateway Protocol
BMP	BGP Monitoring Protocol
CCPA	California Consumer Privacy Act
CI/CD	Continuous Integration / Continuous Deployment
CLI	Command Line Interface
CRM	Customer Relationship Management
CUI	Controlled Unclassified Information
DMS	Document Management System
DNA	Deep Network Analyzer
DNA	Infinera Digital Network Administrator for OLS
DNS	Domain Name System
DOE	Department of Energy
DTN	Data Transfer Node - A specialized server designed to facilitate high-speed and reliable data transfer between different locations over the network.

ESDB	ESnet Database - An internal database that serves as a source of truth for all the physical and logical attributes of the network.
FEC	Forward Error Correction
GAN	General Adversarial Networks
GDPR	General Data Protection Regulation
HITL	Human-in-the-Loop
ID	Identifier
IDE	Integrated Development Environment
IdP	Identity Provider
IID	Independent and Identically Distributed
IP	Internet Protocol
IPAM	IP Address Management
KPI	Key Performance Indicator
LBL	Lawrence Berkeley National Laboratory
LDAP	Lightweight Directory Access Protocol
LIME	Local Interpretable Model-Agnostic Explanations
LLM	Large Language Model
LSTM	Long-Short Term Memory
MAC	Media Access Control
MCP	Model Context Protocol
ML	Machine Learning
MLOps	Machine Learning Operations
MOU	Memoranda of Understanding
MTBF	Mean Time Between Failures

NDA	Non-Disclosure Agreement
NERSC	National Energy Research Scientific Computing center
NLP	Natural Language Processing
NOC	Network Operations Center
OIDC	OpenID Connect
OLS	Open Line System
OWL	Web Ontology Language
PCAP	Packet Capture
PII	Personally Identifiable Information
PIPE	Policy, Innovation, Practices, and Engineering - An ESnet technical talk series.
PM	Polarization Measurement
PO	Purchase Order
R&E	Research and Education
RAG	Retrieval-Augmented Generation
RBAC	Role-Based Access Control
RCA	Root Cause Analysis
RFP	Request for Proposal
RMA	Return Material Authorization
RPA	Robotic Process Automation
SARIMA	Seasonal Autoregressive Integrated Moving Average
SHAP	Shapley Additive exPlanations
SLA	Service Level Agreement
SLE	Service Level Expectation
SNMP	Simple Network Management Protocol

SN	ServiceNow - A cloud-based enterprise technology platform that helps businesses automate and streamline workflows.
SP	Service Provider
SVR	Support Vector Regression
TLP	Traffic Light Protocol
TNMS	Telecom Network Management System
TTR	Time to Resolution
UX	User eXperience
VPN	Virtual Private Network
WAN	Wide Area Network
WG	Working Group
WP	Work-Package

Appendices

Appendix A1. ESnet Data and AI Workshop 2025 Agenda

Day 1 - Feb 25		
8:30am - 9:00am	Opening remarks	Chin Guok
9:00am - 10:55am	Session 1.1: Understanding the questions / problems	Chris Tracy / Bruce Mah
10:55am - 11:15am	Break	
11:15am - 12:30pm	Session 1.2: Understanding the questions / problems	Chris Tracy / Bruce Mah
12:30pm - 1:30pm	Lunch	
1:30pm - 3:00pm	Session 2.1: Understanding our data	Ed Balas
3:00pm - 3:20pm	Break	
3:20pm - 5:00pm	Session 2.2: Understanding our data	Ed Balas
5:00pm - 5:30pm	Day-1 closeout	Chin Guok
6:30pm - 8:00pm	Dinner @Comals - Prioritization of Session 1&2 outcomes	
Day 2 - Feb 26		
8:30am - 9:00am	Recap from Day-1	
9:00am - 10:30am	Session 3.1: Understanding AI - presentations	John Wu / Arpit Gupta
10:30am - 10:50am	Break	
10:50am - 12:30pm	Session 3.2: Understanding AI - Q&A	John Wu / Arpit Gupta
12:30pm - 1:30pm	Lunch	

1:30pm - 3:00pm	Session 4.1: Bringing it together, building work-packages	John MacAuley
3:00pm - 3:20pm	Break	
3:20pm - 5:00pm	Session 4.1: Bringing it together, building work-packages	John MacAuley
5:00pm - 5:30pm	Day-2 closeout	Chin Guok
6:30pm - 8:00pm	Dinner @Angelinas - Prioritization of Session 4 outcomes	
Day 3 - Feb 27		
8:30am - 9:00am	Recap from Day-2	
9:00am - 11:00am	Session 5.1: Where do we go from here	Chin Guok / Chris Tracy
11:00am - 11:20am	Break	
11:20am - 12:00pm	Session 5.2: Where do we go from here	Chin Guok / Chris Tracy
12:00pm - 12:30pm	Closing remarks	Chin Guok

Appendix A2. ESnet Data and AI Workshop 2025 Attendees

ESnet staff:

Edward Balas	ESnet Software Engineering - Measurements & Analysis
Sowmya Balasubramanian	ESnet Software Engineering - Measurements & Analysis
Justas Balcas	ESnet Planning & Innovation - Pilots & Prototypes
Shawn Brown	ESnet Network Services - Site Solutions
Dale Carder	ESnet Network Services - Peering & Upstream Connections
Evangelos Chaniotakis	ESnet Software Engineering - Orchestration & Core Data
Christopher Cummings	ESnet Cybersecurity - Security Engineering
Eli Dart	ESnet Science Engagement
Patrick Dorn	ESnet Network Services Technical Lead
Samir Faci	ESnet Software Engineering - Measurements & Analysis
Sukhada Gholba	ESnet Software Engineering - Measurements & Analysis
Brooklin Gore	ESnet Systems - Business Automation
Chin Guok	ESnet Leadership
Micheal Haberman	ESnet Cybersecurity - Vulnerability & Threat Management
Derek Howard	ESnet Planning & Innovation - Pilots & Prototypes
Dylan Jacob	ESnet Network Services - Core Routing
James Kafader	ESnet Software Engineering - Measurements & Analysis
Ezra Kissel	ESnet Planning & Innovation - Advanced Technologies & Testbeds
Yatish Kumar	ESnet Planning & Innovation - Pilots & Prototypes
Shawn Kwang	ESnet Software Engineering - Platform Engineering
Andrew Lake	ESnet Software Engineering - Measurements & Analysis
John MacAuley	ESnet Planning & Innovation - Pilots & Prototypes
Bruce Mah	ESnet Planning & Innovation - Planning & Architecture
David Mitchell	ESnet Network Services - ENOC
Samuel Moats	ESnet Network Services - Data and Facilities Services
Inder Monga	ESnet Leadership
Samuel Oehlert	ESnet Cybersecurity - Security Engineering
Scott Richmond	ESnet Software Engineering - Orchestration & Core Data
Chris Robb	ESnet Network Services - ENOC
Jason Roeckle	ESnet Software Engineering - Orchestration & Core Data
Cody Rotermund	ESnet Network Services - Site Solutions
Stacey Sheldon	ESnet Planning & Innovation - Pilots & Prototypes
Michael Sinatra	ESnet Network Services - Core Routing
Eric Smith	ESnet Network Services - Engineering Automation & Tools
Joshua Stewart	ESnet Network Services - Peering & Upstream Connections
Garrett Stewart	ESnet Software Engineering - Orchestration & Core Data
Jiachuan Tian	ESnet Planning & Innovation - Planning & Architecture
Christopher Tracy	ESnet Planning & Innovation - Planning & Architecture

Swikritee Wagle	ESnet Business Office
Paul Wefel	ESnet Network Services - Data and Facilities Services
Anne White	ESnet Network Services - Data and Facilities Services
Brendan White	ESnet Planning & Innovation - Planning & Architecture
Andrew Wiedlea	ESnet Science Engagement
John Wu	ESnet Planning & Innovation - Advanced Technologies & Testbeds
Xi Yang	ESnet Planning & Innovation - Pilots & Prototypes
Seyoung Yu	ESnet Planning & Innovation - Advanced Technologies & Testbeds

External guests:

Claudionor Coelho, Jr.	Chief AI Officer, Zscaler
Arpit Gupta	Professor, University of California, Santa Barbara, and Berkeley Lab Faculty Scientist working with ESnet
Satyandra Guthula	PhD student, UCSB College of Engineering
Sangeetha Abdu Jyothi	Assistant Professor, University of California, Irvine
Taghrid Samak	Engineering Manager, Meta
Vyas Sekar	Professor, Carnegie Mellon University
Raghu Subramanian	Founder and CEO, Nonadoo
Walter Willinger	Chief Scientist, NIKSUN

Appendix B1. WP01 (Alerting)

Problem Statement

Today, not all services have complete monitoring nor a consistent methodology for alerting and availability tracking. This creates several problems in operations, where gaps in visibility require humans to watch dashboards to detect problems. Often they use human intuition or ad hoc reasoning to assess indicators of degradation over time. In service planning, a lack of availability statistics makes it difficult to understand the reliability of lower level services, making it challenging to estimate a new service's possible availability target or set improvement goals for existing services. Monitoring inconsistency leads to divergent approaches to alerting across services and teams, often increasing costs of services and training time. Additionally, without understanding the operational interdependencies between services or components, our ability to reason about root cause is impeded. Alerts may not have sufficient actionable information regarding an outage or degradation, requiring extra time for human operators to gather the necessary data to solve the problem.

Broadly speaking there are six classes of datasets that are useful inputs: **network connectivity and performance data** (such as network reachability or instantaneous network throughput); **host- and system-level data** (such as CPU usage or disk utilization); **application-specific performance metrics** (such as Kafka message queue lag); **service instance specific dependency graphs**; a comprehensive set of configuration changes to system components; and a complete data source that defines hosts, nodes, applications etc.

Known Constraints

Some of the data is considered sensitive within a certain context and further analysis to assess sensitivity is needed. The solutions need to be usable even when the network itself is impaired, implying an expectation of > 99.9% uptime and minimal dependence on external services.

Additionally, for some deployment contexts we have to contend with privacy laws that may restrict data collection and usage.

Opportunities and potential solutions

Ideally we could develop a shared pattern across departments and services to build greater understanding in the organization and increase overall effectiveness. A potential solution would involve generating measurement and monitoring configurations through a system that can define collection and monitoring policy declaratively. This would combine with a source of truth to generate monitoring configs and include post-processing of raw alarms that uses dependency metadata to interpret the broader context of related alarms.

Automated (non-human monitored) alerting is sensitive to false positives. There is an opportunity to build analysis middleware to catch and label anomalies in a common manner for monitoring sensors of specific types, such as time series, histograms, scalar values, and pareto graphs. This approach would create a synthesized quality or soft failure metric based on inferring service behavior.

Problems are often identified by a combination of individual sensor measurements (CPU utilization, message rate and downtime notification of a dependent resource). ESnet has an opportunity to build these dependencies using a more generic middleware framework, possibly including AI techniques.

An intelligent system that can reason about service dependency, topology and active alarms would help Incident Responders more quickly observe, orient and prioritize incidents to ensure optimal resolution.

In these proposed solutions, users would first visualize and monitor data from service measurement sensors. They would then automate the analysis of measurements to flag anomalies. Once programmed, the data would be automatically processed with limited human intervention. Finally, users would synthesize the anomaly results from individual sensors to build a more intelligent alerting system.

Gap Analysis

As we continue to evolve and improve our environment, we've identified areas where we can enhance our capabilities. One key aspect is having a comprehensive and up-to-date inventory of services and service instances across ESnet. This would enable us to more efficiently generate configurations, define policies, and analyze issues when they arise. Additionally, we see an opportunity to expand our monitoring and measurement capabilities to get a more complete picture of each service.

We're also working to address some data gaps that are hindering our ability to implement a comprehensive solution. For example, we could benefit from having more complete and readily available datasets, as well as policy-based definitions for monitoring and alerting. Furthermore, having a system to translate monitoring data into actionable insights would be valuable. We're also exploring ways to better represent the complex relationships between services and their interdependencies, which would facilitate more thorough analysis.

Another area we're focusing on is improving our configuration management. Having a centralized dataset of configuration changes across various components, such as routers, hosts, and applications, would allow us to better understand the impact of changes and identify potential correlations with outages. This would also involve integrating with tools like Ansible and NSO, and monitoring key file changes on systems.

Lastly, we're interested in exploring new techniques for anomaly detection that can help us identify potential issues before they become major disruptions. By leveraging advanced analytics and machine learning, we aim to develop more effective automatic detection methods that can alert us to subtle changes in our environment.

Appendix B2. WP02 (Rules Correlation)

Problem Statement

Correlating host-based rules, firewall configurations, network device ACLs, and route tables is highly challenging during connectivity troubleshooting due to fragmented data sources, lack of a unified view, and insufficient contextual insights. This leads to prolonged downtime, inefficiencies, and difficulty in identifying root causes, even when issues originate outside ESnet. It's a very common occurrence to have to iterate through multiple steps owned by multiple groups when setting up a service to make sure it can connect to the proper upstream and downstream services/systems.

The lack of a unified view and automated correlation capabilities results in manual correlation via CLI, spreadsheets, and firewall UI, **which wastes an inordinate amount of engineer time** and is prone to errors. This leads to incomplete metadata, inconsistent data formats, and limited analytics tools for complex network configurations. Furthermore, the current state of data availability is incomplete, with missing traffic logs, outdated topology data, and limited visibility into different datasets depending on which group an engineer is in.

Broadly speaking, there are several classes of datasets that are useful inputs: firewall rules (iptables, ACLs, Panorama), route tables (IS-IS, BGP, static routes), host configurations (IP addresses, interfaces, host-based routing tables for management), traffic logs (Flowdata, packet captures, Zeek conn logs, high touch), network topology (device interconnections, subnets), and blackhole routes (SCRAM). Additionally, historical incident data is required for training models.

Known Constraints

Some of the data is considered sensitive within a certain context, and further analysis to assess sensitivity is needed. The solutions need to be usable even when the network itself is impaired, implying an expectation of > 99.9% uptime and minimal dependence on external services. Technical constraints include fragmented data sources, limited metadata, and hardware limitations.

From a legal and governance perspective, data privacy compliance is a concern. The solution must ensure that sensitive network data is handled in accordance with relevant regulations, such as GDPR and CCPA. This may require additional measures, such as data encryption, access controls, and auditing.

Opportunities and Potential Solutions

Ideally, we could develop a network configuration correlation tool that automates the mapping of network rules, traffic patterns, and device configurations, enabling us to troubleshoot connectivity issues more effectively. Three potential solutions are:

Correlation Engine: This solution involves graph-based modeling of network configurations and anomaly detection for conflicting rules. The correlation engine would analyze data from various sources, including firewall rules, route tables, and traffic logs, to identify potential issues and provide recommendations for remediation.

Automated Rule Validator: This solution uses predictive modeling to flag misconfigured rules and provides a CLI plugin for real-time validation during rule changes. The automated rule validator would help reduce the risk of human error and ensure that network configurations are consistent and accurate.

NLP Reachability Analysis: This solution applies an NLP interface to enable natural language questions, such as "Can host A reach host B on port C?" The NLP reachability analysis would provide a user-friendly interface for network engineers and operators to query the network and quickly identify potential issues.

These solutions would provide a unified, real-time visualization and analysis platform that reduces troubleshooting time and enhances ESnet's value proposition as a trusted network advisor. The platform would enable network engineers and operators to quickly identify potential issues, prioritize remediation efforts, and improve overall network reliability and performance.

Gap Analysis

As we continue to evolve and improve our environment, we've identified opportunities to enhance our capabilities. One area for improvement is having a more integrated and cohesive approach to data management. Currently, our data repository is fragmented, and our metadata could be more consistent. Additionally, we could benefit from more advanced analytics tools to help us better understand complex network configurations and improve data availability.

To make progress in this area, we're exploring ways to collect and normalize data from various sources, such as firewall logs, route tables, traffic captures, and host configurations. We're also interested in developing a system that can translate monitoring data into actionable insights and provide a clearer understanding of the relationships between services and their interdependencies. Furthermore, we're looking into ways to leverage advanced analytics and machine learning to detect potential issues before they become major disruptions.

The development of a network configuration correlation tool is a complex challenge that requires careful consideration of several factors, including data analytics, machine learning, and software development. Such a tool would need to be able to handle large amounts of data, perform sophisticated analytics and modeling, and provide real-time visualization and alerts. It would also need to be highly available, scalable, and secure, with robust access controls and auditing capabilities.

While this is an ambitious project, we believe that the potential benefits are significant. A network configuration correlation tool could help us improve network reliability and performance, reduce downtime, and give our engineers and operators more visibility and control. This, in turn, could free up resources and enable our team to focus on more strategic and innovative work.

Appendix B3. WP03 (Data Quality)

Problem Statement

Security operations are hindered by the lack of structured, normalized, and consistently available data across critical infrastructure. This data fragmentation impedes rapid and accurate incident response. This makes it difficult to answer fundamental operational questions such as: where a host was last seen, who owns a particular IP or MAC address, or which devices are active on specific VLANs or switch ports.

Security teams often rely on manual correlation of disparate systems—such as DNS, LDAP, SN, and ESDB. Each contains inconsistencies, outdated records, or missing entries making it difficult to identify devices no longer present on the network, or validate service ownership and configuration data during investigations. This undermines efforts to create automated detection and response workflows, slows investigations, and creates blind spots that could be exploited during a security event.

Broadly speaking, there are several classes of datasets that are essential for improving incident response capabilities:

- **Network telemetry and traffic data** (Zeek, NetFlow/High Touch, Suricata, OpenTelemetry, Pcaps)
- **System and infrastructure logs** (Syslog, application logs, firewall logs)
- **Host and identity data** (MAC/ARP/ND/IP/VLAN/Port mapping, LDAP, DNS, ESDB)
- **Virtual infrastructure data** (VMware)

While many of these data sources exist—particularly within Splunk or other observability platforms—they often lack interoperability or completeness. Enabling AI/ML models to reason over this data will require improved normalization, correlation, and completeness across these diverse inputs. This is no easy task, but almost any problem that is deemed important enough to solve with AI/ML/LLM will require its data to be pre-processed/normalized.

Known Constraints

It would require a significant amount of resources to solve this problem. Compliance with government mandates such as M-21-31 and M-22-09 (Zero Trust) imposes strict requirements on logging and data visibility. Legal obligations, including GDPR and CCPA, require careful handling of sensitive data, necessitating strong access controls, audit mechanisms, and data minimization practices.

From a technical perspective, many existing data stores suffer from poor metadata tracking and incompatible formats, complicating efforts to unify them into a single, cohesive system. Sensitive contract-related information and mixed Trust Level Protocol (TLP) data—such as TLP:red intermixed with TLP:clear—further increase the risk of unintentional exposure.

Operationally, there are known risks in integrating with legacy systems and overcoming resistance to changes in data governance or incident response workflows.

Opportunities and potential solutions

The primary opportunity lies in improving incident response by auditing, normalizing, and centralizing existing security-related data sources—many of which already reside in Splunk.

While the raw data is largely available, inconsistencies, gaps, and broken integrations limit its usefulness. An initial focus on auditing log coverage and health across sources will help establish a reliable foundation for both automated and manual analysis.

Several potential solutions include:

- **Log Coverage Auditor:** A tool to identify missing or broken data sources across the logging pipeline, ensuring full visibility into hosts, network flows, and service activity. This would enable baseline completeness assessments and help prioritize remediation.
- **Normalization Pipeline:** A system to clean, enrich, and standardize disparate log formats—potentially using tools like Cribl—to ensure consistent structure and semantic alignment across Zeek, Suricata, syslog, firewall logs, and DNS data. This would improve downstream analysis and simplify cross-source queries.
- **Centralized Data Access Strategy:** While some data types (e.g., PCAPs) may reside outside Splunk due to storage or access constraints, there is an opportunity to build indexable metadata or reference stubs within Splunk to improve discoverability and correlation.
- **Flexible Interaction Modes:** Analysts require both automated workflows (such as saved searches, alerts, and dashboards) and ad hoc investigative capabilities. The solution should support both modes, enabling proactive alerting as well as reactive exploration during live incident response.

These solutions would reduce time-to-insight for investigations, ensure higher data fidelity across security workflows, and enable future integration of AI/ML for correlation and anomaly detection.

Gap Analysis

As we continue to refine our security analytics capabilities, we've identified areas where we can improve our data management and incident response processes. While Splunk provides a solid foundation for many of our essential datasets, there are still some gaps that we need to address to achieve comprehensive incident response and proactive security analysis. Specifically, some key data types, such as MAC/ARP/ND/IP/VLAN/Port configurations, are not consistently ingested or structured in a way that allows for easy analysis. Additionally, we're working to improve the integration of PCAP data with Splunk's architecture to unlock its full potential. Furthermore, we're exploring ways to better leverage our service registry, which reflects intended service configurations, to enhance our security analytics.

To make progress in this area, we'll need to focus on auditing and validating our data collection processes, as well as implementing robust normalization procedures to ensure data consistency. Enhancing metadata collection across all data sources will also be crucial for enabling reliable

correlation and analysis. By addressing these areas, we can improve the quality of our data, enhance our incident response capabilities, and lay a solid foundation for leveraging AI/ML in security analytics. This will ultimately help us to better protect our systems and improve our overall security posture.

Appendix B4. WP04 (Lifecycle)

Problem Statement

Organizations struggle to track the relevance and staleness of business documents and contracts, leading to outdated decision-making, compliance risks, and operational inefficiencies. Current document management systems (DMS) lack sufficient metadata, centralized tracking, and visibility into the lifecycle of documents such as contracts, MOUs, purchasing agreements, and business records.

Desired Outcome: Establish a centralized mechanism that tracks lifecycle metrics (e.g., last updated, version, access patterns, approval status), flags stale or outdated documents, enforces accountability, and leverages AI/ML to enhance document relevance scoring.

Target Users:

- Document Managers
- Legal/Compliance Teams
- ESnet Staff (Procurement, NOC, Engineering)
- Business Office & LBNL Procurement
- Archetypes: “Compliance Enforcer,” “Document Curator”

User Experience Goals:

- Immersive or Assistive interfaces depending on user roles
- Interfaces embedded within existing workflows (e.g., Ask Gemini in Google Docs)
- Human-in-the-loop validation for AI-powered recommendations

Relevant Data Sources:

- Document metadata (e.g., author, version, last updated)
- User interaction logs (access frequency, edits)
- Legal/compliance policy referencesBusiness repositories (Google Docs, Confluence, FMS, etc.)
- Jira tickets, Git logs, email/slack content (future expansion)
- Unstructured formats (PDFs, Excel files)

Known Constraints

- Legal & Privacy Compliance: GDPR/CCPA regulations must be followed; documents may contain sensitive or financial data requiring restricted access.
- Technical Limitations:

    - Existing DMS platforms lack robust metadata tracking or document lifecycle awareness.
    - No centralized view of Google Docs and other file repositories.
    - Integration challenges with legacy systems and data silos.
- Organizational Barriers:
    - Resistance to changing established workflows.
    - Fragmented ownership and inconsistent cleanup of outdated materials.
    - Lack of institutional knowledge regarding legacy documents.
- Security Considerations:
    - No unified permissions model for AI/ML document access
    - Risk of over-privileged access (e.g., global superuser solutions for metadata crawling)

Opportunities and Potential Solutions

A more detailed summary is available in the table below. AI/ML Timeline (Projected):

- Year 1: Proof-of-concept on single source (Google Docs or Confluence)
- Year 2: Expansion to broader systems (Git, Slack, Jira, etc.) with per-source integration

Solution Area	Description	Data Needs	User interaction
Metadata Tagging System	Standardize and enrich metadata across systems	Document logs, user access history	UI alerts, editable annotations
Automated Reminder Engine	Notify stakeholders about stale or unapproved documents	Scheduled updates, stakeholder lists	Email/IM reminders
AI Relevance Scoring	Use NLP and ML models to predict staleness or flag irrelevant content	NLP features, historical metadata, and model feedback loop	Visual indicators in DMS
Statistical Summary Techniques	Univariate statistical methods (e.g., thresholds, usage frequency) to identify out-of-date documents	Timestamp metadata, access logs	Dashboard summaries

Document Clustering	LLMs group similar or related documents for better navigation and searchability	Unstructured data, content embeddings	AI-powered “related documents” pane Data Governance Framework
Data Governance Framework	Define retention policies, ownership, and archival strategies	Cross-source metadata, legal flags	Admin interfaces for governance rules
Central Repository and Search	Create federated access/search across Google Docs, Confluence, SharePoint, etc.	API integration and search indexing	Unified query interface

Gap Analysis

Current State	Desired State
Incomplete or missing metadata	Standardized, real-time metadata tagging
No centralized tracking or accountability	Integrated lifecycle monitoring and logging
No link between documents and business context	Contextual awareness via linked tickets, edits, and user roles
Limited document accessibility	Federated search across platforms
Unstructured and redundant content	Enriched and deduplicated via AI/UX enhancements

As we continue to refine our data management processes, we've identified opportunities to enhance our metadata and data governance capabilities. Specifically, we see potential for improvement in the following areas:

- Developing a shared metadata standard to ensure consistency and clarity across our datasets, including versioning, ownership, and approval status.
- Implementing an accountability layer to track edits and reviews, ensuring transparency and accountability.
- Enhancing metadata fields to include important information, such as version history and expiration dates.

- Improving data cleanup processes to ensure that multiple document versions are properly labeled and managed.
- Streamlining access and permissions across tools to simplify data access and management.
- Establishing infrastructure and policy for managing relevance scoring and LLM data access to ensure that our data is accurate and reliable.
- Developing a test harness or validation process for automation and AI recommendations to ensure that our systems are functioning as intended.

In order to realize our vision, we must take into account the following considerations that will shape our approach:

- Understanding user behavior and intent through log analysis and access patterns is an ongoing challenge that requires careful consideration.
- Protecting sensitive or privileged document content while ensuring seamless access is a delicate balance that must be struck.
- Designing intuitive UI components that provide timely and relevant feedback without overwhelming users is crucial for a positive user experience.
- Establishing trust in AI/ML recommendations through iterative human-in-the-loop feedback loops is essential for ensuring the accuracy and reliability of our systems.

Appendix B5. WP05 (Data Catalog)

Problem Statement

Here are two example use cases behind this particular work-package:

(1) As an ESnet staff member I want an **up-to-date** view of what datasets collected by the org are available, (also: and are available to me specifically). This will require ongoing maintenance.

(2) I want to know what type of data is in each collection, the schema, the last updated timestamp, responsible party for data collection and curation, if there is an API or other access method, what that is.

The driver behind this need is that often the data that we need to do our jobs might already exist and we would not know. In addition we will need to know the data we have so that other ML tooling can make use of it.

Known Constraints

Some datasets will contain sensitive information, and the visibility or existence of those datasets itself will need to be controlled.

The mere existence of some datasets may be sensitive information.

Opportunities and Potential Solutions

ESnet proposes leveraging commercially available ML tools to crawl the various data storage locations (structured databases, documentation, servicenow, Google docs, business systems) and produce an inventory of those datasets periodically and on-demand.

The tools will need a starting point, much like the data sources spreadsheet collected right before the summit.

Gap Analysis

Today we do have the aforementioned spreadsheet that was collected, at the expense of some effort, that has probably limited accuracy and is missing important information about how the datasets could be accessed / used by humans or other tools.

ESnet will need more metadata about our datasets, including access policies, schemas, APIs, etc. In the case where these datasets exist in some sort of database system, the metadata could be collected; in cases where the datasets are informal (i.e. wikis, spreadsheets, etc), the metadata will need to be created.

Appendix B6. WP06 (Network Services)

Problem Statement

ESnet provides guaranteed bandwidth as a service toourcustomers. The following are three different usage scenarios:

(1) As **a network engineer**, I am interested in understanding the specifics of these commitments and how the service is being utilized.

(2) As **an ESnet Engagement staff member**, I would like to understand the commitment details and service usage to assist in coordinating the various science projects.

(3) As **an IRI staff member or as senior leadership** (ESnet, DOE, ASCR), I would like to have access to this information to comprehend how the service is being utilized.

In all these scenarios, the following information would be desired:

- The number of customers utilizing the bandwidth service.
- The amount of guaranteed bandwidth and its duration.
- The percentage of the allocation utilized by each customer.
- Whether there have been any service violations (where ESnet was unable to meet the bandwidth commitment). If so, when and by how much?

To address these inquiries, we will primarily require the following types of data:

- A list of business agreements (SLA/SLO) if any, with guaranteed bandwidth commitments made to customers.
- Data from OSCARS and other telemetry sources to monitor current usage and identify any service violations.
    - The OSCARS data can assist in identifying the entities utilizing the service, the number of circuits with guaranteed bandwidth created, and the percentage of the allocation utilized.
    - Queue drops from telemetry data can aid in identifying potential service violations.

Known Constraints

None identified. If IP addresses or some other data is considered sensitive, we may need to find a way to filter it or provide some role based access.

Opportunities and Potential Solutions

There's an opportunity to establish a process and an intake form to handle bandwidth guarantee requests. Once this is in place, the form data can be automatically imported into other systems as needed. We can use Google form or ServiceNow for this purpose.

Next, there's a chance to export the statistics available in OSCARS and make them accessible in a dashboard. For dashboard visualization, we can leverage some existing open source tools like Grafana. API access to the data is also desired, as it will enable integration into other applications. Potentially, the data from OSCARS can also be exported to Stardust. This will enable users to correlate the data from OSCARS with other telemetry data. Stardust also provides an API for accessing the data.

Once the data is available in some common repository, we can focus on analysis. Statistical libraries can be used to analyze the data. Descriptive statistics can be employed to summarize bandwidth consumption trends, time-series models can track usage patterns, and forecasting models can predict usage. Correlation/regression analysis can identify relationships between SLAs, utilization, and performance. If API access is made available, it can be integrated into alerting and other applications.

Finally, there's an opportunity to develop visualizations for all the metrics being collected and analyzed.

Gap Analysis

One of the most significant challenges is the absence of a formal service intake process and a common repository for existing Service Level Agreements (SLAs). Establishing this will be the initial step.

Furthermore, the usage and telemetry data are currently distributed across various sources. To ensure data accessibility and consistency, we must identify a suitable integration method or import data from one source into another. Additionally, implementing role-based access control will be necessary to restrict data access to authorized users only.

Appendix B7. WP07 (Business Ops)

Problem Statement

As a business ops staff, there is a challenge in the mapping of our services/inventory to sub-contracts/POs for invoice validation and procurement renewals. We want a defined database or a centralized dashboard of curated data for the Business Office to cross reference from POs and invoices.

There are various databases but no clear direction on which should be referenced for the most up to date information relevant to the Business Office.

Data is being pulled from the following:

- ServiceNow
- FMS
- BAR
- ESDB
- Google Drive

Known Constraints

None within ESnet.

Opportunities and Potential Solutions

Create a dashboard linking data from multiple sources that are already in place (i.e. ESDB and ServiceNow), e.g., link POs to circuit ID, cross-connects, colocation via information on ESDB (vendor, contact, end date, services)

Gap Analysis

One area for improvement is ensuring data consistency and accuracy across our systems. Currently, we have multiple systems in place, but there is a need for clearer connections and mappings between them. This makes it challenging to understand how information from one source relates to others.

Additionally, we could benefit from establishing processes to ensure data is regularly updated and validated. It would also be helpful to clearly define roles and responsibilities for data management and maintenance, to ensure that everyone knows what is expected of them.

Appendix B8. WP08 (Outage Notification Parsing)

Problem Statement

Planned and unplanned outage notifications from network providers (e.g., Lumen, Internet2, GÉANT) typically arrive as emails. These notifications are manually processed and entered into the relevant systems, such as ServiceNow, ESDB, or an orchestrator application, which can be time-consuming and error-prone. If a provider later updates or cancels the notification, staff must also manually edit the corresponding records in these systems.

For a Network Engineer or NOC Engineer, having these maintenance notifications automatically parsed and entered into ServiceNow (and other inventory systems) would reduce human error and speed up operations. Likewise, any updates received from providers would be efficiently propagated to keep all records in sync. A related need is the ability for other software components, such as orchestrators, to access this parsed dataset of outage and maintenance information. By recognizing devices or circuits that are offline or undergoing maintenance, these systems could proactively reschedule or defer network provisioning tasks.

From a data standpoint, this work-package would rely on:

- Outage notifications from providers.
- Inventory data from ServiceNow, ESDB, and other systems.

Known Constraints

Some providers may consider circuit IDs, outage details, or other information sensitive or covered by NDAs. Consequently, relevant legal and policy considerations need to be addressed. The automation system must ensure that all confidential or proprietary data is processed and stored securely. Otherwise, there do not appear to be major hardware or software constraints, but any new tool must integrate seamlessly with existing ticketing and inventory systems.

Opportunities and Potential Solutions

One proposed solution is to harvest emails (e.g., from trouble@es.net or a Google Groups API feed) and feed them to an AI-based or automated parsing tool. This parsing engine would convert the email content into a structured format, which would then trigger the creation of related change records via the ServiceNow API. Ideally, the parser would also identify and link specific items mentioned in the notification (such as circuits or devices) to their corresponding entries in ServiceNow.

Once records are automatically created and updated, Network Engineers and NOC staff would still have full visibility into ServiceNow, but far less manual work would be required. Additional automation tools could leverage these updates to avoid provisioning or reconfiguring devices known to be offline, thereby reducing failures and improving workflow efficiency.

Gap Analysis

Available data sources appear sufficient for the initial stages of automating outage notifications. ServiceNow and related systems contain enough information to map provider circuit references to the correct configuration items (CIs). Nevertheless, ensuring consistent naming conventions and robust lookups is key for accurate matching.

It is also critical to validate that any changes, extensions, or cancellations from providers can be captured and updated across all relevant systems without delay. In some cases, if the data in ServiceNow or ESDB is incomplete, additional updates or new data flows may be required. Otherwise, the existing data appears adequate to support an automation solution that can significantly reduce manual ticket entry and error rates.

Appendix B9. WP09 (Ticket Resolution)

Problem Statement

ESnet has a long history of troubleshooting a diverse range of unexpected issues on its network and systems footprints. Many of these efforts have their solution documented via a ticket or other documentation sources. For new incidents, ESnet would like to be able to pull from this previous experience to surface potential solutions for staff to attempt instead of beginning the troubleshooting process from scratch. The following are two more examples:

As a **NOC engineer** I would like the ability to **analyze prior solutions to trouble tickets** (issues) allowing me to quickly **determine a resolution** to a new but similar problem.

As a **Network Engineer** I would like the ability to **analyze prior solutions to trouble tickets** (issues) allowing me to quickly **determine a resolution** to a new but similar problem.

These datasets could be used to solve this problem:

- ServiceNow trouble ticket data.
- Jira tickets related to trouble tickets.
- Slack channel discussions.
- E-mail conversations
- Zoom meeting transcripts (when captured)

Known Constraints

The data sources that contain information about troubleshooting may also contain sensitive information about ESnet or its sites infrastructure. Care must be taken to ensure that this information isn't shared externally or pulled into a LLM as training material. Since this is proposed to be a human-in-the-loop activity, the risk of poor guidance is mitigated somewhat, but there is a small chance that more junior staff will potentially go down the wrong path, potentially leading to lengthening the resolution or worsening the effects of the incident.

The ticket export data may overwhelm the AI model and some of our cloud-based systems may limit the amount of data we can get out of it.

Ticket data may contain PII, IP addresses or information about site access that we wouldn't want to broadcast externally. Ticket data may include information for no-show sites that violate our agreement with the site or expose infrastructure information.

We must maintain customer confidentiality, and have special handling constraints for sensitive site data.

Additionally, the issues below could potentially impede or block progress on the efficacy of a new tool:

- Limited unstructured data in the tickets can sometimes make it hard to determine the solution.
- Structured ticket data may not be filled in properly / left unpopulated.
- No formalized mechanism for capturing the context for why we made a change that may have resulted in this problem.
- Decentralized data sources might present a challenge for programmatic analysis by an AI-based tool

Data sources related to the troubleshooting activity may contain sensitive data:

- Tickets and data sources may contain customer sensitive information.

Opportunities and Potential Solutions

The team envisioned an interactive command prompt or customized web interface focused on troubleshooting issues using information from previously solved tickets, live data queries, and sources-of-truth.

- Formalize ServiceNow ticket and Jira issue structure to better capture problem description and solution steps. This information will be the primary source of knowledge for problem resolution.
- The problem space could potentially be reduced by pre-culling or filtering ticket/Jira data to only focus on issues where a solution was actively made by ESnet. (e.g. don't populate with tickets where vendors made a solution and only focus on issues where ESnet flags that they implemented a solution)
- Introduce an audit step into the ticket/Jira lifecycle to validate information populated before allowing the ticket/Jira to be closed. Short term pain for long term gain to verify consistent and complete data is populated.
- Feed ticket/Jira and related slack channel discussions into target solutions to provide knowledge base. Would this give enough context to get reasonable responses to trouble queries? Hand curation of slack data may be needed to identify correct solutions to problems to feed only valid solutions into the model.
- Introduce Model Context Protocol (MCP) type technology to bring AI technologies and provide dynamic query capability onto live data to help drive system responses with additional context. Should help reduce the need to log into multiple GUI to troubleshoot and train systems to understand what dynamic data can help solve a particular problem.
- Vendors may already have AI-based solutions that could assist, though they may be disparate systems that would need an engineer to visit multiple tools and they may be cost prohibitive.

The NOC engineer would like to interact with the system through either a command prompt or web interface tailored to troubleshooting activities.

Gap Analysis

The following gaps will need to be addressed as part of development of knowledge-based query solution:

- Can we define troubleshooting workflows to determine possible optimizations and interdependencies?
- A proof of concept will need to be developed to prove feasibility of the proposed solution.
- Define a common ticket/Jira format with fields that help clearly identify the problem, symptoms, and steps to resolution.
- Introduce procedural steps into teams working on trouble resolution to verify proper documentation was provided before allowing the ticket to be closed.
- Determine what information needs to be fed into the solution to provide the desired results. Train the models with this information.
- Build MCP integrations for the identified troubleshooting information to allow the system to perform dynamic data retrievals.
- Ticket Quality Assurance is a difficult task to do at scale. Additionally, the distributed nature of support at ESnet means that a wide range of staff in different teams need to be aware of the standards with enough accountability to enforce consistency.

All needed datasets are available today, however, some of the datasets are incomplete (ticket resolution information), or inconsistent (hand curated ESDB data). Will need to put an effort in to normalize data field names so they correlate between data sources being ingested.

Appendix B10. WP10 (Correlate Alarms)

Problem Statement

Here is a motivating example for this work-package: As a **NOC enginee**r I want the ability to rapidly **correlate alarms and maintenance notifications** to **real-world (customer-facing/ESnet-facing) impacts**.

More broadly, determining the affected services for an ongoing outage is one of the biggest difficulties, mainly due to data related issues:

- Multiple alarm feeds with correlation issues between feeds.
- Decentralized and inconsistently labeled network topology data sources.
- No process for periodic audit of data for correctness.
- Lack of data structure / incompatible data structures in tools - in some cases can't determine if a customer has protected or unprotected services.
- Severity of alarms / alarm noise
    - Relationships of alarms to other data structures;
    - Grouping based on Incident (Alarm X relates to Circuit Y and Service Z).
- Signal-to-noise ratio issues (high noise) for logging.

datasets needed to address the problem include:

- Alarm data
- Infrastructure documentation that ties back to service impact
- A defined availability model to evaluate outages against for categorization (e.g. Up, Impaired, Down, etc.)

Known Constraints

These issues currently block us from building such a tool:

- No way to programmatically interpret the unstructured information imparted by vendor/site maintenance notifications
- Data inconsistency/missing in network documentation
- Service instance database does not (currently) exist which makes correlation a challenge to implement across the board.

Data sources related to the troubleshooting activity may contain sensitive data.

Opportunities and Potential Solutions

- Existing inputs: Spectrum, TNMS, Syslog, LibreNMS, e-mail notifications, phone calls, self reported
- The newly created monitoring working group may want to consider using AI for some of the alarm correlation logic that David already has built into his auto-ticketing scripts to accept more alarm feeds and use better logic to determine the root cause.
- Ideally any solution will present itself within existing workflow tools (e.g. within ServiceNow) and be available to future tools as yet defined.
- The user should be able to query any generated response to potentially refine the answer or explore the response more fully.

Gap Analysis

While we have a significant amount of data available, there are opportunities to improve its structure, completeness, and accuracy. For example:

- Our network model data may benefit from additional validation and verification to ensure its accuracy and completeness.
- We are working to develop a comprehensive service catalog that will help us better understand the relationships between different components, although some areas may still require refinement.
- Notifications from sites and peers may sometimes lack sufficient detail, which can make it challenging for our systems to quickly identify potential impact sources.

Appendix B11. WP11 (Predict Hardware Failures)

Problem Statement

Networking equipment hardware failures are disruptive events that create significant service impacts. Although completely avoiding hardware failures on operational equipment is unlikely, there are often early warning signs that can predict a hardware failure event, captured in seemingly unrelated logging and telemetry data.

If hardware failure likelihood could be predicted with reasonable accuracy, then proactive steps would be taken to replace at-risk hardware in advance of a failure, reducing unplanned outages and improving overall network stability.

If successful, this work-package would provide the following outcome:

- Network and NOC Engineer users of this solution would be presented with a report of at-risk hardware (line cards, transponders, power supply units, pluggable optics, etc) based on log and telemetry data analysis. The users would then be able to proactively schedule maintenance events to replace at-risk hardware before a failure event occurs, monitor the health over time of network equipment, and make informed decisions to prioritize hardware upgrades based on the estimated life of deployed hardware.

The following datasets would be necessary to fulfill this work-package:

1. Optical Performance Metrics to provide voltage levels to determine if more power is being required to maintain signal integrity.
2. Transponder and transceiver bit error rate (BER) counts used to to detect a change in behavior over time.
3. Forward Error Correction (FEC) state changes indicating an increase in error correction needed to maintain usable connectivity on circuits.
4. Hardware SNMP and Syslog data reporting current equipment state and changes over time, including temperature, voltage, error counts, memory usage, and other important system details. This includes realtime streaming telemetry from all levels of the hardware stack on network and operational facility equipment.

Known Constraints

It is likely that a vendor will not authorize a proactive Return Material Authorization (RMA) before a hardware failure or significant performance degradation has occurred. However, it may be possible to cycle out hardware based on failure predictions to spares pool, or negotiate an RMA agreement with a vendor based on the predicted impacts.

Syslog data can be proprietary, however optical metrics should not be considered sensitive.

Opportunities and Potential Solutions

For a syslog-based solution, event facility and severity will be analyzed. For certain device classes, all log messages will be published for analysis. Algorithms must be developed to analyze these syslog streams for events that precede hardware failures, such as escalating severity logs or specific triggers like increased packet error rates.

For optical equipment, Optical Performance Monitoring (PM) will be utilized to develop an algorithm that detects deterioration based on increasing voltage levels, elevated Bit Error Rate (BER), and changes in Forward Error Correction (FEC) compensation. It is acknowledged that distinguishing between deterioration in an optical fiber versus the laser may be challenging in some situations.

Required datasets:

1. Stardust and LibreNMS for transponder and overall circuit health metrics, as well as overall SNMP data collection
2. Netlog/Syslog data for general hardware state and logs, as well as circuit and service performance logs
3. DNA for Open Line System (OLS) metrics not captured by Stardust and LibreNMS

User Interactions could include a report or dashboard indicating the most likely hardware to fail, with direct links to the data used to perform the analysis, ensuring that the tool “shows its work” and the prediction accuracy can be verified by a user based on historical outage and degradation examples.

Gap Analysis

Further investigation into existing datasets is required to identify key indicators of hardware failure. Algorithms will be developed to analyze these target metrics for predictive purposes. Additionally, a system for capturing, displaying, and linking alarms to source time-series data may be necessary

Further investigation and development may be required to retrieve and store OLS PM Data to a centralized location for analysis. Performance measurements for new hardware types (such as coherent optics) would be required to set a performance baseline before sufficient production data has been collected. Finally, some amount of historical forensics would be required for model training to correlate past failures/degradations with collected data.

Appendix B12. WP12 (Detect External Configuration Anomalies)

Problem Statement

Efficient routing of network traffic between independent network entities depends on the sharing of configuration and network data between autonomous networks or “peers”. However, a misconfiguration of data shared from one network to another can result in network disruptions or degradations as the receiving network “peer” implements routing decisions based on the data obtained from the adjacently connected network.

If successful, this work-package would provide the following outcome:

- Network and NOC Engineers would be presented with an analysis tool that generates a report or a running log of received network configuration data from peers, enumerating data path changes taking place based on the configurations received from the connected peer, highlight routing path changes from the point of view of the network control plane, and notify NOC and Network engineers of detected configuration anomalies and their impact on the current network state.

datasets needed to address the problem include:

1. Received routing updates such as BMP, received routes, and BGP preference updates.
2. External third party tools such as routeviews and PeeringDB to provide a global network peering view.
3. Flow data to detect unexpected or unintended traffic flow changes.
4. SNMP/Netlog data to highlight unexpected or unintended reductions in traffic at the hardware level.
5. Geo-locating IP address to correlate between regions and ensure that peering changes align with best practices for efficient network routing across the world.
6. Configuration validation via Orchestration tools to ensure received configurations generate the expected outcomes in advance of implementation.

Known Constraints

This solution has the potential for broader application across various networks, including potential open-source or commercialization opportunities. While specific data sources may be network-proprietary, deployment in other National Research and Education Networks (NRENs) would provide additional data perspectives and enable cross-correlation of events. Summarized event data, with a lower Traffic Light Protocol (TLP) classification, could facilitate sharing between partner networks. However, differing naming and data conventions between networks may present challenges.

Data sources related to this analysis activity may contain sensitive data.

- Flow data is private and needs to be considered sensitive.

Opportunities and Potential Solutions

Initial implementation will focus on individual data sources, with cross-correlation to be addressed in future iterations. Emphasis will be placed on monitoring ingress traffic to detect external network changes.

Specifically:

1. **Flow Data:** Leverage the existing Stardust ingest pipeline, map flow to ingress interfaces, consider IP address aggregation (/24, /48), and detect traffic shifts between interfaces, generating loggable events.
2. **BGP Data:** Monitor ingress BGP attributes (e.g., AS-path, MED) for changes, maintain a watchlist of specific attributes, and generate events upon detection of changes. Implement mechanisms to manage attribute thrashing. Utilize existing router table reflector logs.
3. **SNMP Data:** Monitor ingress interface counters via Stardust for near-zero counts, triggering events upon threshold crossings.

Required datasets:

1. **Flow Data:** Capture network ingress flow data and associated interfaces. Detect IP traffic shifts between interfaces. Generate structured log entries for event analysis.
2. **BGP Data:** Monitor ingress BGP attributes (e.g., AS path, MED). Generate structured log entries for BGP attribute changes.
3. **SNMP Data:** Monitor interface counters. Generate structured log entries for interface packet count changes.

Gap Analysis

1. **Flow Data Analysis:** Implement application logic to monitor flow data and detect changes in /24 and /48 prefixes per interface.

2. **BGP Property Monitoring:** Develop application logic to monitor and detect changes in specific, predefined BGP properties.

3. **SNMP Threshold Detection:** Implement SNMP threshold crossing detection for ingress port packet counts.

4. **Log Management:** Address the challenge of high-volume, low signal-to-noise "firehose logs" to improve misconfiguration identification.

The majority of the necessary data exists today, and is stored and accessible. Specifically, Flowdata and SNMP interface packet counts are available in Stardust and Clickhouse, Meta-data is available

from ingress ports directly connected to peers, and BGP AS path information and attributes are available on route reflector logs and external peering data bases such as routeviews and PeeringDB.

As an opportunity, control plane updates will need to be ingested and stored in a long term structured format that can be utilized for analysis and training.

Appendix B13. WP13 (Capture Configuration Intent)

Problem Statement

For engineers working on system configuration changes, they would like to associate configuration intent to record context associated with a code or metadata change to speed up the configuration process, and eliminate blind spots.

Target users include system administrators, or network engineers responsible for equipment in their domain. Potentially NOC or on-call staff looking at an incoming trouble ticket and trying to correlate the issue with the change that precipitated and thus the intent for that change.

Slurp in data from the underlying data sources and put it into a frontend tool that can provide a query interface where one could look up a string, a Git tag, a filename, or other context information that would then be used to find all the source of truth that link to the configuration process. The contextual information also includes trouble tickets, Jira issues, and wiki documentation that help build the configuration intent.

The following datasets would be needed:

1. Git in code repositories (code, ansible)
2. LDAP Updates (userdb and hostdb)
3. Jira
4. wiki/confluence
5. Google Drive Files
6. ESDB

We manage configuration of all of our own routers and have a history of the configuration changes.

Known Constraints

Data from Git repos need access control. The same applies to other data sources as well. Generally speaking, the data are sufficiently accessible, as the access is internal only.

One catch is that there is no clear way to associate a ticket with the resulting changes. Fields would need to be added.

Opportunities and Potential Solutions

There is a hierarchy of configuration files that could be examined for developing the tool for understanding the intent. No exact tool for this problem yet. There are research papers on discovering specifications (specification mining) for software development, and this work could borrow from those research:

- https://www.carolemieux.com/texada-ase15_final.pdf

- https://www.usenix.org/conference/nsdi20/presentation/birkner
- https://github.com/batfish/pybatfish/

Other related works:

- https://www.usenix.org/system/files/nsdi20-paper-birkner.pdf
- https://conferences.sigcomm.org/hotnets/2023/papers/hotnets23_sharma.pdf
- https://www.carolemieux.com/texada-ase15_final.pdf
- https://github.com/batfish/pybatfish/
- https://web.cs.ucla.edu/~varghese/research/nsdi20.pdf
- https://raghavan.usc.edu/papers/sage-sigcomm21.pdf
- https://ieeexplore.ieee.org/document/6032596

Gap Analysis

We need to better understand how to describe a configuration “intent”. There is on-going work on correctness and validation of network configurations. We need to figure out what tools are actually usable for ESnet.

We need to develop an ontology of intent for network engineering. Human intent is typically expressed in natural language, therefore LLM and similar tools are likely useful in this context.

This would be an entire application most likely because you would need some sort of query interface and a way to view historical data. Humans are only involved as users, they don’t have to insert data to the system. They can review the quality/results vs. blindly accepting the answer as “the truth”.

Appendix B14. WP14 (Fast Contract Lookup)

Problem Statement

Here are two routine tasks from ESnet that are currently time-consuming:

1. As a **Network Engineer or a DAFT team member**, I would like to look up contract information to determine which contract is applicable to a **specific instance of equipment**.
2. As a **Network Enginee**r or **a business office staff member**, I would like to **perform billing validation** by verifying that all **charges billed by a vendor are valid**. The invoices may include equipment, circuits, cross-connects, remote hands, licenses, and various other services.

To help completing these and similar tasks, we will need information from the following:

- Contract details / Invoices. This is available in Google sheets and some in Lab purchasing systems.
- ESDB for equipment information and some circuit information.

Known Constraints

Data sources related to this analysis activity may contain sensitive data. Therefore, some kind of role-based access may need to be implemented.

Opportunities and Potential Solutions

For the first use-case, we need the contracts and invoices to be ingested into some common database and be made available in a normalized searchable format. Users should be able to search using keywords and look up all the related information. There may be instances where they may also provide the service or equipment name and the relevant contract information will need to be retrieved. The services, equipment and other information should have a normalized identifier that makes it easy to search in other databases or repositories across ESnet. We may be able to leverage existing AI technologies like Gemini to process the invoices and/or contracts and retrieve the relevant information. Since Gemini is already enabled in the lab account, it can be easily leveraged to search files in Google drive and our emails. For other AI tools, we will need to make the data available in a secure manner. A natural language process chatbot can be used as the interface to this application.

ServiceNow also provides a contracts module that would make it easy to associate contracts to specific network elements, devices, vendors and companies (customers). This would require the data that is available in emails, documents and other places to be ingested into servicenow.

For the second use-case, in addition to the information from service contracts, we will also require information from our inventory database like ESDB. As a first step, the information in the database and the contracts need to be normalized. The information in the inventory database also needs to be accurate. We will also need to build a simple application or agent that can retrieve information from the service contracts (maybe some type of AI agent) and the inventory database and present the combined information.

Gap Analysis

Currently, the contracts are available in various formats - emails, google sheets and in lab purchasing systems. All the information needs to be integrated and normalized.

Another potential issue is the accuracy of data in the inventory database. The association of a PO to named/identified network elements (circuits, equipment) etc is also imperfect. Also, inventory data ownership is a gap.

Appendix B15. WP15 (Consistent Data Management)

Problem statement

Today, teams across the organization use different methods and formats for data management and analysis. This inconsistency leads to fragmentation in how systems are monitored, analyzed, and developed. For example, DAFT employs a distinct monitoring approach compared to NS, which itself diverges from the models used by PAG and Research teams. Without a shared and normalized data format, cross-team collaboration becomes more difficult, operational tooling must be customized per team, and core infrastructure components lack a unified view. This inconsistency creates friction in development, limits reusability, and slows the pace of innovation across the organization.

Furthermore, there is currently no common model for describing resources and their components. As a result, teams are unable to easily correlate data across systems, or build advanced applications that depend on a holistic understanding of infrastructure and services. Workstructure is intended to address this challenge by enabling the adoption of an organization-wide, consistent data model.

Broadly speaking, realizing this opportunity will require access to all available organizational datasets. This includes infrastructure metrics, resource inventories, monitoring telemetry, service dependencies, and any domain-specific information necessary to build a unified, composable model of the system.

Known Constraints

Some data sources involved in this analysis may include sensitive information, requiring further review to assess handling requirements. Not all systems will be able to adopt a shared format or ontology due to legacy constraints or technical limitations.

In such cases, exceptions may need to be defined, or data segmented into separate namespaces. Reaching alignment on a unified schema may also raise organization-wide debates, as teams have differing needs and perspectives.

Additionally, variations in access control models and potential automation challenges may introduce complexity that is not yet fully understood.

Opportunities and Potential Solutions

A consistent set of naming conventions, ideally mandated across domains, would provide a strong foundation for semantic alignment. To support this, an extensible ontology framework—such as OWL—could be used to formally define resource types and relationships while allowing for ongoing evolution of the schema.

To maintain flexibility, namespacing strategies could be introduced to accommodate team-specific extensions without disrupting the core model. This would enable parallel development within domains while preserving a shared structure across the organization.

Identifying and empowering a Domain Owner for each area represented in the Workstructure effort would be a critical step toward ensuring broad alignment and long-term sustainability. These individuals would act as stewards of both the model and its adoption, facilitating consensus and guiding schema evolution over time.

Importantly, the solution does not require full ontology coverage from the outset. Instead, an iterative approach would allow the system to evolve incrementally—delivering immediate value while expanding support for additional components and relationships as understanding deepens.

Gap Analysis

As we move forward in this area, we recognize the importance of addressing some key organizational considerations. While the technical aspects of this challenge are relatively well-understood, we've seen that aligning teams and stakeholders can be complex. One key area of focus is clarifying ownership and accountability for this domain, which will help ensure that everyone is working towards common goals.

Another important consideration is developing a framework for access control and data management, which will help us ensure that sensitive or restricted data is handled appropriately. Additionally, we need to establish a clear method for prioritizing areas of misalignment across teams, so that we can focus on the most critical issues first.

By addressing these organizational considerations, we can create a solid foundation for standardizing data and ontology, and ultimately drive greater interoperability and impact. This will require collaboration and alignment across teams, but we believe that by working together, we can make significant progress in this area.

Appendix B16. WP16 (Query All Data)

Problem Statement

ESnet engineers often need to search across a wide range of systems and documentation sources to answer operational or design-related questions. Today, this process is fragmented and time-consuming, requiring manual queries across disparate platforms such as Google Docs, internal wikis, Jira, Lucidchart diagrams, ESDB, Stardust, syslog, and ServiceNow. Each tool has its own interface and search syntax, making it difficult to extract and correlate relevant information efficiently.

The lack of a unified, intelligent search capability leads to duplicated effort, slower decision-making, and reduced situational awareness. Engineers often miss critical context that is buried in unstructured documents or siloed data stores, which can result in incorrect assumptions or redundant work.

To improve operational efficiency and decision quality, ESnet requires a system that allows engineers to query across all available data and documentation sources through a single interface. The results should be intelligently ranked by relevance and augmented with direct links to the most applicable supporting documents, logs, tickets, or diagrams. This would enable engineers to rapidly validate hypotheses, find authoritative references, and act with greater confidence.

Relevant datasets and systems include:

- **Documentation** (Google Docs, Confluence Wiki, Lucidcharts)
- **Operational records** (Jira tickets, ServiceNow, ESDB)
- **Telemetry and logs** (syslog, Stardust, NetFlow)
- **Topology and architecture diagrams**
  This unified query capability would support both routine operations and complex troubleshooting or planning tasks.

Known Constraints

This work-package proposes a unified query system that must operate within strict access and compliance boundaries. It must not access or expose Controlled Unclassified Information (CUI), Personally Identifiable Information (PII), or any other sensitive content that is restricted under federal or organizational policies. Furthermore, the system must not provide data to any external parties or systems outside of ESnet's administrative domain.

To ensure data security and privacy, the solution must support fine-grained access controls, allowing users to only view data and documents they are authorized to access. This necessitates robust user authentication and authorization mechanisms that can integrate with existing access management systems, such as SLP.

From a content readiness standpoint, unstructured documentation must follow consistent formatting standards to support reliable parsing, indexing, and future ingestion. For example, new wiki pages

may need to adhere to a predefined template to be fully queryable and semantically understood by the system.

These legal, technical, and procedural constraints must be addressed to ensure the system remains compliant, secure, and effective in a highly controlled operational environment.

Opportunities and Potential Solutions

There is a strong opportunity to streamline and enhance access to institutional knowledge by building an intelligent query interface powered by Retrieval-Augmented Generation (RAG) and relevance-based ranking (e.g., PageRank). This system would enable ESnet engineers to ask natural-language questions and receive ranked, contextualized answers, along with direct links to relevant documents and records.

A RAG-based approach would combine document embedding and semantic search with generative AI to synthesize responses from distributed, unstructured data. This would dramatically reduce the time engineers spend searching across platforms and help surface non-obvious relationships between historical incidents, configuration notes, operational logs, and service documentation.

The core datasets required include:

- **Documentation repositories** (Google Docs, Confluence Wiki, Lucidcharts)
- **Operational records** (Jira tickets, ServiceNow entries, ESDB entries)
- **Network telemetry/logs** (Stardust, syslog, NetFlow)
- **Historical incident summaries and engineering analyses**

End-users would interact with the system through a query interface capable of handling both keyword and natural-language input. A programmatic API would also support automation workflows, such as triggering queries in response to new incidents, or embedding the search capability within existing tools and dashboards.

For example, an engineer could ask, "Have we seen a similar BGP flap on AS293 months ago?" and receive a ranked list of prior incidents, relevant logs, and linked documentation—all without manually checking each tool.

This solution has the potential to significantly reduce cognitive overhead, increase operational agility, and preserve institutional knowledge for both current and future teams.

Gap Analysis

To realize the proposed solution, several key gaps must be addressed between the current and target state. Most notably, existing ESnet documentation and operational records are distributed across multiple silos—Google Docs, wiki pages, Jira, ServiceNow, and others—without a unified access layer.

This fragmentation makes it difficult to perform cross-platform searches or apply consistent indexing and access controls.

For a retrieval-augmented generation (RAG) system to function effectively, all relevant documents must be vectorized, requiring preprocessing, embedding, and storage in a searchable index. Today, this vectorization has not been performed, and many documents exist in unstructured or inconsistently formatted forms that hinder reliable ingestion.

Furthermore, access controls vary across platforms, and there is no centralized mechanism to enforce fine-grained user permissions or authentication across all data sources. A federated access control layer will be required to ensure that search results respect data sensitivity and user authorization constraints.

In summary, the gaps include:

- Lack of a centralized index across all documentation silos
- Unvectorized and inconsistently structured documents
- Absence of a unified access control framework across systems
- No existing infrastructure for real-time semantic ranking and retrieval

Bridging these gaps will be essential for enabling an intelligent, secure, and effective search experience that empowers ESnet engineers to quickly find and act on institutional knowledge.

Appendix B17. WP17 (Automating Site Deployment)

Problem Statement

Network engineers today **want to streamline site deployment by automating the pre-orchestration process, which includes setting up the device's base configuration, bootstrapping, and performing other essential configuration tasks.** By automating these pre-orchestration steps, Network engineers can reduce manual errors, increase efficiency, and improve overall site deployment reliability.

Known Constraints

Automating site deployment processes can be complex and challenging due to the numerous steps involved. Asynchronous execution of these steps may not always be feasible. Additional restraints may emerge through user interviews that need to be conducted with network engineers who regularly participate in site deployments, and these interviews will also inform the requirements for the solution design.To develop effective solutions for automating site deployments, we must first address the following constraints.

- Inventory Management: Ensuring accurate and up-to-date inventory data to support Bill of Materials (BOM) requirements.
- Staging and Shipping Coordination: Coordinating with ESnet staging and shipping processes to ensure timely and efficient delivery of equipment.
- Inter-Tool Communication: Resolving communication gaps between different tools, systems, and teams involved in the site deployment process.
- Access to Credentials: Securing access to necessary credentials for all relevant sites.
- Spreadsheet-Based Information: Consolidating and updating information currently stored across various spreadsheets to ensure consistency and ease of use.
- Limited Resources: Acknowledging the constraints posed by limited resources that impact the deployment process.
- Configurable Credentials: Noting that credentials located in the base configuration can also be sensitive, underscoring the need to manage access and control.
- Some IP info may be considered sensitive for the host site.

Opportunities and Potential Solutions

The automation of site deployment processes presents a unique opportunity to improve the way network engineers work. By leveraging hybrid AI solutions, we can automate parts of the pre-orchestration process, improve accuracy and efficiency in various tasks, and enable Network Engineers to focus on more strategic activities. Additionally, introducing an AI-assisted deployment agent can help streamline the process by providing real-time guidance and checklists, reducing the

risk of human error. We can provide feedback to this AI agent at the end of each site deployment for further improvements.

To enhance efficiency in site deployment, it is crucial to establish a clear and comprehensive framework. This involves meticulously documenting each step involved in the process, including requirements, dependencies, and constraints. By doing so, we can ensure that all stakeholders have access to up-to-date information, facilitating informed decision-making.

A potential proposed solution is to implement an AI agent for site deployment automation.The agent will need to enumerate the steps involved in the process, which will be based on **having access to high-quality documents that are up-to-date with all the steps**, including requirements, dependencies, and constraints that will be used by the AI agent to help assist network engineers. These steps and documents can be updated if the process changes and the AI assistant will update its workflow based on the updated documents. Having access to all the steps involved in the document, the **AI agent can plan and help the network engineer figure out what steps from the checklist can be performed asynchronously and what steps are currently blocked and automate the reservation of resources**. **It can also take additional actions on behalf of the network engineer that we deem safe.**

Then, we will define a standardized process that network engineers can follow, **identifying steps that might not be possible to automate**. Next, we will evaluate automation options for all the steps that could be automated and define processes for manual steps. By implementing this structured approach, we can develop an AI agent that streamlines site deployment processes, reducing errors and increasing productivity.

Gap Analysis

The current analysis is based on talking to a few network engineers, **to get a clear picture of how this process can be improved we need to conduct more user interviews and understand the gaps more clearly.** This also **seems to be a process issue more than an AI automation issue**. AI and automation can definitely help improve and add more efficiency but **standardization of things like IP Address block assignment** (i.e. loopback, management, etc.) and **Site naming schema** need to be addressed.

Appendix B19. WP19 (AI Sandbox)

Problem Statement

ESnet needs a “sandbox” environment where known ESnet datasets can be used with different AI models/techniques/workflows. Such that ESnet staff can learn about what different models/techniques/workflows are useful for different ESnet-specific datasets and purposes.

User Story: I want to start analysing ESnet data with an AI model, but since I don’t know which model is best, I need a sandbox space to develop ideas and come up with proof-of-concept(s) where we can evaluate different techniques against each other, using real ESnet data.

The data needed exists at ESnet. It needs to be collected and collated for development purposes.

Known Constraints

Some of the data may be sensitive. The data would need to be sanitized, or we would need to take other measures to ensure that ESnet data stays within ESnet. Subject matter experts would need to incorporate the data into the sandbox environment so that appropriate sanitization/redaction occurs.

The team who builds the sandbox would need to provision enough storage to hold the datasets, and a programming or execution environment for running the models against the datasets. Sizing this sandbox space may present challenges.

Opportunities and Potential Solutions

To begin, this might be implemented as:

- A storage system with datasets in a well-known directory structure
- A set of models
- Instructions on the wiki for how to run the models and feed them the data.

This might evolve over time as different models and/or datasets are incorporated into the sandbox environment.

It might be useful to explore whether/how LBL’s CBorg could be used with ESnet data. There may be sensitivity/privacy issues with this.

In the future, ESnet’s NERSC allocation might be used to train models on ESnet data, or do larger runs against larger datasets.

Finally, it would be great to have staff give PIPE talks describing their use of the sandbox and what they learned, to disseminate the knowledge within ESnet.

Gap Analysis

Existing gaps:

- datasets need to be provided by domain experts,

- AI Models need to be incorporated into the environment,
- Systems and storage resources need to be allocated,
- Documentation/instructions need to be written.

Appendix B20. WP20 (RFP/Contract Builder)

Problem Statement

Today, building RFPs and contracts is a time-consuming process that often requires significant manual input and coordination. NS engineers and management teams are frequently required to gather fragmented information from multiple sources in order to assemble complete documents. This leads to inefficiencies, delays in procurement cycles, and inconsistencies in how requirements and expectations are articulated across engagements.

Without a streamlined process or centralized source of truth, teams rely heavily on institutional knowledge or ad hoc decision-making to define technical specifications and scope. This increases the risk of incomplete or misaligned contracts, which can result in delivery delays, unmet expectations, or costly revisions after project initiation.

Broadly speaking, enabling a more efficient approach would require a structured and minimal-input method for generating complete RFPs and contracts. This could involve standard templates, reusable components, auto-populated fields based on service definitions or historical data, and clear workflows for validation and approval—reducing cognitive load and ensuring consistency across procurement activities.

Known Constraints

Some of the source data—such as RFP documents generated by DOE, ESnet, or partner organizations—is considered sensitive and typically classified as TLP:Green, as it is shared only with a limited set of vendors. RFP responses may contain even more sensitive information, particularly pricing details, and are often protected under NDAs and treated as TLP:Red.

Similarly, the output of any system handling this data, including drafted RFPs or contracts, inherits the same level of protection due to the inclusion of internal requirements and vendor-specific considerations.

Any solution in this space must account for these sensitivities and ensure appropriate handling, storage, and access controls.

Opportunities and Potential Solutions

Ideally, we could develop an AI-assisted framework to streamline the creation of RFPs and contracts, enabling teams to generate high-quality documents with minimal manual input while maintaining compliance and consistency. A foundational step would involve training a language model on a curated set of ESnet-sourced RFPs and contracts—particularly those originating from Network Services (NS)—to establish a domain-specific baseline.

To further enhance the system's effectiveness, we could incorporate additional high-quality RFPs and contracts from across LBNL that have proven successful in eliciting the desired vendor responses. These documents would serve as exemplars, reinforcing preferred structure, tone, and scope. The

language model would also be guided by a set of rule-based constraints to ensure adherence to LBNL-required phrasing and terminology—such as the appropriate use of “must,” “should,” and “shall.”

Optionally, and with careful attention to privacy and NDA constraints, the model could be extended with anonymized RFP responses. These would be evaluated in relation to their originating RFPs, allowing the system to learn patterns in successful versus less effective proposals. This feedback loop could refine the model’s understanding of how best to elicit high-quality vendor submissions.

At the user level, a natural language interface would enable ESnet staff or others to generate new RFPs by specifying the type of asset (e.g., router, optical system, server) along with any unique requirements. The system would analyze training data in conjunction with the provided input, assembling a tailored document by combining relevant sections from similar historical RFPs and modifying content where appropriate. For any ambiguous areas, it would proactively request clarification.

The resulting output would be a draft RFP or contract with inline scoring or color-coded indicators reflecting the model’s confidence and completeness for each section. This would guide reviewers in focusing their attention and finalizing the content more efficiently.

To support broader clarity, a structured summary would be generated alongside the document, offering both internal stakeholders and prospective vendors a concise, consistent overview of key requirements and objectives—reducing the risk of misinterpretation or misalignment during the bidding process.

Gap Analysis

Our current environment likely contains sufficient data to enable the development of an AI-assisted RFP and contract generation system—provided LBNL grants approval for its use. The internal repository of RFPs and contracts created by ESnet and NS teams offers a rich source of domain-specific content. However, there remains an open question about the utility of incorporating RFP responses into the training set. While these documents may offer valuable insights, they also introduce significant privacy concerns, particularly regarding sensitive pricing information protected by NDAs.

In scenarios where access to internal RFPs or contracts is limited, publicly available documents from peer institutions or similar agencies could serve as supplementary training material. However, these sources may lack alignment with LBNL-approved legal language and phrasing, reducing their effectiveness as primary inputs.

To increase the accuracy and adaptability of the model, several types of metadata would be highly beneficial. This includes the creation date of each RFP to trace the evolution of institutional language, and outcome-based confidence ratings—identifying whether an RFP resulted in a failed procurement due to unclear language or lack of viable bids, or whether it successfully attracted high-quality responses.

In addition, access to internal documentation—such as wiki pages or institutional guidelines that describe changes in required legal phrasing—would provide important context. These resources would help the system better interpret and apply evolving standards across documents, ensuring future outputs remain compliant with current expectations.

Appendix B21. WP21 (Unified Document Search)

Problem Statement

Currently, ESnet employees are faced with the challenge of searching for information scattered across various documents and knowledge bases. A global search interface would greatly simplify this process, allowing users to quickly find the answers they need. By consolidating these disparate sources into a single search platform, ESnet can improve collaboration, enhance decision-making, and foster innovation.

The ideal solution should be able to handle a wide range of data formats and quality levels, as well as accommodate varying levels of access control and security requirements. By leveraging AI or hybrid solutions, the search agent could learn from user behavior, refine its results, and provide increasingly accurate recommendations over time. By providing a unified search interface that spans ESnet's entire knowledge landscape, the organization can unlock new insights, streamline workflows, and drive progress in time-sensitive work.

Known Constraints

Several known constraints must be considered when developing this search agent. For instance, data-level access control is crucial, ensuring that only authorized personnel have access to sensitive data or documents. Additionally, query-level access control is necessary to limit the types of queries that can be performed on the data. Furthermore, integrating existing rules and restrictions from multiple data sources using RBAC/ACL considerations across systems will require careful consideration.

Additionally as most of this is internal we should not rely on any solution that requires sending our data out. We might want to limit ourselves to something we can self host.This is something we can align with the lab's policy based on legal advice.

Opportunities and Potential Solutions

Various AI methods could be employed to develop this search agent. For example, known AI/ML solutions like Google Gemini, docq.ai, meilisearch, NotebookLM, and RAG could be leveraged. A projected timeline for development and deployment would need to be established, taking into account the time required to generate our own AI solution or utilize off-the-shelf products. Technical challenges, such as data quality, computational resources, and query-level access control, must also be addressed. Furthermore, organizational challenges like staffing to support a chosen direction will require careful consideration.

Gap analysis

The biggest concern is how RBAC/ACLs and general security is handled with a model where multiple users have different access on different systems and how to avoid leaking data to unauthorized parties.

Potential approach. If we are able to get source material for the AI response, we could run everything behind a proxy to ensure that the user is validated and has access to all supporting information. Any response that includes a supporting document that the user does not have access to, is rejected and an AI query is reformulated to ensure the document is excluded.

Another potential approach would be for each underlying system to provide its own query interface and enforce answering queries with respect to the access control inherent in each underlying system. Then a higher level search tool would be making queries to the underlying systems somehow propagating the identity of the user performing the query.

Appendix B22. WP22 (Ticket Summarization)

Problem Statement

ServiceNow tickets often contain large volumes of unstructured and inconsistently formatted information. This data may include manual entries, vendor emails, and responses to those emails, frequently resulting in redundant or duplicated content. Such inconsistencies make it difficult to accurately understand the timeline of events or assess the current status of an incident. The goal of this effort is to generate a concise, organized summary that reflects the current state of the incident and the actions taken so far.

Target users are ServiceNow operators and staff who need situational awareness for incident response or managerial oversight. The expected interaction is through the ServiceNow interface, where users can request and view AI-generated summaries within the same ticket interface.

This functionality is especially useful during shift handoffs or for new engineers picking up ongoing incidents. Summarization could also support better documentation, especially where After Action Reports are created out-of-band and not attached back to original tickets.

Known Constraints

Data Sensitivity: Ticket data may contain personally identifiable information (PII) like names, phone numbers, addresses, and circuit IDs. Both input and output from summarization tools must be protected to avoid information leakage.

Data Quality: While ServiceNow tickets contain the needed information, fields like resolution notes are poorly maintained. There is no consistent enforcement of high-quality documentation.

UX and Integration Limits: Access is generally mixed or role-based. The exact interface granularity depends on the approach taken.

Privacy and Compliance: Any external tools must comply with ESnet's privacy standards and ensure sensitive data remains protected, especially in contracts.

Limited In-House Prompt Engineering Expertise: While ESnet can move data through APIs easily, an effective prompt remains a gap.

Opportunities and Potential Solutions

There are two proposed implementation paths:

1. **ServiceNow Paid Add-On**
    - Easy to deploy and serves as a reference baseline.
    - Unknown Cost and quality.
    - Fast path to be used by end-users.
2. **Custom Implementation**

- Leverages own familiarity with ServiceNow APIs.
- Downloads ticket data and formats it into prompts for LLMs, like CBorg or locally.
- Offers better control over privacy and customization of the experience.
- Can be implemented in parallel with the commercial option. Useful as a staff training opportunity.

Main data input: ServiceNow ticket logs, additional data from ESDB if needed. Users interact with the tool directly in ServiceNow. Statistical methods are **not applicable**. If successful, this project may set the grounds for broader AI/ML use.

Gap Analysis

The largest gap lies in crafting effective prompts to generate useful summaries. This limits the quality of results from even the best models. Despite access to data, fields like resolution notes are inconsistently populated. After Action Reports are rarely linked back to tickets, and missing information for context.

There is a lack of clarity about how tightly integrated the summarization UI should be - whether assistive, embedded, or optional. This affects implementation planning. Any approach involving third-party tools must be reviewed for data security compliance and contractual protections.

Appendix B23. WP23 (Federated Authentication)

Problem Statement

ESnet manages sensitive scientific data and HPC resources, and external collaborators require access to these resources for research. However, DOE labs enforce strict security, compliance, and usability constraints, making it challenging to balance security, productivity, and compliance.

The outcome of this work-package should be a system that can secure data sharing, allowing external collaborators to access ESnet resources without exposing sensitive data to unauthorized parties. This system should also enable streamlined collaboration, allowing researchers to focus on science rather than authentication hurdles, thanks to Single Sign-On (SSO) and Just-In-Time (JIT) provisioning. Furthermore, the system should ensure compliance, with all access controls aligning with government regulations. Additionally, the system should be scalable, allowing for easy onboarding of new collaborators, such as universities and international labs, via federations. Finally, the system should enable rapid detection and mitigation of breaches through advanced monitoring tools, providing effective incidence response.

To achieve this outcome, several datasets are required. These include user data such as affiliation, role, and clearance, as well as ACLs, which specify the resources and rules for access control. Federated Identity Metadata is also necessary, including metadata aggregators and protocol support. Dataset metadata, such as classification and ownership, is also required, along with logs for real-time monitoring and certificates. Finally, agreements, including Memoranda of Understanding (MOUs) and regulatory compliance documents, are also necessary to ensure that the system is properly configured and managed.

Known Constraints

There are several known constraints that need to be considered. From a legal perspective, the system must comply with DOE regulations and federal law, including CISA Directives. Additionally, there are hardware and software limitations that must be taken into account, such as the presence of legacy systems and potential incompatibility between federated Identity Providers (IdPs) and ESnet's Service Provider (SP). Furthermore, the system must also address concerns related to data sensitivity, including the protection of classified information, personally identifiable information (PII), and export-controlled data. These constraints highlight the need for a carefully designed and implemented system that can balance security, compliance, and usability requirements.

Opportunities and Potential Solutions

One potential solution is to enhance InCommon Federation Integration, leveraging the existing InCommon Federation and OIDC to streamline authentication for DOE labs already part of the federation while addressing gaps for non-InCommon labs. This solution would require several key datasets, including InCommon Metadata, which provides a list of participating DOE labs, their identity providers, and public keys. Additionally, User Attributes such as eduPersonPrincipalName, eduPersonAffiliation, and role-based attributes (e.g., researcher, admin) would be necessary, as well as

access logs that contain historical data on authentication attempts, failed logins, and resource access patterns.

The analysis involved in this solution would entail identifying DOE labs not yet part of InCommon and prioritizing onboarding based on collaboration frequency. This would also involve mapping user attributes to ESnet's Access Control policies and auditing existing SAML configurations for compliance with security standards. In terms of end-user interaction, this solution would enable SSO for InCommon labs via their institutional credentials, as well as multi-factor authentication enforcement for labs whose IdPs do not support it. Furthermore, the solution would provide real-time attribute validation to grant or deny access to datasets dynamically, ensuring that access is controlled and secure.

Gap Analysis

There are several gaps between the current state and the desired state. The current state is characterized by inefficiencies, such as manual authentication workflows for non-InCommon labs and inconsistent attribute release leading to access control mismatches. Additionally, there are insufficient logging and analytics capabilities, which hinder the ability to detect suspicious activity or optimize resource allocation, and lack data-driven decisions for policy refinement. Furthermore, there are missing data elements, including lab-specific authentication details, unified role definitions, dataset licensing terms, user feedback channels, and real-time security metrics. The required datasets are not all available today, and some are incomplete or of poor quality. There are also data sensitivity concerns and barriers that prevent the use case, and external access needs to be carefully managed.

To address these gaps, it is essential to collect and normalize data from various sources, develop a system to convert monitoring data into actionable alerts, and provide a well-structured representation of hierarchical topology and service interdependency. Additionally, it is crucial to implement effective automatic techniques to detect multi-variate anomalies that may indicate soft failures before they develop into major service disruptions. This will require a comprehensive approach that takes into account the complexities of the current state and the requirements of the desired state.

Overall, the development of a secure data sharing and collaboration system would be a complex and challenging project, requiring significant resources and expertise. However, the potential benefits of such a system would be substantial, including improved security.

Appendix B24. WP24 (Legacy Code)

Problem Statement

Motivating example: As an engineer I want to be able to remove and rewrite legacy perl scripts and programs that exist at ESnet. Many of these programs are used in Network Services (NS), but there are other scripts floating around as well.

Outcome: Perl programs at ESnet replaced with an equivalent software written in python, or another more-readable language.

Required dataset: The perl software at ESnet.

Known Constraints

The input scripts and program should not be exposed publicly. The resulting software should also not be publicly released, unless it is decided upon after-the-fact.

Opportunities and Potential Solutions

Ask an LLM with coding capabilities to rewrite the perl script(s).

- Input=perl script: rewrite perl script into python.
    - One can choose python, or another readable language.
    - Another possibility is to output “readable” perl that rewrites all the magic variables. This is to help understand what the initial software does.
- Input=perl script : Created unit-test and other tests to be able to validate what the script/program is doing. The goal is to also write unit tests for the converted program (e.g. python test) to verify the conversion.

End-user interaction consists of validating the new software. And then replacing the older software with the new ones.

Gap Analysis

The main gap is that the author(s) of the legacy code are no longer at ESnet. There is no institutional knowledge of exactly what the software did; and while some current staff members may be able to piece it together, that knowledge is fragmented.

There is a lack of a testing harness to validate any software conversion.

Lastly, once the software is converted, there needs to be an understanding as to who owns the new software. And how it will be maintained going forward, as to not repeat this problem in the future.

Appendix B25. WP25 (NLP Interfaces to Systems)

Problem Statement

Our systems provide Web, API and CLI interfaces to the ESnet engineers and customers to fulfill service orders and/or perform certain system functions. All users do not need to know coding, or API flow, or CLI commands to accomplish such tasks. By offering an NLP interface, we are not only able to enable such accessibility but also better understand users who sometimes may not even know what to ask.

The following datasets are needed:

1. System configuration and API manuals and documents (Vendors and ESnet)
2. SoT data from ESDB
3. ESnet wiki pages for related areas
4. History of configurations in gitlab etc.
5. Configuration automation scripts like Ansible playbooks in gitlab etc.
6. Example user intents in NL and configurations and scripts
7. ServiceNow users tickets for related topics
8. Select external configuration examples from GitHub etc.

Known Constraints

Risks:

- The users who require an NLP interface often may not be able to judge if their intent was correctly and precisely interpreted.
- It is unclear how a mechanism to enforce AAA policies for admission control can be integrated into the NLP interface.
- Training datasets pose a risk of leaking sensitive information through the tuned NLP model.
- The ambiguity of natural language may lead to uncertainty in interpretation when involved in legal issues.

Software Limitations:

- We are dealing with diverse types of systems and interfaces. Some are commodity while others are ESnet specific. The amount of work for model fine tuning is unclear.
- May require a large amount of work for training data preparation (documents and labeling).

Data sensitivity:

- Same as the risk for policy enforcement. System configuration data may be leaked via the NLP interface if users can interact with it with lots of questions and inquiries.

Opportunities and Potential Solutions

Opportunities for engineers:

- This saves a lot of training time for network engineering. Fluence with system configuration structure and/or programming to system interfaces takes a long time.
- Interaction with an NLP interface and iteration to refine the intent can increase clarity and reduce errors.

Opportunities for customers:

- This is a welcome addition to the existing web portal and API interfaces. It reduces the training time and helps clarify the user intent.
- An NLP interface may also serve as a single point of touch interface to connect users to the actual interface based on users' expressed intent in NLP.

Some use scenarios:

- A network engineer asks to create a VLAN interface or a VRF etc. in an NL description that is vendor agnostic.
- A network engineer asks to summarize ISIS or BGP link states across a region of the network without typing in a complex set of commands.
- A network engineer asks for creating an Ansible script to interact with the NSO.
- A scientist wants to reserve a network path for data transfer between network endpoints A and Z by providing an NL description of the intent.
- A scientist wants to connect their lab network to a public cloud through ESnet. They do not know where to start and just ask the NLP interface by providing some initial description and iterate with detailed asks until a service is composed, reviewed and committed.

Gap Analysis

- Where to start? What LLM product? How to fine tune?
- How to anonymize the training datasets?
- Documents may be missing or not structured for the fine tuning, especially for some ESnet's home brewed products.
- Mechanism for integration of user admission control and data access policy enforcement has yet to be investigated.

Appendix B26. WP26 (Information Architecture)

Problem Statement

Here are the example scenarios behind this work-package:

1. As **an engineer**, I can understand what systems produce data and which systems consume data so that I can improve data accuracy and synchronization between systems without duplication.
2. As **an engineer**, I can understand the context of the existing design before making changes so that I can avoid making changes that break dependencies.
3. As **an engineer**, I can understand the source of data so that when changes of data are detected or needed, I know where to effectively update data.
4. As **a software engineer**, I can understand who I need to work with to make schema changes so that I can add features or deprecate fields.
5. As **an engineer**, I can understand the taxonomy of each system so that we effectively communicate or translate without error.

The following information/datasets are needed to help addressing the above requests:

1. the list of existing datasets produced by different systems at ESnet. This includes the taxonomy of each system. For example, in some systems, we call an entity a Location, in others we call it a Site, or a Pop
2. For each system (or collection of systems) we need to know:
    a. What service the system provides
    b. Who authorizes changes to the system
    c. What are the inputs and outputs
    d. What are the dynamics of data updating (eg: API's, Manual, etc)
    e. Freshness requirements of workflows
3. We also need to know how data flows through each system to other systems.

Known Constraints

None identified.

Opportunities and Potential Solutions

To address the identified problem, several potential solutions have been proposed. One key solution is to develop a directed graph that illustrates data flow between systems. In such a directed graph, each node would represent a system, and each edge represents data relationships and flows, including metadata about how these interactions occur, such as initiation and transportation.

In addition, we need to maintain a detailed spreadsheet listing systems of record and their respective data sources and fields. Each system could expose its list of data sources, and all its fields. This will aid in auto-generation and ensure data integrity.

Thirdly, we need to implement a change control process for managing the addition, removal, and updating of systems and datasets.

Lastly, we can provide this document to large language models (LLMs) in understanding where to look for data and in what order is another important solution. Initially, low-tech solutions like an embedded wiki diagram can be used to map out data flow, before considering the development of more complex user experiences.

Gap Analysis

As we continue to refine our data documentation processes, we've identified opportunities to build on previous efforts. While there have been attempts to document datasets and their flows, we recognize that these efforts are incomplete and could benefit from further development. To take this initiative forward, it would be beneficial to designate a specific individual or team to own and maintain this documentation process. Additionally, having a management sponsor to provide support and guidance would help drive the initiative forward.

Another key aspect is ensuring that all necessary datasets, along with their supporting metadata, are available and complete. While spreadsheets have been used to document this information, we see an opportunity to consolidate and formalize these efforts. By doing so, we can create a more comprehensive and sustainable data documentation process that meets our needs.

Appendix B27. WP27 (Requirements Management)

Problem Statement

Over the past few years, ESnet has tried a variety of workflows to support requirements development and management of resources, via Engineering Design Reviews, etc. These processes have focused on identifying specific project needs for ESnet resources and to coordinate fulfillment across multiple ESnet teams.

These single project requirements review processes are not, however, integrated into an overall environment that facilitates an ESnet-system level view of project requirements - although we do have Jira supporting system level understanding of resources and schedules, along with use of Confluence to track requirements for some portions of ESnet. This tracing between Confluence and Jira is not uniform, however, and in any case does not provide a clear, consistent, and easy way to communicate requirements across ESnet groups.

The outcome of this work-package would be a common schema, processes for requirements evaluation, and queryable repository across ESnet. This repository could serve as the basis for follow-on capabilities such as requirements dashboards, natural language query, and resource forecasting.

This would allow engineers a definitive source for clear, complete and accurate requirements so supporting more efficient customer need fulfillment. It would also support engineer work planning and improved consistency of requirements across the organization, as well as enhanced communication and sharing of these requirements across ESnet.

The datasets needed to address the problem include:

- Measurements of the resource utilization of our existing system
- Timelines
- Target and baseline metrics
- Security requirements
- Repository of requirements and a process to populate this (e.g. standardized intake process)

Known Constraints

None

Opportunities and Potential Solutions

Organization processes that:

- Normalize how and when we collect and provide requirements
- Provide a transparent process for prioritizing work

- Integrate into a common requirements management system. This system should provide a verification and validation capability
- Ensure there is a responsible party for collecting, facilitating, and maintaining a requirements register.
- Provide forward looking roadmaps for effort, budget and resource prioritization and planning.

Gap Analysis

We're currently exploring ways to improve our requirements gathering and prioritization processes. While different groups within the organization have their own approaches, we recognize the value of creating a more streamlined and coordinated approach.

Our current process for determining priorities and scheduling work across the organization could be more transparent. Making this process more visible and accessible is important.

The data we need to address this challenge is available, but it's scattered across the organization in various formats and levels of completeness. To take advantage of this data, we should establish a centralized repository or standardized process for generating, sharing, and managing requirements across different organizations within ESnet.

Appendix B28. WP28 (Mission Support Management)

Problem Statement

Unlike other user facilities, such as HPC or Light Sources, ESnet users are not tied to specific grants or facility resource allocations, and do not need to interact with ESnet staff in order to obtain services. As a result, unlike these other User Facilities, ESnet does not have an automated or scalable mechanism by which we can track who we support, what they need from us, and what we are doing to support their science. As a result, ESnet is going to have to develop and build a material/workflow solution so that we can gather and understand this information, and better serve science program data mobility needs.

If successful, this work-package would have the following outcomes:

- Science Engagement users of this solution would have the ability to track what programs and Principal Investigators or other Points of Contact are making use of ESnet, and update information on what support these programs need, what support ESnet is providing, and other pertinent information on the science such as science program key milestones and dates, activities, etc. This would enable Science Engagement to map ESnet's contribution to science programs as well as help ensure our engagement outreach efforts are better synchronized with scientific needs.
- Network Engineering users of this solution would have the ability to look up information about science projects when responding to tickets or requests for resources, so that they can better and more quickly understand the context for science program resource requests. Staff would also be able to better forecast future resource needs as well as more quickly assess the impacts of network outages, or operational changes, upon science requirements.

The following data-sets would be necessary to fulfil this work-package:

- A census of scientific end-users and their activities - some of this information could be extracted from existing requirements reports, but much would need to be culled from DOE program managers, budget documents, and other sources.
- A way to tie flow-IDs to science programs, and/or a way to include science program labels to flows at our endpoints.
- A way to extract information from our existing ticketing system and categorize it by science program and present it in reverse chronological order in a way that provides a quickly readable summary of interactions and program needs/activities.

Known Constraints

Collecting all of this data together may either involve PII, or create PII through agglomeration. There may be sensitivity among some research programs with sharing information on status, activities, or data flows. In these cases, the solution will have to be flexible and able to accommodate incomplete or missing information, as data availability for many flows and activities will never be complete.

Opportunities and Potential Solutions

The solution for this work-package will probably require stitching together a number of different components.

1. Data from existing external data transfer schedules, such as Slurm at NERSC, to capture what projects are sending data across ESnet. This will involve collaboration with external user facilities, and possibly work with our current DTN software to bring this information at time of transfer into the system envisioned by this work-package.
2. Building an external registry, which users can provide information on their project either with or without the help from SET, perhaps as part of the Requirements Review process.
3. Some kind of CRM system which can pull tickets and other information from engagements and help organize it in a form that can be ingested for analysis, query, display and combined with flow data
4. Perhaps some kind of AI model that can work with these data streams to write up summaries of what flows, ESnet activities and engagements are underway supporting a science project, or respond to questions from staff for related questions.

End Users will interact with the data in primarily three ways:

1. Via natural language processing, it would be desirable to be able to ask questions like:
    a. “What HEP programs have transferred data via ESnet in the past week”
    b. “Who is the PI for the DUNE project and when did we last contact them?”
    c. “What projects have the greatest percent increase in ESnet traffic this year?”
2. Via graphing displays and Stardust output - allowing us to construct data portals supporting customer relationships
3. Via specific API or search queries upon underlying data or constructed data summaries

Gap Analysis

The largest gap that needs to be solved is how to get programmatic data associated with flows from external User Facilities. This will require collaboration and involvement with DOE PM, as well as other facilities’ leadership.

The rest of the capabilities in this work-package are more or less off-the-shelf - and would require integration of data from sources within ESnet.

Appendix B29. WP29 (Dataset Unified Query)

Problem Statement

A variety of data sources are available within ESnet, there is a need to simplify the data accesses with a unified query interface. Here are a few examples:

- As a network engineer, I can find the ESDB page, Stardust metrics, relevant documentation, and DNS entries associated with an interface so that I don't have to use multiple interfaces to fetch data
- As an LLM, I can find the relevant data sources to fetch more data from given a vague query so that I can answer user questions without hallucinating as much.
- As a software engineer, I have examples of how to fetch data from every system at ESnet so that I can build my own integrations without reinventing the wheel.
- As a network engineer I can correlate dependencies between different datasets so that I can understand the impact of changes throughout our system.
- As a network engineer I can check that data is synchronized between systems so that decisions are made with accurate information.

The datasets already exist, we need client libraries and a universal search index.

Known Constraints

Access control issues - should have a mechanism to enforce user access control so that a search will only display results the user is allowed to see.

A potential data sensitive issue is that the dataset index could have greater access than many end users if not done locally.

Opportunities and Potential Solutions

We should build client libraries for all ESnet APIs. Then, we should use these client libraries to ingest the high-level documents from all systems into a single index. We can build MCP tools for each client library and the search index itself so that LLMs can effectively use the search tool itself and the results.

There needs to be sufficient metadata associated with each data source to allow an LLM to infer what types of information can be retrieved from them (e.g. docstrings for MCP tools with the Python MCP SDK).

Gap Analysis

We need better examples of using all ESnet APIs. We should use these examples to inform client development (e.g. common patterns for authentication, wrapping common use cases with functions

to simplify request building) and then make these accessible to LLMs via MCP and by ingesting top-level documents into a search index (which itself is also accessible via MCP tools).

All the needed data exists, but is not accessible to LLMs because we do not have tools they can use. We need much more API documentation and examples across ESnet to build these tools.