Knowledge Graph-Enabled Pre-Claim Review Solution: Medicare Domain Pilot

Statement of Work

Proposed for

Careplus

27.05.2025

Executive Summary

Careplus, a leading provider of home health services in Texas, currently allocates significant clinical resources to its Medicare pre-claim review process. Each submission demands approximately two hours of a clinical manager's time – skilled nursing professionals whose expertise would be better applied to direct patient care rather than largely administrative duties, such as checking dates, verifying signatures, ensuring diagnosis code consistency, and organizing documents according to Medicare's stringent requirements. This represents both a misallocation of skilled clinical resources and a significant operational cost.

This proposal outlines a focused 6-week engagement to develop and implement a pilot Knowledge Graph (KG) solution. This represents a strategic evolution from previously considered approaches, driven by the critical need for deterministic answers, deep contextual understanding, and auditable results within the Medicare environment. The limitations inherent in other information retrieval methods, such as their probabilistic nature and potential for generating plausible but incorrect information, are particularly unsuitable for Careplus's requirements where precision and reliability are paramount. Knowledge Graphs, by contrast, offer a structured, ontology-driven framework that enables precise, verifiable, and contextually aware information processing.

The 6-week Medicare focused pilot will concentrate on a high-impact aspect of Careplus's Medicare pre-claim review process, serving as both a proof-of-value and a foundational step. This targeted approach will demonstrate the core capabilities of KG technology to:

- Significantly reduce document processing time for the selected use case
- Improve accuracy in validation through deterministic checks based on defined rules and relationships
- Enhance compliance by providing auditable and transparent reasoning paths
- Establish a clear path towards scaling these benefits across broader operational areas

Importantly, this Medicare pilot is envisioned not merely as a standalone solution but as the crucial first phase in establishing a comprehensive, scalable Knowledge Graph platform. Such a platform holds the potential to serve multiple domains within Careplus, transforming data into actionable intelligence and fostering data-driven decision-making throughout the organization. This SOW details the scope, approach, and deliverables for this pivotal 6-week engagement. Development will be executed by a lean team of experts on a time and material basis with an indicative cost of $26,200.

Current Process Analysis

The Medicare Pre-Claim Documentation Challenge

Careplus's clinical managers currently dedicate approximately two hours per submission to navigating Medicare's intricate pre-claim review process. This essential function, while vital for ensuring compliance and securing payment, diverts highly skilled nursing professionals from their primary role of patient care into predominantly administrative tasks. This diversion represents a significant misallocation of valuable clinical resources.

Current Workflow Components

To manage Medicare's complex documentation requirements, Careplus has structured its pre-claim review process into three primary task groups.

1. Task Group 1: Initial Eligibility Documentation: This includes face-to-face encounter documentation, physician progress notes, primary diagnosis validation, and hospitalization history verification
2. Task Group 2: Care Plan & Evaluation Documentation: This encompasses Plan of Care (485) documentation, initial therapy evaluations, OASIS assessment documentation, and supporting clinical documents. Key challenges include validating service frequencies against orders, ensuring care plan alignment with face-to-face findings, and verifying signatures and dates on therapy evaluations
3. Task Group 3: Recertification & Updates: This involves updated care plans, provider transition documentation, continuing care orders, and additional service authorizations

Manual Processing Requirements

Across all task groups, clinical managers are required to perform several time-intensive manual tasks:

1. Document Collection & Organization: Retrieving documents, organizing files per Careplus's system, and ensuring all required signatures and dates are present
2. Cross-Document Validation: Comparing primary diagnoses, verifying date consistency, checking provider credentials, and validating service codes
3. Medicare Compliance Verification: Ensuring all required elements are present, validating against Medicare rules, and organizing documents for submission

Impact on Clinical Operations

This manual process culminates in several operational challenges for Careplus [1]:

1. Resource Misallocation: Clinical managers spend extensive time on administrative tasks, diverting skilled nursing time from patient care at high hourly rates
2. Process Inefficiencies: Multiple manual review passes, duplicate information checking, and time-consuming document organization lead to inefficiencies
3. Compliance Risks: Manual verification of complex requirements increases the risk of human error, potentially leading to missed inconsistencies and non-compliance

Opportunity for Knowledge Graph-Driven Transformation

The aforementioned challenges, particularly those related to cross-document validation, consistency assurance, and compliance adherence, highlight an opportunity for a more profound transformation using Knowledge Graphs. While the initial pilot will target a specific component, the KG methodology itself is designed to create a holistic and interconnected understanding of data. This capability addresses the root causes of current inefficiencies and risks in a more fundamental manner than approaches focused solely on text extraction and surface-level comparison.

The systemic issues identified within Careplus's current process—such as manual cross-checking of diagnosis codes across multiple documents, ensuring consistency in medical terminology, and the inherent risk of human error in document matching—are particularly well-suited to a KG-based solution. Knowledge Graphs excel at solving complex puzzles by understanding the intricate interplay and critical, structured logic connecting disparate pieces of information across various documents. By establishing a single source of truth that evolves with new information and changing regulations, a KG can provide a robust foundation for ongoing compliance and operational excellence.

Solution Architecture & Approach (Medicare Knowledge Graph Pilot)

Rationale for Knowledge Graph Approach

The decision to employ a Knowledge Graph (KG) for this pilot, and as a foundation for future enhancements, is based on its inherent advantages over other information retrieval and processing methodologies, particularly for the complex and rule-intensive domain of Medicare pre-claim review.

Knowledge Graphs provide deterministic, not probabilistic, answers. In an environment where compliance and accuracy are non-negotiable, the ability to deliver verifiable results based on explicitly defined rules and relationships is crucial. Unlike systems that retrieve information based on semantic similarity, which may yield plausible but potentially incorrect or out-of-context responses, KGs operate on a structured representation of knowledge, ensuring precision.

Furthermore, KGs enable a deep understanding of concepts and their interrelationships, moving beyond surface-level text matching. They capture the underlying meaning and context, allowing for more nuanced and accurate analysis. For instance, a KG can understand not just that a patient has a certain diagnosis and receives a specific medication, but also how that medication relates to the diagnosis, potential contraindications with other conditions the patient might have, and how these factors align with Medicare coverage guidelines.

This deep understanding leads to clear explanations and trust. The reasoning path of a KG in arriving at a conclusion is traceable and auditable, which is essential for validating compliance and building confidence in the system's outputs. This transparency contrasts with the “black box” nature of some AI systems where understanding the basis for a particular answer can be challenging.

Finally, KGs are inherently better suited for complex, multi-hop queries and rule-based validation, which are characteristic of Medicare processes.

Core Principles of the Pilot

The 6-week Medicare Knowledge Graph pilot will adhere to the following core principles:

- **Deterministic & Precise**: The primary objective is to demonstrate the KG's capability to deliver accurate and reliable validation based on a defined ontology and explicit rules

- **Hyper-Focused Use Case**: The pilot will concentrate on a specific, high-value validation point within the Medicare pre-claim review process. This targeted scope allows for the rapid demonstration of tangible benefits and establishes a clear proof-of-value
- **Rapid Value Delivery**: The 6-week timeline is designed for swift execution, enabling Careplus to quickly observe the core KG capabilities and their potential impact
- **Foundation for Scalability**: The architecture, methodologies, and learnings from this pilot will be directly applicable and will inform the strategic development of a larger, enterprise-wide Knowledge Graph platform at Careplus

Medicare Pilot: Hyper-Focused Use Case Definition

The pilot will address a critical consistency check that is currently time-consuming and prone to error. The selected use case is: "*For patients with a specific diagnosis (e.g., Congestive Heart Failure), ensure the Face-to-Face (F2F) encounter documentation confirms this diagnosis, was conducted by an allowed physician type within the Medicare-stipulated timeframe (e.g., 90 days prior to or 30 days after the start of care), and that the Physician's Order aligns with this diagnosis and the F2F encounter timeframe.*"

The key documents involved in this use case will be:

- Face-to-Face (F2F) Encounter Notes
- Physician's Order
- Relevant sections of OASIS assessment documentation (if applicable to the specific diagnostic validation)

Medicare Pilot: Initial "Mini-Ontology"

To support the hyper-focused use case, an initial "mini-ontology" will be developed. This mini-ontology is a simplified, yet formal, model representing the key concepts (entities), their properties (attributes), and the critical connections (relationships) relevant to the chosen validation scenario. This structured representation is fundamental to the KG's ability to perform precise reasoning.

The proposed core entities and key relationships for the pilot are detailed in Table 1. Each entity will also have defined attributes; for example, the F2FEncounter entity will have attributes such as {date, physician_name, patient_identifier, relevant_narrative_snippet}.

Table 1: Medicare Pilot Mini-Ontology Definition

Category	Element Name	Description	Example Attributes / Connections
Entities	Patient	An individual receiving care.	Patient ID, Date of Birth
	Physician	A healthcare provider involved in the patient's care or orders.	Physician Name, NPI Number
	F2FEncounter	A face-to-face encounter event.	Encounter Date, Physician Name (performing), Patient Identifier, Relevant Narrative Snippet
	PhysicianOrder	An order for services or care initiated by a physician.	Order Date, Ordering Physician, Patient Identifier, Ordered Diagnosis
	Diagnosis	A medical diagnosis code or description.	ICD-10 Code, Description
	ClinicalFinding	Specific clinical information or observations documented.	Finding Description, Date of Observation
	ServiceDate	A date associated with a service or encounter.	Date Value
Relationships	hasDiagnosis	Connects Patient to Diagnosis.	Patient --hasDiagnosis--> Diagnosis

	performedBy	Connects F2FEncounter to Physician.	F2FEncounter --performedBy--> Physician
	documents	Connects F2FEncounter to ClinicalFinding.	F2FEncounter --documents--> ClinicalFinding
	ordersServiceFor	Connects PhysicianOrder to Patient.	PhysicianOrder --ordersServiceFor--> Patient
	relatedToEncounter	Connects PhysicianOrder to F2FEncounter.	PhysicianOrder --relatedToEncounter--> F2FEncounter
	hasDate	Connects F2FEncounter (or PhysicianOrder) to ServiceDate.	F2FEncounter --hasDate--> ServiceDate
	confirmsDiagnosis	Connects F2FEncounter to Diagnosis (indicating confirmation).	F2FEncounter --confirmsDiagnosis--> Diagnosis
	alignsWithOrder	Connects F2FEncounter to PhysicianOrder (indicating temporal/diagnostic alignment).	F2FEncounter --alignsWithOrder--> PhysicianOrder

Document Scope for Pilot

The pilot will focus on processing the following specific document types, as they pertain to the defined use case:

1. Face-to-Face (F2F) Encounter documentation
2. Physician's Order (focusing on diagnosis codes, physician details, and dates)
3. Relevant excerpts from OASIS assessments, if directly pertinent to validating the specific diagnosis and its consistency with F2F documentation

Targeted Information Extraction & Validation Logic

The process will involve targeted information extraction from the scoped documents. Instead of ingesting entire documents indiscriminately, the focus will be on precisely extracting specific snippets of information that correspond to the entities, attributes, and relationships defined in the mini-ontology. This precision is key to building an accurate and efficient KG.

Extraction will be driven by the use of LLMs/VLMs like Google's Gemini series of models. Advances in LLM technology have made their use for this purpose much more deterministic, faster and cheaper than using traditional libraries like Tesseract. In order to ensure the LLMs do not hallucinate at this stage, guardrails will be put in place at both the system prompt as well as the structured output level. In addition, each document will get an extraction confidence score that will be persisted throughout the lifecycle of the document. This enables manual QA work to happen at a later stage, both from an engineering perspective as well as from a subject matter expert point of view.

Validation logic will be implemented as graph queries (e.g., Memgraph). These queries will be designed to traverse the populated graph, checking for adherence to the rules defined by the use case and the relationships in the mini-ontology. An example query could be: "For a given Patient, find their F2FEncounter. Does this F2FEncounter confirm Diagnosis X? Was this F2FEncounter performed by an allowed Physician type? Is the F2FEncounter's hasDate within 90 days prior to or 30 days after the start_of_care_date derived from the related PhysicianOrder for Diagnosis X?".

Pilot Solution Components

The pilot solution will comprise the following key components:

- **Lightweight Graph Database**: Setup of an open-source graph database (Memgraph) to store and manage the graph data for the pilot.
- **Data Ingestion & Transformation Scripts**: Custom scripts will be developed to parse the extracted structured data (from documents) and transform it into the graph format (nodes representing entities, edges representing relationships) for loading into the graph database
- **Document Processing Utilities**:

    - Capabilities to process PDF documents, including both text-based and scanned formats using LLMs/VLMs
    - A secure mechanism for Careplus to upload PDF documents (e.g., via a designated AWS S3 bucket)
- **Basic Query Interface**: A simple mechanism or tool to execute the developed graph queries against the pilot KG and view the results

Pilot Output & Validation

The primary output of the pilot will be a populated knowledge graph containing the extracted entities and relationships from the sample documents, along with the results of the consistency checks performed by the graph queries. These results will explicitly indicate whether the reviewed document sets meet the defined validation criteria or highlight identified discrepancies.

A critical aspect of the pilot is the validation of its outputs by Careplus domain experts. This collaborative validation is not merely a quality check; it is an integral part of the knowledge capture process. Outputs from both the extraction and validation stages will be presented as spreadsheets (using Google Sheets) for easier evaluation. The ability for technical users to query the databases directly to validate any information will be available regardless.

The pilot serves as an initial exercise in codifying Careplus's domain expertise and Medicare rules into a formal, machine-understandable structure. The accuracy and relevance of the mini-ontology, the extraction logic, and the validation queries depend heavily on this iterative feedback loop with Careplus SMEs. This collaborative approach ensures the KG accurately reflects the operational realities and compliance requirements, building a shared understanding and a robust foundation for future development.

Optional: Pure LLM Benchmarking for Comparative Analysis

While the primary approach leverages Knowledge Graphs for their deterministic nature and long-term platform benefits, there is initial value in establishing empirical evidence of this approach's better performance over pure LLM implementations. This comparative analysis would:

1. Validate technology choice by providing data-driven proof that the Knowledge Graph approach delivers equal or better results than a pure LLM implementation
2. Provide better platform intelligence by generating insights for future domain expansions where pure LLM approaches might be more suitable

If selected, this optional component would involve:

- Developing a parallel pipeline using pure LLM calls (without Knowledge Graph)
- Processing the same document sets through both approaches
- Measuring and comparing key metrics:
    - Human agreement rates (target: KG must achieve ≥ pure LLM)
    - Token usage and associated costs

The Knowledge Graph approach must demonstrate performance that is equal to or better than the pure LLM approach, validating the architectural decision and investment in graph technology.

We estimate an additional 2 weeks of effort over the original timeline. Resourcing and budgets are provided in the Project Resourcing & Costs section below.

Implementation Timelines

The 6-week Medicare Knowledge Graph Pilot is structured to deliver rapid insights and a functional proof-of-value. The timeline is broken down into weekly phases, each with specific activities and key milestones, drawing from the proposed pilot stages.

Table 2: 6-Week Medicare Pilot Implementation Timeline

Week	Key Activities	Key Milestones/Deliverables
Week 1	Project Kick-off & Joint Planning Session. Finalize hyper-focused use case details with Careplus SMEs. Refine and collaboratively finalize the mini-ontology (entities, attributes, relationships). Establish secure document transfer mechanisms. Initial technical environment setup.	Signed-off Use Case Definition. Finalized Mini-Ontology Document. Receipt of initial sample documents from Careplus.
Week 2	Develop and configure tools/scripts for targeted information extraction from pilot documents (F2F, Physician Orders, OASIS excerpts) based on the finalized mini-ontology. Focus on precision for defined entities and attributes. Initial unit testing of extraction modules on	Core information extraction scripts/modules functional for key entities and attributes. Initial extraction test results.

	sample documents.	
Week 3	Set up the chosen open-source lightweight graph database. (Memgraph) Develop scripts to transform extracted structured data into graph format (nodes and edges). Perform initial ingestion of extracted data from a subset of sample documents into the graph database.	Pilot graph database instance operational. Data transformation scripts complete. Graph database populated with initial data set.
Week 4	Develop specific graph queries (e.g., Cypher for Neo4j or an equivalent language for the chosen database) to perform the consistency checks defined in the use case (e.g., diagnosis match, F2F timeliness, physician type validation). Execute initial queries against the populated graph. Conduct first-pass validation of query results with Careplus SMEs.	Core validation queries developed and tested. Initial consistency check results generated. SME feedback on initial query outputs.
Week 5	Refine information extraction scripts and graph queries based on Careplus SME feedback and testing	Pilot system demonstrating improved accuracy on test data.

	results. Conduct more comprehensive testing with a broader set of provided sample documents. Begin drafting documentation: mini-ontology, extraction process, graph schema, query logic.	Iterative refinements implemented. Draft pilot documentation.
Week 6	Conduct User Acceptance Testing (UAT) with Careplus SMEs. Address any critical issues identified during UAT. Finalize all pilot documentation.	Successful UAT completion by Careplus. Final pilot system demonstrating use case validation. Finalized pilot documentation package.

Note: If a parallel system of extraction and querying via a traditional LLM method is to be built for quality and metrics comparison, we estimate an additional 10 business days of work. If this additional system is agreed upon, then the final estimates in terms of time and budget will be provided during the project kickoff phase, with details entered into Rocketlane (the project management tool that will be used).

Assumptions & Risks

Assumptions

The successful and timely completion of this 6-week Medicare Knowledge Graph Pilot relies on the following assumptions:

- Document Accessibility & Quality:
  - Careplus will provide timely access to a representative set of relevant Medicare documents (Face-to-Face Encounter notes, Physician's Orders, and applicable OASIS assessment excerpts) for the defined pilot use case. These documents are assumed to be in standard PDF or common image formats
  - The quality of scanned documents will be sufficient for effective Optical Character Recognition (OCR) to extract textual data
- Careplus SME Availability:
  - Careplus subject matter experts (e.g., clinical managers, compliance specialists) will be available for scheduled workshops, ontology validation sessions, query validation, and User Acceptance Testing (UAT) as outlined in the implementation timeline. An estimated commitment of 4-8 hours per week during key validation weeks (Weeks 1, 4, 6) is anticipated
- Mini-Ontology Stability:
  - The core entities, attributes, and relationships defined for the pilot's mini-ontology are expected to be largely stable following the collaborative refinement process in Week 1. Significant changes to the ontology mid-pilot could impact the timeline and deliverables
- Scope Focus:
  - The scope of the pilot will remain strictly focused on the agreed-upon hyper-focused use case (and questions) and the defined mini-ontology. Scope expansion requests will be addressed through a separate change control process

Risks

The following potential risks have been identified for the Medicare Knowledge Graph Pilot:

- Technical Risks:
  - Information Extraction Accuracy: The precision of extracting specific entities, attributes, and relationships from varied document structures using NLP/ML techniques or rule-based systems may require more iterative refinement than allocated to achieve the desired level of accuracy

    - Ontology Mismatches: The initial mini-ontology, despite collaborative refinement, might not perfectly capture all necessary nuances of the Medicare rules or document content for the use case. Adjustments discovered later could necessitate changes to extraction logic and graph queries. A core challenge in any KG development is ensuring the graph accurately represents the intended meaning and rules of the domain (''semantic fidelity''). If the ontology is flawed or incompletely captures the domain's semantics, or if the extraction process populates it incorrectly, the KG's deterministic answers could be deterministically wrong.
    - Graph Database Performance (Minor for Pilot): While less likely for the limited scale of a pilot, unforeseen performance issues with the chosen lightweight graph database could arise, though this is considered a low probability
- Operational Risks:
    - SME Unavailability: Insufficient or delayed availability of Careplus domain experts for crucial validation and feedback sessions could impede progress, delay timelines, and potentially impact the accuracy of the KG's domain representation
    - Timeline Constraints: The 6-week timeline is aggressive. Unforeseen technical complexities, delays in receiving sample documents, or extended feedback cycles could pose a risk to completing all activities within this timeframe
- Data Risks:
    - Unexpected Document Variation: If the sample documents provided exhibit significantly greater variation in format, structure, or language than initially anticipated, the complexity of the information extraction task could increase
    - Data Quality Issues: Pre-existing errors, ambiguities, or inconsistencies within the source documents themselves could be inadvertently propagated into the knowledge graph or lead to misinterpretation if not identified and handled appropriately during the extraction and validation phases

Deliverables (Medicare Knowledge Graph Pilot)

Upon completion of the 6-week Medicare Knowledge Graph Pilot, Moative will provide Careplus with the following deliverables:

1. **Populated Pilot Knowledge Graph:**
    - An instance of the selected lightweight, open-source graph database (e.g., Memgraph/GraphRAG) populated with entities, attributes, and relationships. This data will be extracted from the agreed-upon set of Careplus sample documents relevant to the defined pilot use case
2. **Mini-Ontology Documentation:**
    - A comprehensive document detailing the finalized mini-ontology developed for the Medicare pilot use case. This will include definitions of all entities, their attributes, and the relationships between them, as outlined in Table 1
3. **Information Extraction Scripts & Configuration:**
    - The set of scripts (e.g., Python scripts) and any associated configuration files developed and used for the targeted information extraction from the source documents to populate the knowledge graph
4. **Graph Query Set:**
    - The collection of graph queries (e.g., written in Cypher for Neo4j or an equivalent language for the chosen database) developed to perform the specific consistency checks and validations defined by the pilot use case
5. **Validation Report:**
    - A report summarizing the results of the knowledge graph's consistency checks performed on the sample document set. This report will include metrics on accuracy (as validated by Careplus SMEs) and any identified discrepancies or patterns.

Success Metrics (Medicare Knowledge Graph Pilot)

- Accuracy of Entity & Relationship Extraction:
    - Precision and recall for key entities (e.g., Patient ID, Diagnosis Codes, Physician Name, Encounter Dates) and critical relationships
    - Target: ~95% precision and recall for core elements from the document set
- Accuracy of Consistency Checks:
    - Correct identification of consistencies and inconsistencies for the defined use case rules (e.g., diagnosis match between F2F and Order)
    - Target: ~80% agreement rate between graph recommendations and SME decisions (measured at the case level, not individual extraction level)
- Query Performance Time:
    - Time taken for the defined graph queries to execute and return results
    - Target: Average query execution time <5 seconds per document set
- Careplus SME Validation:
    - Confirmation from Careplus SMEs that the pilot system correctly identifies relevant information and that its outputs are understandable
    - Target: Successful validation and formal acceptance of the pilot's findings at the end of UAT

RACI Matrix

Activity/Deliverable	Moative	Careplus
Use Case Definition & Finalization	R/A	R/C
Mini-Ontology Design & Initial Draft	R/A	C
Mini-Ontology Validation & Finalization	R/C	R/A
Document Provision & Access Management	C	R/A
Information Extraction Development	R/A	C
Graph Database Setup & Configuration	R/A	I
Data Ingestion & Transformation Script Development	R/A	I
Graph Query Development	R/A	C
Graph Query Logic Validation	R/C	R/A
User Acceptance Testing	C	R/A

Legend:

- R: Responsible (Does the work)
- A: Accountable (Owns the outcome)
- C: Consulted (Provides input)
- I: Informed (Kept updated)

Notes:

- The active involvement of Careplus Clinical SMEs as 'Responsible/Accountable' for validation tasks (e.g., Mini-Ontology Validation, Query Logic Validation, UAT) is crucial for ensuring the pilot KG accurately reflects domain knowledge and meets Careplus's needs

Project Resourcing & Costs

		Indicative Hours					
Role	**Rate**	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**
Project Manager	$75.00	8	8	8	8	8	8
AI/ML Engineer 1	$50.00	40	40	40	40	40	40
AI/ML Engineer 2	$50.00	32	32	32	32	24	24
Solution Architect	$150.00	2	2	2	2	2	2

Total indicative consultant costs: $26,200.

If the optional LLM based pipeline activity is also taken up, then we estimate the following:

- 2 weeks of work by one AI/ML engineer (80hrs)
- 8hrs of project management work (doubling up as QA)

The indicative costs for the 2-weeks would be $4400.

Please note the proposed development costs do not include cloud or other infrastructure expenses. Moative will provide these costs separately based on the final solution architecture and data and processing requirements. We estimate this to be around $2500 based on current pricing for AWS and Google Gemini as the LLM.

Billing

The time estimates proposed above are indicative. Moative will keep track of the time spent by its team and bill on actuals based on the rate card.

Billing will be at the end of each week during the entire duration of the project.

Next Steps

Upon a go-ahead from Careplus, Moative shall incorporate this SOW into a draft 'Master Services Agreement.' Once mutually executed, Careplus and Moative will jointly decide the kick-off date for the 6-Week Medicare Knowledge Graph Pilot.

Following the successful completion and review of this pilot engagement, Moative proposes scheduling a follow-up strategic discussion to review the pilot outcomes in detail and collaboratively define the roadmap for the ''Path-to-Platform,'' leveraging the insights and foundational work from this initial phase to build out a comprehensive Knowledge Graph capability for Careplus.

Appendix A: The Knowledge Graph Platform Vision

The 6-week Medicare Knowledge Graph Pilot, detailed in this Statement of Work, serves as a strategic initial step. The learnings, foundational ontology work, validated extraction techniques, and demonstrated benefits from this pilot are intended to pave the way for a significantly more comprehensive, enterprise-grade Knowledge Graph (KG) platform.

Core Components of a Scalable Knowledge Graph Platform

Component Layer	**Key Functionalities**	**Example Technology/Approaches**
Data Ingestion Layer	Processing of diverse document types (PDFs - text & scanned, Word, etc.). OCR capabilities for image-based documents. Connectors for structured data sources (databases, APIs, data warehouses). Batch and real-time data ingestion pipelines. Secure, scalable document and data staging (e.g., cloud storage).	Apache NiFi, Kafka; Tesseract OCR, AWS Textract, Google Vision AI; Python (PyPDF, python-docx); AWS S3
Knowledge Extraction & Modeling Layer	NLP/ML models for automated entity extraction, relationship extraction, and attribute population from unstructured text. Ontology management tools: creation, editing, versioning, mapping, and alignment of ontologies/taxonomies. Support for importing and integrating standard healthcare ontologies (e.g., SNOMED CT, LOINC, ICD-10-CM). Tools for semantic modelers and domain experts to define and refine domain-specific knowledge structures. Potential use of Large Language Models (LLMs) for advanced extraction and knowledge enrichment.	spaCy, Stanford CoreNLP; Protégé, TopBraid Composer; LLMs (GPT, Gemini, Claude) via APIs.

Graph Database & Storage Layer	Scalable graph database for storing nodes, edges, and their properties. Support for transactional updates (CRUD operations) and complex graph traversals. Potentially a Vector Database for storing and querying embeddings to enable semantic search and similarity calculations alongside graph structures. Robust data persistence, backup, disaster recovery, and high availability mechanisms.	Neo4j, Memgraph, Amazon Neptune, JanusGraph, TigerGraph; Faiss, Milvus, Pinecone.
Querying, Analytics & API Layer	Support for standard graph query languages (e.g., Cypher for Neo4j, Gremlin for TinkerPop-enabled databases, SPARQL for RDF graphs). RESTful APIs for programmatic access to the KG (data ingestion, querying, updates, administration). Graph visualization and exploration tools for business users and analysts. Potential integration with AI/ML models for advanced querying (e.g., natural language to graph query) or KG-informed predictions.	FastAPI/Flask (Python) for APIs; Neo4j Bloom, Gephi, Cytoscape; Tableau, Power BI connectors.
Platform Management & Governance Layer	Comprehensive security model: authentication, authorization, fine-grained access control. HIPAA-compliant (or equivalent) data handling practices, encryption at rest and in transit. Data quality management framework: validation rules, data cleansing, lineage tracking. Version control for ontologies, graph schemas, and potentially graph data snapshots.	OAuth2/OIDC; KMS, HashiCorp Vault; ELK Stack/Splunk for logging; Prometheus, Grafana; Apache Atlas.

Constructing the Platform: A Phased Evolution

The development of the full Knowledge Graph platform is envisioned as an iterative, phased

evolution, building upon the successes and learnings of preceding stages. This approach allows for incremental value delivery, risk mitigation, and adaptation to evolving needs.

- **Phase 1: Medicare Pilot & Foundational Learning (This SOW)**
    - Focus: Execute the 6-week Medicare pre-claim review pilot.
    - Outcomes: Validated mini-ontology for a core Medicare process, functional KG for the use case, documented extraction and query logic, initial performance benchmarks, and key learnings regarding Careplus data and processes. This phase establishes the initial "seed" for the larger platform.
- **Phase 2: Expand Medicare Use Cases & Core Ontology Development**
    - Focus: Broaden the scope within the Medicare domain.
    - Activities: Incorporate additional document types and validation rules from other Medicare Pre-Claim Task Groups (e.g., Care Plan & Evaluation, Recertification). Systematically map and integrate core elements of standard healthcare ontologies like ICD-10-CM and LOINC (particularly for OASIS data elements) into the growing Medicare KG, aligning them with the pilot's mini-ontology and expanding it. Refine and scale the data ingestion and extraction pipelines for Medicare documents.
    - Outcomes: A more comprehensive Medicare Knowledge Graph, enriched with standard terminologies, capable of supporting a wider range of pre-claim review validations.
- **Phase 3: Develop Core Platform Services & Infrastructure**
    - Focus: Build out the robust, reusable components of the enterprise KG platform.
    - Activities: Engineer and deploy scalable data ingestion pipelines adaptable to various data sources. Implement sophisticated ontology management tools and workflows. Develop standardized APIs for accessing and interacting with the KG, as identified in Table 5. Establish robust security, governance, and monitoring frameworks. Select and implement the enterprise-grade graph database.
    - Outcomes: A foundational KG platform with core services ready to support multiple use cases and domains.
- **Phase 4: Extend to New Domains & Use Cases**
    - For each new domain, develop or adapt domain-specific ontologies, working

closely with domain experts. Configure ingestion pipelines for new data sources.

Operationalizing the Knowledge Graph Platform

A Knowledge Graph platform is not a static system; it is a dynamic, evolving ecosystem that requires ongoing management, governance, and expertise to deliver sustained value. Operationalizing the platform involves establishing the right team roles, processes, and tools.

Role	Key Responsibilities	Required Skills
Platform Architects/Engineers	Design, build, and maintain the reusable core platform infrastructure and services. Ensure scalability, performance, and reliability of the KG platform. Define technical standards and best practices.	Expertise in graph databases, data integration technologies, cloud platforms, API design, and system architecture.
Data Engineers	Develop, maintain, and optimize adaptable ETL/ELT pipelines for ingesting data from diverse sources into the KG. Implement data quality checks, transformation logic, and data validation processes.	Strong skills in data modeling, SQL/NoSQL databases, scripting languages (Python), data pipeline tools (Apache Airflow, NiFi), and data quality frameworks.
AI/ML Engineers	Develop, train, fine-tune, and deploy NLP/ML models for automated entity extraction, relationship extraction, and attribute population from unstructured data (documents, notes). Explore and integrate LLMs for advanced knowledge extraction and querying capabilities.	Expertise in machine learning, natural language processing, deep learning frameworks.
Semantic Modelers / Ontologists	Design, develop, and maintain core and domain-specific ontologies and taxonomies. Work closely with domain experts to capture and formalize business rules and knowledge.	Strong understanding of semantic technologies, ontology languages (OWL, RDF/S), data modeling, knowledge representation principles, and experience with ontology editing tools (e.g., Protégé).

Key Processes for Operationalization

These will evolve as the initiative progresses, milestones are hit and lessons learned. But broadly, the key processes are:

- Ontology Lifecycle Management: Processes for developing, reviewing, updating, versioning, and retiring ontologies, including integration of updates from standard terminologies
- Data Governance & Quality Assurance: Establishing clear ownership, policies, and procedures for maintaining data accuracy, consistency, and compliance within the KG. Regular data quality audits and remediation workflows
- Change Management: Managing updates to the graph schema, extraction models, and platform components with minimal disruption
- Performance Monitoring & Optimization: Continuously monitoring platform performance, query execution times, and resource utilization to identify and address bottlenecks