

AI + Metadata: An Agentic Approach to Standards-Based Automation

PHUSE US Connect 2026 - ET02

Kevin Burges

Director of Automation Architecture

VeraMed

“2025 is the year of agents”

Greg Brockman, OpenAI Co-founder

“Embrace the rise of AI agents and catapult humanity into an era of unprecedented cognitive augmentation and creative synergy!”

Emmanuel Apetsi, CEO, SISU AI

2025

The Agentic AI Moment

“Agents are going to... bring about the biggest revolution in computing since we went from typing commands to tapping on icons”

Bill Gates, Founder, Microsoft

“Agentic AI is the self-driving car of Large Language Models.”

Steve Salvin, Founder, Aiimi

The Original Vision



Protocol → CRF → SDTM → ADaM → TLF

Generate machine-readable protocols from PDF/Word

Generate draft CRFs aligned to standards

Generate SDTM metadata and mappings

Generate SDTM transformation code

Generate ADaM and TLFs from study data

The Reality Check





Trust

The Trust Challenge



Internal survey data (n=233):

- 75% require verification before trusting AI outputs
- 53% don't trust AI for regulated activities
- Top concern: AI accuracy/reliability in regulated environment

Yet:

- 73% see opportunities for AI to free up time
- 72% are optimistic or excited about AI
- 39% fall into early adopter categories



Regulatory

AI vs Traditional GxP - The Tensions



AI Strength	Traditional GxP	Tension
Probabilistic reasoning	Deterministic, validated behaviour	How do we ensure reproducibility?
Learns and improves	Controlled, versioned systems	Self-learning challenges change control
Handles complex unstructured data	Traceability of results	"Black box" vs audit trail
Reduces human workload	Accountability, qualified oversight	Who is responsible for AI decisions?

Regulatory Guidance Is Evolving

Recent and upcoming guidance supports thoughtful AI adoption:

- **GAMP 5 Second Edition (2022):**
 - Risk-based validation
 - Critical thinking to determine appropriate approaches
 - Not “one size fits all”
- **EU GMP Annex 22 (draft, 2026):**
 - Explicitly addresses AI in GMP environments
 - Differentiates between “critical” and “non-critical” use cases
 - Can extend these principles to GCP environments

Key Messages

- Validation effort should scale based on risk to patient and data integrity
- Regulation is explicitly evolving to support the use of AI



Policy

A Risk-Based Governance Framework for AI Tools



- Structured approach
- Determines appropriate controls for each AI use case
- Human oversight at all decision points
- Applies to internally developed and third-party solutions

Use Case → Usage Profile → Risks → Controls → Risk Zone Assignment

A Risk-Based Governance Framework for AI Tools



Use Case example

“Claude Code for developing assistive tools”

Usage Profile example

- Autonomy: “Agentic”
- Platform: “Anthropic-hosted Team plan”
- Models: Claude (Opus, Sonnet, Haiku)
- Action: “Create/Modify non-GxP content”
- Data: “Company Confidential”
- Users: “Technical Solutions Group”

Risks example

- R-0181: “Agent workflow orchestration governance”
- R-0277: “AI systems making consequential decisions without human input”

Controls example

- OPS-027: “Prompt and Context Quality Assurance”
- OPS-023: “Human Validation Before Downstream Use”

Risk Zone Assignment example

“Amber: Restricted use”

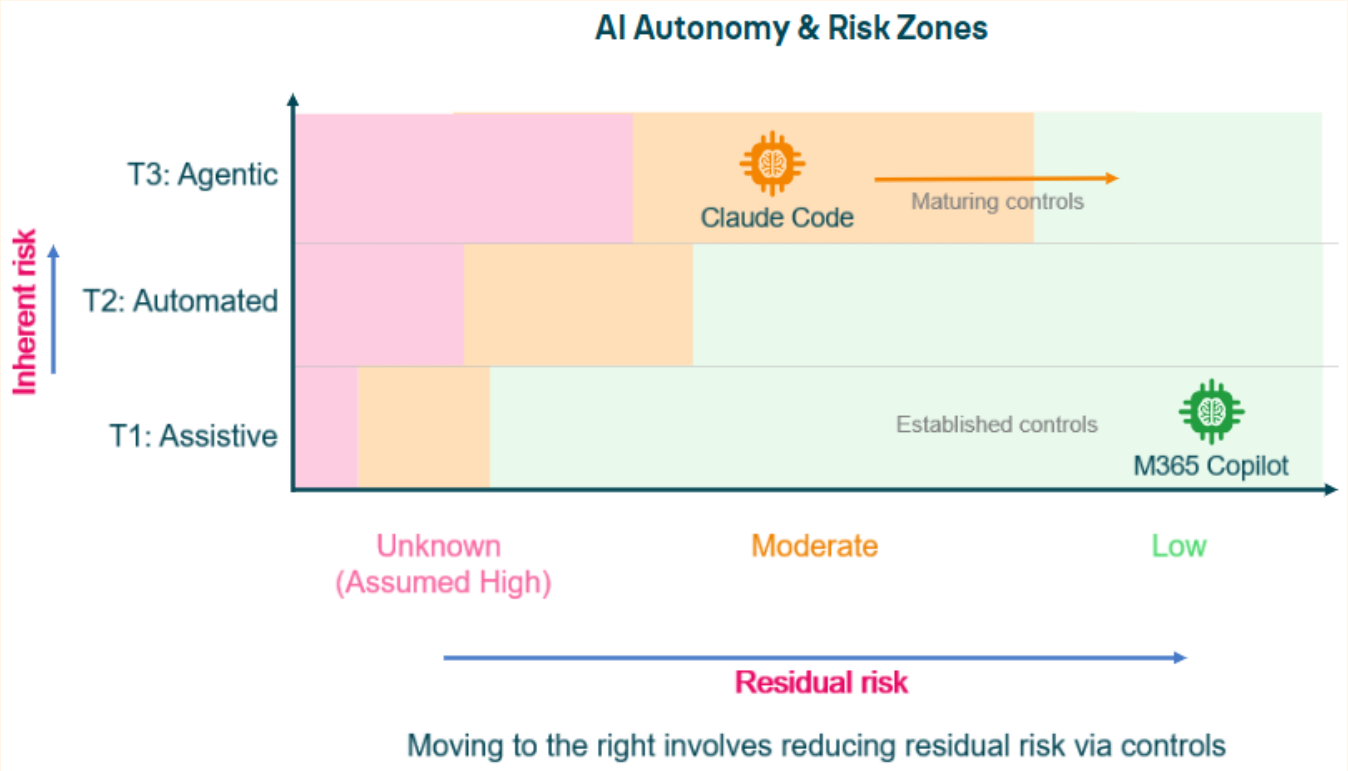
Usage Profile – Tiers of Autonomy



Tier	Description	Examples
T1: Assistive	Human-initiated, single-task chatbot interactions. Output is suggestion only.	Prompt Optimiser, CDASH Expert agent, Research Assistant
T2: Automated	Defined workflows with AI components. Human oversight at checkpoints.	Low-code workflow automation, document processing pipelines
T3: Agentic	Autonomous goal pursuit with tool access. Minimal human intervention during execution.	Coding agents, complex multi-step analysis

Risk Zones

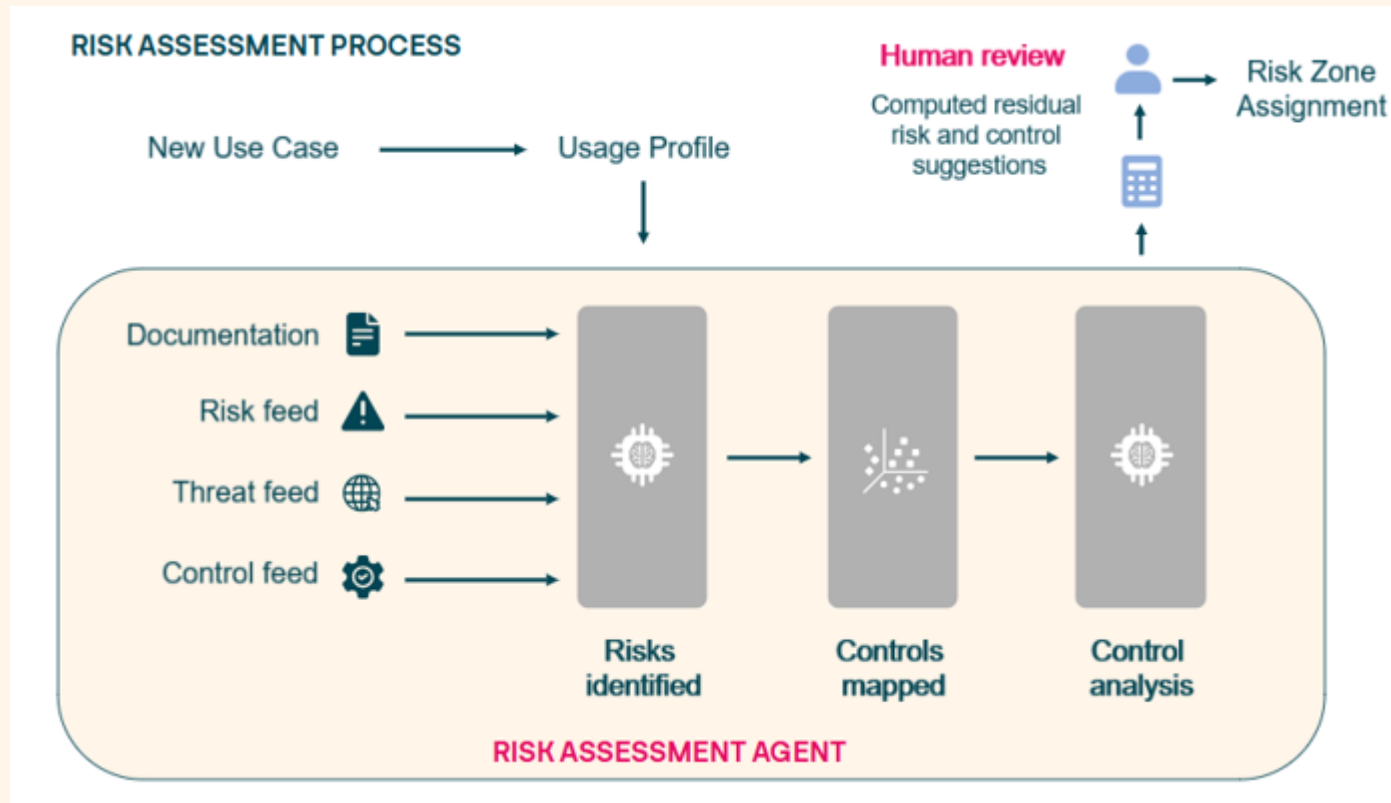
Risk Zone assigned based the “residual risk” remaining after the identified controls have been put in place.



Zone	Meaning	Mandated Controls
Red	Default entry state for unanalysed use cases. Residual risk unknown.	Sandboxed environments until formal assessment completed
Amber	Risk analysis conducted, residual risk reduced but not to threshold permitting widespread deployment. May remain sustainably in Amber.	Regular Governance Committee review, enhanced monitoring
Green	General organisational use permitted. Residual risk reduced to acceptable minimum.	Established, typically automated controls

AI-Assisted Risk Assessment

- Rapid increase in AI tools
- Use cases scale with organisational AI adoption
- AI-assistant increases human capacity to identify, assess, and monitor risks
- More comprehensive and consistent assessment than achievable manually at scale



Critical Safeguards:

- **Advisory capacity only**
- Humans can override, modify, or reject recommendations
- Governance Committee makes final zone determination
- Preserves accountability and allows judgement



Technical

Reliability: Metadata and Standards are Critical



“What domains are in the SDTM 3.4 Interventions class?”

Correct	AG	CM	EC	EX	ML	PR	SU	
Copilot response #1		CM	EC	EX		PR	SU	Missing AG, ML
Copilot response #2		CM		EX	ML	PR	SG	Missing AG, EC Extra SG

AI Without Metadata and Standards is Guessing

Layers of Standards



Standards exist in multiple layers, each adding value:

Layer	What It Provides	Examples
Format Standards	Common way to express content; interoperability between tools and organisations	CDISC ODM, Define-XML, Dataset-XML, Dataset-JSON
Content Standards	What to collect and how to structure it; industry consistency	CDASH, SDTM, ADaM, NCI Controlled Terminology
Semantic Standards	Meaning and relationships; enable genuine AI reasoning about studies	USDM, Biomedical Concepts, Dataset Specialisations
Organisational Standards	Organisation-specific implementations; consistency and data quality within an organisation	Organisational CRF standards, Organisational SDTM/ADaM standards

The Semantic Future

Emerging CDISC standards unlock end-to-end automation:

- **USDM (Unified Study Definitions Model):** Machine-readable protocol defining study goals, structure, endpoints, activities, eligibility
- **Biomedical Concepts:** Semantic definitions of what we're measuring (e.g., "blood pressure" with all its attributes and relationships)
- **Dataset Specialisations:** Mappings from biomedical concepts to SDTM variables, enabling deterministic transformation

Vision:

Together, these provide a machine-understandable backbone for an entire study.

An AI system equipped with this metadata could reliably generate CRFs, SDTM mappings, and transformation code - because it understands what the study means, not just what the data looks like.

Reality Check

We're not there yet.

Industry adoption is early.

But this is the direction, and **organisations building toward it now will be well-positioned.**

Deterministic-First Principle

“Deterministic-first, with human-verified probabilistic fallback”

Example: SDTM Classification of CRFs

1. Apply deterministic rules derived from standards metadata (e.g., CDASH, SDTM-IG classification rules encoded in Python)
2. If match found:
Return classification with high confidence and rule-based justification
3. If no match:
Fall back to probabilistic LLM inference with documented reasoning
4. Human reviews and approves (Human-in-the-Loop, HITL), especially for lower-confidence classifications

Key Messages:

- Every classification is backed by a justification and confidence level
- Humans know where to focus their review
- The audit trail shows exactly how each decision was made.
- Humans retain responsibility

Technical Enabler: Model Context Protocol (MCP)



MCP adds a deterministic layer to LLMs

- Connect to external tools and knowledge sources
- Query authoritative sources deterministically



Using our earlier example:

1. User asks LLM:
"What domains are in the SDTM 3.4 Interventions class?"
2. LLM recognises it needs authoritative metadata
→ identifies the CDISC Library MCP tool
3. LLM calls the MCP tool
→ MCP queries CDISC Library API directly
4. MCP returns the authoritative answer
→ AG, CM, EC, EX, ML, PR, SU
5. LLM responds to user with correct answer and source citation

Technical Enabler: Agent Skills



Agent Skills represent a modular pattern for extending LLMs

- Transform a general-purpose LLM into a domain specialist
- Progressive context loading (retains focus, reduces cost)
- Supports deterministic and probabilistic behaviour

Technical Enabler: Agent Skills



A proposed Skill hierarchy for SDTM Classification:

SDTM Classification Orchestrator

- Orchestrates the high-level process flow

CRF Parser

- Reads CRF metadata (ODM, Excel) into common format (**deterministic**)

GOC Classifier

- Classifies into a General Observation Class (GOC) using rules (**deterministic-first**)
- Falls back to **probabilistic inference** based on metadata-derived rules
- Loads the minimum context required to classify into a GOC.

<GOC-type> Classifier (multiple)

- Classifies into a specific Domain using rules (**deterministic-first**)
- Falls back to **probabilistic inference** based on metadata-driven rules
- Loads the minimum context required to classify the Domain in the relevant GOC

Technical Enabler: Agent Skills



The Skill hierarchy:

- Minimises token usage
- Provides justification and confidence levels for human review
- Uses structured metadata to provide a defined pathway from input to output

Structured input metadata

- CRF input formats converted to standard internal metadata format

Structured output metadata

- Classifications stored as structured metadata, with confidence levels and justifications for human review

Technical Enabler: Agent Skills



Filesystem structure

```
cs_sdtm-classifier/
├── skills/                # Claude Skills (SKILL.md files)
│   ├── orchestrator/     # Orchestration of deterministic and probabilistic classification
│   ├── goc-classifier/   # Classify into a GOC
│   └── {findings|events|interventions|special-purpose}-classifier/ # Classify into a Domain
├── src/                  # Python modules (deterministic)
│   ├── parsers/          # ODM, Define-XML, Excel
│   ├── classifiers/      # GOC, domain
│   ├── confidence/       # Scoring, thresholds
│   └── utils/
├── reference/            # Reference metadata (plugin-like)
│   ├── config/thresholds.json
│   ├── goc-definitions/{goc}/[metadata.json, patterns.json]
│   └── domain-definitions/{sdtm-ig|custom}/{goc}/{DOMAIN}/[metadata.json, patterns.json]
├── tests/                # Unit, integration, benchmarking
└── docs/                 # Documentation
```

Technical Enabler: Agent Skills

Classification metadata

```
{
  "classifications": [
    {
      "id": "550e8400-e29b-41d4-a716-446655440000",
      "source": {
        "id": "VS.FORM",
        "type": "odm",
        "name": "Vital Signs"
      },
      "goc": {
        "name": "Findings",
        "confidence": 0.95,
        "method": "llm-reasoning",
        "reasoning": "Form contains test-result-unit structure...",
        "rules_matched": [],
        "contra_indicators": []
      }
    },
    {
      "domain": {
        "name": "VS",
        "confidence": 0.98,
        "method": "rules",
        "reasoning": "Variables VSTESTCD, VSORRES, VSORRESU...",
        "rules_matched": [],
        "contra_indicators": []
      },
      "domain_alternatives": [],
      "needs_review": false,
      "metadata": {
        "token_count": 1250,
        "processing_time_ms": 145,
        "reference_files_loaded": [...]
      }
    }
  ]
}
```

Technical Enabler: Agent Skills



Benefits:

- Deterministic where possible, probabilistic where necessary
- Can handle novel CRF content (probabilistic fallback)
- Full justification for every classification - auditable
- Progressive context loading (increases reliability, decreases cost)



Learnings

What We've Learned So Far

Governance first, technology second

Deploying AI tools without a clear governance framework leads to either paralysis (“we can’t approve this”) or shadow IT (“I’ll just use my personal account”). The framework enables, not constrains.

People are more ready than you think, but need clarity

In our survey, 73% of staff see opportunities for AI. The major barrier is unclear guidance, not resistance. Give people guardrails and they’ll run.

Start simple

Build confidence with low-risk use cases (Tier 1/Tier 2, non-client data) before tackling GxP-critical applications. Quick wins create momentum.

Approval and permissions can be harder than the technology

Getting tools approved for evaluation can take longer than the evaluation.

A Note on Probabilistic Behaviour



Remember: People are probabilistic too!

- Ask a person to classify something multiple times
- Ask multiple people to classify something

We already handle this through review, documentation, and quality processes.

Using probabilistic AI doesn't inherently add NEW risk

It can be managed within existing quality frameworks - and often with better documentation than human-only processes provide.

Key Takeaways



- **AI in GxP is possible**
 - Regulatory guidance supports risk-based approaches
- **Metadata and standards are the key**
 - They remove the guesswork and increase clarity
- **Deterministic-first with human-in-the-loop**
 - A practical pattern for compliance
- **Start the journey now**
 - Iterate toward the agentic semantic future



Thank you!