From Open Data to Operations - Building AI Governance Infrastructure

Noel McLoughlin
Principal Engineer TD

DISCLAIMER - My presentation and demo is based on personal Open-Source contributions and collaboration outside FINOS

Personal

https://github.com/noelmcloughlin/

Community

https://github.com/linkml/linkml

Industry Experiments

https://github.com/lmodel/

My Open-Source interest – Model driven Technology

LinkML Community Open Source Project

JOURNAL ARTICLE

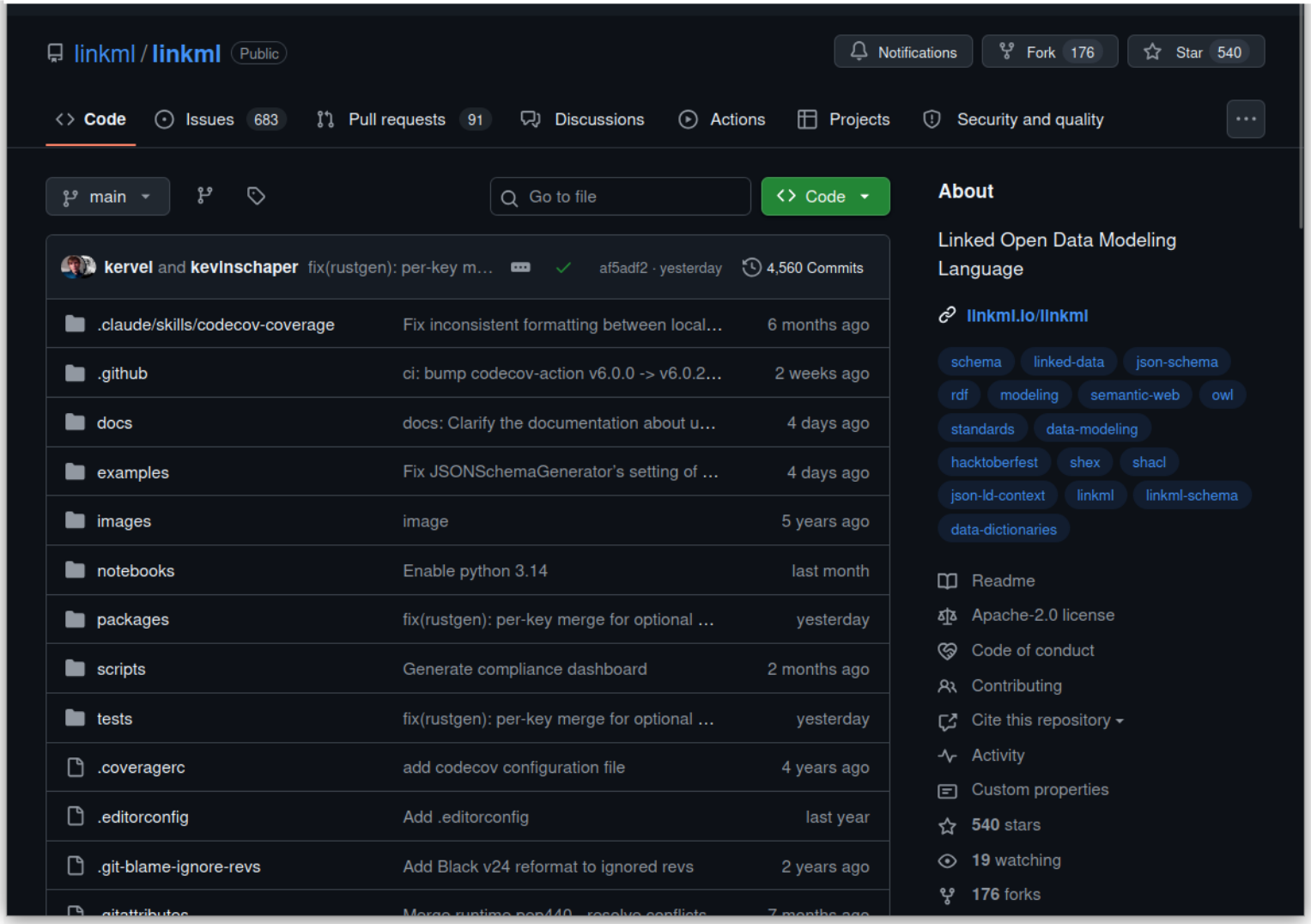
LinkML: an open data modeling framework

Sierra A T Moxon ✉, Harold Solbrig, Nomi L Harris, Patrick Kalita, Mark A Miller, Sujay Patil , Kevin Schaper, Chris Bizon, J Harry Caufield, Silvano Cirujano Cuesta ... Show more

GigaScience, Volume 15, 2026, giaf152, https://doi.org/10.1093/gigascience/giaf152

Published: 12 December 2025 **Article history ▾**

LinkML Community Members (listed in alphabetical order):
Richard M. Bruskiewich, Seth Carbon, Eric Cavanna, John-Marc Chandonia, Shreyas Cholia, Ben Dichter, Emiley A. Eloe-Fadrosh, Vincent Emonet, Shahim Essaid, James A. Fellows Yates, Joseph Flack, Satrajit S. Ghosh, Damien Goutte-Gattat, Dorota Jarecka, Dazhi Jiao, Marcin P. Joachimiak, Vlad Korolev, Volodymyr Lapkin, Noel McLoughlin, Sierra D. Miller, Michael Milton, Josh Moore, Moni Muñoz-Torres, B. Nolan Nichols, Tim E. Putman, Justin T. Reese, Victoria Savage, Philip Stroemert, Jeremy Teoh, Anne Thessen, Isaac To, Puja Trivedi, Vincent Vialard, Trish Whetzel



https://www.statnews.com/2026/05/19/ai-helped-find-treatment-newborn-ultra-rare-disease/

HEALTH

How AI helped treat a newborn's ultra rare disease. 'It was almost like a light switch.'

An AI tool, Biomedical Data Translator, helped doctors at Mayo Clinic find a treatment for Jorie Kraus

Get Alerts | Email | Share

Two-year-old Jorie Kraus with her father, Dave Kraus, at the STAT Breakthrough Summit West in San Francisco on Tuesday. Jack Simpson for STAT

By Casey Ross May 19, 2026
Chief Investigative Reporter, Data & Technology

SAN FRANCISCO — In the first, tenuous weeks of her life, Jorie Kraus and her parents faced her possible death repeatedly. Muscles throughout her tiny body simply didn't work properly. Her heart. Her legs. Her larynx. Even the involuntary action of breathing was labored, and constantly faltering.

Behind this "AI" headline was LinkML …

AI depends on Data:

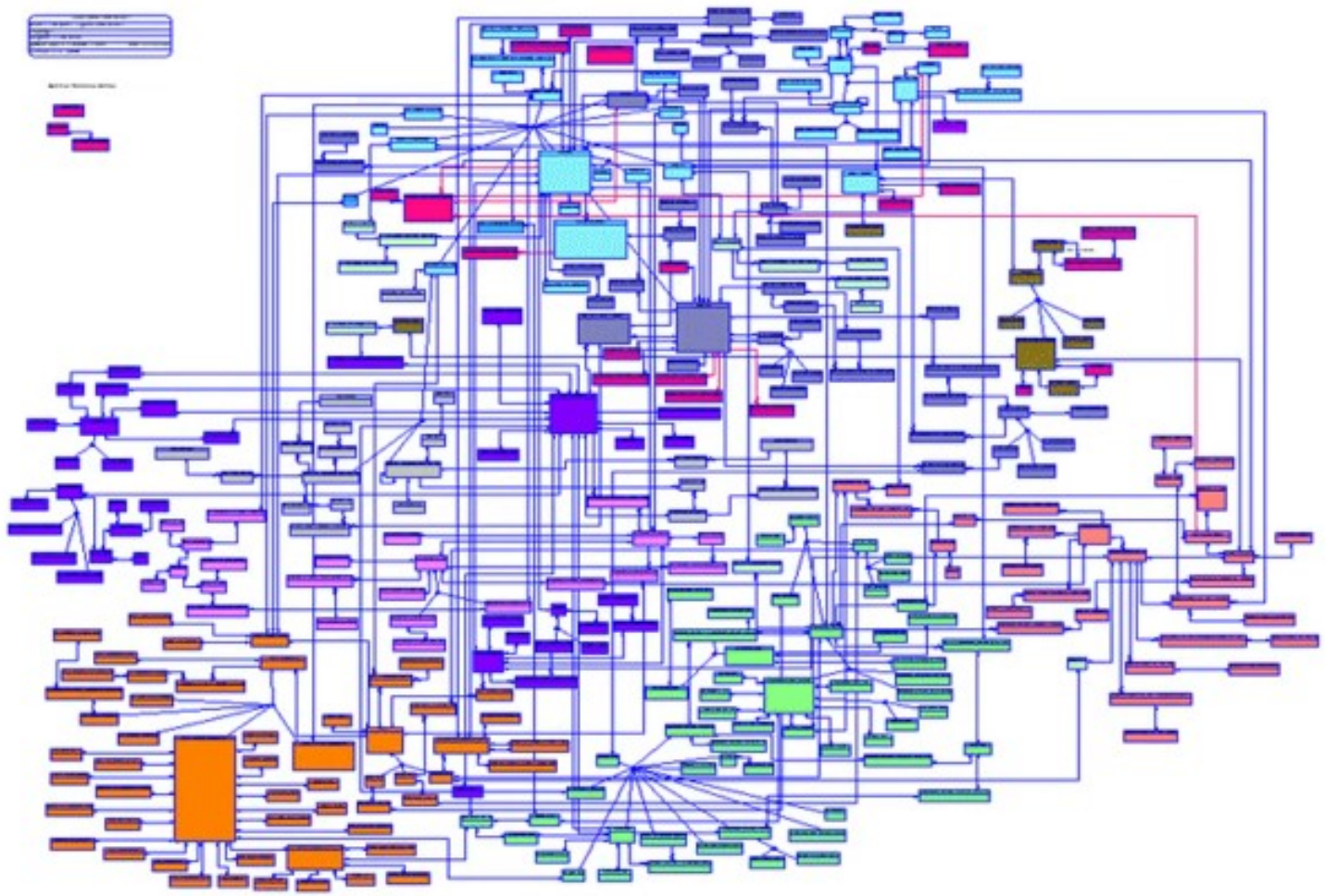
- with comprehensive contextual metadata
- being labelled with standardized labels
- being linked

LinkML is foundation for AI ready data

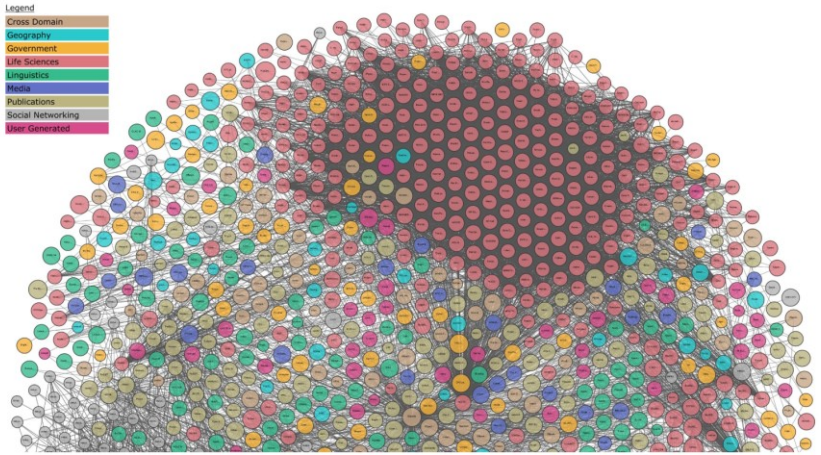
- Social glue
- Schema Language, Tools
- Data Management and extraction

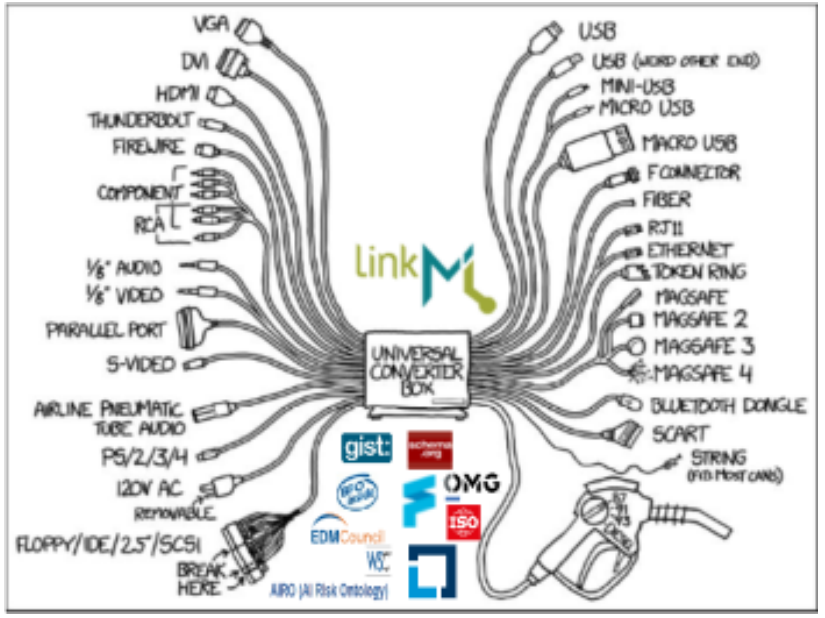
AI-assisted workflows for integrating data using LinkML

And our institutional data looks like this *



Can we retrofit Linked Data semantics?





* See "**What is Data-Centric Architecture"**, Dave McComb, DCAF 2021, creator of https://datacentricmanifesto.org

The Linked Open Data Cloud: https://lod-cloud.net/

AI-assisted workflows for integrating data using LinkML

Stretch Goal: Compliance at speed of AI

Open Standards, Obligations, Incidents, and Whitepapers

NIST, CISA, EU, FED, ISO, IEC, W3C, etc.

causes

influences

Open Curation

TODAY'S STATE: Compliance Doom Cycle

Regulators START LEFT (of Industry) → New Obligations → Large Compliance remediation → Compliant → Doom Cycle repeats

IDEAL FUTURE STATE: **Start Left** Compliance Joy *

enables

START LEFT → Automated Compliance → New Obligations → Small Compliance remediation → Joy Cycle repeats

** Standards provide common structure & language & defensibility - identifiers, reusable Ontologies, Curation pathways.*

NIST reciprocity: https://csrc.nist.gov/glossary/term/reciprocity

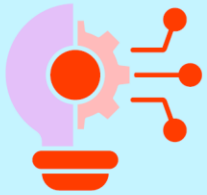
Linked Data Standards

Linux Foundation Pathways to Open Data

"Open Data definition - Data infrastructure with the technical and legal requirements in place to make the data freely accessible for universal use, reuse, and redistribution".

Data silos hamstring research & innovation, and have become increasingly onerous alongside growing data needs to train AI models.

Open data is freely accessible for universal use, leading to **new avenues for innovation, greater reliability, & increased trust.**



The unique qualities of data as compared to software — such as **maintenance, quality, privacy, & license diversity** — make its openness challenging.

Significant human resources are required for cleaning, standardizing, & maintaining a dataset.

The financial & resource **costs of dataset maintenance** engender a **tradeoff between** the **quality** of the data & the **cost** of accessing it.

While a lack of standardization makes datasets unusable, AI tools offer opportunities to better manage unstructured data.

Data privacy concerns stem from compliance with regulations such as GDPR, which create **a climate of risk aversion.**

Proprietary control over data gives companies greater certainty around compliance & quality while **reducing the fear of losing competitive advantage.**

"Semi-open" data platforms allow for collaborators to share best practices and other pre-competitive data while maintaining their competitive advantage.

Overture Maps Foundation has built an **open, agnostic, & standardized geospatial data platform** for data owners & service providers to leverage.

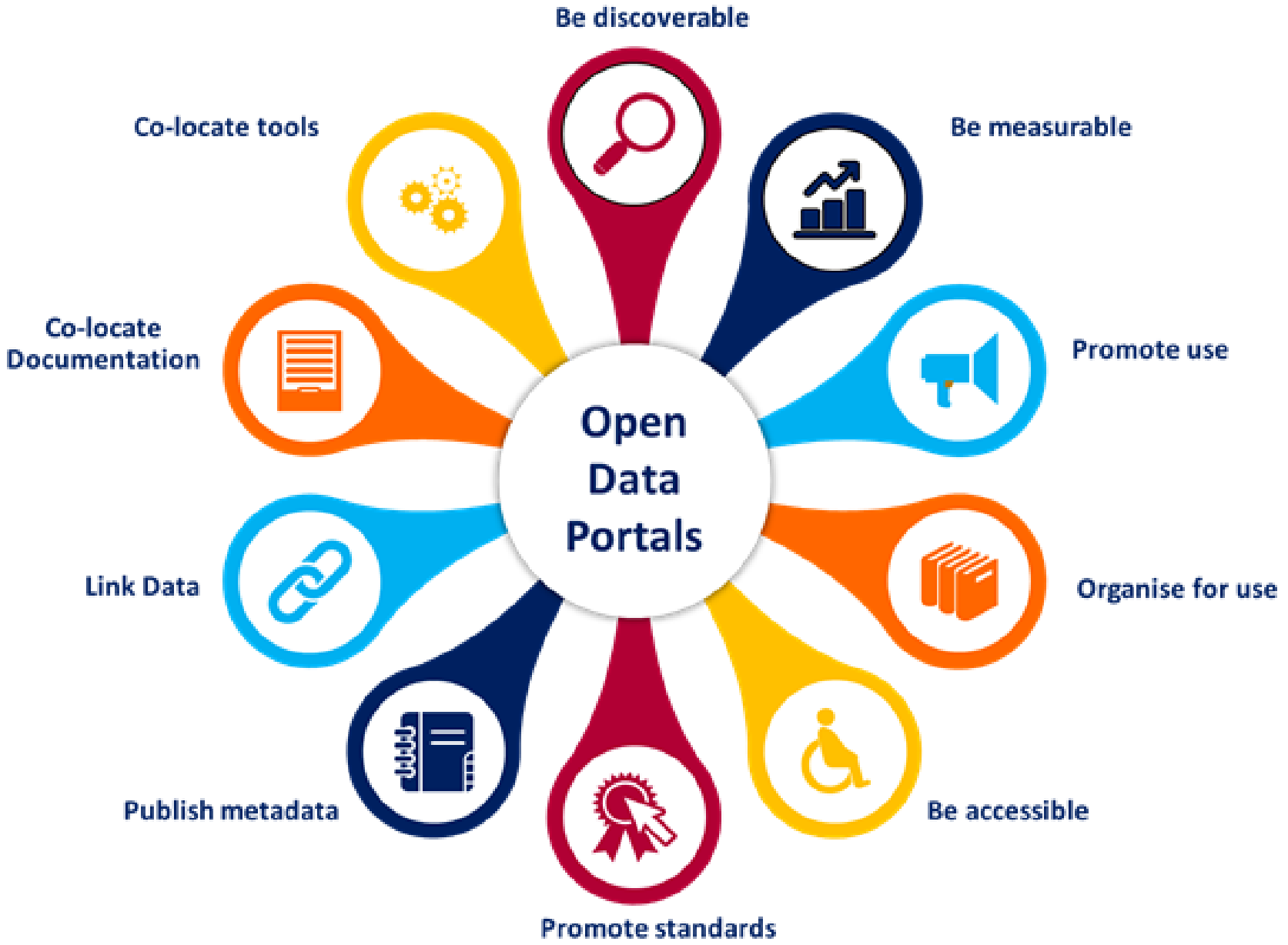
Building open data infrastructure requires a **reworking of current data collection & sharing processes.**

Open data requires **incentivizing collaboration** around a pre-competitive layer while **incorporating checks & balances** in the governance structure.

Hermansen, A., & Wiegmann, P. M. (2025). Pathways to Open Data: Findings from the 2024 World Open Innovation Conference Challenge Session. The Linux Foundation. https://doi.org/10.70828/XBBS2621

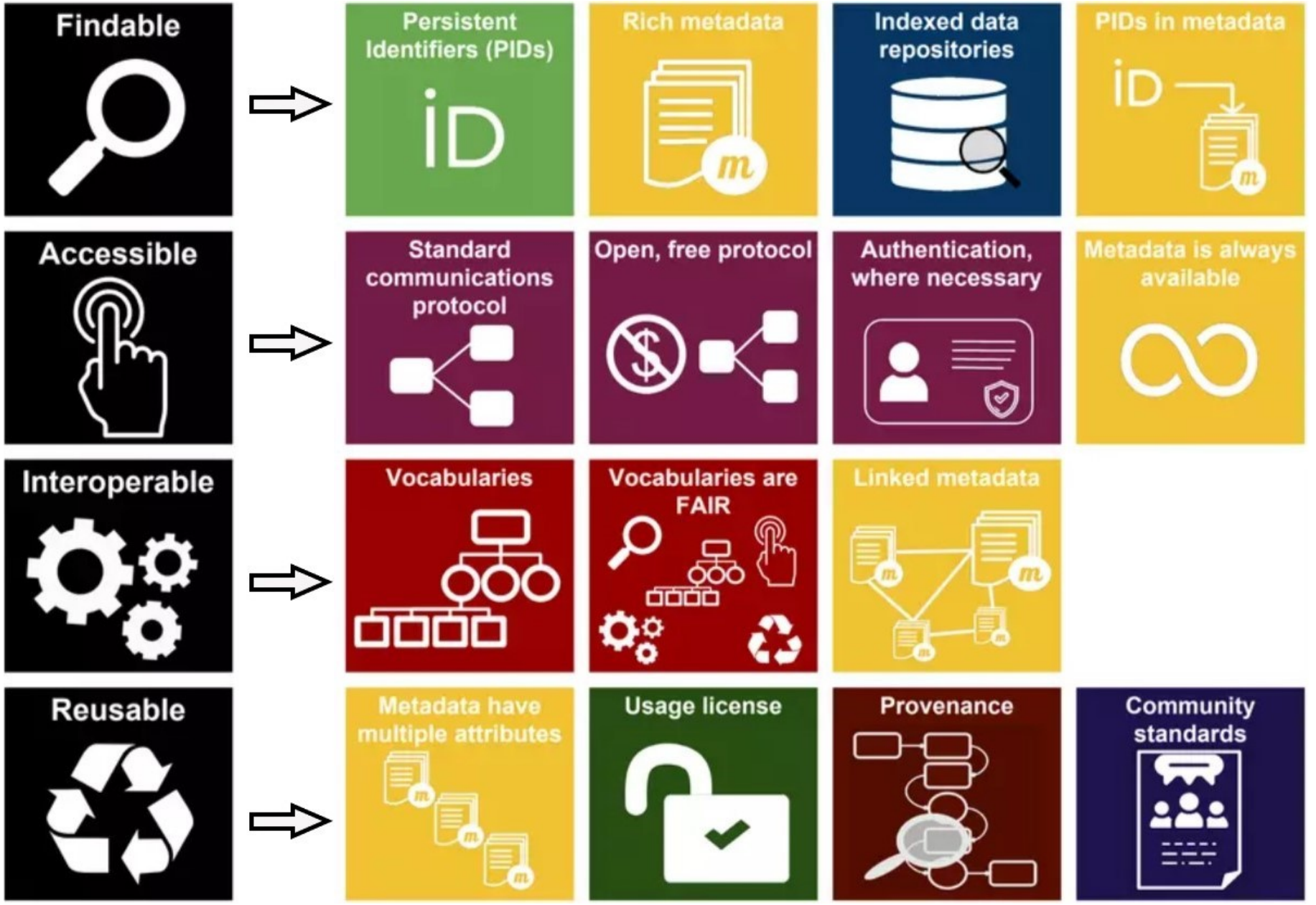
Portals to Open Data

"Ten ways portals can organize for sustainability and added value by examining what is required on the other side of the equation - meeting and provoking demand".



The future of open data portals: https://data.europa.eu/doi/10.2830/879461 and Digital Public Goods https://www.digitalpublicgoods.net/

FAIR Open Data

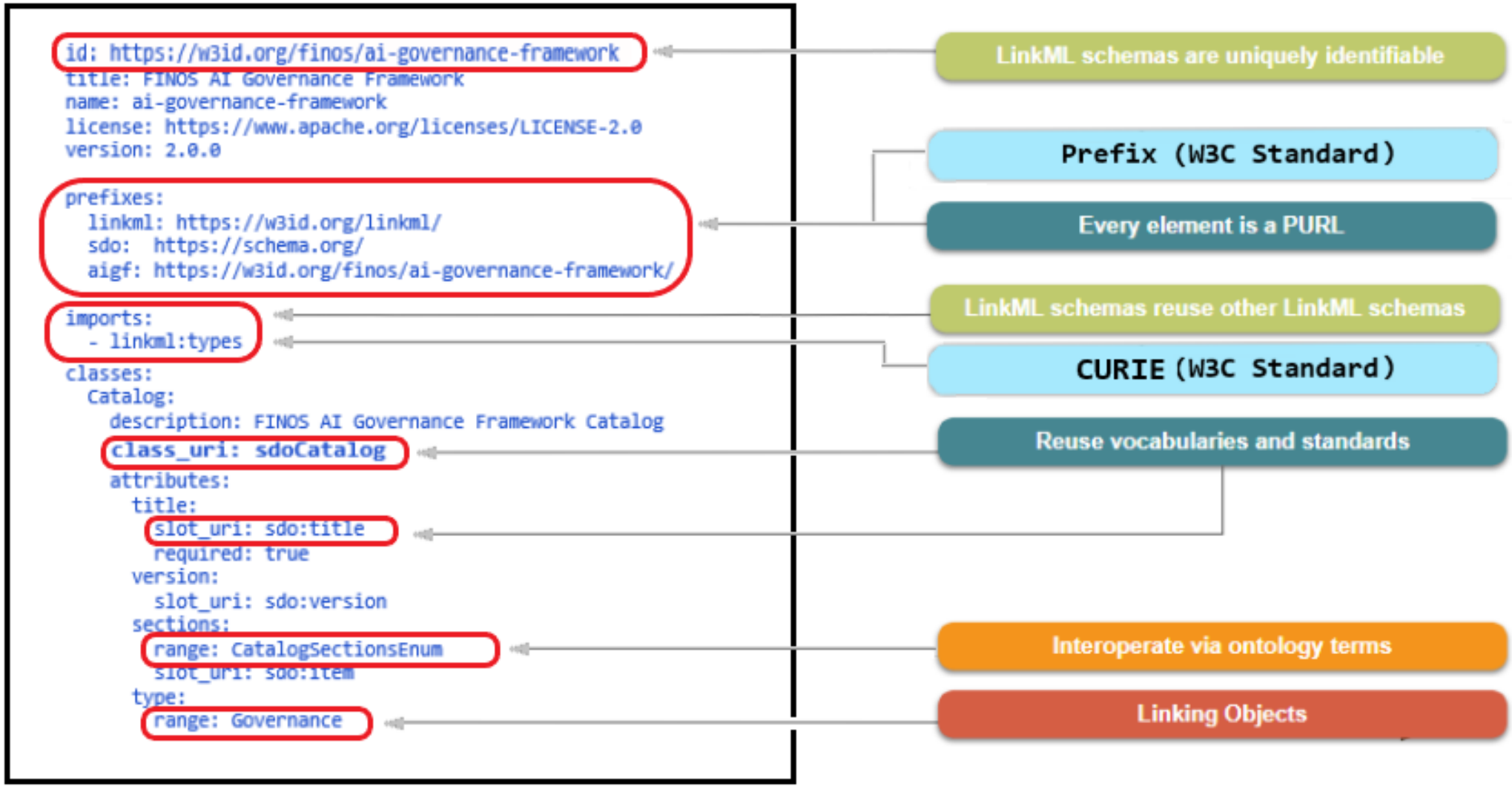


Principles

Findable
Accessible
Interoperable
Reusable

https://www.go-fair.org/fair-principles/

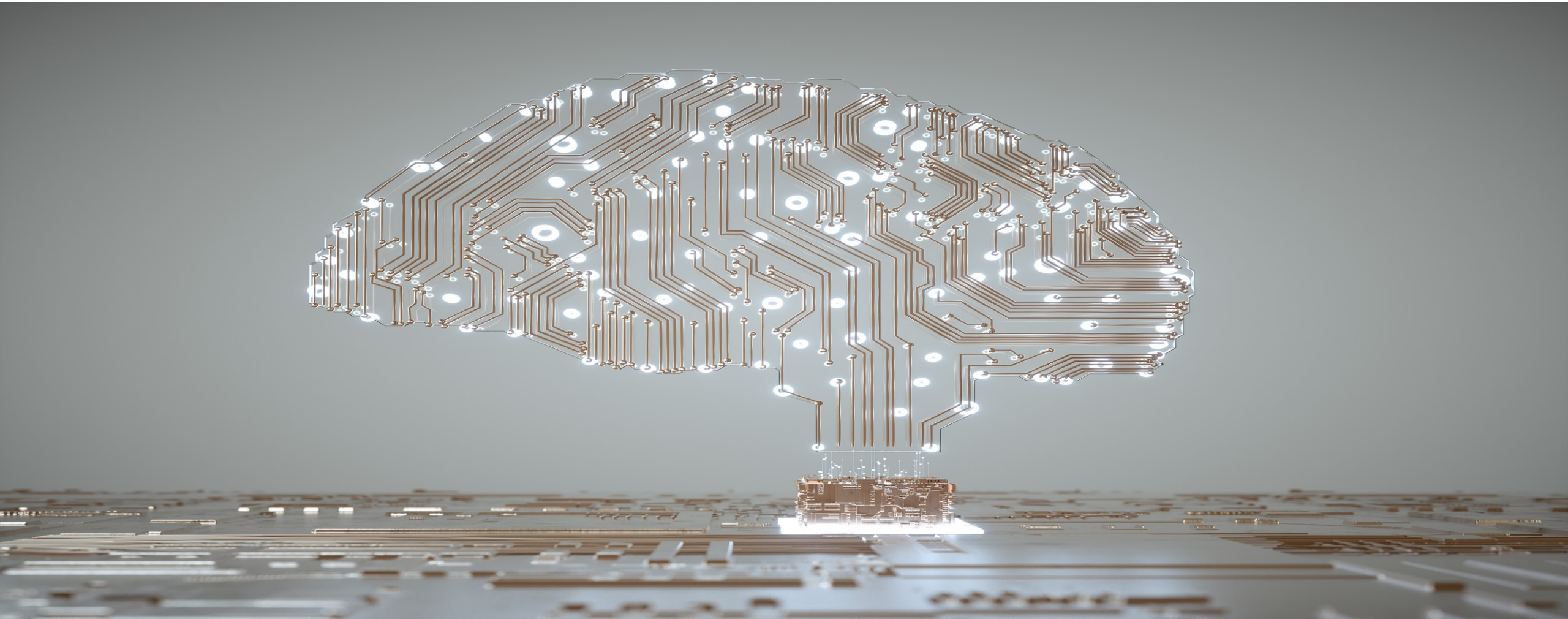
LinkML is a FAIR Schema Language



https://air-governance-framework.finos.org/ with Linked Data Sidecar: https://github.com/lmodel/ai-governance-framework

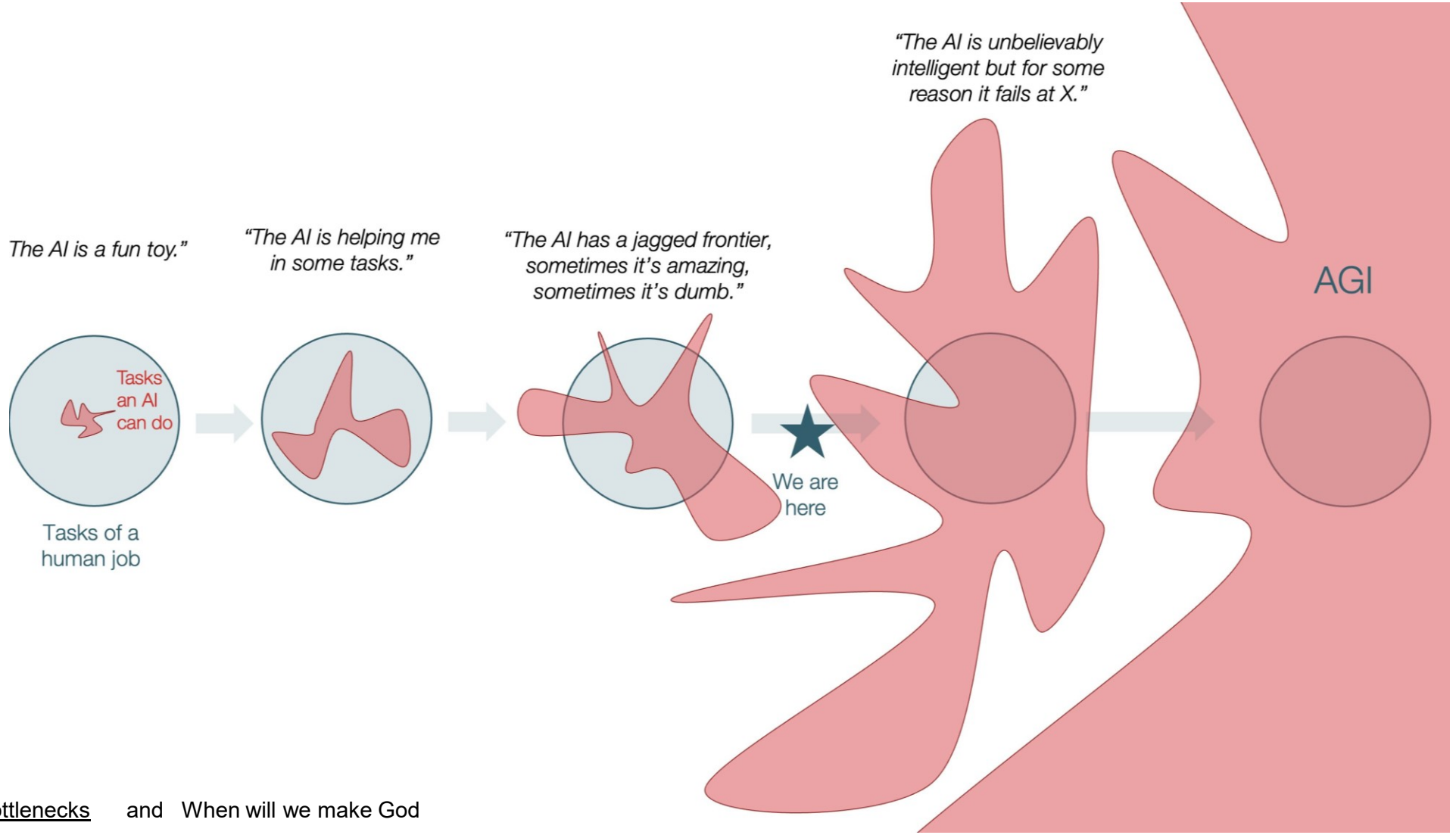
AI-assisted workflows for integrating data using LinkML and CURIES and Prefix mappings

AI Risk Model



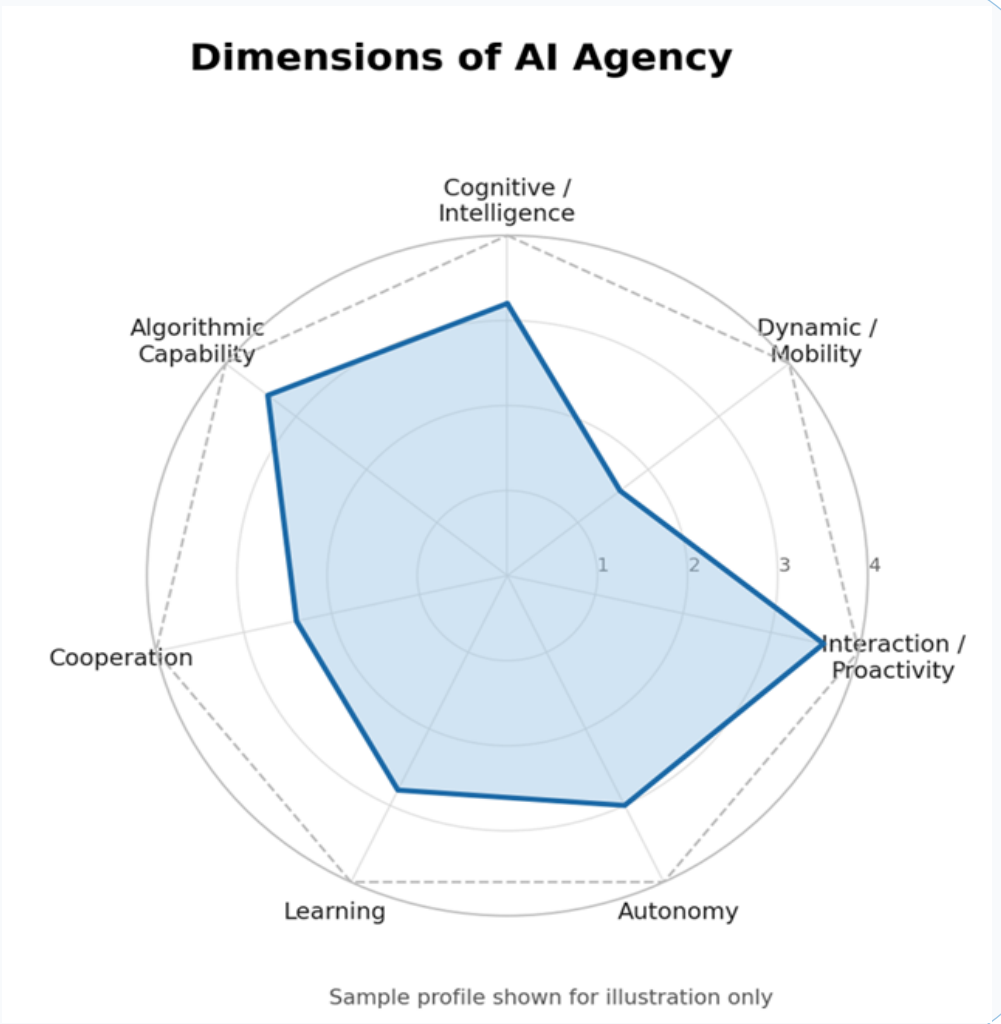
Artificial General Intelligence (AGI)

"Bottlenecks can create the impression that AI will never be able to do something, when progress is held back by a single jagged weakness. When that weakness becomes a reverse salient, and AI labs suddenly fix the problem, the entire system can jump forward … What does it matter if AI is relatively bad at running a vending machine, if the AI still becomes better than any human?"



The Shape of AI Jaggedness Bottlenecks and When will we make God

AI AUTONOMY CHANGES THE RISK MODEL (TODO)



Use these definitions when scoring an AI system (1-4) on the radar chart.

Cognitive / Intelligence: From reflex response to model-based, inference, prediction, and evaluation-driven behavior.

Dynamic / Mobility: Whether the agent itself, its sensors, or its effectors can move or reconfigure.

Interaction / Proactivity: Inactive -> reactive -> deliberative/proactive behavior in response to environmental change.

Autonomy: Controlled -> dependent -> autonomous operation with increasing independence from external control.

Learning: Conservative -> remembering -> learning agent, based on whether capability improves through experience.

Cooperation: Individualistic -> competitive -> collaborative interaction with other agents or teams.

Algorithmic Capability: Underlying computational power available to the agent; foundational across all other dimensions.

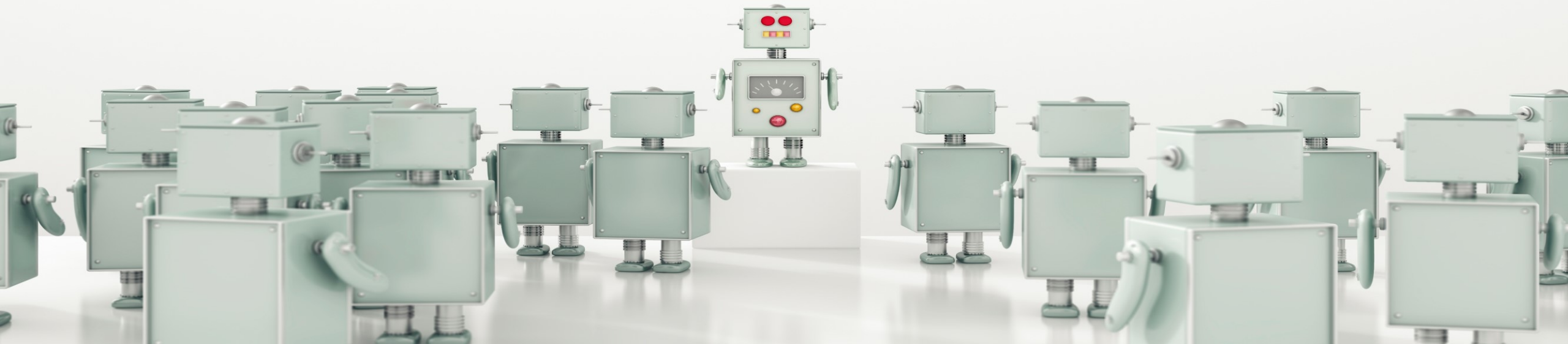
We MUST govern AGI – not other way around

Humans are rising!

*With a "START LEFT" mindset, DevSecOps posture, and the power of "NIST Reciprocity", they are building something bigger than compliance – **an Open GRC movement**.*

*Open Curation (digital schemas, obligations, mappings, crosswalks, test data) builds Secure Linked Data Supply Chain fueling each companies own "**Secure Software Supply Chain**" with structure, traceability, and confidence.*

***Human oversight moves higher – into harmonization, curation, and trust**.*



Open GRC Curation

Reciprocity as trust-substrate for enablement

reciprocity

Definitions:

Mutual agreement among participating organizations to accept each other's security assessments in order to reuse information system resources and/or to accept each other's assessed security posture in order to share information.

Sources:

NIST SP 800-39 under Reciprocity

Mutual agreement among participating enterprises to accept each other's security assessments in order to reuse information system resources and/or to accept each other's assessed security posture in order to share information.

Sources:

CNSSI 4009-2015

NIST SP 800-12 Rev. 1 under Reciprocity

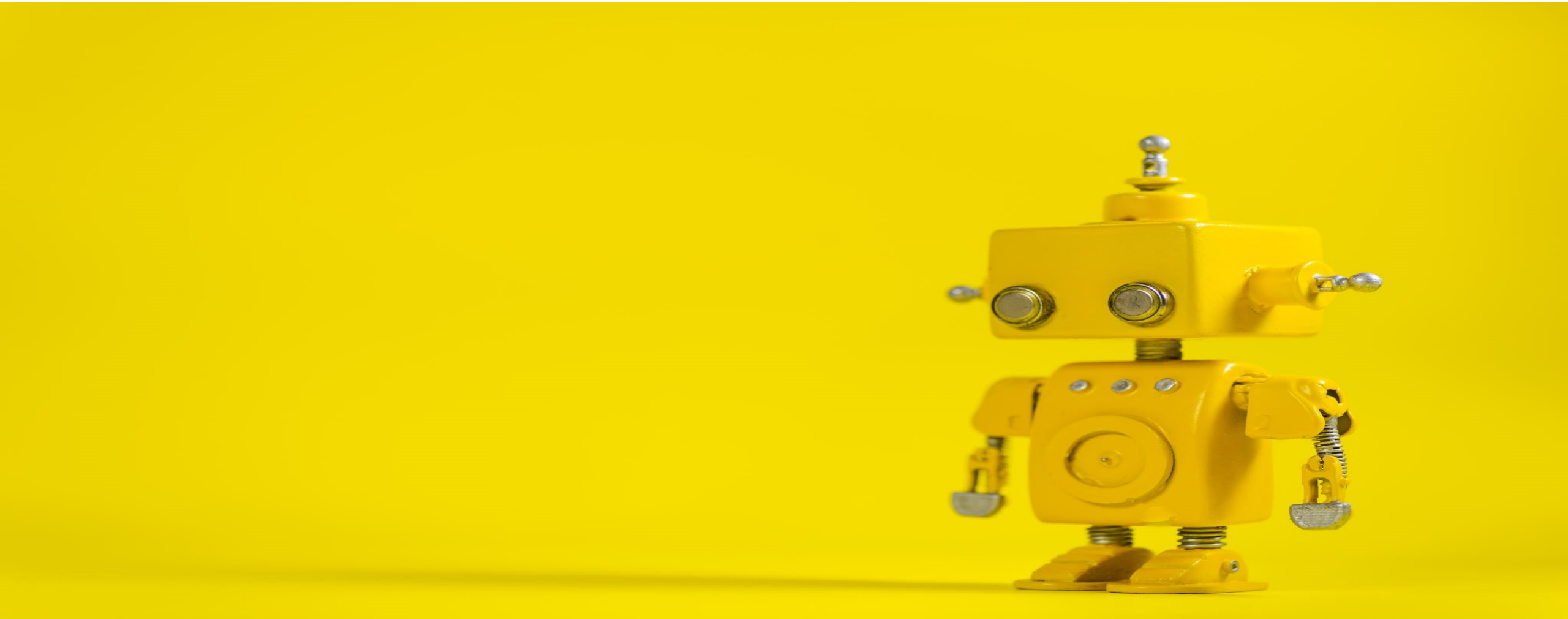
Reciprocity

Accept each other's security assessments to reuse system resources and/or to accept each other's assessed security posture to share information.

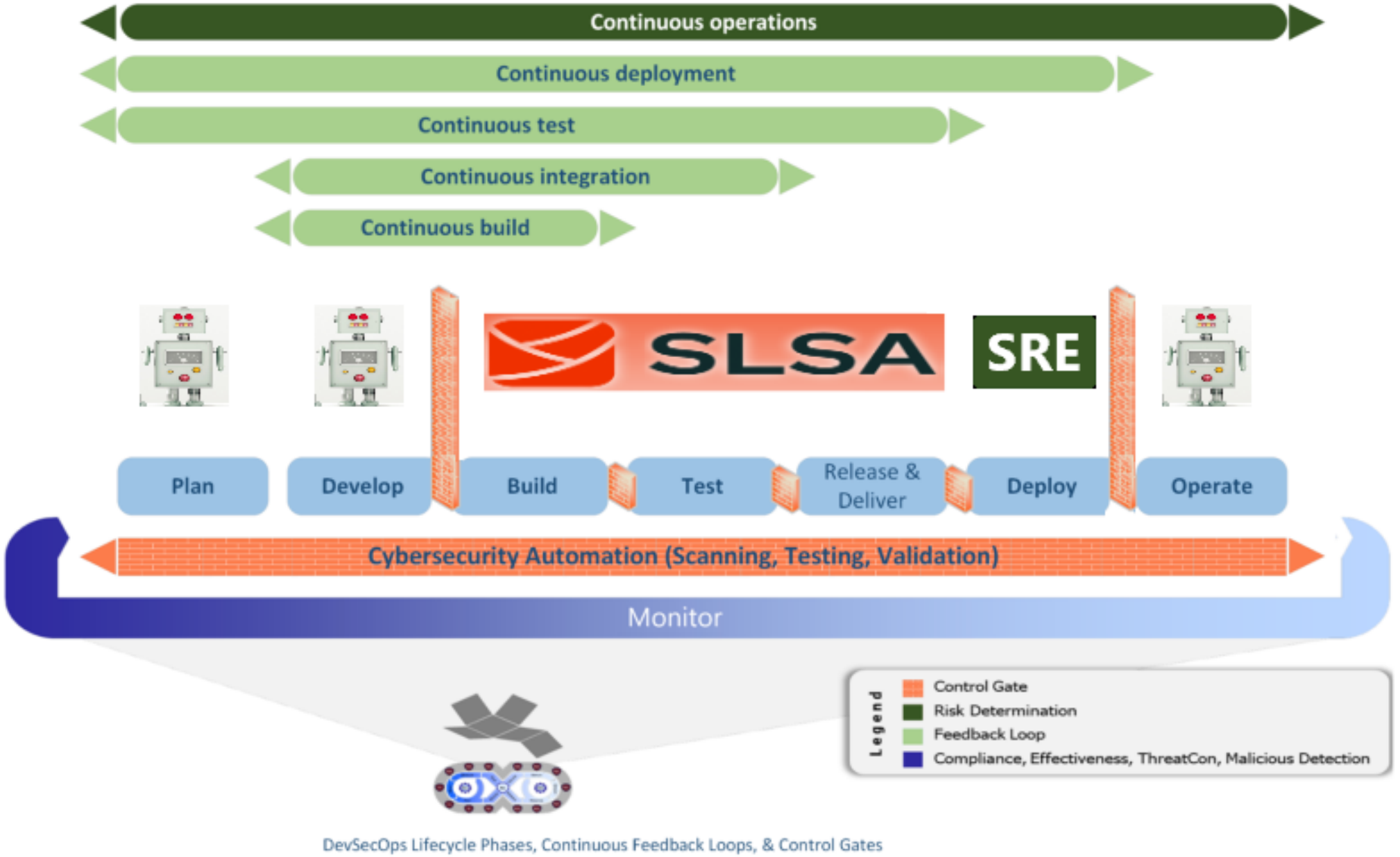
Link: NIST Reciprocity

See CSRMC (Cyber Security Risk Management Construct) and CSRMC Strategic Tenets

DevSecOps Architecture: AI and Supply Chains



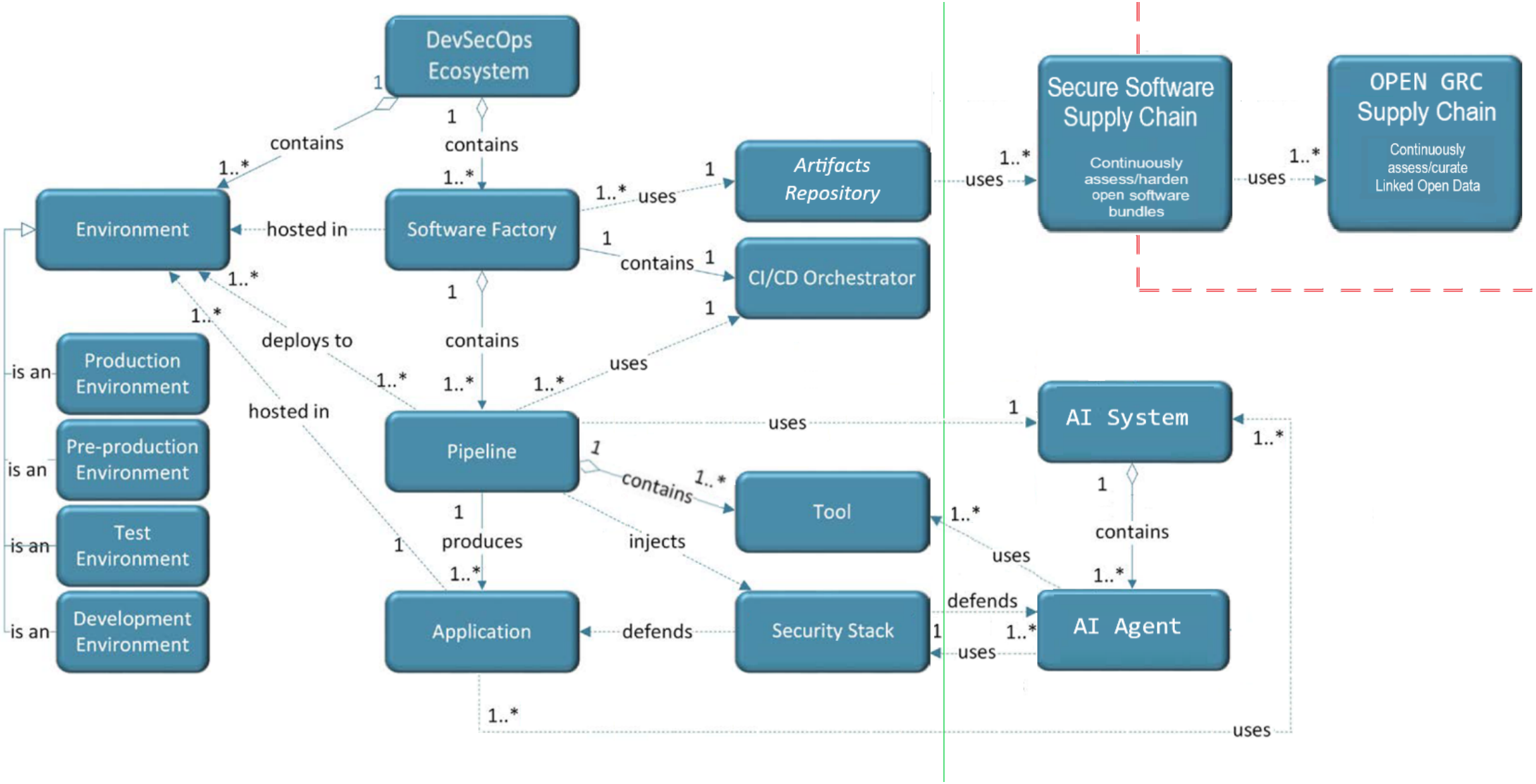
DevSecOps Infinity Loop unfolded – Where does AI fit?



Modified version of image published by US DoD DevSecOps Reference Design

See also CNCF Secure Software Factory and FRSCA

DevSecOps Reference Architecture: Conceptual – What adjustments are needed?

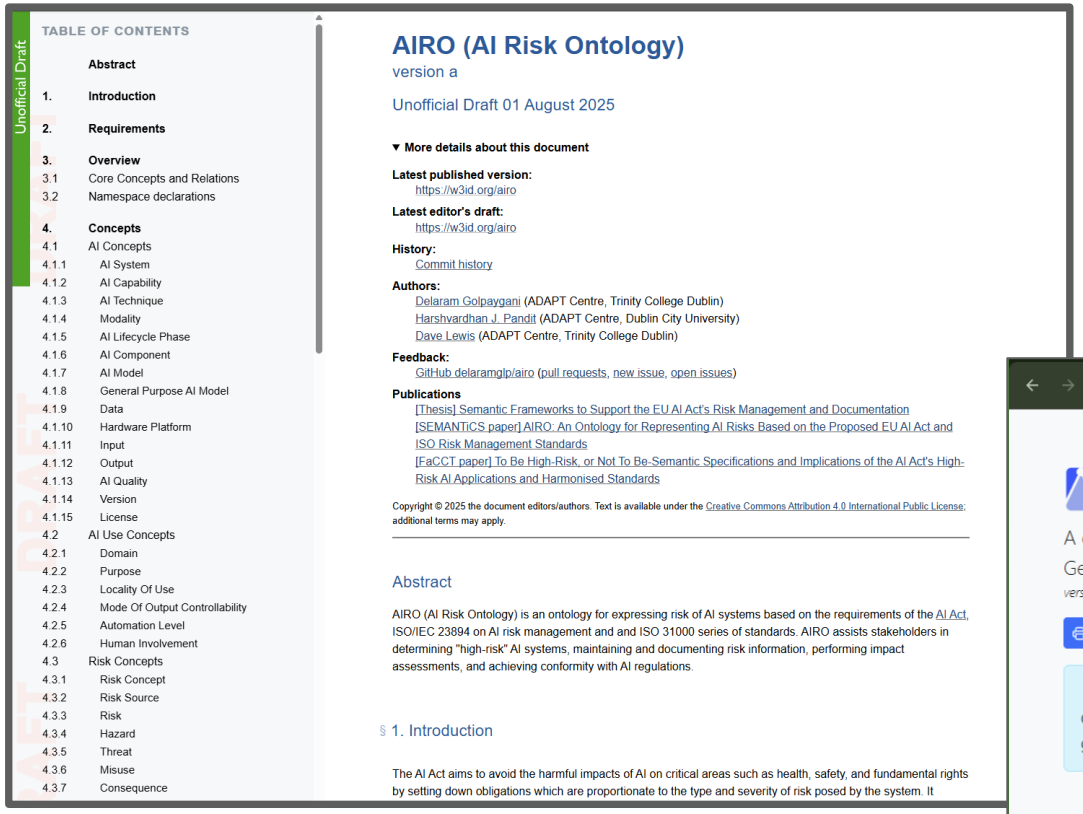


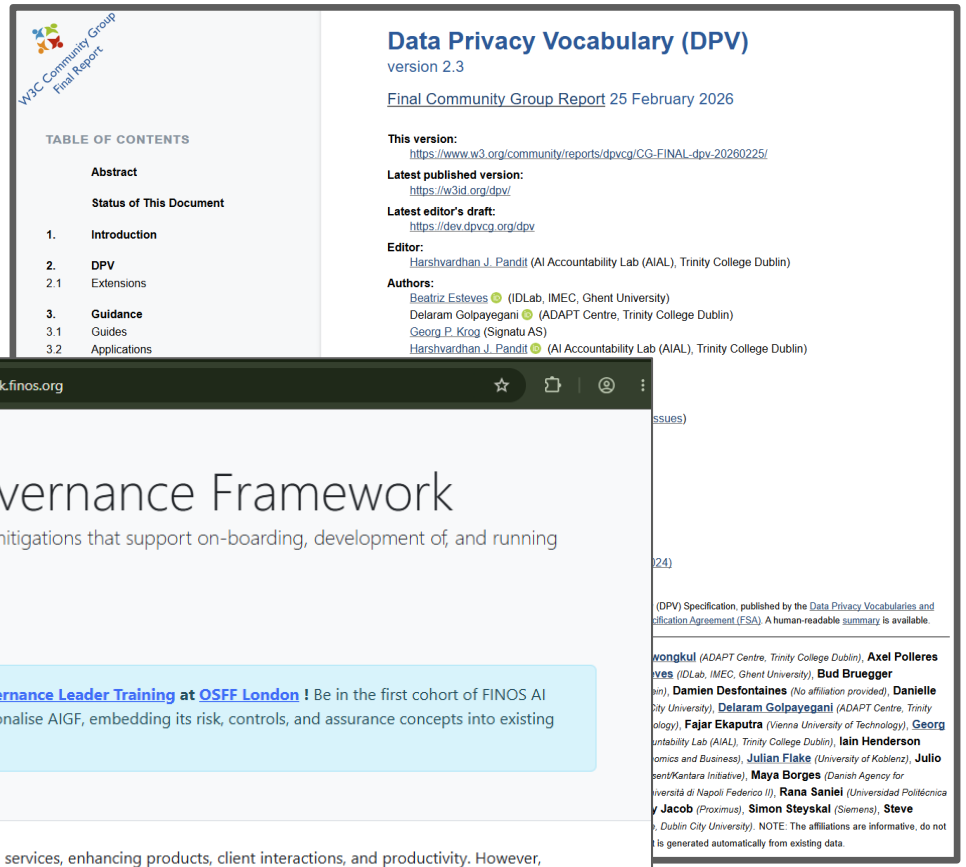
Modified version of image published by US DoD DevSecOps Reference Design

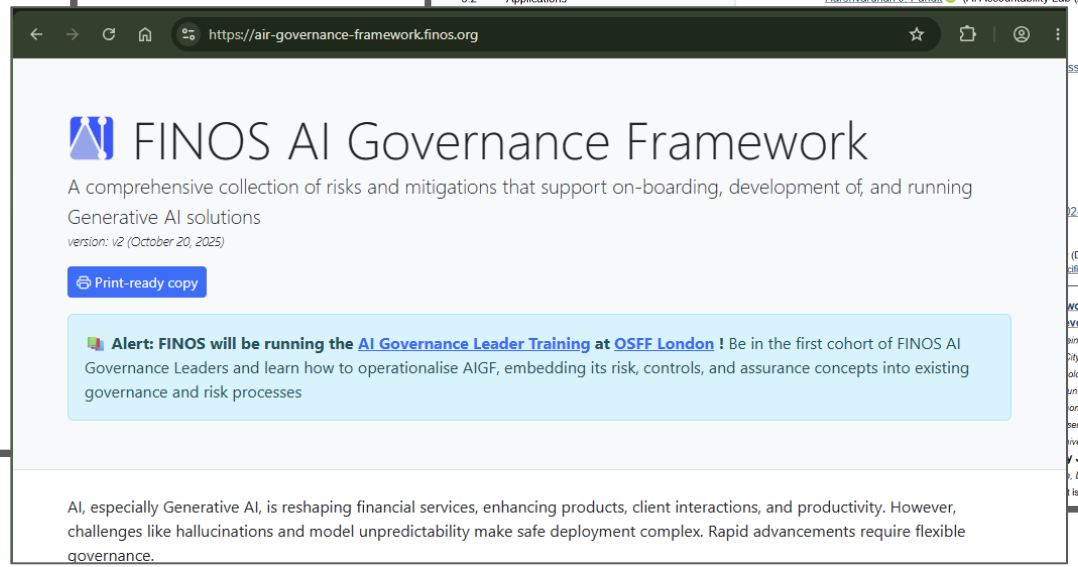
Solution: AI Governance Knowledge Graph

Open-Source Approach

Build AI Governance Knowledge Graph using W3C standards – with seamless integration of FINOS AI Governance Framework







https://w3id.org/airo

https://w3id.org/dpv/

AI Governance Knowledge Graph

JOURNAL ARTICLE

LinkML: an open data modeling framework

Sierra A T Moxon ✉, Harold Solbrig, Nomi L Harris, Patrick Kalita, Mark A Miller, Sujay Patil , Kevin Schaper, Chris Bizon, J Harry Caufield, Silvano Cirujano Cuesta ... Show more

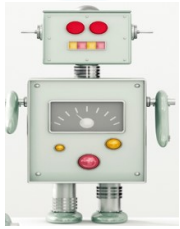
GigaScience, Volume 15, 2026, giaf152, https://doi.org/10.1093/gigascience/giaf152

Published: 12 December 2025 **Article history** ▾

Many real-world projects have adopted LinkML to leverage its core features. For example, the Biolink Model [22] is built using LinkML to define a comprehensive, ontology-aligned schema that connects concepts across diverse biological domains, enabling interoperability among tools, datasets, and knowledge graphs. The National Center for Advancing Translational Science Biomedical Data Translator [23] project builds on this by using the Biolink Model to

Connect concepts across domains

FINOS AI Governance Framework, CDM, Fluxnova, CALM, Common Cloud Controls; Linux Foundation MCP; NIST, ISO, OWASP, ATT&CK, SOX, NIS2, CRA, AI ACT, GDPR, etc



https://doi.org/10.1093/gigascience/giaf152

AI Atlas Nexus – Knowledge Graph of AI Things – with API

https://ibm.github.io/ai-atlas-nexus/

- 🏗️ An ontology has been provided, that combines the AI risk view (taxonomies, risks, actions) with an AI model view (AI systems, AI models, model evaluations) into one coherent schema
- 🧩 AI Risks collected from IBM AI Risk Atlas, IBM Granite Guardian, MIT AI Risk Repository, NIST Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, the AI Risk Taxonomy (AIR 2024), the AILuminate Benchmark, Credo's Unified Control Framework, and OWASP Top 10 for Large Language Model Applications
- 🔗 Mappings are proposed between the taxonomies and between risks and actions
- 🐍 Use the python library methods to quickly explore available risks, relations and actions
- 🚨 Use the python library methods to detect potential risks in your usecase
- 📥 Download an exported graph populated with data instances
- ✨ Example use-case of auto-assistance in compliance questionnaires using CoT examples and AI Atlas Nexus
- 🔧 Tooling to convert the LinkML schema and instance data into a Cypher representation to populate a graph database

AI Ontology
LinkML Schema

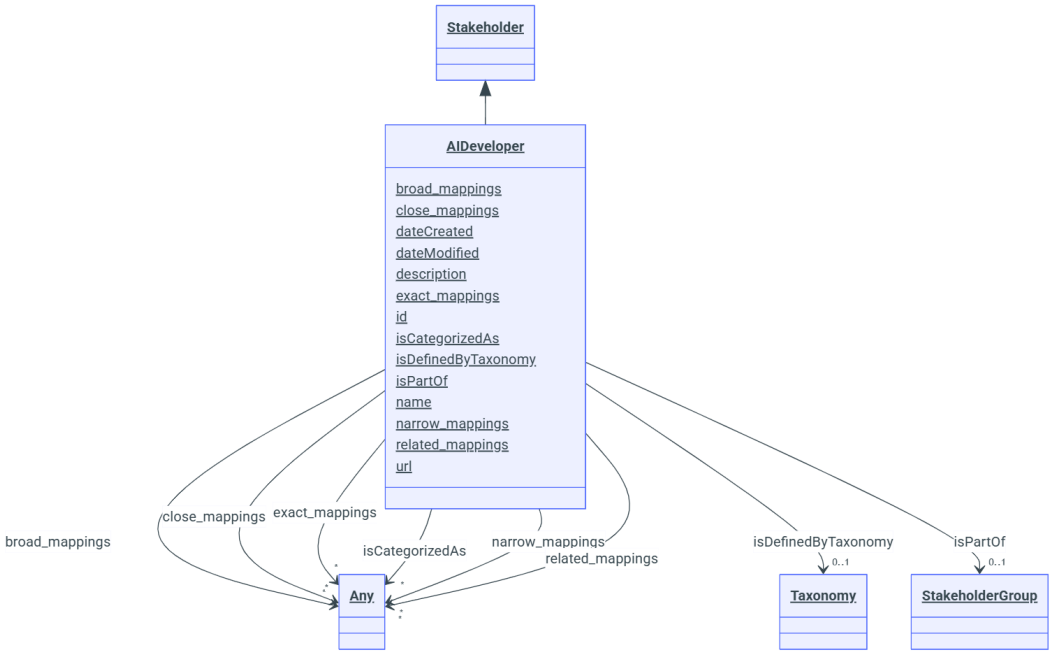
Persistent Identifier

https://w3id.org/ai-atlas-nexus

Class: AIDeveloper

An organisation or entity that is concerned with the development of AI services and products.

URI: airo:AIDeveloper



Inheritance

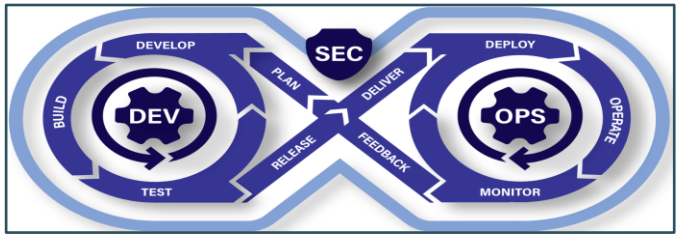
- Entity
- Stakeholder

DEMO

https://github.com/noelmcloughlin/ai-linkmo

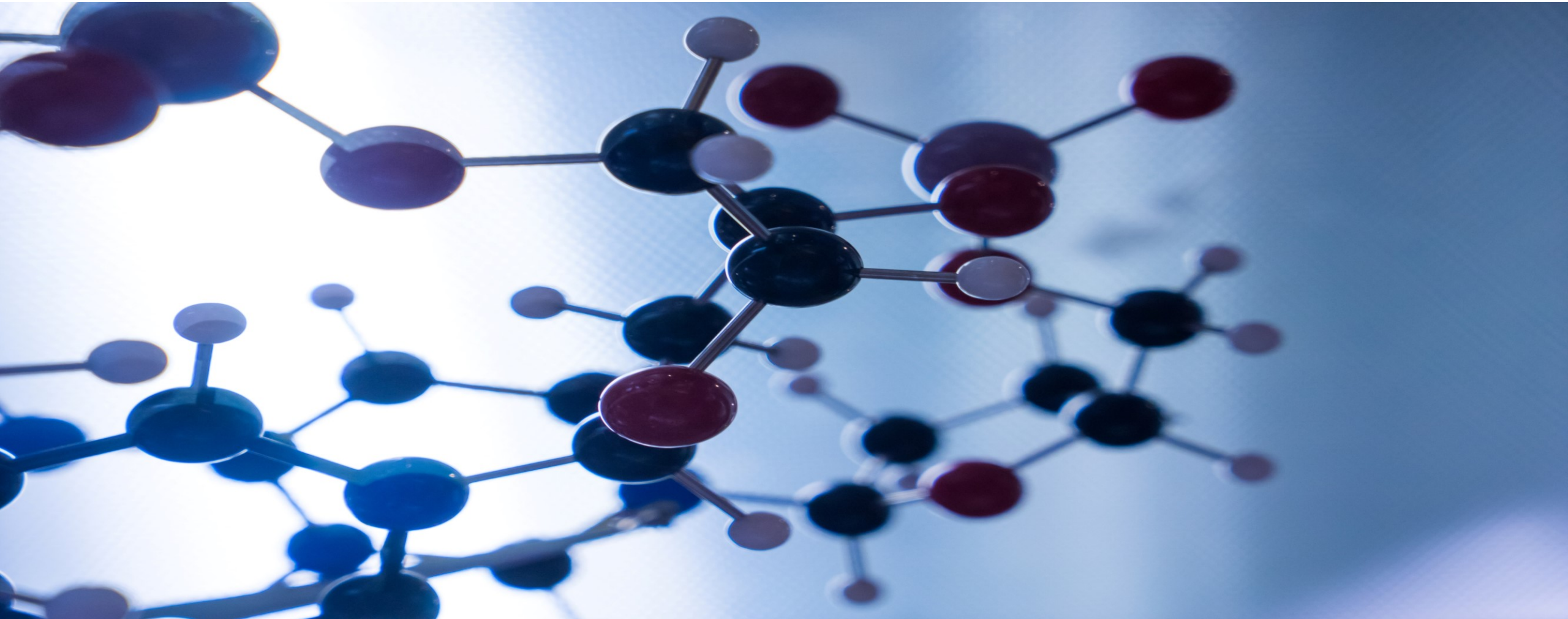
Demo Architecture

Schema-first -> single source of truth

	**Command Line Interface (CLI)**	**Python** **Automation and CI/CD Integration**
	**Application Program Interface (API)**	**Fast API** **Interoperability and Automation**
WWW	User Interface (UI)	**Svelte Typescript** **Exploration**
	Graph DB	**Neo4J** Relationship analysis and Regulatory Crosswalks

Questions / Discussion

Knowledge Graph Mappings Standards



Mappings standard exchange format

The Simple Standard for Sharing Ontological Mappings (SSSOM) is a community-driven standard designed to facilitate the exchange and integration of semantic entity mappings. As data interoperability becomes increasingly crucial across various domains, SSSOM provides a standardized format to share mappings, enabling researchers and developers to more easily connect and utilize diverse datasets. By establishing a common framework, SSSOM enhances the consistency, quality, and discoverability of mappings, thereby supporting more effective data integration and analysis.

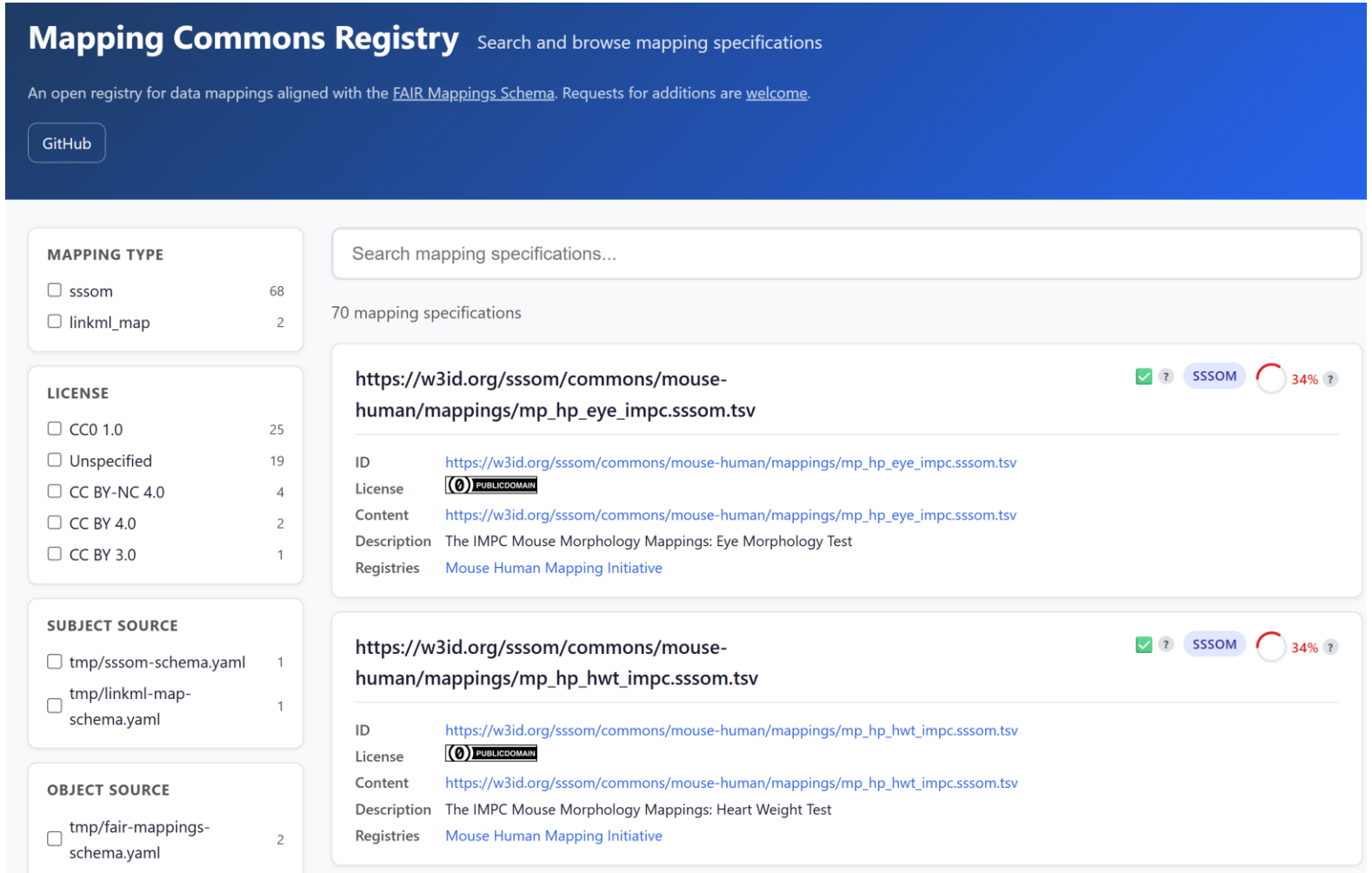
- **Standardization**: SSSOM provides a unified format for representing semantic, or ontological, mappings, making it easier for different systems and organizations to exchange mapping

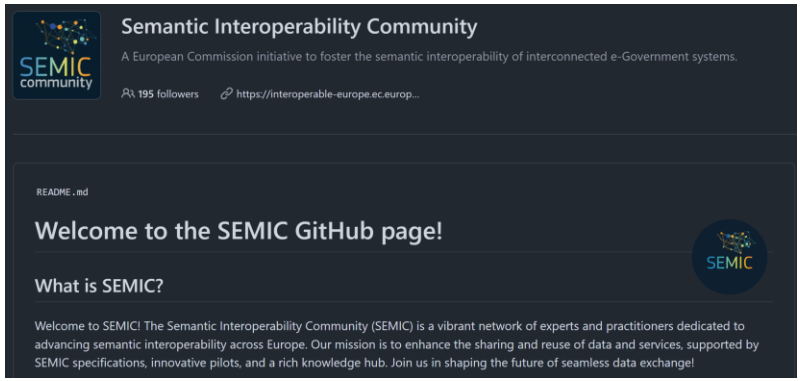
Persistent Identifier

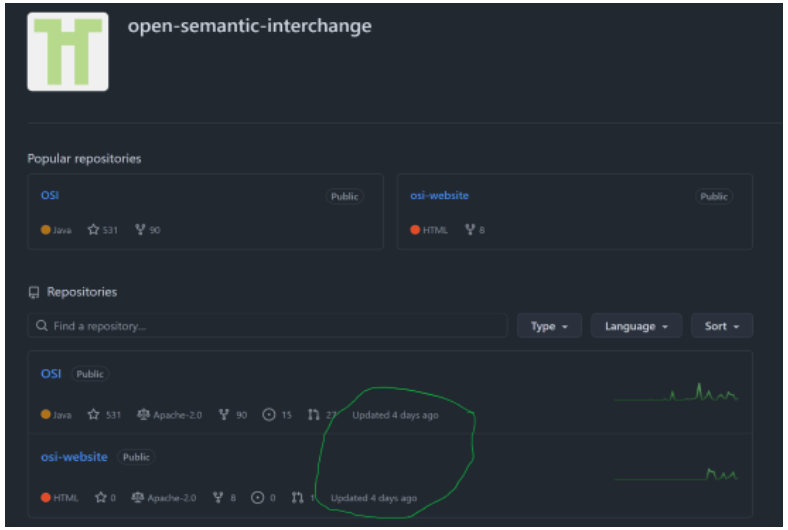
http://w3id.org/sssom

To simplify Governance, a FAIR open mappings registry for storing, curating, managing, and sharing Open GRC artefacts with provenance would help – think "DockerHub for governance"

Potential Consortia: Mapping Commons Community (Corporate Social Responsibility), EU SEMIC Community, OSI







See https://mapping-commons.github.io/ and https://github.com/SEMICeu

See also https://github.com/open-semantic-interchange

What processes, entities, and agents were involved in curation of mappings?

Open Curation advantages:

- Transparency and quality (many eyes)
- "Human in the loop" provenance
- lowers cost of systemic Cyber Resiliency

Persistent Identifiers

http://w3id.org/semapv
http://w3id.org/semapv/vocab/semapv.owl
And for every vocabulary

Semantic Mapping Vocabulary

Metadata

IRI

http://w3id.org/semapv/vocab/semapv.owl

Title

Semantic Mapping Vocabulary

Creator

https://orcid.org/0000-0002-7356-1779

Contributor

https://orcid.org/0000-0002-6095-8718
https://orcid.org/0000-0003-4423-4370

License

https://creativecommons.org/publicdomain/zero/1.0/

Version Iri

http://w3id.org/semapv/vocab/releases/2026-06-10/semapv.owl

Description

The Semantic Mapping Vocabulary provides and defines terms used for creating and maintaining semantic mappings, in particular mapping metadata.

Classes

LLM-based matching process

IRI https://w3id.org/semapv/vocab/LLMBasedMatching

Description

Where an LLM is used only to produce embeddings that are then compared by similarity, the process is an embedding-based (transformer-based) matching process rather than an LLM-based one. The classifier is how the model is used, not which model. Multi-step, tool-using, autonomous workflows driven by an LLM are captured by semapv:AgentBasedMatching.

A machine learning-based matching process in which a large language model determines matches between entities, typically through single-step or few-shot prompting, or through fine-tuning the model to judge mapping candidates or generate mappings directly.

Example

OLaLa prompts a large language model with the labels and surrounding context of two candidate entities and uses the model's yes/no decision as the match outcome.

Source https://doi.org/10.1145/3587259.3627571

Sub Class Of machine learning-based matching process

AI Governance Mappings - Semantic Knowledge Organization System

related_mappings
broad_mappings
close_mappings
exact_mappings
narrow_mappings

Persistent Identifier

https://www.w3.org/TR/skos-reference

W3C Recommendation

SKOS Simple Knowledge Organization System Reference

W3C Recommendation 18 August 2009

This version:
http://www.w3.org/TR/2009/REC-skos-reference-20090818/
Latest version:
http://www.w3.org/TR/skos-reference
Previous versions:
http://www.w3.org/TR/2009/PR-skos-reference-20090615/
Editors:
Alistair Miles, STFC Rutherford Appleton Laboratory / University of Oxford
Sean Bechhofer, University of Manchester

Please refer to the **errata** for this document, which may include some normative corrections.

See also **translations**.

Abstract

This document defines the Simple Knowledge Organization System (SKOS), a common data model for sharing and linking knowledge organization systems via the Web.

Many knowledge organization systems, such as thesauri, taxonomies, classification schemes and subject heading systems, share a similar structure, and are used in similar applications. SKOS captures much of this similarity and makes it explicit, to enable data and technology sharing across diverse applications.

The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organization systems. It may be used on its own, or in combination

Example Semantic Mappings: finos_to_nist_ai_600_1.sssom.tsv

```
# curie_map:
#   finos: https://air-governance-framework.finos.org/risks/
#   nist_ai_600_1: https://w3id.org/lmodel/nist-ai-600-1/
#   semapv: https://w3id.org/semapv/vocab/
#   skos: http://www.w3.org/2004/02/skos/core#
# license: https://www.apache.org/licenses/LICENSE-2.0.html
# mapping_date: '2026-06-04'
# mapping_set_description: FINOS AI Governance Framework risks (ri-N) to NIST AI 600-1 GenAI Profile risks.
# mapping_set_id: https://github.com/finos/ai-governance-framework/tree/main/linkml/src/ai_governance_framework/mappings/finos_to_nist_ai_600_1.sssom.tsv
# mapping_set_version: 0.1.0
```

subject_id	subject_label	predicate_id	object_id	object_label	mapping_justification	author_id	mapping_date	comment
finos:AIR-RC-001	Information Leaked To Hosted Model	skos:relatedMatch	nist_ai_600_1:2-4	2.4. Data Privacy	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-RC-001	Information Leaked To Hosted Model	skos:relatedMatch	nist_ai_600_1:2-9	2.9. Information Security	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-SEC-002	Information Leaked to Vector Store	skos:relatedMatch	nist_ai_600_1:2-4	2.4. Data Privacy	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-SEC-002	Information Leaked to Vector Store	skos:relatedMatch	nist_ai_600_1:2-9	2.9. Information Security	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-OP-004	Hallucination and Inaccurate Outputs	skos:relatedMatch	nist_ai_600_1:2-2	2.2. Confabulation	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-OP-004	Hallucination and Inaccurate Outputs	skos:relatedMatch	nist_ai_600_1:2-8	2.8. Information Integrity	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-OP-005	Foundation Model Versioning	skos:relatedMatch	nist_ai_600_1:2-8	2.8. Information Integrity	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-SEC-009	Data Poisoning	skos:relatedMatch	nist_ai_600_1:2-8	2.8. Information Integrity	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-SEC-009	Data Poisoning	skos:relatedMatch	nist_ai_600_1:2-12	2.12. Value Chain and Component Integration	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-OP-016	Bias and Discrimination	skos:relatedMatch	nist_ai_600_1:2-6	2.6. Harmful Bias and Homogenization	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-OP-019	Data Quality and Drift	skos:relatedMatch	nist_ai_600_1:2-8	2.8. Information Integrity	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-RC-022	Regulatory Compliance and Oversight	skos:relatedMatch	nist_ai_600_1:2-9	2.9. Information Security	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:AIR-RC-023	Intellectual Property (IP) and Copyright	skos:relatedMatch	nist_ai_600_1:2-10	2.10. Intellectual Property	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:UC-002	Autonomous Wealth Management	skos:relatedMatch	nist_ai_600_1:2-2	2.2. Confabulation	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:UC-002	Autonomous Wealth Management	skos:relatedMatch	nist_ai_600_1:2-4	2.4. Data Privacy	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	
finos:UC-002	Autonomous Wealth Management	skos:relatedMatch	nist_ai_600_1:2-8	2.8. Information Integrity	semapv:LLMBasedMatching	https://github.com/finos/ai-governance-framework	2026-06-04	

- Maximum accessible preferred (i.e. TSV) but other formats possible
- Mapping justification uses SEMAPV terms (i.e. LLMBasedMatching = awaiting Human Curation)
- The Mapping predicate is SKOS relatedMatch (pending Human Curation)
- **LLM infer meaning from semantics in this file -> burns less Tokens**

https://github.com/lmodel/ai-governance-framework/blob/main/linkml/src/ai_governance_framework/mappings/finos_to_nist_ai_600_1.sssom.tsv

Persistent Identifier Standards

Standardized Identifiers for common data exchange

Persistent Identifier (PID) format

https://w3id.org/**finos**
https://w3id.org/**linuxfoundation**

Persistent prefixes examples

https://w3id.org/**linuxfoundation**/id
https://w3id.org/**linuxfoundation**/data/id
https://w3id.org/**linuxfoundation**/data/mappings/spdx-to-eu-act.sssom.tsv
https://w3id.org/**lmodel**/iso27001/schema
https://w3id.org/**lmodel**/common-domain-model/schema/cdm_observable_event
https://w3id.org/**lmodel**/fluxnova-bpm-platform/schema/provenance/fluxnova_bpm_ai_provenance

Persistent identifier examples

https://training.galaxyproject.org/training-material/topics/fair/tutorials/fair-persistent-identifiers/tutorial.html

EU - D7.1.3 - Study on persistent URIs

Further reading

Recommended reading

Supervising the Unsleeping: Overseeing Agentic AI When it Cannot be Fully Human

Thursday, 4 June | 09:30 - 10:30 GMT | Liffey Hall 2, Level 1 | Beginner level

BREAKOUT SESSION · AI GOVERNANCE · AI LITERACY · AI AND MACHINE LEARNING · BENCHMARKING · CUSTOMER TRUST AND EXPECTATIONS · FRAMEWORKS AND STANDARDS · RISK MANAGEMENT · FINANCE AND BANKING · LEGAL · TECHNOLOGY

Download the presentation slides

The EU AI Act mandates human oversight for high-risk AI systems. But oversight did not start with GenAI. Financial services had a head start with mature model risk management and validation culture due to the regulatory scrutiny and risk it faces. Here, oversight grew up in traditional data science and MRM: with dashboards and "management by exception" alerts routing anomalies to humans for review. This paradigm has been strained by GenAI, where open-ended workflows with probabilistically generated unstructured data breaks the tidy oversight metrics. Manual human review has hit the walls of scale, speed and consistency. Now, agentic AI exacerbates this further: agents plan, act and iterate at machine pace. This makes human "watch-every-output" oversight impossible, and we lose the focus point of human accountability. We unpack the core paradox: agentic aims to automate, but its autonomy creates failure surfaces that demand detailed governance, risking the benefits it promises. We lay out current best practices for governing agents: defense-in-depth controls, human-in-command checkpoints and automated monitors that prove control at scale.

What you will learn:

- A concise history of human oversight: from quantitative, thresholder "management by exception" to agentic AI with oversight as part of the system.
- Banking's regulatory driven enhancements: mature MRM/validation, immutable audit trails and AI-assisted evaluation.
- Concrete governance patterns for oversight of agentic AI.

Featured in this session

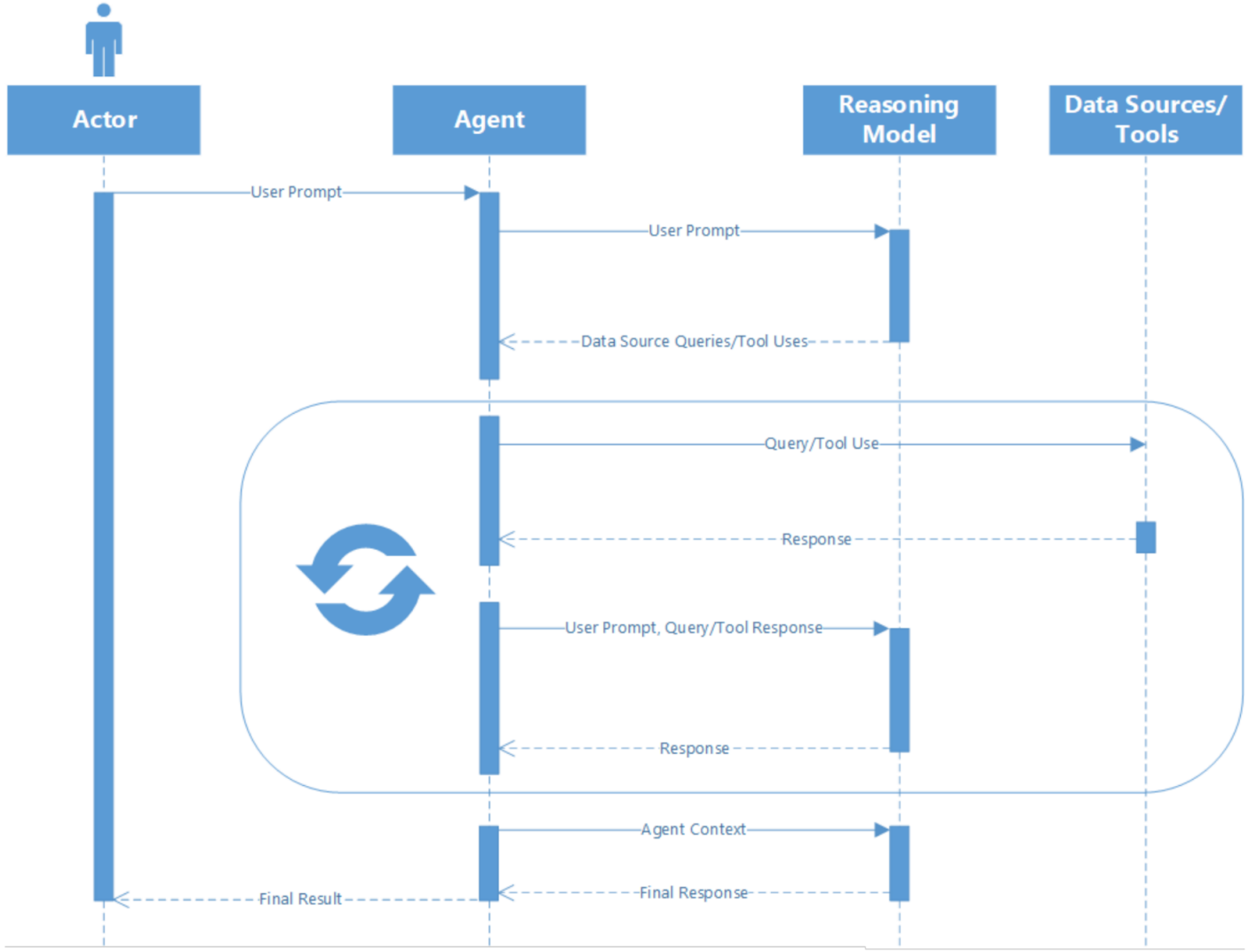
Andrew Harrison
AIGP, CIPP/E
Senior AI Governance Specialist
Booking.com
View bio

Andrew Sutton
Visiting Fellow
Oxford Martin School AI Governance Initiative
View bio

Nicole Onuta
Lead AI Risk
ING
View bio

https://iapp.org/conference/iapp-aigg-europe/agenda/aigg26-supervising-the-unsleeping-overseing-agentic-ai-when-it-cannot-be-fully-human

AI Agent Identity and Authorization – NIST Standard coming …



https://www.nccoe.nist.gov/sites/default/files/2026-02/accelerating-the-adoption-of-software-and-ai-agent-identity-and-authorization-concept-paper.pdf