



Pharma General Ontology (PGO)-Terminology Building a Semantic Backbone for FAIR and AI- Ready Knowledge Graphs

Giovanni Nisato

Pistoia Alliance PGO-Terminology Project Manager
BioIT World 2026 | Knowledge Graphs session | Boston, May 2026



The Pistoia Alliance

A member-led pre-competitive organisation

- Global not-for-profit organisation, incorporated 2008.
- **Mission:** lower barriers to innovation through pre-competitive collaboration.
- Member-led enabling pharma & life science companies to work together without additional legal agreements.

By the numbers

- 200+ member companies
- 30+ active projects
- 1,000+ participants

Some of Our Members



Agenda

- The challenge: a pharma semantic interoperability gap
- Introducing PGO-Terminology: purpose, scope, process
- Where we stand: Phase 1 & 2
- PGO-T as backbone for AI-ready knowledge graphs
- Industry voices [video — 3 min]
- Getting involved

Thanks to the funders of the project:



PGO-terminology steering group

Thanks to:

- Ben Gardner, Philippe Rocca-Serra / AstraZeneca
- Firas Al Laham, Elena Businaro / Chiesi Group
- Peter McQuilton / GSK
- Markus Hartmann / Merck KgaA
- Saritha Kuriakose, Darko Hric, Preethi Elangovan/Novo Nordisk
- Martin Romacker, Nick Perry/ Roche
- Birgit Meldal / Pfizer
- Christine Memmott / Vertex Pharmaceuticals



The Challenge: A New Tower of Babel?

The pharma ecosystem generates vast amounts of data — across R&D, manufacturing, clinical, regulatory, and commercial domains. Yet information exchange remains slow, costly, and error-prone.

Three structural problems:

- **Similar concepts, different words** — 'compound', 'molecule', 'substance', 'drug' used inconsistently across domains and organizations
- **Deep but disconnected domain ontologies** — each domain builds its own, creating semantic silos
- **No cross-pharma lingua franca** — no shared, community-governed reference set of core concepts

Result: teams spend enormous time on semantic reconciliation before any real work begins. Ontologies proliferate. Interoperability remains aspirational.

“With limited semantic interoperability we are creating semantic silos, slowing the path towards data centricity while increasing costs.” — from the PGO-T Charter



[https://en.wikipedia.org/wiki/Tower_of_Babel#/media/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_\(Vienna\)_-_Google_Art_Project_-_edited.jpg](https://en.wikipedia.org/wiki/Tower_of_Babel#/media/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg)

Ontology-space activities at Pistoia Alliance

– Pharma General Ontology-Terminology (PGO-T)

Shared terminology & semantic reference for cross-organisation interoperability

– IDMP Ontology

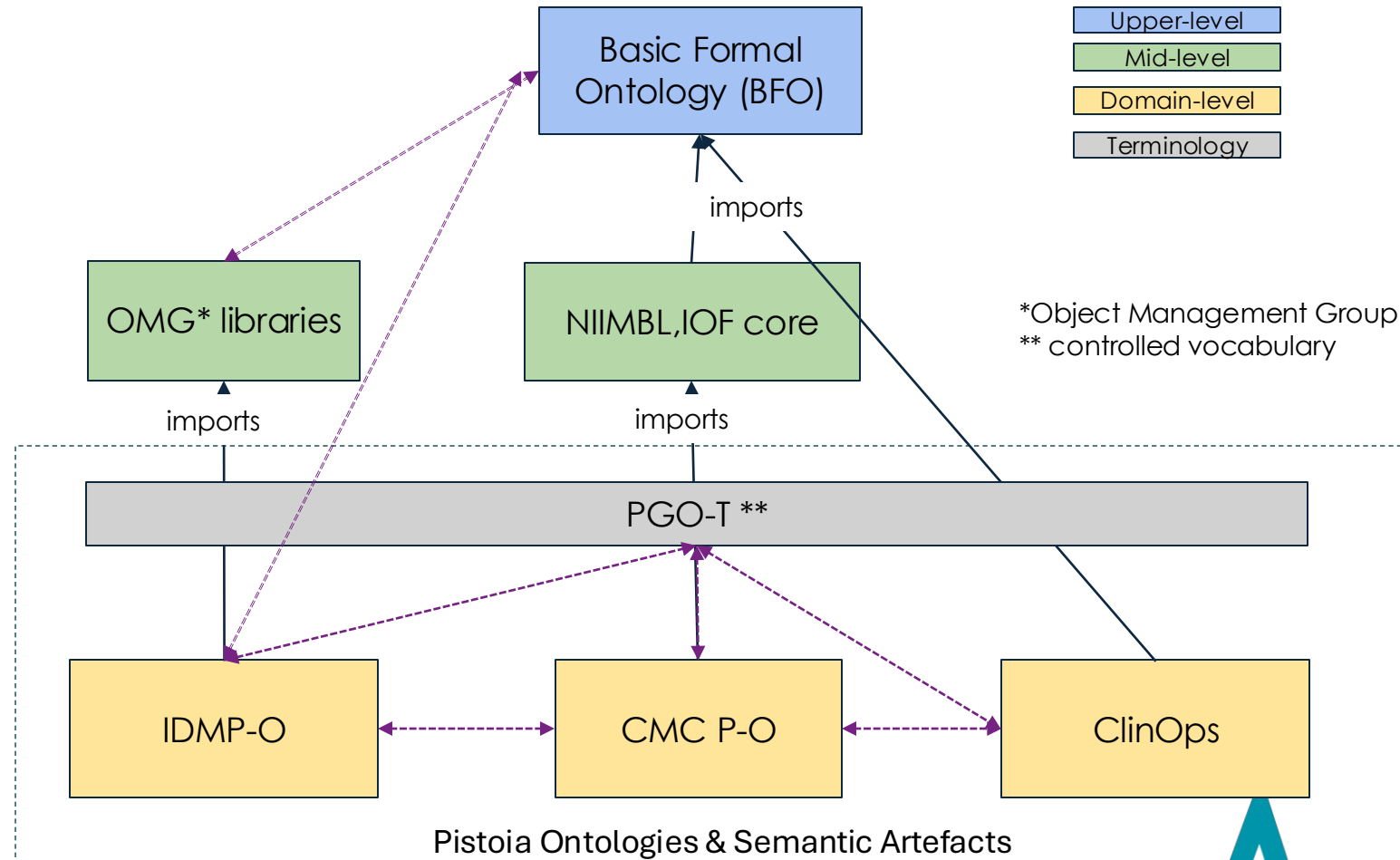
Regulatory identification of medicinal products

– CMC Process Ontology

Chemistry, Manufacturing & Controls process semantics

– ClinOps Ontology

Clinical operations shared vocabulary



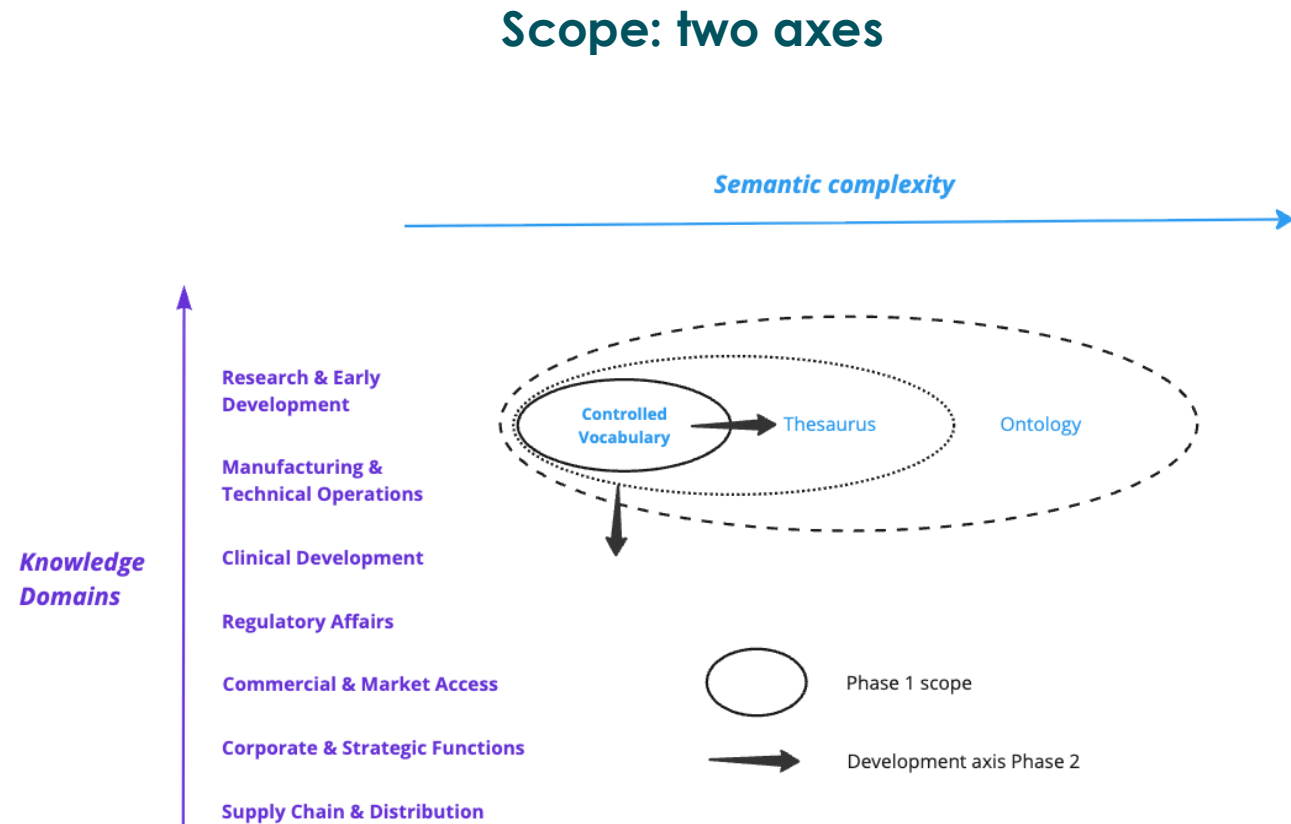
Pharma General Ontology (PGO)-Terminology

What PGO-T is

- A community-governed reference set of core pharmaceutical concepts
- A “lingua franca” — declares clearly which terms are used and why they were chosen
- Open access : public GitHub, CC by 4.0 / MIT

What PGO-T is not

- Not a replacement for MedDRA, SNOMED, CHEBI...
- Not a top-down mandate: community alignment via Steering Group
- Not domain-specific: bridges all pharma domains
- Not an ontology at this point in time



Design Principles & Community Governance

Governance model

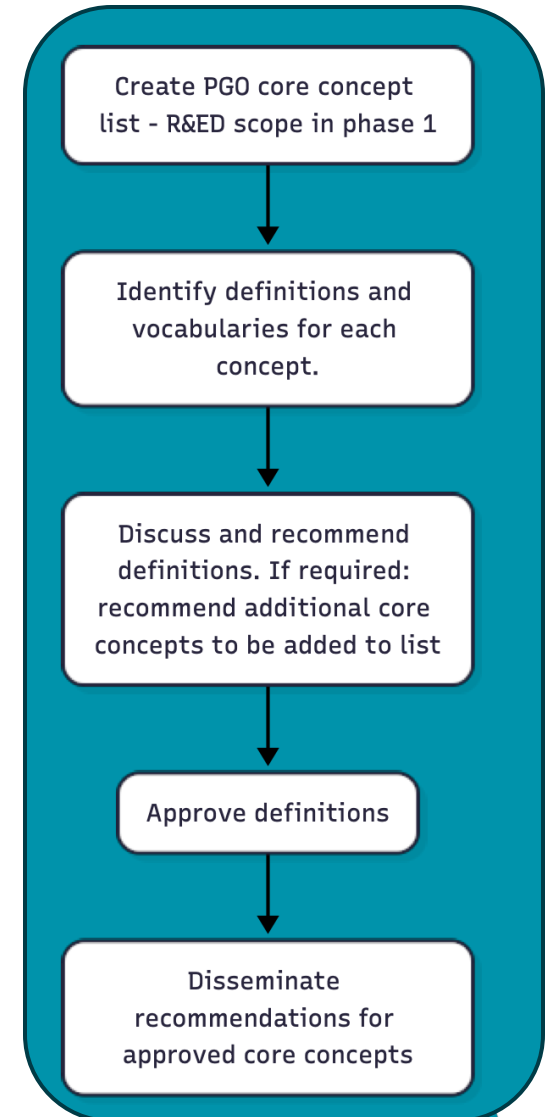
- **Steering Group (sponsors)** — pharma company members who fund and direct the project; decide scope, approve expert proposals
- **Expert Group** — ontologists, curators, and domain experts who **chose existing definitions**
- **Project Manager** — alignment, documentation, and delivery

Definition rules — every concept must:

- Not re-invent the wheel: ref. existing definitions from known sources
- Be human-readable and understandable across domains
- Have a publicly accessible, resolvable URL/IRI
- Phase 2+: Be machine-readable (e.g. OWL)

Process

- Community-driven: no single company decides
- Iterative alignment cycles: expert review → Steering Group acceptance → GitHub publication



Initial modeling with LinkML

The Pistoia Alliance Pharma General Ontology: Experience Using LinkML in Pharma.

Joshua Valdez^{1†}, Philippe Rocca-Serra^{2,8†*}, Markus Hartmann³, Darko Hric¹, Asiyah Yu Lin⁴, Birgit Meldal⁵, Peter McQuilton⁶, Riccardo Mariani⁷, Martin Romacker⁸, Giovanni Nisato⁴ and Ben Gardner².

¹ Novo Nordisk, Copenhagen, Denmark

² AstraZeneca, 1 Francis Crick Avenue Cambridge Biomedical Campus Cambridge CB2 0AA UK.

³ Merck Group, Germany.

⁴ Pistoia Alliance, US.

⁵ Pfizer, Cambridge, UK.

⁶ Glaxo Smith Kline PLC, Stevenage, UK.

⁷ Chiesi Farmaceutici, 26/A, Via Palermo, 43122 Parma – Italy

⁸ Hoffman la Roche, Basel, CH.

⁹ Oxford e-Research Centre, University of Oxford, OX1 3QG, Oxford, UK.



https://ceur-ws.org/Vol-4196/paper_3.pdf

<https://github.com/Pistoia-Alliance-Inc/Pistoia-Alliance-PGO/>



Where We Stand: Phase 1 Delivered, Phase 2 Underway

Phase 1 (2024-2025) — Completed

- 17 core concepts at Controlled Vocabulary level
- Focus: Research & Early Development domain
- Publicly available on GitHub and Pistoia Alliance website

Selected Phase 1 concepts

- Gene | Protein | Cell /Species
- Assay | Biomarker | Biospecimen
- Substance | Compound | Drug
- Disease | Indication | Clinical_study

Find us:

<https://github.com/Pistoia-Alliance-Inc/Pistoia-Alliance-PGO/>

Phase 2 (2025–2026) — In progress

Expanding beyond R&D:

Additional concepts include : Adverse Event, Excipient, Instrument, Study Site... (> 20 concepts)

Growing horizontally:

- Implementing schemas for machine-readable output
- Generating mappings to external ontologies

Sustainability model:

PGO as permanent industry reference data product



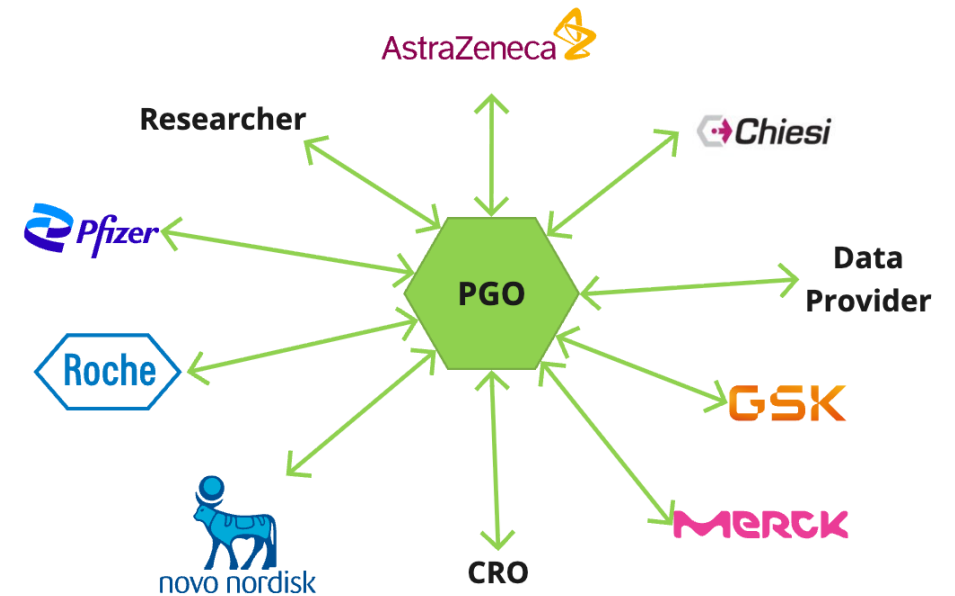
PGO-T: Long-Term Vision (beyond current phase)

What the industry will see (tangible)

- Machine-readable Core-terminology on e.g. EMBL-EBI OLS / BioPortal pointing to public definitions
- Declaration / endorsement: who is implementing PGO-T in internal terminology management systems
- Community resource implementing W3C standards

What the industry will be able to do

- Quicker decisions on recommended reference source
- Tag data assets with PGO URIs — linkable and reusable
- Map core-concepts in datasets to one unified PGO-T — enabling cross-company data exchange
- Exchange data across CROs, partners, and regulators while maintaining internal URI patterns



PGO-T: Semantic Backbone for AI-Ready Knowledge Graphs

Knowledge graph challenge

- Knowledge graphs are only as good as their underlying semantics
- Ambiguous nodes undermine graph traversal, inference, and LLM grounding
- Cross-company KGs require shared concept identifiers without them, graph federation fails

How PGO-T may help enable AI-friendly knowledge graphs

- **URIs**: resolvable identifiers for every node/concept, queryable by machines
- **Entity disambiguation** — “indication” in one company maps unambiguously to “indication” in another
- **Cross-graph federation** — organisations exchange graph fragments that are semantically interoperable
- **RAG & LLM grounding** — “FAIR-annotated” graph data reduces hallucination risk in AI pipelines

Practical example: vision

A CRO delivers a data product tagged with PGO-T URIs.

The receiving pharma company's knowledge graph ingests it directly — no manual mapping, no reconciliation meetings etc.



Funders' Voices: Why PGO-T Matters for AI



**AI requires a
common data
language**

Getting Involved

PGO-Terminology is a community project — it grows with your participation

Use it — open access

- **Browse concepts:** github.com/Pistoia-Alliance-Inc/Pistoia-Alliance-PGO
- **Visit:** pistoiaalliance.org/project/pharma-general-ontology
- **Cite PGO –T** in your data systems, products and knowledge graphs
- **Let us know** when you use PGO-T and how we can improve

Contribute — Pistoia Alliance members

- **User** — provide feedback on adoption, usage, and improvements
- **Expert** — ontologists & curators: provide input on definitions to the Steering Group
- **Sponsor** — fund and direct Phase 3 and beyond

Contact

Giovanni Nisato — Project Manager, Pistoia Alliance
giovanni.nisato@pistoiaalliance.org



Thank you!

Contact: giovanni.nisato@pistoiaalliance.org

