A Multi-Level Architecture for Reusable Materials Ontologies — The OntoCrafter Ceramics Ontology (OCO) as Reference Implementation

Thomas Pannek[a], Wolfgang Grond[a,*]

[a]*Numberland — Dr. Wolfgang Grond, Lohfeld 20, 95326, Kulmbach, Germany*

Abstract

The Materials Science and Engineering ontology landscape is fragmented along multiple axes simultaneously. Horizontally: a recent survey identified 94 ontologies of which over 40 are structurally incompatible; each new application domain — ceramics, polymers, batteries, smart materials — typically restarts ontology design from scratch. Vertically: EU regulation (CSRD, CSDDD, PPWR, CBAM, R2R, AI Act, ESPR) forces material, manufacturing, supply-chain, and lifecycle data into integrated digital product passports, leaving ontologies that only address horizontal fragmentation incomplete for any contemporary consumer. And mechanistically: a vocabulary that records that BNT-BT has $d_{33} \approx 580\,\mathrm{pC/N}$ stores a fact but cannot surface *why* — Bi-6s^2 lone-pair stereo-activity, anomalous Born effective charges, soft modes, defect chemistry — without a systematic explanation skeleton. We propose a multi-level modular architecture with two independent classification axes — level of abstraction (**L0** bridges, **L1** material-agnostic laboratory-notebook, **L2** material-class-specific, **L3** categorical reasoning) and consumer audience (material vs. compliance) — in which the material-specific level is internally organised by a seven-tier mechanistic-explanation skeleton (Symmetry, Energy/DFT, Thermo/CALPHAD, Kinetics, Microstructure, Defect chemistry, Bonding) applicable to any crystalline ionic oxide. The level-and-audience modularity dissolves the horizontal fragmentation, the compliance audience absorbs the vertical regulation pressure, and the seven-tier organisation of Level 2 delivers the mechanistic explanation depth. We instantiate the architecture as the OntoCrafter Ceramics Ontology (OCO v0.94): 5 196 classes

---

[*] Corresponding author.
Email address: `grond@numberland.com` (Wolfgang Grond)

across 44 modules; 167 348 OWL axioms (40 454 logical) including 5 920 reified Neumann tensor constraints; 1 674 properties; 829 cross-ontology bridge mappings across 40 sections; 1 172 SHACL shapes; 163 published competency questions with 52 executable SPARQL tests (52/52 pass). Functional ceramics are the reference material domain with BNT-BT as the end-to-end pilot demonstrated across all seven explanation tiers; ferritic high-performance ceramics are in active development as the second material instance. v0.94 is the release engineered to enter productive practice; v1.0 is reserved for the state after that practice has fed back corrections.

Keywords: materials ontology, multi-level architecture, functional ceramics, LIMS/ELN integration, competency questions, BFO, PMDco, interoperability, reasoning

1. Introduction

The design of an industrial materials ontology faces three simultaneous challenges. None is new in isolation; what is new is that an ontology entering productive use today must answer all three at once.

The first challenge is the *horizontal fragmentation of the materials-science ontology landscape*. Norouzi et al. (2024) identify 94 ontologies in the field, of which over 40 are structurally incompatible (Rajamohan et al., 2025). Within the German Platform MaterialDigital (PMD) consortium[1], more than a dozen projects — KnowNow (Ben Hassine and Stark, 2024), Mieller-Ferrit (Mieller et al., 2024b), SmaDi (Maas et al., 2024), KupferDigital, GlasDigital, StahlDigital, DiProMag, iBain (Bekemeier et al., 2025b) — each developed its

---

[1] The PMD Phase-1 publication portfolio sets the practical context for this work: (Bardehle et al., 2025; Bekemeier et al., 2025a; Raabe et al., 2022; Gramlich et al., 2024; Chakraborty et al., 2024; Tomkovic et al., 2026; Schilling et al., 2026; BeygiNasrabadi et al., 2025; Leipner et al., 2025; Gumbsch et al., 2025; Bayerlein et al., 2024b; Bekemeier et al., 2025b; BeygiNasrabadi et al., 2024d; Schilling et al., 2024a; Shoghi et al., 2024; Schilling et al., 2024b; Roters et al., 2024; Fliegener et al., 2024; Nerella et al., 2024; Eisenbart et al., 2024; Sajjad et al., 2024; BeygiNasrabadi et al., 2024c; Nahshon et al., 2024; Bjarsch et al., 2024; Mieller et al., 2024b,a; Chen et al., 2024a; Luger et al., 2024; Aschemann et al., 2024; Nebel et al., 2024; Beran et al., 2024; Maas et al., 2024; Özçep et al., 2024; Bayerlein et al., 2021; Gogula et al., 2024; Pan et al., 2023; Arendt et al., 2024; Bayerlein et al., 2024c; BeygiNasrabadi et al., 2024a,b; Bayerlein et al., 2024a; BeygiNasrabadi et al., 2024e; Bayerlein et al., 2024d; Meng et al., 2024; Agrawal et al., 2024; Unger et al., 2023; Blum et al., 2023; BeygiNasrabadi et al., 2023; Schmidtke et al., 2023; Valdestilhas et al., 2023; Mertens et al., 2024; Chen et al., 2024b; Mutz et al., 2022; Garabedian et al., 2024; Rego et al., 2022; Lizarazu et al., 2024; Rogge et al., 2024).

own material-specific ontology, repeatedly re-modelling laboratory workflows, equipment, and measurement methods that are essentially identical across all of them. Only a fraction of the modelled knowledge is material-specific (the actual class hierarchy of ceramics, polymers, batteries, ...); the largest part — workflow provenance, equipment, methods, identifier schemes — is material-agnostic and reusable.

The second challenge is the *vertical convergence pressure driven by EU regulation*. The Corporate Sustainability Reporting Directive (CSRD/ESRS), the Corporate Sustainability Due Diligence Directive (CSDDD), the Packaging and Packaging Waste Regulation (PPWR), the Carbon Border Adjustment Mechanism (CBAM), the Right-to-Repair Directive (R2R), the EU AI Act, the Construction Products Regulation (CPR), and the Ecodesign for Sustainable Products Regulation (ESPR) all require digital product passports (DPPs) that combine material with manufacturing and value-chain information. A materials ontology that addresses only horizontal fragmentation (per material class) without this vertical extension is incomplete for any consumer operating under contemporary EU regulation. The convergence is visible in the structure of the PMD programme itself: PMD Phase-3 explicitly incorporates the *Wertschöpfungskette* (value chain) as a programme requirement, anticipating integration with Manufacturing-X.

The third challenge is *mechanistic explanation depth*. A materials ontology that records that BNT-BT has $d_{33} \approx 580\,\mathrm{pC/N}$ at its morphotropic phase boundary stores a fact. An ontology that can also surface *why* — Bi-6s^2 lone-pair stereo-activity, anomalous Born effective charges, soft-mode at the Brillouin-zone centre, R3c↔P4mm domain coexistence, $V_O^{\bullet\bullet}$ defect chemistry, Hall-Petch grain-boundary mediation, Mn-acceptor pinning of domain-wall mobility — provides authoritativeness that a description-only vocabulary cannot. The challenge is to attach mechanistic explanation classes systematically without inflating the vocabulary into a textbook.

We answer these three challenges with a single integrated design: a *multi-level, modular ontology with two independent classification axes plus a seven-tier mechanistic-explanation skeleton internal to the material-specific level*. The first axis is the **level of abstraction** (**L0** bridges, **L1** material-agnostic laboratory-notebook, **L2** material-class-specific, **L3** categorical reasoning). The second is the **consumer audience** (material versus compliance). The material-specific level (**L2**) is internally organised by a **seven-tier mechanistic-explanation skeleton** applicable to any crystalline ionic oxide: Symmetry, Energy/DFT, Thermo/CALPHAD, Kinetics, Microstructure, Defect chemistry, Bonding. The two axes are independent — a module is placed on each at the level appropriate to its content, and consumers select which subsets they need — while the seven-tier skeleton structures the depth of mechanistic content within Level 2. The level-and-audience modularity dissolves the horizontal fragmentation, the compliance audience absorbs the vertical EU-regulation pressure, and the seven-tier organisation of the material-specific level delivers the mechanistic explanation depth — three challenges, three architecturally distinct answers, without collapsing them onto a single axis where they would compete.

We instantiate this design as the OntoCrafter Ceramics Ontology (OCO v0.94), the release engineered to enter productive use. OCO ships as a class-bearing module set with bridge mappings, SHACL shapes, a competency-question catalogue with executable SPARQL tests, and architecture decision records; the full metric inventory and method-of-measurement is in Section 4. Functional ceramics are the reference material domain, with BNT-BT as the end-to-end pilot demonstrated across all seven explanation tiers. v0.94 enters productive practice; corrections surfaced by that practice are the planned input for v1.0.

This paper contributes:

1. A formal description of the four-level modular architecture, driven by ten engineering design principles (modularity, adaptability, interoperability, purpose, equality, compatibility, functionality, authoritativeness, facetedness, portability), of which five go beyond the Norouzi REQ-canon (Section 2.4, Section 3). The same modular pattern addresses horizontal material-domain fragmentation and vertical EU-regulatory convergence with one structural primitive.
2. The seven-tier mechanistic explanation skeleton formalised as the internal organising principle of Level 2 (material-specific knowledge), generic for crystalline ionic oxides, with cross-tier SHACL constraints that make the explanation chain auditable (Section 3.4.2).
3. A reference implementation as the OntoCrafter Ceramics Ontology (OCO v0.94), realising the architecture across class-bearing material modules; a Neumann tensor engine that generates reified symmetry constraints algorithmically over the 32 crystallographic point groups; SPARQL-based phase-state coupling that derives the active point group from sample temperature and composition; a coordination-polyhedron module (`oco-localstructure`) that closes the mechanistic chain from composition through local geometry to tensor symmetry; multi-axis parameter classification along role, reference, and material-abstraction layer; an external-cache pattern for high-volume reference data (Shannon ionic radii, IUCr bond-valence parameters, pyxtal Wyckoff positions,

Materials Project DFT corpus, Pauling electronegativities); and a disciplined reuse-before-invention bridge policy (Section 4).

4. A vertical extension of the same architecture into the compliance and value-chain stack: modules for Life Cycle Assessment (with ecoinvent and EN 15804+A2 bridges), CSRD/ESRS reporting, supply-chain due diligence (CSDDD), packaging (PPWR), carbon-border adjustment (CBAM), right-to-repair (R2R), the EU AI Act, ODRL/trust, Manufacturing-X identifier and traceability infrastructure (AAS IEC 63278, Catena-X CX-0010/CX-0146), and laboratory automation (SiLA). The same modular level discipline absorbs these without restructuring the material core (Section 4.8).
5. Validation against the Norouzi quality-requirement canon (REQ1–REQ9, all nine met), the OOPS! pitfall audit (Poveda-Villalón et al., 2014), a published catalogue of area-tagged competency questions with executable SPARQL tests, and a SHACL validator suite (counts in Section 4; reasoner/audit results in Section 5).
6. Discussion of limitations and the v1.0 roadmap, framed by the corrections expected from productive practice — including a second ceramic material system (ferritic high-performance ceramics) currently in active development (Section 6).

The remainder of the paper is organised as follows. Section 2 surveys the MSE ontology landscape, positions the multi-level architecture against three established schools of design (BFO-aligned mid-level, EMMO-centric, and bottom-up domain-specific), and lays out the ten design principles that drive OCO (Section 2.4). Section 3 presents the four-level architecture together with the material/compliance audience axis as the second classification dimension, plus the seven-tier mechanistic-explanation skeleton as the internal organisation of Level 2. Section 4 describes OCO as reference implementation — the material modules (including the Neumann engine, phase-state coupling, multi-axis parameter classification, the coordination-polyhedron module, and the seven-tier explanation skeleton with its external-cache pattern), the bridge inventory, and the vertical extension into the compliance and value-chain stack. Section 5 reports the validation results — all nine Norouzi requirements met. Section 6 discusses strengths, limitations, and the v1.0 horizon shaped by productive practice. Section 7 summarises and outlines the roadmap.

2. Background and Related Work

2.1. *The State of MSE Ontologies*

Norouzi et al. (2024) provide the most comprehensive recent survey of

MSE ontologies, evaluating 94 semantic artifacts (4 top-level, 8 mid-level, 60 domain-level, 2 application-level, plus 20 not openly evaluable) against nine quality requirements (REQ1–9) and an OOPS! pitfall audit. Their key findings frame the work presented here:

- Only 9 of 94 ontologies (10 %) publish competency questions.
- No ontology adopts user stories or personas as design anchors.
- Critical OOPS! pitfalls (P19 multiple-domain/range, P40 namespace-hijacking, P31 incorrect `equivalentClass`, P11 missing domain/range) are widespread, even in established ontologies such as QUDT (217 P11 hits).
- BFO is the most-reused top-level (16×); EMMO follows (12×); PMDco is reused only 1× directly despite being deployed in 13 BMBF projects.

2.2. Three Architectural Schools

Schule A — BFO-aligned Mid-Level. PMDco (Bayerlein et al., 2024d), MWO (Beygi Nasrabadi et al., 2025), MSEO, NFDIcore. Strict ISO/IEC 21838-2 conformance, mid-level material-science vocabulary above BFO, formal reasoner-friendly. Cost: high BFO learning curve, "portion-of" constructs awkward for non-philosophers.

Schule B — EMMO-centric. BattINFO, CHAMEO (Del Nostro et al., 2022), GPO, EMMO Crystallography, `domain-electrochemistry`. Adopted in EU projects (EMMC, BIG-MAP, OntoTrans). Cost: each subdomain imports the full EMMO top-level; many subdomains are at v0.x maturity.

Schule C — Bottom-up domain-specific. KnowNow (Ben Hassine and Stark, 2024), MaterialDigital project ontologies, NanoMine, MatKG. Pragmatic and fast, but inter-project interoperability weak.

2.3. Specific Gaps in the Current Landscape

Beyond the methodological pitfalls flagged by Norouzi et al. (2024), two content gaps emerge from a systematic reading of the recent literature:

No deep functional-ceramics modeling. KnowNow is restricted to LTCC multi-layer components; Mieller et al. cover only NiCuZn ferrite (Mieller et al., 2024b); BattINFO addresses batteries; SmaDi (Maas et al., 2024) models four smart-material classes but at the level of constitutive parameters, not microstructure or defect chemistry. The combination of 230 space groups, Kröger-Vink defect notation (Kröger and Vink, 1956), Newnham composite connectivity (Newnham et al., 1978), and a full coupled-effects family (piezoelectric, pyroelectric, magnetostrictive, . . .) is absent from all surveyed ontologies.

Integrity constraints not delivered alongside the TBox. OWL's open-world semantics interprets missing triples as unknown rather than as violations — a deliberate and correct design choice for an open-world reasoning language, but one that leaves enforceable integrity constraints (mandatory fields, value ranges, cardinality caps) outside what TBox axioms can express. The standard remedy is SHACL, which adds closed-world shape validation as an independent artifact type alongside the TBox; Özçep et al. (2024) demonstrate the pattern with their SHACL-OBDA validator. The gap in the current MSE ontology landscape is not the Open-World Assumption itself — it is the correct semantics for what OWL is for — but that most surveyed MSE ontologies ship a TBox without accompanying SHACL shapes, leaving consumers without a standard way to validate ABox compliance.

2.4. Positioning of OCO

OCO's architecture is governed by ten design principles drawn from engineering practice (platform strategy, construction catalogues), quality management (SIPOC), and the materials-science domain itself. These principles, not the Norouzi requirement canon, are the *primary* design drivers; the canon is a verification surface that the principles must clear, not a brief from which they derive. Two emphases run through the principle set: a *lessons-learned* dimension that distills what worked and what did not in the PMD Phase-1 ontology cohort (Section 2), and a *forward-looking* dimension that anticipates the EU-driven convergence of MaterialDigital with Manufacturing-X — the second axis of fragmentation introduced in Section 1. We list the principles here because they explain the architectural choices that follow in Section 3, and because five of them go beyond what REQ1–9 demand.

2.4.1. Ten Design Principles

Figure 1 organises the ten principles below as a causal chain: each principle cluster enables the next phase of what an industrial materials ontology must deliver, from consumer-side modular choice all the way through to sister-domain extension. The detailed list that follows is grouped in the same order.

1. **Modularity.** A materials ontology should follow the same logic as engineering platform strategy: classes and properties are organized into

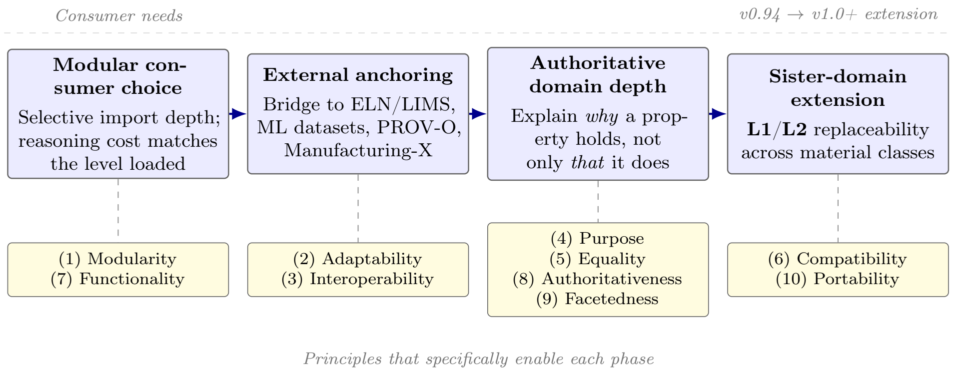

Figure 1: Causal-chain view of OCO's ten design principles. The four phases (top row) trace the consumer-need-to-extension flow that the architecture must answer; each phase is enabled by a specific cluster of principles (bottom row). Numbers refer to the principle list below.

reusable modules that can be combined per consumer, analogous to the construction catalogues used by automotive platforms. A consumer who needs only equipment and method vocabulary imports two modules, not the whole ontology.

2. **Adaptability.** Bridges to external ontologies are isolated in a dedicated level (**L0**), and the imported external sources are versioned and cached locally. When an external ontology releases a new version, only the bridge file changes — no class definitions in OCO break, and the cache makes the build reproducible. The same mechanism extends *forward*: standards that are still consolidating (EU Construction Products Regulation 2024/3110, Critical Raw Materials Act, evolving DPP-XBRL schemas) can be tracked via provisional bridges that mature alongside the upstream release, without forcing rework in the material core.
3. **Interoperability.** An ontology is not a free-standing artifact but a vocabulary embedded in a workflow. For a materials ontology this requires concrete anchoring in the open-source materials stack: ELN/LIMS, machine-learning datasets (Croissant), materials databases (OPTIMADE), and provenance tooling (PROV-O).
4. **Purpose.** Vocabularies without a stated purpose accumulate dead classes. OCO binds its scope to a published catalogue of competency questions, each tagged with an *area* (tensor symmetry, phase state, route, lifecycle, . . .) and accompanied by an executable SPARQL test where possible.
5. **Equality.** Experimental documentation is never complete; any record

is an SIPOC fragment — who was the supplier of the inputs, what was the process, what was the output, who is the customer. Scientific publications are treated identically: a paper is an SIPOC fragment of one or more experiments. The same provenance pattern applies to both, which removes the artificial line between raw experimental data and published results.

6. **Compatibility.** OCO is the architectural successor of two earlier PMD-Phase-1 ontologies (KnowNow for LTCC multi-layer ceramics, SmaDi for smart materials including piezo-ceramic) and the EU-regulatory peer of the Manufacturing-X ecosystem (AAS asset administration shells, Catena-X BPN identifiers, ESPR-driven sector DPPs). Backward compatibility — ingesting Phase-1 ontologies via bridges rather than remodelling — and forward compatibility — anchoring in Manufacturing-X identifier and traceability infrastructure via bridges to the same kind of pattern — are the same requirement under the same kit (Section 4.8).
7. **Functionality.** Beyond reusable modules, OCO is organized into four *functional layers* — **L0** (bridges), **L1** (laboratory-notebook level, material-agnostic, ELN/LIMS-ready), **L2** (material-specific knowledge), **L3** (reasoning axioms) — each addressing a different consumer concern. Maintenance becomes localized: a PMDco update touches only **L0**; a new material adds only **L2**.
8. **Authoritativeness.** For functional ceramics the *full mechanistic-explanation chain* — from local structure and bonding through symmetry, energy landscape, thermodynamic phase regions, kinetics, microstructure, and defect chemistry to the resulting tensor-component symmetry — is load-bearing knowledge. The seven-tier skeleton (Section 3.4.2) is the systematic carrier of that chain. A central traversal through it is *phase diagram* → *active crystallographic phase* → *tensor-component symmetry*: without that traversal an ontology cannot answer "which piezoelectric coefficients are non-zero for BNT-BT at room temperature near the morphotropic phase boundary?" But this is one path of several — defect chemistry to Mn-acceptor pinning of domain-wall mobility, soft modes at the Brillouin-zone centre to incipient ferroelectricity, microstructure to Hall-Petch toughening — and a vocabulary that cannot deliver any of them along its central domain questions is not authoritative. OCO models the entire seven-tier chain explicitly, with the Neumann tensor engine, phase regions, and point-group bindings carrying the symmetry-to-tensor edge as one example implementation. The chain is not ceramics-specific: Neumann's principle and the layered explanation pattern hold for every crystalline material,

so the same mechanism governs metals (martensitic transformations in shape-memory alloys, soft-magnetic alloys across their Curie transition) and crystalline organics (pharmaceutical polymorphs, organic semiconductors). A sister-domain **L2** replacement (Section 5.4) inherits this entire infrastructure unchanged — only the concrete phase-region populations and the domain-specific seven-tier content differ.

9. **Facetedness.** A material parameter is not adequately described by a single classification axis. OCO classifies every parameter along three independent axes simultaneously: **role** (state, response, transport, structure, statistical, topological, fit), **reference** (fundamental, material, interface, defect, measurement, specimen, process, environment), and **layer** (atomic, crystalline, microstructural, mesoscopic, macroscopic). The axes are populated by reasoner-classified `rdfs:subClassOf` relations, so SPARQL queries can navigate any subset of facets without bespoke join logic.
10. **Portability.** The architecture must support replacement of the material-specific level without touching **L0**, **L1**, or **L3**. A metallurgy ontology adopting the same **L0**/**L1**/**L3** need only contribute its own **L2** module to deliver a fully reasoner-ready ontology for its domain.

2.4.2. Where OCO Goes Beyond the Landscape Canon

The Norouzi REQ1–9 canon emphasizes reasoning, modularity, bridges, and documentation — necessary but not sufficient for a working materials ontology. Five of the ten principles above are *not* captured by the canon and reflect requirements that surfaced from real consumer use cases rather than from the survey methodology:

- **Adaptability** is not explicitly in REQ1–9. The canon recognizes the importance of reuse but treats external version drift as a documentation matter. OCO's strategy (bridge-isolation plus external version cache) is an engineering response to the empirical observation that bridge files are the *only* component routinely broken by external releases.

- **Equality** is absent from the canon. The canon implicitly distinguishes experimental data from literature; OCO unifies both as `prov:Activity` instances with the same provenance pattern. This is the architectural prerequisite for the SIPOC-Extractor companion tool and for cross-paper aggregation without forcing premature reconciliation of conflicting reports.

- **Compatibility** with predecessor projects (KnowNow, SmaDi) addresses a fragmentation problem the canon identifies but does not solve: even within a single consortium, every project re-models the same laboratory concepts. OCO's bridge level ingests both predecessors verbatim, allowing their concepts to coexist with OCO's deeper modeling.
- **Authoritativeness** in the functional-ceramics sense (the full seven-tier mechanistic-explanation chain, of which phase → tensor coupling is one central traversal) is a domain requirement that the canon — being domain-neutral — cannot generate. Without it, a ceramics ontology cannot stand behind its own competency questions about coupled effects.
- **Facetedness** generalizes what the canon calls "rich semantics" into a concrete multi-axis classification with reasoner-supported navigation. The closest the canon comes is requiring "more than `rdfs:label`" — a much weaker bar.

The remaining five principles (modularity, interoperability, purpose, functionality, portability) align with REQ1–9 but are realized with engineering rather than compliance intent: modularity is platform strategy, not a list of files; functionality is level separation by consumer concern, not just file count; purpose is bound to executable tests, not free-text statements.

2.4.3. Architectural Consequence

The four-level architecture that the remainder of this paper describes is the synthesis of these ten principles, not a direct answer to REQ1–9. Modularity, functionality, and portability dictate the multi-level structure itself; adaptability dictates the bridge-level isolation; interoperability and compatibility dictate the bridge inventory; purpose dictates the area-tagged competency catalogue; equality dictates the unified `prov:Activity` treatment of experiments and publications; facetedness and authoritativeness dictate the depth of the material-specific level. The Norouzi canon is satisfied by this architecture (Section 5), but as a verification rather than a brief — and the architecture extends to requirements the canon does not raise. The reference implementation for ceramics demonstrates that the pattern is viable; the modular structure invites sister-domain implementations (metallurgy, polymers, batteries) that would share **L0** and **L1** with OCO while replacing **L2**.

3. The Multi-Level Architecture

The architecture introduced in Section 1 comprises two independent classification axes plus a mechanistic-explanation skeleton internal to the material-specific level. The first and primary axis is the **level of abstraction** — **L0** bridges, **L1** material-agnostic laboratory-notebook level, **L2** material-class-specific knowledge, **L3** categorical reasoning — governed by the architectural principles in Section 3.1 and detailed level-by-level in Sections 3.2 to 3.5. The second axis is the **consumer audience** (Section 3.6), which separates the material core from the compliance and value-chain modules without restructuring either. The two axes are independent: every module is placed on each at the level appropriate to its content, and consumers select the subsets they need. Within **L2** (the material side), the content is organised internally by the **seven-tier mechanistic-explanation skeleton** (Section 3.4.2), which lets the ontology surface *why* a property holds, not only that it does. Compliance modules have no analogous internal structure — they are regulatory rather than mechanistic — which is why the seven-tier skeleton is the internal organisation of one specific level rather than a third independent axis.

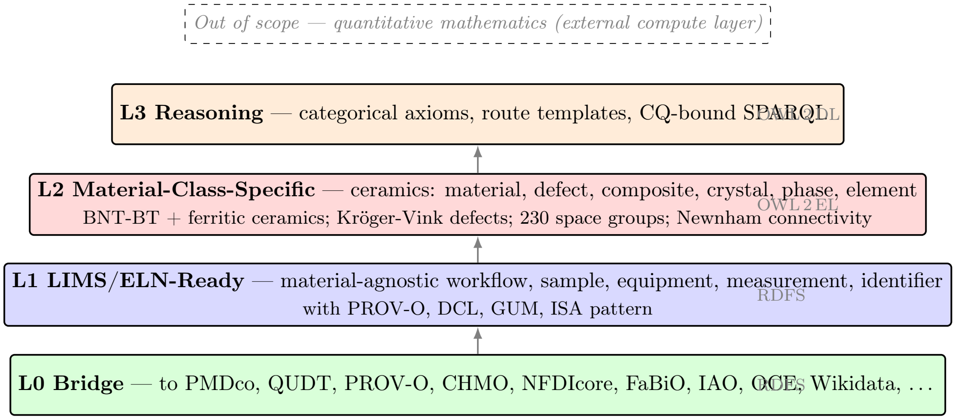

Figure 2: The four-level architecture for reusable materials ontologies. Consumers select their import depth: **L0** alone for bridge-only integration, **L0**+**L1** for material-agnostic LIMS/ELN, **L0**+**L1**+**L2** for full ceramics knowledge, and the opt-in **L0**+**L1**+**L2**+**L3** bundle for OWL-DL reasoning. Quantitative mathematics is explicitly out of scope for this release (Section 6.4).

3.1. Architectural Principles

The architecture rests on five design principles:

1. **Level separation.** Each level has a clearly bounded scope: **L0** bridges to external standards, **L1** carries the material-agnostic vocabulary, **L2** the material-specific vocabulary, **L3** the categorical reasoning. Consumers select their import depth.
2. **Modularity within levels.** Each level decomposes into independently maintainable modules. Consumers can import single modules (e.g., `oco-equipment` alone) without pulling the entire level.
3. **Opt-in reasoning.** Categorical reasoning (**L3**) is not part of the default distribution. Consumers without reasoning needs do not pay the OWL-DL reasoner cost.
4. **Bridge localization.** External version changes (PMDco v3.1, QUDT v3.3) are confined to a single level (**L0**). Internal modules reference only `bridge:*` classes, which remain stable under our control.
5. **Cache pattern for external reference data.** Large externally-curated bodies of reference data — mid-level ontology versions consumed by **L0** bridges, structural and thermodynamic constants consumed by the mechanistic tiers of **L2** — are not embedded in OCO's TBox but cached locally as versioned snapshots, pinned via SHA in `bridge/external_versions.yaml`. This separates reference-data scale from ontology scale: the TBox stays compact, while consumers query large reference corpora (Materials Project DFT entries, IUCr bond-valence parameters, pyxtal Wyckoff positions, Shannon ionic radii, Pauling electronegativities) with reproducible version pinning. Concrete instances appear in Section 3.2 (external ontologies) and Section 3.4.2 (mechanistic-tier reference data).

These five principles separate the architecture concretely from the two dominant schools introduced in Section 2. PMDco offers five distributions (*full*, *base*, *simple*, *minimal*, *main*), but these are vertical slices through the same mixed vocabulary — modular by class count, not by architectural level. The **L1**/**L2** split that isolates "laboratory workflow without material classes" from "material description without workflow classes" along the consumer-concern axis cannot be expressed in PMDco's distribution scheme. EMMO sub-domain ontologies (BattINFO, CHAMEO, GPO, the EMMO crystallography and electrochemistry modules) inherit the full EMMO top-level (700+ classes) and pay the corresponding reasoner cost regardless of which classes a given consumer actually uses; the opt-in reasoning principle realized here via a separate **L3** distribution cannot be retrofitted onto an EMMO sub-domain without breaking its EMMO inheritance. The contrast is therefore not over rival mid-level vocabularies but over the modularity *axis* (class count versus consumer concern) and the reasoning *posture* (mandatory inheritance versus opt-in distribution).

3.2. Level 0 — Bridge to External Standards

L0 is the level at which OCO's interoperability claim is made concrete. Its content is a curated set of `bridge:*` classes that act as named anchors between OCO's internal vocabulary and external mid-level ontologies. Internal modules reference only these anchors, never the external IRIs directly — this isolates external-version churn (a PMDco v3.1 release, a QUDT update) to a single level that is small enough to audit class-by-class.

The selection of bridge targets is not opportunistic: a target is admitted only if it satisfies all of the following criteria.

1. **OWL/RDF format.** The target must be available as a machine-readable ontology in a standard W3C format. Industrial classification schemes published only as proprietary XML or catalog formats are excluded.
2. **Mid-level granularity.** Targets must occupy the *mid-level* between abstract foundations and instance data. Direct alignment to upper-level foundations is achieved transitively via the mid-level targets that already commit to them (e.g., BFO via PMDco), avoiding redundant philosophical commitments at the OCO surface.
3. **Established adoption.** The target must be in active use within the materials-science or research-data-management community, with a versioned release history.
4. **Bounded import cost.** A bridge entry adds at most a handful of anchor classes to OCO. Targets whose useful adoption requires importing hundreds of unrelated upper classes are inadmissible by this criterion — it would defeat the modular multi-level design.
5. **License compatibility.** Targets must permit redistribution under terms compatible with OCO's CC-BY/CC-BY-SA scheme.

Table 1 lists the eleven bridge targets with substantial class-level coverage (each anchored to $\geq$14 OCO classes). A further six targets (IAO, m4i, SOSA, DCAT, W3C-Org, IUCr-Wiki) are present as minimal anchors (1–2 classes each), and the explicit cross-ontology mapping inventory in `bridge_mappings.yaml` contributes 467 further mappings, including bridges to the predecessor PMD ontologies KnowNow and SmaDi, to OPTIMADE, Croissant and the ELN-Filetype standard, and to the EMMO sub-modules ISQ, chemistry, and materials (Section 4.7).

Table 1: **L0** bridge targets with substantial coverage. *Anchors* counts the distinct OCO classes that map to the target via `owl:equivalentClass`, `rdfs:subClassOf`, or `skos:exactMatch`/`closeMatch`, measured on the merged distribution `oco_merged.ttl` (2026-05-25). Sorted by anchor count.

Target	Role in OCO	Anchors
ChEBI (Hastings et al., 2016)	chemical entities (compounds, salts, reagents)	163
CHMO (Batchelor, 2017)	characterization methods (XRD, SEM, EIS, . . .)	151
QUDT QuantityKinds (Hodgson et al., 2018)	physical quantities (kinds and units)	141
Wikidata Q-IDs (Vrandečić and Krötzsch, 2014)	universal entity anchor (eponyms, elements, materials)	136
PMDco v3.0.0 (Bayerlein et al., 2024d)	materials mid-level (process, sample, property, equipment)	105
PROV-O (Lebo et al., 2013)	provenance of activities, agents, entities	68
NFDIcore (NFDI Section ELSA and NFDIcore Working Group, 2024)	NFDI cross-domain RDM concepts	32
DOI namespace (International Organization for Standardization, 2022)	citation infrastructure (literature, datasets)	22
OBI	biomedical-investigations vocabulary (assays, instruments)	20
schema.org	web vocabulary (organizations, products, identifiers)	15
FaBiO (Peroni and Shotton, 2012)	bibliographic-resource typing (Article, Patent, Standard)	14
Total (substantial only)	many OCO classes carry anchors to more than one target, so the column sums to more than the count of distinct bridged OCO classes	**867**

Targets explicitly not bridged, and why. A second list of candidate targets has been screened and rejected. The rejection categories are themselves part of the architectural argument.

- *Foundational ontologies* (BFO (Arp et al., 2015), UFO-A (Guizzardi, 2005)). OCO is BFO-aligned transitively via PMDco; re-asserting BFO at the OCO surface would duplicate that commitment without adding classes. UFO-A's rigid Kind/Role/Phase distinctions would force the reclassification of all existing OCO classes into a meta-typology that the materials community has not adopted — a change with high cost and no demonstrated benefit for our competency questions.

- *Centralist mid-level competitors* (EMMO (Ghedini et al., 2019) and its sub-domain ontologies such as the EMMO quantity, electrochemistry, and atomistic modules). EMMO sub-domain ontologies require importing the full EMMO core (700+ classes), which contradicts the bounded-import-cost criterion above and forecloses the opt-in reasoning principle: every consumer pays the full EMMO load regardless of need. We discuss the centralist-versus-multi-level contrast further in Section 6.

- *Mid-level MSE competitors to PMDco* (Fraunhofer MSEO). PMDco v3 has emerged as the consortial convergence target for the BFO-aligned mid-level MSE space, with active maintenance and broad German-ecosystem adoption. Maintaining a second mid-level MSE anchor in parallel would reintroduce the dual-anchor ambiguity in the very level whose purpose is to provide a unique external anchor per concept; the concepts MSEO covers are reachable via PMDco.

- *Quantity/unit competitors to QUDT* (OM (Rijgersberg et al., 2013), EMMO `domain-quantity`). PMDco has standardized on QUDT as its unit anchor. Maintaining a second quantity bridge would introduce ambiguity in the very level whose purpose is to provide a unique external anchor per concept.

- *Non-OWL industrial classification schemes* (ETIM (ETIM International, 2024), eCl@ss (eCl@ss e.V., 2024), ISO 15926-14 (International Organization for Standardization, 2020), NAMUR NE 100 (NAMUR Working Group, 2018)). These are widely adopted in industry but are distributed only as BMEcat-XML, proprietary catalogs, or non-OWL formats. No semantic mapping path exists that does not first require the external community to publish an OWL representation.

- *ELN/LIMS tool ontologies* (Chemotion, eLabFTW, openBIS, Kadi4Mat, NOMAD-Oasis, ...). These systems are *consumers* of OCO's **L1** schema, not peer ontologies: their concepts (user sessions, folder permissions, signature workflows) are deliberately outside OCO's modeling scope.

One known gap in the present roster. **MeSH** (Medical Subject Headings, U.S. National Library of Medicine) is published in SKOS/RDF and qualifies under our criteria; it is not bridged because the pharmaceutical sister domain is not currently a consumer, but no architectural obstacle exists.

The discipline of writing down both the inclusion criteria and the explicit rejections — and of naming the known gaps — turns **L0** into the documented contract between OCO and the surrounding ontology landscape, rather than a loose collection of opportunistic mappings.

3.3. Level 1 — Material-Agnostic LIMS/ELN Level

L1 contains all concepts that are independent of the specific material under investigation: a spray dryer dries pharmaceutical suspensions as well as ceramic slurries; an XRD measures metals as well as ceramics; a poling station works on PVDF polymers as well as on BNT-BT. The level decomposes into fourteen pure-**L1** modules plus the **L1** fragments of two modules whose classes span **L1** and **L2** via the multi-axis classification (Section 4.6). Consumers import only the parts they need — an asset-management tool may load `oco-equipment` alone; an LIMS may bundle `oco-investigation`, `oco-process`, `oco-equipment` and `oco-measurement` and skip the rest. Table 2 lists the modules and the architectural reason for each.

A core L1 pattern: literature and laboratory data as semantically equivalent sources. OCO treats published-literature extractions and own-laboratory experiments as first-class equivalent ABox contributors. Both are modelled as `prov:Activity` instances and share the same provenance, quality, and uncertainty annotations — distinguished only by their activity class (`Investigation` vs. `LiteratureExtraction`) and their verification status, not by a separate vocabulary. A SPARQL query asking "what is the d_{33} of BNT-BT?" returns both validated lab measurements and extracted literature values, ranked by data-confidence level, without the schema forcing one source to be preferred a priori. This unification is the operational consequence of the equality principle (Section 2.4, principle 5) and is the architectural prerequisite for using large-scale literature corpora alongside in-house experimentation; it also enables the SIPOC-fragment data model (Section 4.9) to accommodate incomplete records uniformly across both source types.

Table 2: The **L1** modules (all under the `oco-*` namespace). *Classes* counts the **L1**-level classes per module in the merged distribution; modules marked † also carry **L2** classes (concrete materials, tensor-component sub-properties — their **L2** portion is reported in Table 3).

Module	Classes	Why this module is in **L1**
`measurement`	554	Characterization-method taxonomy (XRD, SEM, EIS, d_{33}, impedance spectroscopy, . . .); methods are material-agnostic, what differs is the sample.
`property`†	429	Physical-quantity backbone for laboratory parameters (coercivity, remanent polarization, piezoelectric coefficients d_{ij}, . . .), classified by property nature. Cannot be replaced by QUDT alone. The **L2** portion (Table 3) holds the tensor-component sub-properties.
`process`	263	ProcessStep hierarchy (Mixing, Calcination, Sintering, Poling, Patterning, . . .); workflow primitives are domain-stable across material classes.
`investigation`	224	Investigation, Study, Assay, Sample, Batch, SampleState, Provenance — the ISA lab-notebook skeleton, plus DCL 0–5 and ExperimentStatus lifecycle (planned → in-progress → completed → archived).
`equipment`	218	Apparatus (furnaces, ball mills, spray dryers, XRD diffractometers, . . .). Cross-domain anchor for asset-management and procurement tools.
`supplier`	137	Supplier and Specification — provenance of feedstock, commercial identifiers, lot numbers, certificate-of-analysis anchors. Material-agnostic procurement vocabulary.
`representation`	89	Data-shape vocabulary: `ScalarValue`, `Vector`, `Matrix`, `Tensor` (with rank-specific subclasses), `TimeSeries`, `ProcessCurve`, `ThermalProfile`. **L2** instantiates these shapes with concrete axes (e.g. piezoelectric hysteresis loops, phase diagrams).
`instrument`	82	Analytical instruments and their detection components (`NMR_Spectrometer`, laser-flash apparatus, Raman spectrometer, diffraction detectors, AFM tips, . . .). Split from `equipment` so that instrument-level modelling — with manufacturer, signal chain, and detector granularity — is separable from the broader durable-asset taxonomy.
`identifier`	63	External-identifier types (CAS, InChI, ICSD, Pearson, Wikidata, lot number, . . .). A three-tier hierarchy (mandatory / recommended / optional, ∼45 types in total) is SHACL-validated per material class.
`simulation`	40	Simulation methods (DFT, MD, phase field, FEM, CALPHAD) and software anchors. Treated as a sibling of `process`: a simulation step is a step in the workflow, not an afterthought.
`documents`	29	Document taxonomy spanning lab/industry artefacts (`TechnicalManual`, `ProductDataSheet`, `SafetyDataSheet`, `CertificateOfAnalysis`, `CalibrationCertificate`) plus the publication subtypes bridged to FaBiO/CiTO. Required so that LiteratureExtraction can be modelled as a first-class `prov:Activity`.
`consumables`	22	Single-use lab consumables, dominated by crucibles and boats (alumina, zirconia, platinum, . . .). Material- agnostic procurement vocabulary that does not fit the durable-asset framing of `equipment`.
`diffraction`	17	Diffraction-specific concepts (`DiffractionReflection`, `OrientationReflection`, `ResidualDensityAnalysis`, `StructureRefinement`, `ReflectionStatistics`). Kept separate from `measurement` because the diffraction sub-tree carries its own analytic methods and refinement metadata beyond the generic measurement-step vocabulary.
`environment`	17	Process and storage atmospheres (air, inert, argon, nitrogen, . . .) annotated with composition. Cross-cutting and material-agnostic, so factored out of `process`.
`material`†	9	Material superclass plus four substance-based subclasses (Metal, NonmetallicInorganic, Organic, Composite) as the agnostic taxonomy frame; concrete materials (BNT-BT, KNN, Al_2O_3, ferritic ceramics, . . .) live in **L2** under these branches. Process roles are modelled separately in the `oco-role` module (Table 3).
`rationale`	5	DesignRationale + RationaleConfidence: meta-modelling for *why* a sample/process/specification was chosen — captures decision provenance, not just data provenance.
Total L1	**2 198**	

The level is explicitly designed to be ELN-/LIMS-importable. Beyond the module structure, four cross-cutting features make **L1** operational for laboratory-data systems:

- **Data confidence levels (DCL 0–5):** a mandatory annotation on every `MeasurementResult` or `ExtractedValue`, ranging from *Unknown* (DCL 0) to *Validated* (DCL 5). This separates raw entries, peer-reviewed values, and gold-standard data within the same knowledge graph.
- **ExperimentStatus lifecycle:** planned → in-progress → completed → archived, with a separate `StatusArchived` terminal state. Required for any LIMS that distinguishes intent from execution.
- **GUM uncertainty fields:** CombinedStandardUncertainty with the five datatype properties prescribed by JCGM 100:2008, attached to existing `MeasurementUncertainty` classes rather than introduced as a parallel hierarchy.
- **LiteratureExtraction as first-class `prov:Activity`:** extraction from publications and patents is a tracked activity with its own VerificationStatus, so SIPOC-fragment data (Section 4.9) carries its own provenance trail rather than being silently merged with laboratory data.

Substitutability. The level's existence as a separate artefact is the architectural lever that distinguishes OCO from a monolithic ceramics ontology: a sister project (metallurgy, polymer science, pharmaceutics) imports **L1** unchanged together with its own **L2**, replacing only the material-class-specific content. This is the structural mechanism by which the "∼70 % workflow duplication across PMD-consortium projects" problem identified in Norouzi et al. (2024) is dissolved.

3.4. Level 2 — Material-Class-Specific Level

L2 contains the material-class-specific knowledge — the part of OCO that a sister project (metallurgy, polymer science, pharmaceutics) would replace while keeping **L0**+**L1** unchanged. For ceramics it has two internal facets: the *material modules* themselves (Section 3.4.1), and the *seven-tier mechanistic-explanation skeleton* (Section 3.4.2) that structures how the material content can be queried for cause-chain explanations rather than only for property values.

3.4.1. Material Modules

For ceramics, **L2** decomposes into nine architecturally significant modules (Table 3) that together cover the structure-property-defect chain that distinguishes functional-ceramic modelling from generic materials description. Four additional smaller modules (`dataquality` 17, `dimension` 14, `application` 12, `calc` 10) carry support vocabulary for the same level and total 53 further classes. A fifth module, `oco-role`, holds the abstract `Role` class plus 14 named individuals (`solvent`, `binder`, `precursor`, `reactant`, `dopant`, `plasticizer`, ...) that are attached to process steps rather than to material classes — the pattern mentioned alongside the `material` module (Table 2). The `oco-localstructure` module listed below additionally contributes 7 **L1**-frame classes (CoordinationPolyhedron root + family abstracts Tetrahedron / Octahedron / Cuboctahedron) alongside its 12 **L2** polyhedra.

A cross-cutting feature spans **L2** and **L1**: the **coupled-effect family** (32 effect classes: piezoelectric, pyroelectric, electrostriction, magnetostriction, magnetoelectric, thermoelectric, Verdet, Cotton-Mouton, plus 24 diagonal/cross effects following Nye's convention (Nye, 1985)). The effect classes themselves are physical quantities and live in `oco-property` at **L1**; the symmetry constraints that connect each effect to its allowed crystallographic point groups live at **L2** via the `oco-tensor` module and the reified `oco:NeumannConstraint` instances generated programmatically by the Neumann tensor engine (Section 4.5). This split keeps the universal physical-quantity hierarchy in the agnostic level while the material-bound symmetry constraints stay where they belong — and replaces what would otherwise be hundreds of manually curated cross-axioms with a single algorithm.

Substitutability. The nine principal modules in Table 3 are the surface that a sister project replaces. The ferritic-high-performance-ceramics implementation currently in development is the first concrete test of this substitutability within the ceramics family; cross-domain substitutability (metallurgy, polymers) is on the validation roadmap (Section 6).

3.4.2. Seven-Tier Mechanistic-Explanation Skeleton

The material-specific content of **L2** is organised internally by a seven-tier mechanistic-explanation skeleton that distinguishes an authoritative materials ontology from a vocabulary that only describes. Every material property recorded in OCO sits within a chain of physical mechanisms; for industrial use that chain must be queryable from the same knowledge graph as the property value itself. OCO formalises this chain as a fixed skeleton of seven explanation tiers, each answering one fundamental materials-science question and contributing dedicated modules and (where applicable) an external

Table 3: The principal **L2** modules for ceramics (all under the `oco-*` namespace). *Classes* counts the **L2**-level classes per module in the merged distribution. The modules marked † also carry **L1** content (Table 2). A sister project for a different material class would replace these modules while leaving **L0**+**L1** untouched.

Module	Classes	Why this module is in **L2**
`property`†	1 018	Material-specific tensor-component sub-properties (Voigt components d_{ij}, s_{ijkl}, π_{ij}, …) and their multi-axis classification (Section 4.6). Generated from family-templates plus outlier-override; the **L1** portion (Table 2) holds the laboratory backbone.
`crystal`	317	Crystal structure: 230 space groups, 32 point groups, Bravais lattices, Pearson symbols, plus functional-ceramic tags (centrosymmetric, polar, piezoelectric, pyroelectric, enantiomorphic). Space groups are universal physics, but the tags that connect symmetry to functional behaviour are ceramics-flavoured. A polymer ontology would replace the tag set with chain-conformation descriptors.
`material`†	259	Concrete material classes (BNT-BT, KNN, BCZT, Al_2O_3, ZrO_2, ferritic high-performance ceramics, …) and their composition relationships. The skeleton sits in **L1**; the catalogue that a sister project replaces with its own metals/polymers/actives lives here.
`tensor`	191	A `MaterialTensor` root plus 190 tensor-class sub-roots (Voigt families, polarization, conductivity, elasticity, piezo-, pyro-, magnetoelectric, …), each annotated with tensor order, symmetry class, and axial/polar flag. The structural anchor for the Neumann engine (Section 4.5).
`element`	135	Periodic-table classes with element-specific properties (atomic number, electronegativity, common oxidation states, ionic radii). Bridged to Wikidata. Elements sit in **L2** because element subsetting is material-specific: ceramic-relevant elements (Ti, Zr, Pb, Bi, Na, K, …) differ from those that drive metals (Fe, Al, Cu, Ni) or polymers (C, H, O, N).
`defect`	91	Defect chemistry. Kröger-Vink notation for ionic solids (Kröger and Vink, 1956) ($V_O^{\bullet\bullet}$, Mg_{Si}'') and Brouwer-diagram modelling of defect equilibria. A metallurgy ontology would replace this wholesale with Burgers-vector dislocation notation; a polymer ontology with chain-defect descriptors.
`phase`	73	Phases and phase transitions: solid solutions, morphotropic phase boundaries, ferroelectric/paraelectric transitions, plus concrete phase regions for BNT-BT (four with (T, x) bounds) and NiCuZn ferrite (Curie transition). The substrate that the phase-state coupling (Section 4.5) consumes.
`composite`	13	Composite topology. Newnham connectivity (Newnham et al., 1978) (0-0, 0-3, 1-3, 2-2, 3-3, …) for two-phase functional-ceramic composites. Small but architecturally important: the connectivity notation is what makes piezoelectric-composite reasoning possible at all.
`layer`	6	The five material-abstraction layers (Atomic, Crystalline, Microstructural, Mesoscopic, Macroscopic) plus their Layer root. Each tier is a distinct tuple of *state variables*, *governing laws*, and *characteristic length scale* — not a spatial partition (Section 4.6).
`localstructure`†	12	Concrete coordination polyhedra (TiO_6, BiO_{12}, BaO_{12}, SiO_4, ZnO_4, NbO_6, FeO_6, NiO_6, …) inheriting from the **L1** family abstracts. Closes the mechanistic chain *composition* → *local geometry* → *symmetry* → *allowed tensor components* (Section 4.5).
Total (principal	**2 115**	

reference-data cache (Table 4).

Why "tiers" — five material-abstraction layers vs. seven mechanistic-explanation tiers. We deliberately call the seven mechanistic-explanation stages *tiers*, not layers, to keep them terminologically distinct from the five *material-abstraction layers* (Atomic, Crystalline, Microstructural, Mesoscopic, Macroscopic) carried by the `oco-layer` module listed in Table 3. The two concepts answer different questions. The material-abstraction layers describe the *scale and governing-law regime* at which a property is observed — each is a distinct tuple of state variables and governing laws used as a classification axis for parameters (Section 4.6). The mechanistic-explanation tiers describe the *causal mechanism stack* that links composition through symmetry, energy, thermodynamics, kinetics, microstructure, defects, and bonding to the observed property. *Microstructure* appears in both lists by name only: in the material-abstraction sense (layer) it is a description scale (grain-sized features); in the mechanistic-explanation sense (tier) it is the cause-chain stage that supplies polycrystalline averaging, Hall–Petch coupling, and texture effects.

Generic skeleton, not material-specific. All seven tiers apply to every *crystalline ionic oxide* — the central material class for functional ceramics (ferroelectrics, magnetics, ion conductors, high-T_c superconductors). BNT-BT is the pilot material on which the skeleton was concretised; it is not the reason the skeleton has these seven tiers. The same skeleton applies without restructuring to NiCuZn ferrite, yttria-stabilised zirconia, $BaTiO_3$, and any related material domain — only the per-tier instance content changes.

External reference-data caches. Each tier that draws on a large external reference corpus does so through the cache pattern formalised as Principle 5 in Section 3.1: a versioned local snapshot rather than embedded TBox classes. The per-tier cache totals appear in the right-most column of Table 4; all caches are SHA-pinned in `bridge/external_versions.yaml`. The OWL TBox stays compact; the externally curated data scales independently.

Cross-tier validation. The chain of explanation is auditable. Cross-tier references are formalised as seven SHACL NodeShapes (`shapes/m72_cross_layer_shapes.ttl`) that flag incomplete annotations: a morphotropic phase boundary (Tier 3 Thermo) must reference a domain-state set (Tier 1 Symmetry); a soft mode (Tier 2 Energy/DFT) must reference a subgroup relation (Tier 1 Symmetry); a Brouwer diagram (Tier 6 Defect chemistry) must reference a chemical-potential diagram (Tier 3 Thermo); an aging-kinetics

Table 4: The seven mechanistic-explanation tiers in OCO, with implementing modules and external reference-data caches. Each tier answers one fundamental question about a crystalline ionic oxide. Caches are pinned to upstream versions in `bridge/external_versions.yaml` per the cache-pattern principle (Section 3.1).

#	Tier	Question answered	Implementing modules	External cache
1	Symmetry	Which crystal structure, which symmetry operations? Point-group theory (Neumann's principle) governs which tensor components are allowed; subgroup-relations drive phase transitions; Wyckoff positions give dopant and defect sites.	`oco-symmetry` (11) atop `oco-crystal`, `oco-tensor`	1 731 Wyckoff positions (pyxtal); 1 934 bond-valence parameters (IUCr Brown 2020)
2	Energy / DFT	Which electronic structure, which phonon modes, which Born effective charges? First-principles ground for intrinsic properties without empirical fitting; soft-mode analysis identifies phase-transition mechanisms.	`oco-energy-dft` (33 classes + 7 properties)	∼155 000 Materials Project DFT entries
3	Thermo / CALPHAD	Which phases are stable at (T, p, x)? Sublattice models and Redlich–Kister polynomials govern Gibbs-energy surfaces; sintering windows, miscibility gaps, and morphotropic phase boundary locations follow.	`oco-thermo` (25, incl. MPB + chemical-potential classes) atop `oco-phase`	—
4	Kinetics	How fast do diffusion, reaction, sintering, switching, and aging proceed? Fick, Arrhenius, JMAK, Hillert grain growth, Cahn–Hilliard spinodal decomposition, domain-wall mobility.	`oco-kinetics` (Fick, Arrhenius, JMAK, Hillert, Cahn–Hilliard, DW mobility, aging)	—
5	Microstructure	Which grain size, texture, and grain-boundary character? Polycrystalline averaging through Voigt–Reuss–Hill, Mori–Tanaka and Hashin–Shtrikman bounds; Hall–Petch coupling to mechanical response.	`oco-microstructure` (26: VRH, MT, HS, Hall–Petch, CSL grain boundaries)	—
6	Defect chemistry	Which point defects, in which concentrations, with which charge compensation? Kröger–Vink notation, Brouwer diagrams, donor / acceptor / iso-valent / amphoteric doping strategies.	`oco-defect` extended (+48 Kröger–Vink, Brouwer, doping classes)	—
7	Bonding chemistry	Which electron configuration, which hybridisation, which bonding character (ionic / covalent / metallic)? Pauling rules, HSAB classification, crystal-field theory, lone-pair stereo-activity.	`oco-bonding` (28: HSAB, lone-pair, hybridisation, Pauling rules, crystal field)	497 Shannon ionic radii; 91 Pauling electronegativities

model (Tier 4 Kinetics) must reference a Kröger–Vink defect (Tier 6 Defect chemistry); a Hall–Petch model (Tier 5 Microstructure) must reference a grain-boundary classification (Tier 6 Defect chemistry, via grain-boundary defect types); a Born effective charge (Tier 2 Energy/DFT) must reference a bonding-character annotation (Tier 7 Bonding chemistry); a spinodal decomposition (Tier 3 Thermo) must reference a Cahn–Hilliard model (Tier 4 Kinetics). The constraints are deliberately lightweight (`sh:Warning` severity, not `sh:Violation`) because cross-tier annotation grows evolutionarily; missing references should attract modeller attention, not block the pipeline.

End-to-end pilot. The BNT-BT pilot ABox (`examples/abox_pilots/bnt_bt_pilot.ttl`) carries concrete instances on each of the seven tiers with explicit cross-tier annotations: a soft mode at the Γ point linked to the R3c→P4mm subgroup relation; a Born effective charge $Z^*_{33}(\mathrm{Ti}) \approx +7.2$ explained by Bi-6s^2 lone-pair stereo-activity; a morphotropic phase boundary at $x_{\mathrm{BT}} \approx 0.06$ linked to the R3c/P4mm/Cm domain-state set; an aging-kinetics model linked to the $\mathrm{V}_{\mathrm{O}}^{\bullet\bullet}$ / $\mathrm{Mn}_{\mathrm{Ti}}''$ defect-dipole chemistry; a Hall–Petch parameterisation with $\sigma_0 \approx 3\,\mathrm{GPa}$, $k_{HP} \approx 0.8\,\mathrm{MPa}\sqrt{\mathrm{m}}$ at $d = 2\,\mu\mathrm{m}$; and a complete DFT-workflow provenance (PBEsol functional, PAW pseudopotentials, $8 \times 8 \times 8$ Monkhorst–Pack mesh, 600 eV cutoff). The pilot demonstrates that the skeleton holds together operationally, not only conceptually — a query for "why does BNT-BT achieve $d_{33} \approx 580\,\mathrm{pC/N}$ at the MPB" traverses all seven tiers along the cross-tier annotation chain.

3.5. Level 3 — Categorical Reasoning Level

L3 carries the categorical (boolean-logical) reasoning that crosses module boundaries — the truths that hold *always*, *never*, or *under condition X*, and that have no natural home in any single **L1**/**L2** module because they connect classes from several. Quantitative mathematics (functions, scaling laws, calibration curves) is *not* part of **L3**; it lies outside this release entirely (Section 6.4). **L3** ships as two TTL distributions, `oco-domain.ttl` (categorical axioms and compatibility tables) and `oco-routes.ttl` (ordered process-step templates and their state-sequence entries), compiled from YAML axiom sources under the `axioms/` directory. Their combined contribution to the merged distribution, measured as the ROBOT `measure` (Jackson et al., 2019) delta between `oco_merged_l12.ttl` (**L1**+**L2**) and `oco_merged.ttl` (**L1**+**L2**+**L3**), is 1 400 OWL axioms / 404 logical axioms / 248 ABox axioms. The content decomposes into four families (Table 5).

A characteristic example is the symmetry-effect axiom

$$\mathit{Pyroelectricity} \sqsubseteq \mathit{requires\_pointgroup}\ \mathit{only}\ \mathit{PolarPointGroup}, \qquad (1)$$

Table 5: The four record families in **L3**. *Records* counts source-level entries; each record typically emits several OWL axioms or reified-constraint triples in the compiled distribution.

Family	Records	Why this family is in **L3**
Domain axioms	74	Categorical constraints that link classes from two or more modules. Symmetry-effect axioms (Pyroelectricity requires a polar point group), lifecycle constraints (PoledBody may transition only to AnnealedBody, AgedBody, or FatiguedBody), property-disjointness facts (`agrees_with` ⋈ `disagrees_with`), and partition-disjointness axioms over the 10 material-abstraction layers (Table 3) and the 21 crystal-system root classes. These do not belong in any one module because they cross Property×Crystal, Process×State, or Property×Property boundaries.
Route templates	14	Ordered process-step sequences per (synthesis route × application domain), e.g. SolidState × Piezoelectric prescribes Mixing → DryMilling → Calcination → . . . → Poling as the canonical chain. Required to validate whether a recorded ABox process chain is a legal instance of a known route, and to answer comparability questions across publications. Each template unfolds into ordered step and state-sequence records (129 steps and 43 state-sequence entries in total).
Compatibility entries	56	Status classification of the route × domain matrix (*standard*, *rare*, *incompatible*), e.g. ThinFilmRoute × Structural is incompatible. Lets the ontology distinguish "unusual but valid" from "rule violation".
L3-justifying CQs	144 of 163	Each **L3** axiom is justified by at least one published competency question. Of the full CQ catalogue (count in Section 4), 144 entries map to **L3** axioms; the remainder cover **L1**/**L2** patterns. 52 CQs carry an executable SPARQL test against gold-standard ABoxes (52/52 PASS). Questions are tagged with one of ten reasoning *areas* (tensor symmetry, phase state, route, lifecycle, multi-axis, interoperability, quality, crystal structure, BNT-BT, and the NCZF placeholder), making the reasoning scope auditable per area.
Total	**288**	source-level entries; their compiled contribution to `oco_master_full.ttl` is 1 400 OWL axioms / 404 logical (ROBOT-measured delta over **L1**+**L2**)

which encodes the categorical fact that a material can show pyroelectricity only if its crystallographic point group is one of the ten polar groups. This is the kind of axiom whose evaluation requires the universal restriction (`only`) construct — which is exactly the construct that lifts the OCO-with-**L3** profile from OWL 2 EL to OWL 2 DL (Table 8). Consumers without this reasoning need neither load **L3** nor pay the OWL 2 DL reasoner cost.

The decisive architectural feature of **L3** is not the axioms themselves but their *separation* from the rest of the schema and their binding to published competency questions. This makes the scope of the ontology — which questions it intends to answer — explicit, auditable, and testable: a CQ that loses its SPARQL test is a visible regression, and a proposed new axiom that cannot point to a CQ has no architectural justification for being admitted.

3.6. The Material / Compliance Audience Axis

A second classification axis is independent of the level hierarchy: every module carries an `audience` marker, with value `material`, `compliance`, or both. Twenty-nine modules constitute the material-audience-only set — the materials-science vocabulary proper, including `oco-localstructure` (the mechanistic bridge from composition through coordination polyhedra to symmetry, Section 3.4.2). Eleven modules form the compliance-audience-only set and carry the EU-regulatory and value-chain vocabulary detailed in Section 4.8: Life Cycle Assessment, CSRD/ESRS reporting, supply-chain due diligence (CSDDD), packaging (PPWR), carbon-border adjustment (CBAM), right-to-repair (R2R), the AI Act, Manufacturing-X identifier and traceability infrastructure, regulated substances, recycling, and Safe-and-Sustainable-by-Design. Four further modules are *dual-audience* — they serve both sides simultaneously: `oco-format` (industrial data formats), `oco-odrl` (W3C ODRL policy and Verifiable Credentials trust layer), `oco-time-event` (temporal extents per W3C Time), and `oco-automation` (SiLA 2.0 laboratory automation). These appear in Table 9 marked with †.

The two audiences are independent of the four levels. Every compliance module has its own **L1**/**L2**/**L3** internal structure, just as the material modules do. A consumer who needs only the material core can ignore the compliance audience entirely; the `audience` marker drives module-selection profiles in the distribution. A consumer building a Manufacturing-X data pipeline loads both audiences. The independence is what makes the same architectural primitive — modular layering — absorb a second class of requirements without restructuring the material core.

4. Reference Implementation: OCO

4.1. Quantitative Description

The OntoCrafter Ceramics Ontology (OCO) is the reference implementation of the architecture introduced in Section 3 — two independent classification axes (level and audience) plus the seven-tier mechanistic-explanation structure of the material level. Metrics for the v0.94 release — the version entering productive practice — are measured with ROBOT 1.9.10 (Jackson et al., 2019) and `rdflib` on the merged distribution `distribution/oco_merged.ttl` extended by the Neumann-constraint sub-distribution:

- **Classes:** 5 196 named, distributed across 44 class-bearing modules (45 in the namespace registry). 29 modules carry the material audience and 15 carry the compliance audience, four of them dual-audience (Section 4.8). Every module carries its own **L1**/**L2**/**L3** internal structure independent of the audience axis, and every material-audience module sits on one or more of the seven mechanistic explanation tiers (Section 3.4.2).

- **Properties:** 1 674 total — 574 ObjectProperties, 1 051 DatatypeProperties, and 49 AnnotationProperties.

- **Axioms:** 167 348 OWL/RDFS axioms in the merged distribution, of which 40 454 are logical axioms (the remainder are annotations and bare entity declarations); per ROBOT `measure` (Jackson et al., 2019). The largest single contribution is the 5 920 reified `oco:NeumannConstraint` instances — one per (tensor × point-group) pair, Section 4.5.

- **Bridges:** 11 **L0** targets with substantial class-level coverage ($\geq$14 anchors each, Table 1); a further 829 explicit cross-ontology mappings across 40 sections in `bridge_mappings.yaml` and `mwo_mappings.yaml` cover the material-side standards (PMDco, PROV-O, NFDIcore, MADO, OBI, CIF Core, EMMO sub-modules, QUDT, Croissant, KnowNow, SmaDi), the compliance-side standards (EN 15804, EU PEF, ecoinvent, BONSAI, EFRAG XBRL, CRMA, NZIA, PACT, Catena-X, CSDDD-UNGP, CBAM, R2R, AI Act, AAS, . . .), and the policy/trust layer (W3C ODRL, W3C Verifiable Credentials, SiLA) — Section 4.7, Section 4.8.

- **Competency questions:** 163 published, each tagged with one of ten reasoning *areas* (tensor symmetry, phase state, route, lifecycle, . . .); 52

carry executable SPARQL test queries against gold-standard ABoxes (52/52 PASS).

- **SHACL shapes:** 1 172 mandatory shapes validated via `pyShacl`; dominated by the machine-generated Neumann (768) and CIF-Core (270) subsets, complemented by 7 Cross-Tier NodeShapes (Section 3.4.2) and per-domain hand-curated sets for material-class identifiers, MPB coexistence, and the compliance audience.
- **Architecture decision records:** 132, including explicit negative decisions for rejected bridge targets and rejected modelling alternatives.
- **External reference-data caches:** 5 productive caches under the seven-tier explanation skeleton (Section 3.4.2), pinned to upstream versions in `bridge/external_versions.yaml`: Shannon ionic radii (497), IUCr bond-valence parameters (1 934), pyxtal Wyckoff positions (1 731), Materials Project DFT corpus (∼155 000), Pauling electronegativities (91).
- **Expressivity:** the full distribution sits in OWL 2 DL; the **L0**+**L1**+**L2** bundle (without **L3**) reduces to OWL 2 EL — Table 8 gives the underlying ROBOT measurements.

Comparison with PMDco and EMMO. To place these metrics in landscape context, Table 6 compares OCO v0.94 against PMDco v3 (OCO's primary mid-level bridge target, Section 3.2) and EMMO Crystallography (the EMMO sub-module closest in scope to ceramics). The base-metric values for the comparators are taken from Norouzi et al. (Norouzi et al., 2024), Tab. 9. OCO is roughly an order of magnitude broader than PMDco in classes and properties and almost two orders of magnitude broader in axioms; this is expected because PMDco is a mid-level vocabulary that OCO extends downward into a full material domain. EMMO is federated across 40+ sub-modules, each with its own per-module metrics; we report the Crystallography sub-module as the most analogous single point of comparison.

4.2. Distribution Variants

The OCO distribution is structured so that consumers select only the depth they need (see Table 7). Both bundle files — `oco_master.ttl` and `oco_master_full.ttl` — are thin `owl:Ontology` hulls carrying only `owl:imports` directives; the actual content lives in 17 per-module `oco-<modul>.ttl` files plus `bridge.ttl`. The two master variants are *alternatives*, not additive: importing the full variant subsumes the standard one.

Table 6: Metric comparison OCO v0.94 against PMDco v3 and EMMO Crystallography, two reference points from the MSE ontology landscape. Class, property, axiom, and annotation-axiom counts for PMDco and EMMO Crystallography are taken from Norouzi et al. (2024), Tab. 9; "—" marks metrics not reported as release-time statistics in that survey.

Metric	OCO v0.94	PMDco v3	EMMO Cryst.
Classes	5 196	264	61
Object Properties	574	36	5
Data Properties	1 051	9	1
Total Axioms	167 348	2 154	357
of which Annotation Axioms	126 894	1 454	175
Reified Neumann constraints	5 920	—	—
Cross-ontology bridges	829	—	—
SHACL shapes	1 172	—	—
Competency questions / SPARQL tests	163 / 52	—	—

Table 7: OCO distribution choice as a function of consumer profile. Reasoner column gives the minimal sufficient profile for the axioms actually present (see Table 8).

Consumer profile	Files to load	Reasoner
PMDco / QUDT integration only	`bridge.ttl`	RDFS
LIMS/ELN, no material focus	selected **L1** modules + `bridge.ttl`	RDFS
Ceramics materials scientist	`oco_master.ttl` (**L0**+**L1**+**L2**)	OWL 2 EL
Routes / symmetry / lifecycle reasoning	`oco_master_full.ttl` (adds **L3**)	OWL 2 DL

The reasoner-profile assignment is empirical, not assumed: ROBOT classifies the description-logic constructs actually used in each bundle (Table 8). The decisive contrast is that adding **L3** introduces universal restrictions (∀, `owl:allValuesFrom`, used in route-template completeness constraints) and complex concept negation, both of which lift the profile from OWL 2 EL into full OWL 2 DL. The **L0**–**L1**–**L2** bundle stays in EL-tractable territory — a polynomial-time reasoner (e.g. ELK) is sufficient. This is the multi-level-architecture promise made operational: *reasoning cost is a consequence of the level chosen, not a property of the ontology.*

Table 8: Description-logic constructs (ROBOT `measure` output) per bundle. The bold tokens in `oco_master_full.ttl` are the **L3** additions that lift the profile out of OWL 2 EL.

Bundle	Expressivity (constructs in use)
`oco_master.ttl`	`RRESTRERIF(D)` — role hierarchy + restrictions, ∃, inverse, functional, datatypes
`oco_master_full.ttl`	`RRESTRCUNIVRESTRERIF(D)` — all of the above, plus ∀ **(universal restriction)** and **complex concept negation**

4.3. Bilingual Definitions as Mandatory Field

Each class carries both a German (`definition_de`) and an English (`definition_en`) definition. This is enforced as a hard schema-validator constraint. Among the surveyed MSE ontologies (Norouzi et al., 2024; Zhang et al., 2024; Jalali et al., 2023), none mandates bilingual definitions; OCO is to our knowledge the only MSE ontology that does so as a hard constraint. This matters specifically for German-speaking industrial partners, BMBF projects, and the NFDI-MatWerk consortium.

4.4. Source Lifecycle and Quality Assurance

The distribution is generated from human-readable YAML sources rather than maintained directly as Turtle, which keeps source files git-diffable and review-friendly. Consistency of the distribution is verified against standard tooling: HermiT for OWL-DL reasoning, SHACL validators for shape constraints, and the OOPS! pitfall scanner (Poveda-Villalón et al., 2014) for ontology-engineering hygiene. All build steps and quality checks are reproducible from the source repository, ensuring that the published distribution can be regenerated and audited independently.

4.5. Authoritative Reasoning: Neumann Engine and Phase-State Coupling

The authoritativeness principle (Section 2.4, principle 8) operationalizes the chain *phase diagram* → *active crystallographic phase* → *tensor-component symmetry* as a two-step inference. Two implementation components carry it.

Neumann tensor engine. A dedicated module `oco-tensor` carries 190 material-tensor root classes (Voigt families, polarization, conductivity, elasticity, piezo-, pyro-, magnetoelectric, ...), each annotated with tensor order, symmetry class, and the axial/polar flag. A subprocess engine implements Neumann's principle as a stabilizer-subgroup computation in `sympy` over the 32 crystallographic point groups, generating for every (tensor × PG) pair the set of independent non-zero components. The engine is validated against Nye (1985, Table 9) and IEEE Standard 176 on a 30-case gold standard (30/30 pass) and emits 5 920 reified `oco:NeumannConstraint` instances together with 768 `sh:NodeShape` validators. A hash-keyed cache keeps regeneration cheap: a full re-run after an engine fix completes in 5:30 min wall-time (cache-warm) versus 46 min for the cold first run.

Phase-state coupling. A symmetry library is only useful when the active point group can be *derived*, not assumed. Phase regions are therefore modeled as sub-classes of `oco-phase:CrystalPhase` with temperature and composition bounds expressed in the QUDT `QuantityValue` pattern, and each region carries a `has_point_group` restriction. Two reference phase diagrams are populated: BNT-BT (four regions, including the morphotropic phase boundary as a 3m+4mm coexistence region) and NiCuZn ferrite (ferrimagnetic / paramagnetic across the Curie transition). A SPARQL-`INSERT` pass over the sample ABox infers the active region from (T, x), propagates the point group, and triggers the Neumann constraints — without OWL reasoning, which cannot express numeric intervals. End-to-end inference is gold-standard tested on six cases (6/6 pass). Together, the two components answer the central authoritativeness competency question — "which piezoelectric coefficients are non-zero for BNT-BT at room temperature near the morphotropic phase boundary?" — without any manual point-group assignment by the user.

4.6. Multi-Axis Parameter Classification

Implements the facetedness principle (Section 2.4, principle 9). Every parameter class in OCO now carries three independent classification axes simultaneously, each encoded as `rdfs:subClassOf` so that an OWL 2-DL reasoner navigates them without bespoke join logic:

- **Role** (what the parameter does physically): StateVariable, ResponseParameter, TransportParameter, StructureParameter, StatisticalParameter, TopologicalParameter, FitParameter.
- **Reference** (the semantic anchor): FundamentalConstant, MaterialParameter, InterfaceParameter, SpecimenParameter, ProcessParameter, MeasurementParameter, EnvironmentParameter.
- **Layer** (abstraction level at which the parameter is defined): AtomicLayer, CrystallineLayer, MicrostructuralLayer, MesoscopicLayer, MacroscopicLayer — five classes housed in a separate module `oco-layer`. Each carries its own *state variables* (e.g. lattice parameters at the crystalline layer; effective permittivity at the mesoscopic layer) and *governing laws* (Schrödinger / DFT at the atomic layer; Maxwell-Garnett / Bruggeman at the mesoscopic layer). Layers are *not* spatial partitions but coexisting descriptive views of the same material.

A bulk classification pipeline populates all three axes for the ∼950 `MaterialParameter` subclasses through family-default plus outlier-override rules. The tensor-rank and symmetry-class annotations from `oco-tensor` (Section 4.5) provide a fourth independent hook that connects the parameter taxonomy to the Neumann engine. The resulting query granularity is illustrated by competency questions of the form "which response parameters at the crystalline layer are non-zero for the trigonal point group 3m?", which become single `rdfs:subClassOf` traversals.

4.7. Extended Interoperability and Predecessor Bridges

Realizes the adaptability, interoperability, and compatibility principles (Section 2.4, principles 2, 3, 6) through a disciplined bridge-expansion process. The discipline is captured in a *reuse-before-invention* checklist applied before any new class or property is proposed: (a) does OCO already have a fitting class? (b) external identifier? → subclass of `oco-identifier:Identifier`; (c) provenance? → PROV-O; (d) unit? → QUDT `QuantityValue`. In the most recent interoperability wave, applying this checklist reduced eighteen candidate new properties to zero and fifteen candidate classes to five.

The expanded bridge inventory now contains 829 mappings across 40 sections in the `bridge_mappings.yaml` and `mwo_mappings.yaml` single sources of truth:

- **Materials databases:** OPTIMADE, via a lookup table for the structure-search-API field vocabulary.

- **Machine-learning datasets:** Croissant (MLCommons) for ML-ready dataset metadata.
- **Electronic lab notebooks:** the ELN-Filetype standard, a cross-vendor RO-Crate-based export format adopted by 14 ELN/LIMS systems — bridging to the format covers the entire set without per-vendor mappings.
- **EMMO sub-modules:** ISQ (the ISO/IEC 80000 quantity vocabulary, 93 mappings), chemistry, and materials, each in its own bridge section so that sub-module versioning can be tracked independently.
- **Predecessor PMD ontologies:** KnowNow (Ben Hassine and Stark, 2024) contributes 14 process- and property-class mappings (LTCC-multilayer-specific classes are deliberately excluded, as they belong to a material-specific route rather than the agnostic **L1**); SmaDi (Maas et al., 2024) contributes 15 mappings restricted to the piezoelectric-ceramic subset, with the shape-memory and elastomer concepts excluded as out of ceramic scope.

Three thin application profiles — `oco_eln_profile.ttl`, `oco_materials_db_profile.ttl`, and `oco_ml_profile.ttl` — bundle the modules each consumer class typically needs as `owl:imports` wrappers, avoiding the maintenance load that subset-extracted distributions would incur. External sources are versioned and cached locally, so that a release change in any external target breaks only the corresponding bridge file — the adaptability principle made operational.

4.8. Compliance and Value-Chain Modules

The audience axis introduced architecturally in Section 3.6 carries the value-chain and EU-regulatory vocabulary on the compliance side. The same construction-kit absorbs the vertical-convergence pressure of Section 1 — EU-driven integration of materials, manufacturing, supply-chain, and sustainability data — without restructuring the material core or the agnostic laboratory level. Figure 3 captures the architectural relationship: while the material-science audience is organised *by* level, the compliance/value-chain audience cuts *across* levels — each compliance module carries classes on several of **L1**, **L2**, and **L3** simultaneously (Table 9).

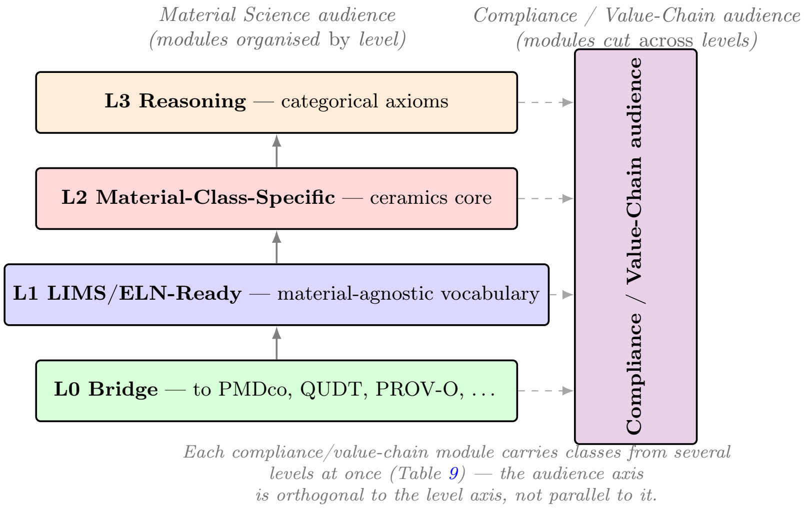

Figure 3: The compliance/value-chain audience as a cross-cutting strip independent of the **L0**–**L3** level hierarchy of the material-science core. The level hierarchy on the left (cf. Figure 2) is the organising principle for the material modules; the compliance/value-chain audience on the right is a single independent column that touches every level. Per-module level distributions are in Table 9.

Modules. Table 9 lists the fifteen modules carrying the compliance audience: eleven compliance-pure modules (LCA, CSRD, CSDDD, PPWR, CBAM, R2R, AI Act, Manufacturing-X, regulated substances, recycling, SSbD) and four dual-audience modules (`format`, `odrl`, `time-event`, `automation`) shared with the material side. Each module has its own **L1**/**L2**/**L3** internal structure — the two axes (audience and level) are independent.

Modules. The compliance-audience modules cover the EU-regulatory wave that will reshape materials-data practice over the next five years.

The joining concept: Digital Product Passports. Across the new modules, the architecturally load-bearing concept is the Digital Product Passport. The `oco-lca` module contributes *material-DPPs* (per-material passport profiles attached to specific material classes such as BNT-BT); the `oco-mfgx` module contributes *sector-DPPs* (per-industry passport profiles tied to specific EU regulations: Battery DPP, Construction-Products DPP under CPR 2024/3110, Generic-ESPR DPP). A concrete product instance inherits from *both* via multiple-inheritance: a BNT-BT actuator in a battery assembly is simultaneously a BNTBT-DPP and a Battery-DPP. This is the same

Table 9: Modules carrying the compliance audience in OCO. Eleven are compliance-pure; four (marked †) carry both the material and the compliance audience and are therefore shared between this table and the material modules. Each module has its own **L1**/**L2**/**L3** structure (column *L1/L2/L3*). The DPP column flags modules that contribute classes to the Digital Product Passport joining pattern.

Module	Cls.	L1/L2/L3	DPP	EU-regulatory anchor / role
`lca`	83	8 / 59 / 16	●	Life Cycle Assessment per ISO 14040/14044/14067, EN 15804+A2, EU PEF 2021/2279; bridges to ecoinvent, BONSAI. Material-DPP profiles attach here.
`csrd`	39	25 / 14 / –		Corporate Sustainability Reporting Directive (ESRS standards, double-materiality, reporting containers); EFRAG XBRL bridge.
`mfgx`	37	9 / 25 / 3	●	Manufacturing-X integration — AAS asset administration shells, Catena-X BPN identifiers, IPCC AR climate factors, sector-DPPs (Battery / Construction CPR / Generic ESPR).
`csddd`	25	13 / 12 / –		Corporate Sustainability Due Diligence Directive (six-stage diligence process, supplier tiers); UNGP + ILO bridges.
`ppwr`	15	7 / 8 / –		Packaging and Packaging Waste Regulation (packaging function disjointness, sub-classes per use case).
`regulation`	14	3 / 11 / –		Cross-cutting regulated-substance vocabulary (recycling-inhibitor list, REACH-style anchors).
`cbam`	10	4 / 6 / –		Carbon Border Adjustment Mechanism (covered material categories: iron/steel, cement, electricity, . . .).
`r2r`	8	4 / 4 / –		Right-to-Repair Directive (covered product categories: smartphone, tablet, large/small appliance).
`aiact`	8	4 / 4 / –		EU AI Act (risk tiers, conformity-assessment anchors).
`recycling`	6	4 / 2 / –		Recycling routes and recyclate streams.
`ssbd`	1	1 / – / –		Safe-and-Sustainable-by-Design framework (Commission Recommendation 2022/2510), skeleton.
Subtotal (compliance-pure)	**246**	86 / 141 / 19		
`format`†	42	5 / 37 / –		Industrial data-format vocabulary spanning ELN, simulation, LCA, Manufacturing-X, and document formats — referenced by many of the compliance modules and by the material pipelines.
`odrl`†	21	5 / 16 / –		W3C ODRL 2.2 usage-policy vocabulary and W3C Verifiable Credentials — the trust and policy layer that compliance reporting and value-chain exchange rest on.
`time-event`†	15	15 / – / –		W3C Time-Ontology-aligned temporal extents, instants, and events — required by reporting periods (CSRD/ESRS), validity windows, and audit-trail provenance.
`automation`†	6	1 / 5 / –		SiLA 2.0 laboratory-automation standard — shared between material instrumentation provenance and compliance traceability.
Subtotal (dual-audience)	**84**	26 / 58 / –		

multi-axis pattern used for material parameters (Section 4.6), applied across the material/regulation divide.

The supplier module as Wertschöpfungs-Akteur anchor. The `oco-supplier` module is the connecting tissue between the material side and the compliance side: every supply-chain tier (CSDDD), every BPN identifier (Manufacturing-X), every CoA reference (material side) routes through this single module. `oco-supplier` is delivered as a public stub (skeleton classes for cross-module referencing) plus a proprietary body (full catalogues, vendor specifications); see Section 6.2.

Reified-constraint pattern generalized. The reified `oco:NeumannConstraint` pattern (Section 4.5) and the reified `oco:RouteTemplate` pattern (Section 3.5) generalize to `oco:LCA_Result_Reified` in the LCA module: domain constraints that are not OWL axioms but ABox instances of named constraint classes, SPARQL-evaluable but not reasoner-evaluable. This pattern — named constraint instances as SPARQL-queryable knowledge anchors — is what makes the construction-kit absorb domains (Neumann tensor symmetry, route templates, LCA results) whose semantics escape OWL 2 DL while keeping the categorical TBox clean.

4.9. ABox Data Model: SIPOC Fragments

A central pragmatic decision: OCO does not assume that ABox data are complete experiment descriptions. Instead, data are ingested as **SIPOC fragments** (Suppliers / Inputs / Process / Outputs / Customers).

The reason is empirical: real-world experiment documentation is *never* complete. With own experiments, completeness can at least be aspired to. With published literature and patents, incompleteness is often deliberate (competitive or patent strategy). A schema that mandates complete fields would either reject most real-world data or force the fabrication of placeholder values — both unacceptable. The SIPOC granularity, combined with the three-tier identifier hierarchy (mandatory/recommended/optional), accommodates incomplete data without breaking the schema. This makes OCO practically usable for the two dominant data sources — own lab data *and* literature extraction.

4.10. Companion Software (out of scope of this paper)

Two independent software components work with OCO; they are *not the subject of this paper*:

- **OCO-Workbench** (Pannek): GUI for data import, process modeling, and ABox population. Operative interface for laboratory personnel.
- **SIPOC-Extractor** (Grond): Pipeline for SIPOC-fragment extraction from scientific literature and patents. Augments own experiments with published-bestand coverage.

Both tools use the OCO TBox as schema; the modular **L0**–**L3** granularity allows them to operate against sub-bundles rather than the full distribution. There is no publication schedule for the tools at this time — this paper focuses exclusively on the architectural pattern.

5. Validation

5.1. Coverage of the Norouzi Quality Requirements

Table 10 maps the nine quality requirements of Norouzi et al. (2024) against the OCO implementation.

REQ4, REQ6, and REQ9 are over-delivered: REQ4 by the disciplined reuse-before-invention bridge policy (Section 4.7); REQ6 by the combination of data-confidence levels, GUM uncertainty, and ADR-documented architectural decisions; REQ9 by the two independent classification axes together with the seven-tier internal organisation of **L2**, which give consumers a four-dimensional subset choice (level, audience, mechanistic depth, and per-module composition) that a one-axis modular ontology cannot match.

The availability dimension shared by REQ1, REQ2, REQ4, and REQ9 is met for the publicly distributed portion of OCO (**L0** and **L1** excluding `oco-supplier`; Section 6.2). The proprietary modules (`oco-supplier`, **L2**, **L3**) contribute to the metrics reported here but are not externally inspectable in the present release; the architectural pattern itself remains fully described in Section 3 and Section 2.4 and is independent of access to those modules.

5.2. Quality Audit

OOPS! self-audit. All twelve OOPS! pitfalls have been audited. Critical pitfalls (P19, P40, P31, P05, P29, P27) are absent. Important pitfalls (P11, P24, P30) are absent or addressed through documented architecture decisions (P11 eliminated by the CIF-Core domain-class consolidation). Two minor pitfalls (P10 class disjointness, P20 cross-source annotation style) are *explicitly accepted* with documented rationale rather than reflexively eliminated, because the alternative would compromise either reasoner performance (P10) or annotation comprehensibility for human readers (P20). Minor pitfall

Table 10: Coverage of Norouzi REQ1–REQ9 in OCO. All nine requirements are met; REQ5, REQ7, and REQ8 — which the survey identifies as branch-wide weakly addressed — were completed in the most recent development wave.

REQ	Requirement	OCO status
REQ1	Comprehensive MSE taxonomy	✓ for ceramics plus EU-regulatory stack: 5 196 classes in 44 modules
REQ2	Experimental *and* simulation data	✓ via separated `oco-investigation` and `oco-simulation` modules with PROV-O provenance plus the `oco-energy-dft` tier (Section 3.4.2) bridged to the Materials Project DFT corpus
REQ3	Accurate inter-concept relations	✓: 1 674 properties + 1 172 SHACL shapes including 7 Cross-Tier NodeShapes (Section 3.4.2); ROBOT-DL + Pellet pipeline, 0 violations
REQ4	Compliance with existing standards	✓✓: 829 bridge mappings across 40 sections to material-side and compliance-side standards plus extensive class-level `external_refs` (Table 1, Section 4.7, Section 4.8)
REQ5	Settings + outcomes + high-throughput + literature	✓: `HighThroughputInvestigation`, `CombinatorialLibrary`, and `LibrarySpot` for batch and spatially-resolved studies
REQ6	Trustworthy + verifiable quality management	✓✓: DCL 0–5, GUM uncertainty, `oco-dataquality` (17), 180 sourced references, 132 ADRs, negative-test ABoxes
REQ7	ML-querying-specific structures	✓: `oco_ml_profile.ttl` application profile, 12 Croissant bridges, `MachineLearningDataset` class
REQ8	ML-predicted-value representation	✓: `MLPredictionActivity`, `MLModel`, `has_prediction_confidence`, via PROV-O pattern (no separate `PredictedValue` class)
REQ9	Modularity beyond primary application	✓✓✓: 44 modules + 3 application profiles + two independent classification axes (**L0**–**L3** level choice + material/compliance audience) + seven-tier internal organisation of **L2** (Section 3)

P08 (naming convention) is addressed by an explicit architecture decision documenting eight deliberate lower-case exceptions for domain notation (`pH`, `pT`, `pO`$_2$, `nPropanol`, `d33`).

Reasoner consistency. The full **L0**+**L1**+**L2**+**L3** bundle — together with the Neumann-constraint sub-distribution — classifies in Pellet with zero unsatisfiable classes; ROBOT `validate-profile` reports no OWL 2 DL violations on the merged distribution.

Competency-question execution. 163 competency questions are published, each tagged with one of ten reasoning *areas* (tensor symmetry, phase state, route, lifecycle, multi-axis, interoperability, quality, crystal structure, BNT-BT, and an NCZF placeholder for the second ceramic material). 52 carry an executable SPARQL test against gold-standard ABoxes; 52/52 PASS. The remaining questions are either meta-discourse without an axiomatic answer requirement, or their test is pending the corresponding gold-standard ABox.

SHACL validation. The mandatory shape suite (count in Section 4) is validated via `pyShacl`, dominated by the machine-generated Neumann (768) and CIF-Core (270) subsets and complemented by per-domain hand-curated sets: 28 in the core `oco_shapes.ttl` (raw-material identifier sets, sintering atmosphere, process-step sample states, literature extraction, material-class identifiers), 25 for morphotropic phase-boundary coexistence regions, 7 Cross-Tier NodeShapes (Section 3.4.2) that audit the seven-tier explanation-chain annotations, plus the compliance-side shape sets for LCA, CSRD, CSDDD/PPWR, AI Act / CBAM / R2R cross-cutting, Manufacturing-X traceability, regulated substances, SSbD, and ODRL policy. Adopting the SHACL-OBDA validator of Özçep et al. (2024) for relational-database back-ends is on the roadmap.

5.3. Architecture-Pattern Validation

The architecture has been operationally validated along both axes plus the seven-tier internal organisation of **L2**:

- **Level axis** (**L0**–**L3**): both distribution variants (`oco_master.ttl` and `oco_master_full.ttl`) parse and reason within their designed reasoner profiles (OWL 2 EL for L0+L1+L2, OWL 2 DL with L3 added). Individual modules of **L1** (e.g., `oco-equipment`) can be imported without pulling the full bundle. **L0** bridge re-emissions during development have not propagated changes into **L1**–**L3** modules, demonstrating the version-localisation promise.

- **Audience axis** (material / compliance): three application profiles (`oco_eln_profile.ttl`, `oco_materials_db_profile.ttl`, `oco_ml_profile.ttl`) load only the modules each consumer class typically needs as `owl:imports` wrappers — the material-audience modules are reachable without the compliance modules, and conversely. The four dual-audience modules (`format`, `odrl`, `time-event`, `automation`) appear in both profile families without duplication.

- **Seven-tier internal organisation of L2** (the material-specific level's internal mechanistic-explanation skeleton): the BNT-BT end-to-end pilot (`examples/abox_pilots/bnt_bt_pilot.ttl`) instantiates concrete classes on every one of the seven layers with explicit cross-tier annotations; the seven Cross-Tier SHACL NodeShapes pass with zero violations on the pilot's 452-triple ABox, demonstrating that the explanation chain is operationally auditable rather than only conceptual.

5.4. Reusability — Work in Progress

A second ceramic material system, **ferritic high-performance ceramics**, is in active development as a second **L2** instance on top of the unchanged **L0**+**L1**. Strategically, this choice tests architectural reusability across structurally distinct ceramic families: BNT-BT populates the perovskite branch of the `oco-crystal` hierarchy and exercises the piezoelectric/pyroelectric subset of the coupled-effect family; ferritic ceramics populate the spinel/garnet/hexaferrite branches and exercise the magnetic subset (magnetostriction, magnetoelectric coupling, Verdet effect). **L1** remains unchanged across both; **L2** `oco-crystal`, `oco-phase`, and `oco-element` are largely reused, while `oco-material` and `oco-defect` are extended with ferrite-specific classes. This is genuine **L1**-reuse and **L2**-extension validation *within the ceramics family.*

We emphasize what this is *not*: it is not cross-domain validation. Validating that **L1** truly transfers to metallurgy, polymers, or batteries requires sister-project implementations that share **L0**+**L1** with OCO — that work has not been undertaken and is left as an explicit roadmap position (Section 7).

A parallel reusability claim is open at the mechanistic-explanation axis: the seven-tier skeleton is designed as a *generic* framework for crystalline ionic oxides (ferroelectrics, magnetics, ion conductors, high-T_c superconductors), not as a BNT-BT-specific construction. The skeleton has been concretised on BNT-BT; its claimed universality for the broader oxide class will be empirically established by the ferritic high-performance-ceramics second instance, whose magnetic phenomenology (Bloch / Néel walls, superexchange, magnetic anisotropy) exercises layers the BNT-BT pilot does not.

6. Discussion

6.1. *What the Architecture Delivers*

Compared with a flat ontology that bundles workflow, material, regulation, and reasoning concerns in a single TBox, the architecture (Section 3) — with its two independent classification axes plus the seven-tier internal organisation of **L2** — delivers five concrete benefits.

1. **Fine-grained consumer choice.** Each consumer imports only the depth they need. An ELN/LIMS integrator without material focus pays neither the ceramics-specific overhead nor the OWL-DL reasoner cost; a regulatory-reporting consumer loads the compliance audience without the materials-physics depth; a materials physicist loads **L2** on the material audience and selects the mechanistic-explanation tiers (Symmetry alone, all seven, or any subset relevant to a specific question). The independence of the two axes *plus* the explicit seven-tier organisation of **L2** permits subset choices along four dimensions (level, audience, mechanistic depth, and per-module composition) that a one-axis modular ontology cannot offer.
2. **Sister-project reuse.** A metallurgy or polymer ontology can replace **L2** while sharing **L0**+**L1**, avoiding the duplication of laboratory, equipment, and measurement vocabulary that currently happens across PMD-consortium projects. The mechanistic-explanation skeleton (Section 3.4.2) is designed to remain unchanged through this substitution — only the per-tier instance content changes.
3. **Localized version updates.** An external-standard release (PMDco v3.1, an EFRAG XBRL Taxonomy update, an updated Catena-X aspect model) requires changes only in **L0** or the relevant bridge section — the material and compliance modules remain untouched.
4. **CQ-bound reasoning scope.** **L3** makes the ontology's intended question scope auditable: a published catalogue of competency questions with executable SPARQL tests (counts in Section 4) documents what the ontology is meant to answer, separately from what it can describe. Cross-Tier SHACL constraints (Section 3.4.2) extend the same audit discipline to the explanation chain.
5. **Mechanistic-explanation depth without TBox inflation.** The external-cache pattern (Section 3.4.2) keeps the TBox compact while letting consumers query against the Materials Project DFT corpus, the IUCr bond-valence parameters, the pyxtal Wyckoff positions, the Shannon ionic radii, and the Pauling electronegativities (cache totals

in Section 4). A flat-vocabulary alternative would either embed those reference data as TBox classes (intractable for the reasoner) or leave them outside the knowledge graph altogether (loss of provenance).

6.2. Open- vs. Closed-Source Mix as Architectural Argument

A non-obvious benefit of strict level separation: the architecture *enables* a mixed open/proprietary distribution model that flat ontologies cannot offer. In the present release this is realized as follows. **L0** is released under CC-BY 4.0, mirroring the licenses of the bridge targets it anchors to (PMDco, EMMO, PROV-O, and FaBiO are CC-BY; QUDT is Apache-2.0) and maximizing bridge adoption in third-party stacks. The publicly released portion of **L1** — which excludes the `oco-supplier` module and the L1 portions of the compliance and value-chain modules listed in Table 9 (`oco-csrd`, `oco-mfgx`, `oco-lca`, `oco-csddd`, `oco-ppwr`, `oco-cbam`, `oco-r2r`, `oco-aiact`, `oco-regulation`, `oco-recycling`, `oco-ssbd`, plus the four dual-audience modules `oco-format`, `oco-odrl`, `oco-time-event`, `oco-automation`) — is released under CC-BY-SA 4.0, with dual-licensing to CC-BY available on request for commercial integrators that cannot adopt copyleft. The `oco-supplier` module, the L1 portions of the compliance and value-chain modules, the entire **L2**, and the entire **L3** are developed under project confidentiality and remain proprietary in the present release. The same modular boundaries that make this split clean would, in a future configuration, also enable a per-module choice: a sister project working under different commercial constraints could open or close a different subset without restructuring.

This mix is structurally impossible in monolithic ontologies (everything-or-nothing open) and in purely proprietary industrial schemas (no external adoption). The architecture being modular *along two independent axes plus the seven-tier internal organisation of **L2*** is exactly what makes the licensing choice fine-grained. The quantitative metrics and validation results in Section 4 and Section 5 are measured on the full distribution; the architectural contribution of this paper — the two-axis modular pattern, the ten design principles, the seven-tier mechanistic-explanation skeleton organising **L2**, the Neumann tensor engine, the phase-state coupling, the multi-axis parameter classification, and the external-cache pattern for reference data — is independent of which particular modules a given consumer can access.

6.3. Limitations

- **Equipment module rests partly on textbook anchors.** While the equipment module now bridges to PMDco, CHMO, OBI, and schema.org, a substantial portion of its class definitions still rest on textbook rather than formal-norm anchors. A systematic norm-anchor sweep is open work.

- **Neumann engine validated on a 30-case gold standard.** The 5 920 (tensor × point-group) constraints (Section 4.5) have been checked against Nye (1985, Table 9) and IEEE Standard 176 on a 30-case sample (30/30 PASS). This covers the principal piezoelectric and elasticity cases; a larger reference set with independent textbook verification across all tensor-class × point-group combinations is open work.

- **Phase-state coupling instantiated for two materials.** The Sample → Region → point-group inference (Section 4.5) has been populated only for BNT-BT and NiCuZn ferrite (six gold-standard cases, 6/6 PASS). Extending the pattern to a third material with substantially different phase phenomenology — e.g. a martensitic or order–disorder transition — is the next coverage test.

- **Multi-axis classification is heuristic-based.** The bulk-classification pipeline (Section 4.6) reaches ∼99 % coverage via family-default plus outlier-override rules; the residual is conservatively over-approximated as `ResponseParameter`. Targeted manual review is open work.

- **L1 cross-domain transferability is architecturally argued, not empirically proven.** See Section 5.4; only within-ceramics-family **L2** substitutability has been tested so far.

- **Seven-tier skeleton universality not yet empirically validated across the oxide class.** The mechanistic-explanation skeleton (Section 3.4.2) is designed as generic for crystalline ionic oxides and has been concretised on BNT-BT. The ferritic high-performance-ceramics second instance (Section 5.4) is the first test of whether tiers such as Magnetic-Microstructure (Bloch / Néel walls) and Magnetic-Bonding (superexchange, Jahn–Teller distortion) can be added as Tier-5 / Tier-7 extensions without restructuring the skeleton itself; that test is in progress.

- **External-cache provenance carries a maintenance contract.** The five reference-data caches (Section 3.4.2) are pinned to upstream versions via SHA. Upstream releases of the Materials Project, IUCr `bvparm`, Shannon compilations, pyxtal, and Pauling revisions will require periodic re-sync and re-validation. This is operational overhead the cache pattern explicitly takes on in exchange for keeping the TBox compact.

6.4. Scope Boundary and Roadmap Position — OCO v0.94

The architecture defines what OCO covers and, with equal discipline, what it does not. Within OCO the boundary is internal: **L3**'s categorical reasoning is opt-in (Section 3.5), and even the bilingual definitional discipline (Section 3.3) is a feature consumers can ignore if they wish. The *external* boundary — between OCO and the surrounding materials-data stack — is the architecturally load-bearing one, and we state it explicitly here.

The categorical/quantitative split. Quantitative mathematics (calibration curves, scaling laws, arbitrary numerical functions) lies outside this release entirely. This is a principled separation, not an oversight. OWL 2 DL can evaluate subsumption, existence, and restriction, but it cannot evaluate a function such as $d_{33}(x) = a + b\,x$ on the interval $[0.04, 0.08]$ for the BNT-BT solid-solution series; SHACL can perform range checks but is equally not a function evaluator. Embedding such functions into OWL classes produces a TBox in which neither categorical nor quantitative reasoning works correctly: the reasoner cannot reason about the function, and the function cannot consult the rest of the OWL graph. Whatever takes responsibility for quantitative mathematics in the surrounding stack must therefore be a separate level with its own evaluator, coupled to OCO classes as anchors but not embedded in them.

Today's workaround, tomorrow's level. Consumers that already need quantitative computation handle this externally: a Python/SymPy/MATLAB compute level holds the equations and queries OCO classes by IRI. This works in practice and is the intended pattern for the present release. A future OCO release will formalize a quantitative-mathematics level above the present categorical levels; the architectural commitment is that the new level will respect the categorical/quantitative split rather than violating it. The categorical levels (**L0**–**L3**) will not change in the process.

The v0.94 framing. This paper describes **OCO v0.94**: the release engineered to enter productive practice. The version number is deliberate. v0.94 means: the architecture is complete along both classification axes, the categorical levels are populated, the seven-tier mechanistic-explanation skeleton organising **L2** is in place, the compliance modules cover the EU-regulatory wave currently in force, the test suite passes, and the BNT-BT end-to-end pilot demonstrates the skeleton operationally. What v0.94 does not yet have is the corrective feedback that only productive use can deliver — conflicts between intended modelling and what actual data exposes, modules that turn out to need refactoring, bridge mappings that upstream releases break, cache contracts that real consumer queries strain. The v1.0 release is reserved for the state after that feedback cycle. Until then v0.94 is the deliberate label for a complete-and-ready-for-practice release that has not yet been corrected by practice.

Other deliberate non-inclusions. For clarity, the following are also outside the scope of this release:

- **Workflow execution level.** Executable workflow orchestration (e.g. pyiron, SimStack) is platform work, not ontology work; OCO provides class anchors that such systems can target.
- **Own ML/AI classes.** OCO does not introduce a machine-learning vocabulary of its own; bridging to existing community efforts (MDO, CMSO) is the chosen path when the need arises.
- **Companion software tools.** The OCO-Workbench and SIPOC-Extractor (Section 4.10) use OCO as their schema but are outside the scope of this paper.

7. Conclusion and Outlook

We have argued that an industrial materials ontology entering productive use today must answer three simultaneous challenges: horizontal fragmentation across material domains, vertical convergence pressure driven by EU regulation, and mechanistic explanation depth that lets the ontology surface why a property holds, not only that it does. The architecture proposed here answers all three with a single integrated design — a multi-level, modular ontology with two independent classification axes (the level of abstraction, from **L0** bridges through **L3** categorical reasoning; and the consumer audience, material versus compliance, with four dual-audience modules at the join) in which the material-specific level (**L2**) is internally organised by a seven-tier mechanistic-explanation skeleton applicable to any crystalline ionic oxide. The level-and-audience modularity dissolves the horizontal fragmentation, the compliance audience absorbs the vertical regulation pressure, and the seven-tier organisation of **L2** delivers the mechanistic-explanation depth.

The OntoCrafter Ceramics Ontology (OCO v0.94) instantiates the architecture for functional ceramics with BNT-BT as the end-to-end pilot: 5 196 classes across 44 class-bearing modules; 167 348 OWL axioms (40 454 logical, per ROBOT) including 5 920 reified Neumann tensor constraints; 1 674 properties; 829 cross-ontology bridge mappings across 40 sections; 1 172 SHACL shapes including 7 Cross-Tier NodeShapes that audit the explanation chain; 163 published competency questions with 52 executable SPARQL tests (52/52 PASS); 132 architecture decision records; and an external-cache pattern that lets the compact TBox query against approximately 155 000 Materials Project DFT entries, 1 934 IUCr bond-valence parameters, 1 731 Wyckoff positions, 497 Shannon ionic radii, and 91 Pauling electronegativities without inflating the OWL vocabulary itself.

The publicly available portion of the distribution (**L0** under CC-BY 4.0, the public modules of **L1** under CC-BY-SA 4.0) is hosted at `https://w3id.org/oco/`. The supplier module, the full **L2**, and **L3** are developed under project confidentiality, as discussed in Section 6.2.

The v0.94 designation is deliberate: the release is engineered to enter productive practice. The v1.0 release is reserved for the state after that practice has fed back corrections — conflicts between intended modelling and what actual data exposes, modules that turn out to need refactoring, bridge mappings that upstream releases break, cache contracts that real consumer queries strain. The shape of the v0.94→v1.0 roadmap, stated without promises:

- **Practice-feedback cycle.** Productive use of OCO in materials-research and compliance-reporting pipelines, with structured intake of corrections through the existing ADR process. Decisions taken on the basis of that feedback are the substance of v1.0.

- **Completion of the second ceramic material instance** (ferritic high-performance ceramics), validating **L1** reusability and **L2** extension within the ceramics family (Section 5.4). The ferritic instance also stresses the seven-tier skeleton on the magnetic side (Bloch / Néel walls, superexchange, Jahn–Teller distortion) that the BNT-BT pilot does not exercise.

- **Discoverability channels.** Listing on MatPortal, IndustryPortal, and OSF, to enter the channels surveyed by Norouzi et al. (2024).

- **Open invitation to sister-domain projects** (metallurgy, polymers, batteries, pharmaceuticals) to share **L0**+**L1** with OCO and contribute their own **L2** instance — the first genuine cross-domain validation of the architecture.

The deeper claim of this paper is not that OCO is the right ontology for ceramics. It is one possible implementation of a broader idea: that the materials-science community can stop re-modelling the same workflow, equipment, measurement, compliance, and mechanistic-explanation concepts in every new domain ontology. The architecture proposed here — two independent axes plus the seven-tier internal organisation of **L2** — is one way to do that. We hope others follow.

Author Contributions (CRediT)

T. Pannek and W. Grond (both): Conceptualization (ontology architecture), Methodology (bridge stack, pipeline), Validation, Writing.

T. Pannek (additionally): OCO-Workbench (companion software for data import and ABox population).

W. Grond (additionally): Materials Science (**L2** ceramic material modeling for BNT-BT and ferritic high-performance ceramics), SIPOC-Extractor (companion software for literature extraction).

Both authors appear on all Numberland publications related to the OCO suite (ontology and two companion software tools), reflecting the close collaboration between the materials-science and architecture/implementation domains.

Acknowledgments

The work underlying this publication was funded by the Federal Ministry of Education and Research (BMBF) under grant number 13XP5228C. The responsibility for the content of this publication lies with the authors.

Data and Code Availability

The publicly available portion of the OCO distribution — **L0** (CC-BY 4.0) and **L1** excluding the `oco-supplier` module (CC-BY-SA 4.0) — is hosted at `https://w3id.org/oco/`. Dual-licensing of the public **L1** modules to CC-BY 4.0 is available on request for commercial integrators that cannot adopt copyleft. The `oco-supplier` module, **L2**, and **L3** are developed under project confidentiality and are not available under an open license in the present release; the rationale and architectural consequences are discussed in Section 6.2. Quantitative metrics and validation results in this paper are reported from internal validation against the full distribution. Academic-collaboration and commercial-license enquiries can be addressed to the corresponding author.

References

Agrawal, et al., 2024. Concrete Mixture Structural Design Optimization. Data-Centric Engineering doi:10.1017/dce.2024.18. pMD Phase-1 paper #45.

Arendt, et al., 2024. Glass Property Predictions Ab Initio. Journal of the American Ceramic Society doi:10.1111/jace.19904. pMD Phase-1 paper #37.

Arp, R., Smith, B., Spear, A.D., 2015. Building Ontologies with Basic Formal Ontology. Technical Report. MIT Press. BFO 2.0 specification, ISO/IEC 21838-2:2021.

Aschemann, et al., 2024. Smart Rubber Extrusion Sensor Techniques. Advanced Engineering Materials doi:10.1002/adem.202401316. pMD Phase-1 paper #29.

Bardehle, et al., 2025. Rubber Compounds Mathematical Models. Advanced Engineering Materials doi:10.1002/adem.202503201. pMD Phase-1 paper #01.

Batchelor, C., 2017. CHMO: Chemical methods ontology. OBO Foundry. http://purl.obolibrary.org/obo/chmo.owl.

Bayerlein, et al., 2021. Perspective Digital Knowledge Representation MSE. Advanced Engineering Materials doi:10.1002/adem.202101176. pMD Phase-1 paper #34.

Bayerlein, et al., 2024a. Semantic Integration Orowan Strengthening. Scientific Data doi:10.1038/s41597-024-03169-4. pMD Phase-1 paper #41.

Bayerlein, et al., 2024b. Semantically Accessible Materials Data Space. Advanced Engineering Materials doi:10.1002/adem.202401092. pMD Phase-1 paper #11.

Bayerlein, et al., 2024c. Semantically Accessible Materials Data Space 2024. Advanced Engineering Materials doi:10.1002/adem.202401092. pMD Phase-1 paper #38.

Bayerlein, B., et al., 2024d. PMD core ontology: Achieving semantic interoperability in materials science. Materials & Design doi:10.1016/j.matdes.2023.112603.

Bekemeier, et al., 2025a. Magnetocaloric Materials Digitalization. doi:2511.19053. pMD Phase-1 paper #02.

Bekemeier, S., Caldeira Rêgo, C.R., Mai, H.L., Saikia, U., Waseda, O., Apel, M., Bayerlein, B., et al., 2025b. Advancing digital transformation in material science: The role of workflows within the MaterialDigital initiative. Advanced Engineering Materials 27. doi:10.1002/adem.202402149.

Ben Hassine, S., Stark, R., 2024. Ontology based knowledge system for ceramic multi-layer components, in: Joint Proceedings of RCIS 2024 Workshops and Research Projects Track, Guimarães, Portugal. URL: https://ceur-ws.org/Vol-3674/RP-paper4.pdf.

Beran, et al., 2024. Structured Data Domain Knowledge Battery. Advanced Engineering Materials doi:10.1002/adem.202401813. pMD Phase-1 paper #31.

Beygi Nasrabadi, A., et al., 2025. NFDI MatWerk ontology (MWO). GitHub repository. URL: http://purls.helmholtz-metadaten.de/mwo.

BeygiNasrabadi, et al., 2023. Digital Materials Mechanical Testing Lab. Computers in Industry doi:10.1016/j.compind.2023.104016. pMD Phase-1 paper #48.

BeygiNasrabadi, et al., 2024a. Digital Representation Tensile Stress Relaxation. Key Engineering Materials doi:10.4028/p-xSmHN2. pMD Phase-1 paper #39.

BeygiNasrabadi, et al., 2024b. Mechanical Testing Dataset Cast Copper. Data in Brief doi:10.1016/j.dib.2024.110687. pMD Phase-1 paper #40.

BeygiNasrabadi, et al., 2024c. Semantic Low-Cycle Fatigue Testing Data. Advanced Engineering Materials doi:10.1002/adem.202400675. pMD Phase-1 paper #22.

BeygiNasrabadi, et al., 2024d. Upper-Level Ontologies Evaluation MatSci. Advanced Engineering Materials doi:10.1002/adem.202401534. pMD Phase-1 paper #13.

BeygiNasrabadi, et al., 2024e. Vickers Hardness Knowledge Graph. Key Engineering Materials doi:10.4028/p-k8Gj2L. pMD Phase-1 paper #42.

BeygiNasrabadi, et al., 2025. NFDI MatWerk Ontology MWO BFO. Advanced Engineering Materials doi:10.1002/adem.202502331. pMD Phase-1 paper #08.

Bjarsch, et al., 2024. Ontologies FAIR Data Additive Manufacturing. Advanced Engineering Materials doi:10.1002/adem.202401528. pMD Phase-1 paper #24.

Blum, et al., 2023. OTTR-centric Ontology Engineering Methodology. doi:2309.13130. pMD Phase-1 paper #47.

Chakraborty, et al., 2024. Steel Recycling Phase Transformation MnSteel. Steel Research International doi:10.1002/srin.202400841. pMD Phase-1 paper #05.

Chen, et al., 2024a. Ontology Digital Infrastructure Glass GlasDigital. Advanced Engineering Materials doi:10.1002/adem.202401560. pMD Phase-1 paper #27.

Chen, et al., 2024b. Ontopanel Visual Ontology Development Tool. Integrating Materials and Manufacturing Innovation doi:10.1007/s40192-022-00279-y. pMD Phase-1 paper #52.

Del Nostro, P., Goldbeck, G., Toti, D., 2022. CHAMEO: An ontology for the harmonisation of materials characterisation methodologies. Applied Ontology doi:10.3233/AO-220262.

eCl@ss e.V., 2024. eCl@ss classification and product description standard. eCl@ss e.V., Cologne. https://www.eclass.eu/.

Eisenbart, et al., 2024. KupferDigital Ontology Copper Life Cycle. Advanced Engineering Materials doi:10.1002/adem.202401735. pMD Phase-1 paper #20.

ETIM International, 2024. ETIM classification standard for the electrotechnical sector. ETIM International, Brussels. https://www.etim-international.com/.

Fliegener, et al., 2024. Digital Methods Fatigue Assessment Steels. Advanced Engineering Materials doi:10.1002/adem.202400992. pMD Phase-1 paper #18.

Garabedian, et al., 2024. FAIR Research Data Experimental Tribology. Scientific Data doi:10.1038/s41597-022-01429-9. pMD Phase-1 paper #54.

Ghedini, E., et al., 2019. EMMO: An ontology for applied physics and materials modelling. JOM doi:10.1007/s11837-019-03613-5.

Gogula, et al., 2024. Optical Castability Glass Melting. Glass Europe doi:10.52825/glass-europe.v2i.1359. pMD Phase-1 paper #35.

Gramlich, et al., 2024. Recycling Copper Contamination 42CrMo4. Steel Research International doi:10.1002/srin.202400844. pMD Phase-1 paper #04.

Guizzardi, G., 2005. Ontological Foundations for Structural Conceptual Models. Ph.D. thesis. University of Twente. Introduces UFO/UFO-A.

Gumbsch, et al., 2025. Guest Editorial Digitalization MatSci. Advanced Engineering Materials doi:10.1002/adem.202500484. pMD Phase-1 paper #10.

Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P., Steinbeck, C., 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Research 44, D1214–D1219. doi:10.1093/nar/gkv1031.

Hodgson, R., Keller, P.J., Markovic, J., 2018. QUDT: Quantities, units, dimensions and data types ontologies. W3C Submission. URL: https://qudt.org.

International Organization for Standardization, 2020. ISO 15926-14:2020 industrial automation systems and integration — industrial plant lifecycle information — part 14: Data model specification in web ontology language. ISO Standard. https://www.iso.org/standard/75949.html.

International Organization for Standardization, 2022. ISO 26324:2022 information and documentation — digital object identifier system. ISO Standard. https://www.iso.org/standard/81599.html.

Jackson, R.C., Balhoff, J.P., Douglass, E., Harris, N.L., Mungall, C.J., Overton, J.A., 2019. ROBOT: A tool for automating ontology workflows. BMC Bioinformatics 20. doi:10.1186/s12859-019-3002-3. source: https://github.com/ontodev/robot.

Jalali, M., Mail, M., Aversa, R., Kübel, C., 2023. MSLE: An ontology for materials science laboratory equipment. arXiv preprint arXiv:2308.07325 URL: https://arxiv.org/abs/2308.07325.

Kröger, F.A., Vink, H.J., 1956. Relations between the concentrations of imperfections in crystalline solids, in: Solid State Physics. Academic Press. volume 3, pp. 307–435. doi:10.1016/S0081-1947(08)60135-6.

Lebo, T., Sahoo, S., McGuinness, D., 2013. PROV-O: The PROV ontology. W3C Recommendation. URL: `https://www.w3.org/TR/prov-o/`.

Leipner, et al., 2025. Ultrasonic Transducers S-Parameters. Mechanical Systems and Signal Processing doi:10.1016/j.ymssp.2025.113172. pMD Phase-1 paper #09.

Lizarazu, et al., 2024. Additively Manufactured Mild Steel Monotonic. Progress in Additive Manufacturing doi:10.1007/s40964-020-00111-z. pMD Phase-1 paper #56.

Luger, et al., 2024. Ontology Digital Twin FRP Structures. Advanced Engineering Materials doi:10.1002/adem.202401437. pMD Phase-1 paper #28.

Maas, J., Leemhuis, M., Mertens, J., Schmidtke, H., Courant, R., Dahlmann, M., Stark, S., Böhm, A., Pagel, K., Hinze, M., Pinkal, D., Wegener, M., Wagner, M.F.X., Sattel, T., Neubert, H., Özçep, Ö.L., 2024. A novel digitalization approach for smart materials: Ontology-based access to data and models. Advanced Engineering Materials 27. doi:10.1002/adem.202302208.

Meng, et al., 2024. Wissensbasierte Digitalisierung Beton. ce/papers doi:10.1002/cepa.2955. pMD Phase-1 paper #44.

Mertens, et al., 2024. Digitalized Data Access DE Models OBDMA. doi:10.1117/12.2661222. pMD Phase-1 paper #51.

Mieller, et al., 2024a. Automatized Simulation Powder Pressing SimStack. Advanced Engineering Materials doi:10.1002/adem.202400872. pMD Phase-1 paper #26.

Mieller, B., Ben Hassine, S., Töpfer, J., Priese, C., Bochmann, A., Capraro, B., Stark, S., Partsch, U., Fresemann, C., 2024b. Ontology-based data acquisition, refinement, and utilization in the development of a multilayer ferrite inductor. Advanced Engineering Materials 26. doi:10.1002/adem.202401042.

Mutz, et al., 2022. Li-Ion Battery Ontology Production Structure. Energy Technology doi:10.1002/ente.202200681. pMD Phase-1 paper #53.

Nahshon, et al., 2024. Semantic Orchestration Dataspace Steel Copper. Advanced Engineering Materials doi:10.1002/adem.202401448. pMD Phase-1 paper #23.

NAMUR Working Group, 2018. NAMUR NE 100: Use of lists of properties in process control engineering workflows. NAMUR e.V., Leverkusen. https://www.namur.net/.

Nebel, et al., 2024. Ontology Battery Production DigiBatMat. Advanced Engineering Materials doi:10.1002/adem.202401540. pMD Phase-1 paper #30.

Nerella, et al., 2024. Automated Phase-Field Simulations Bainite. Advanced Engineering Materials doi:10.1002/adem.202400905. pMD Phase-1 paper #19.

Newnham, R.E., Skinner, D.P., Cross, L.E., 1978. Connectivity and piezoelectric-pyroelectric composites. Materials Research Bulletin 13, 525–536. doi:10.1016/0025-5408(78)90161-7.

NFDI Section ELSA, NFDIcore Working Group, 2024. NFDIcore: A cross-domain mid-level ontology for the NFDI. NFDI. https://nfdi.fiz-karlsruhe.de/ontology/.

Norouzi, E., Waitelonis, J., Sack, H., 2024. The landscape of ontologies in materials science and engineering: A survey and evaluation. arXiv preprint arXiv:2408.06034 URL: https://arxiv.org/abs/2408.06034. workshop SeMatS at SEMANTiCS 2024, Amsterdam.

Nye, J.F., 1985. Physical Properties of Crystals: Their Representation by Tensors and Matrices. Reprint ed., Oxford University Press.

Özçep, Ö.L., Leemhuis, M., Winkler, M., Courant, R., Sattel, T., Maas, J., 2024. Modeling and accessing smart materials with integrity constraints in the shapes constraint language and ontologies — the SmaDi way. Advanced Engineering Materials 27. doi:10.1002/adem.202401017.

Pan, et al., 2023. Genome Mining Glass Chemistry Ion Conductivity. Advanced Science doi:10.1002/advs.202301435. pMD Phase-1 paper #36.

Peroni, S., Shotton, D., 2012. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Journal of Web Semantics 17, 33–43. doi:10.1016/j.websem.2012.08.001.

Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A., 2014. OOPS! (Ontology Pitfall Scanner!): An on-line tool for ontology evaluation. International Journal on Semantic Web and Information Systems 10, 7–34. doi:10.4018/ijswis.2014040102. service: https://oops.linkeddata.es/.

Raabe, et al., 2022. Circular Steel Decarbonization Upcycling Scrap. Annual Review of Materials Research doi:10.1146/annurev-matsci-080222-123648. pMD Phase-1 paper #03.

Rajamohan, B.P., Harding Bradley, A.C., Tran, V.D., Gordon, J.E., Caldwell, H.W., Mehdi, R., Ponon, G., Tran, Q.D., Dernek, O., Kaltenbaugh, J., Pierce, B.G., Wieser, R., Yue, W., Lin, K., Kambo, J., Lopez, C., Nihar, A., Savage, D.J., Brown, D.W., Sharma, H., Giera, B., Tripathi, P.K., Wu, Y., Li, M., Davis, K.O., Bruckman, L.S., Barcelos, E.I., French, R.H., 2025. MDS-Onto: Materials data science ontology. Scientific Data 12. doi:10.1038/s41597-025-04938-5.

Rego, et al., 2022. SimStack Intuitive Workflow Framework. Frontiers in Materials doi:10.3389/fmats.2022.877597. pMD Phase-1 paper #55.

Rijgersberg, H., van Assem, M., Top, J., 2013. Ontology of units of measure and related concepts. Semantic Web 4, 3–13. doi:10.3233/SW-2012-0069.

Rogge, et al., 2024. Jahrestagung DAfStb Forschungskolloquium. doi:10.26272/opus4-61338. pMD Phase-1 paper #57.

Roters, et al., 2024. StahlDigital Ontology Workflows Steel. Advanced Engineering Materials doi:10.1002/adem.202402148. pMD Phase-1 paper #17.

Sajjad, et al., 2024. Environmental Impact Copper-Alloy Semantic. Advanced Engineering Materials doi:10.1002/adem.202401702. pMD Phase-1 paper #21.

Schilling, et al., 2024a. FAIR Structured Data Tensile Testing Ontology. Advanced Engineering Materials doi:10.1002/adem.202400138. pMD Phase-1 paper #14.

Schilling, et al., 2024b. Seamless Science Mechanical Testing Lab Data. Advanced Engineering Materials doi:10.1002/adem.202401527. pMD Phase-1 paper #16.

Schilling, et al., 2026. FAIR Data Polymers OntoFNCT. Results in Engineering doi:10.1016/j.rineng.2026.110549. pMD Phase-1 paper #07.

Schmidtke, et al., 2023. Intelligent Environments Smart Materials. doi:10.1109/IE57519.2023.10179095. pMD Phase-1 paper #49.

Shoghi, et al., 2024. Workflow-Centric FAIR Data Microstructure. Advanced Engineering Materials doi:10.1002/adem.202401876. pMD Phase-1 paper #15.

Tomkovic, et al., 2026. DOE Modeling Species Transport GaN HVPE. Advanced Engineering Materials doi:10.1002/adem.202600007. pMD Phase-1 paper #06.

Unger, et al., 2023. Data Provenance Experimental Simulation Models. Materials Today: Proceedings doi:10.1016/j.matpr.2023.08.081. pMD Phase-1 paper #46.

Valdestilhas, et al., 2023. Semantic Web Materials Science Intersection. Advanced Intelligent Systems doi:10.1002/aisy.202300051. pMD Phase-1 paper #50.

Vrandečić, D., Krötzsch, M., 2014. Wikidata: A free collaborative knowledgebase. Communications of the ACM 57, 78–85. doi:10.1145/2629489.

Zhang, Y., Chen, F., Liu, Z., Ju, Y., Cui, D., Zhu, J., Jiang, X., Guo, X., He, J., Zhang, L., Zhang, X., Su, Y., 2024. A materials terminology knowledge graph automatically constructed from text corpus. Scientific Data 11. doi:10.1038/s41597-024-03448-0.