

Table of Contents

Evidence-Weighted Software Architecture for Integrative Natural Products and Biomarker Research

Design Principles, Scoring Semantics, Regulatory Positioning, and an Evaluation Protocol

Thomas Schüller · Syntagma Forschung, Vienna, Austria forschung@tschueller.com · ORCID 0009-0003-9799-6747 Independent researcher. Privately funded, no institutional affiliation, no conflicts of interest.

Version 0.3 — preprint — 2026-09-04 · License CC BY 4.0

Abstract

This paper describes the design of a research platform (BioLogic × FBS) that joins natural products research with individual biomarker data in one evidence-weighted knowledge model, and it states the conditions under which that design can be evaluated or refuted. The system ingests structured records from open scientific sources (PubMed, ChEMBL, UniProt, LOTUS, ClinicalTrials.gov, FDA FAERS, Reactome, Open Targets and others), normalises chemical identity via InChI/SMILES, codes biomarker readings via LOINC, and links entities in a property graph with four primary node types and nine relationship types. Every relationship carries an evidence record: a class-derived base weight, multiplicative quality modifiers, a derived uncertainty level, and provenance.

Version 0.3 revises three positions taken in v0.1. First, the scoring framework is no longer described as an adaptation of GRADE: GRADE states that it is not a quantitative system and rates a body of evidence per outcome, whereas this system produces a numeric weight per record. The framework is therefore restated as GRADE-informed and explicitly non-conformant, with the differences enumerated. Second, the paper adds a preregisterable calibration protocol with falsification criteria, which converts the largest limitation of v0.1 from an admission into a testable plan. Third, the paper adds a regulatory positioning section covering MDR/IVDR software qualification under MDCG 2019-11 Rev.1, the applicability status of Regulation (EU) 2024/1689 (AI Act) after the 2026 amendment, GDPR Article 9, and the European Health Data Space timetable, each with the explicit trigger that would change the classification.

The system remains a research instrument. It produces associations with uncertainty labels, not diagnoses, treatment recommendations, or causal claims. Twelve limitations are documented. No empirical calibration has yet been performed; all scores in the current version are to be read as ordering heuristics, not as probabilities.

Keywords: evidence weighting, natural products, biomarkers, knowledge graph, provenance, reproducibility, software qualification, research software

1. Introduction

1.1 Problem

Integrative natural products research requires simultaneous handling of heterogeneous record types: chemical structures, protein targets, pathways, preclinical and clinical studies, adverse event reports, individual biomarker readings, and legal-regulatory context. Open databases cover each of these domains separately and well. What is missing is not data but a defensible way of holding records of unequal evidential standing in one queryable structure without silently levelling them.

Two failure modes motivate the design. The first is *flattening*: an in-vitro IC50 and a randomised controlled trial become adjacent rows in the same table and are read as equivalent. The second is *transfer*: a property demonstrated for a standardised extract is carried over to an isolated compound, or an effect shown in a rodent model is spoken of as a human effect. Both failure modes are structural, not accidental — they follow from data models that store the claim and drop the conditions under which the claim was produced.

1.2 Contribution

1. A domain model spanning natural products chemistry, targets, pathways and individual biomarker readings, with substance type as a first-class, non-transferable attribute (extract / isolated compound / derivative / theoretical construct).
2. An evidence weighting framework with deterministic, auditable, versioned scores per relationship, together with an explicit statement of what the numbers do *not* mean (§4.1, §4.7).
3. A four-stage quality gate with quarantine instead of deletion, so that rejected records remain inspectable.
4. A calibration protocol with predefined success and failure criteria (§5), which is the part of the work that can actually refute the scoring design.
5. A regulatory positioning with named trigger conditions for each classification (§7), replacing the disclaimer-only approach of v0.1.
6. An evaluation and beta protocol with data minimisation and stopping rules (§8).
7. Twelve documented limitations (§9).

1.3 What changed from v0.1

Area	v0.1 (2026-05-14)	v0.3 (2026-09-04)
Scoring framing	“GRADE-adapted quantitative framework”	GRADE-informed, explicitly non-conformant; differences listed (§4.1)
Unit of analysis	not stated	stated: record-level, not body-of-evidence (§4.2)

Area	v0.1 (2026-05-14)	v0.3 (2026-09-04)
Calibration	limitation L1, no plan	preregisterable protocol with falsification criteria (§5)
Regulatory	one disclaimer sentence	qualification analysis with trigger table (§7)
Evaluation	51 unit tests	tests plus external evaluation/beta protocol (§8)
Limitations	6	12
References	12	28, verified 2026-09-04
Languages	EN only	EN + DE parity

1.4 Scope and ethical boundary

The system supports research, hypothesis generation and knowledge exploration. It does not diagnose, does not recommend treatment, and does not assert causation. Every output carries an uncertainty label. Military, weapons and defence applications are excluded across the whole project, without exception and independent of funding.

2. Related work and delimitation

The evidence hierarchy is informed by GRADE [1] and by the Cochrane risk-of-bias framework [2, 3]. Chemical identity follows InChI [4] and SMILES [5]; laboratory observations follow LOINC [6]; provenance follows W3C PROV-DM [7]; data handling follows the FAIR principles [8].

In the natural products domain, LOTUS [9] provides open compound–organism links, ChEMBL [10, 11] curated bioactivity data, Reactome [12] pathway knowledge, Open Targets [13] aggregated target–disease associations with its own internal scoring, and UniProt [14] protein reference data. Pharmacovigilance signals come from FDA FAERS, whose known reporting biases are handled as a modifier, not as evidence of effect [15].

Delimitation. The distinguishing property of this system is not coverage — every source used is public — but that a single weighting and uncertainty semantics is applied across population-level and individual-level records, and that the semantics is written down in a form that can be attacked.

Property	ChEMBL	Open Targets	Reactome	LOTUS	FHIR/LIS systems	This system
Curated bioactivity	yes	via ChEMBL	no	partial	no	imported
Target–disease scoring	no	yes (own model)	no	no	no	yes (own model)

Property	ChEMBL	Open Targets	Reactome	LOTUS	FHIR/LIS systems	This system
Natural-product provenance	flags	no	no	yes	no	yes
Individual biomarker readings	no	no	no	no	yes	yes
Uniform uncertainty label per record	no	partial	no	no	no	yes
Substance type non-transferability enforced	no	no	no	partial	n/a	yes
Scores empirically calibrated	n/a	partially	n/a	n/a	n/a	no (\$5)

The last row is the honest one. Open Targets weights evidence by data type in a documented model [13]; this system does the same in a different domain but has not yet shown that its weights track anything.

3. System architecture

3.1 Sources

Fifty-five sources across ten domains are documented in the accompanying source register, each with access method, licence, update frequency, and a quality rating A (peer-reviewed or standard), B (curated, established), C (community or aggregator), D (deprecated or restricted). The register records for each source whether it was verified directly in the current review round or carried over from v0.1; carried-over entries are marked and are not counted as verified.

3.2 Domain model

Twenty-four entity types, in two groups.

Natural products / discovery side (FBS, 12): Compound (InChIKey-indexed), Organism (NCBI Taxonomy), Target (UniProt-indexed), Pathway, Assay, Study, Observation, AdverseEvent, Patent, Quarantine, IngestionLog, and evidence metadata carried on every relationship.

Individual side (BioLogic, 12): User, UserProfile, LabReport, BiomarkerReading (LOINC-coded), BiomarkerTimeseries, Hypothesis, HypothesisEvidence, LabPartner, Sample, GeneratedReport, DataProvenance, AuditLog.

Substance type is an attribute of Compound and of every relationship originating from it. Rule C3 (§4.6) forbids propagation of extract-derived evidence to isolated compounds; the rule is enforced at write time, not at display time.

3.3 Knowledge graph

Four primary node types (Compound, Target, Pathway, Biomarker/Analyte) plus supporting types (System, Nutrient, Gene, Mechanism, Publication). Nine relationship types. Every edge carries: `evidence_class`, `base_score`, `applied_modifiers[]`, `final_score`, `uncertainty`, `source_reference`, `config_version`, `provenance_timestamp`. An edge without a complete evidence record cannot be written; the ingest layer rejects it into quarantine.

3.4 Technology and reproducibility

FastAPI (Python 3.12+), PostgreSQL 16, Neo4j 5.x, Celery/Redis, React/TypeScript, RDKit, MeiliSearch, containerised via Docker Compose. Scoring configuration is a versioned file; the version is stamped on every computed score, so any score can be recomputed and compared against the configuration under which it was produced.

3.5 Module boundaries

Module boundaries are drawn along the regulatory fault line described in §7, following the modular assessment logic of MDCG 2019-11 Rev.1 [16]:

Module	Function	Regulatory character (§7)
M1 Ingest & quality gate	fetch, validate, quarantine	infrastructure
M2 Knowledge graph	storage, traversal	infrastructure
M3 Evidence weighting	score, uncertainty, conflicts	research method
M4 Lab import	parse laboratory reports, LOINC coding	personal data (GDPR Art. 9)
M5 Biomarker context	display reading against reference range	boundary module
M6 Hypothesis workspace	user-side annotation and linking	research method
M7 Export & rights	Art. 15/17/20 operations, audit log	personal data

M5 is the module that decides the regulatory character of the whole system, which is why it is isolated and why its intended purpose is written down separately (§7.2).

Abb. 1 — Systemarchitektur und Modulgrenzen / System architecture and module boundaries

BioLogic x FBS v0.2 — FORSCH-PROV0427 — CC BY 4.0

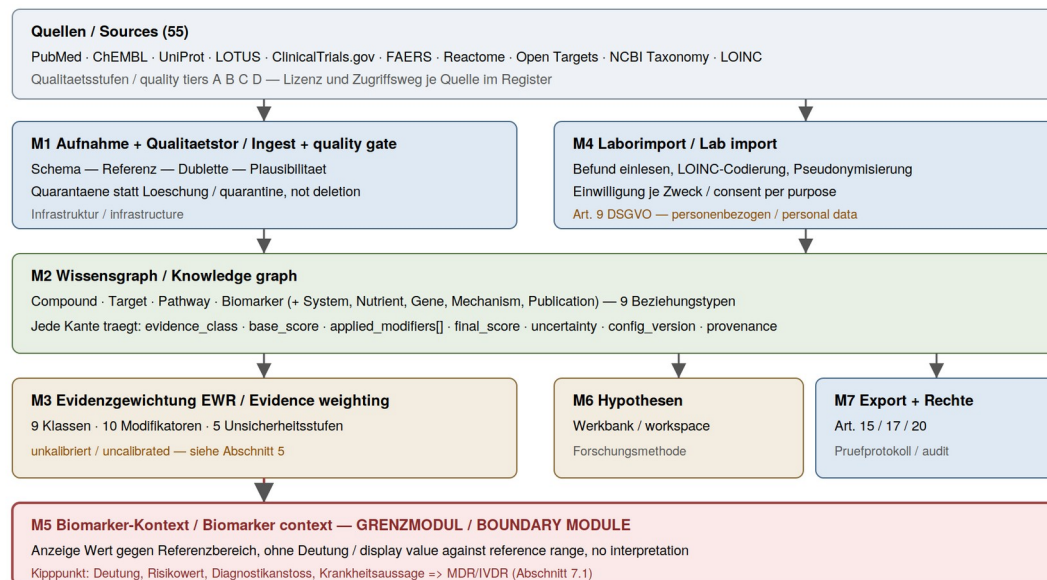


Figure 1 — System architecture with module boundaries (M1–M7). M5 is the boundary module of §7.1.

4. Evidence weighting

4.1 Relation to GRADE — and the correction to v0.1

Version 0.1 described the framework as “a GRADE-adapted quantitative framework”. That description is withdrawn. The GRADE handbook states that GRADE is not a quantitative system for grading the quality of evidence, and that the factors for up- and downgrading reflect continua rather than discrete categories requiring judgement [1, 17]. Three differences follow and are stated here rather than in a footnote:

1. **Unit.** GRADE rates the certainty of a body of evidence *for a specific outcome*. This system attaches a weight to a *single record or relationship*. These are different objects; a high weight here says nothing about the certainty of a body of evidence.
2. **Arithmetic.** GRADE downgrading is a structured judgement, not a product of coefficients. The multiplicative composition used here (§4.4) is an engineering convenience with the properties and defects listed in §4.7.
3. **Recommendation strength.** GRADE separates certainty of evidence from strength of recommendation. This system issues no recommendations at all, so the second half of GRADE has no counterpart and must not be inferred.

The framework is therefore named **EWR (Evidence Weighting for Records)** and described as GRADE-informed. Where GRADE terminology appears in the interface, it appears as a mapping hint, labelled as such.

4.2 Unit of analysis

The scored object is the tuple (source record, asserted relationship, extraction event). Aggregation to a claim level — “what is the state of evidence for compound X on target Y” — is a separate operation (§4.6, gap analysis) and deliberately does not reduce to the mean of edge scores, because edges are not independent samples: multiple papers from one laboratory, or one paper indexed in three databases, would otherwise inflate the aggregate.

4.3 Evidence classes

Rank	Evidence class	Base weight	GRADE hint (non-binding)
1	Systematic review / meta-analysis	1.00	high
2	Randomised controlled trial	0.90	high
3	Controlled clinical trial	0.80	moderate
4	Observational study	0.70	low
5	Preclinical in vivo	0.50	very low
6	Preclinical in vitro	0.35	very low
7	In silico	0.15	n/a
8	Theoretical	0.05	n/a
9	Traditional use, unvalidated	0.02	n/a

The numeric spacing is an ordering device. It encodes that a meta-analysis outranks an in-vitro result; it does not claim that a meta-analysis is 2.86 times “better” than an in-vitro result. §5 tests whether the ordering survives contact with expert judgement.

4.4 Composition

Modifier	Factor	Rationale
Retracted	0.00	absorbing state; record retained, weight zeroed
Risk of bias high	0.50	RoB 2 / ROBINS-I domain judgement [2, 3]
Risk of bias unclear	0.75	insufficient reporting
Small sample	0.60	below a class-specific threshold recorded in config

Modifier	Factor	Rationale
No blinding	0.70	applies where blinding is feasible
Single laboratory	0.70	no independent site
Preprint	0.50	no completed peer review
Conflict of interest	0.75	funder or author interest declared or detected
Standardised protocol	1.10	named, published protocol
Independent replication	1.30	separate group, comparable design

4.5 Uncertainty

Uncertainty is derived, never entered by hand:

- **none** — rank ≤ 2 and final ≥ 0.7 and replicated and no conflict
- **low** — rank ≤ 4 and final ≥ 0.5 and no conflict
- **moderate** — rank ≤ 6 and final ≥ 0.3
- **high** — rank ≤ 7 or conflict present
- **critical** — rank > 7 or retracted or no substantive evidence

The label, not the number, is what surfaces first in the interface. The number is available on demand with its full derivation.

4.6 Conflict handling

C1 Higher-ranked evidence outranks lower-ranked; the lower record is retained and remains visible. C2 Same-rank contradictions are flagged, both records retained, manual review requested. No automatic winner. C3 Extract evidence does not transfer to isolated compounds, and the reverse transfer is equally blocked. C4 Animal data does not constitute human evidence; the species field is part of the edge key. C5 In-silico predictions are hypothesis generators only and can never raise an aggregate above moderate uncertainty.

Gap analysis inverts the same rules: where the highest available rank for a claim is 6 or worse, the claim is reported as an evidence gap rather than as a weak finding, which is a different statement.

4.7 Formal properties and known defects

Stating these is part of the contribution, because they bound what \$5 can establish.

- **Monotonicity holds within a class.** Adding a downgrading modifier never increases the weight.
- **Cross-class monotonicity does not hold.** An RCT with three downgrades ($0.90 \times 0.5 \times 0.6 \times 0.7 = 0.189$) falls below an in-vitro result with an upgrade ($0.35 \times 1.1 = 0.385$). Whether that inversion is desirable is an empirical question, and it is the first thing \$5.4 tests.

- **Multiplicative composition assumes independence between modifiers.** Small sample and no blinding are not independent in practice; the product therefore over-penalises. A additive-in-log-space penalty model with capped total downgrade is the pre-specified alternative in §5.5.
- **Upgrade saturation.** Two upgrades ($1.10 \times 1.30 = 1.43$) can only be realised for base ≤ 0.70 , because of the clamp. Upgrades are therefore structurally weaker than downgrades. This is intentional but untested.
- **Non-identifiability.** Several modifier combinations map to the same final weight. The stored `applied_modifiers[]` array, not the scalar, is the audit object.
- **No propagation of uncertainty along paths.** A two-hop traversal (compound $\rightarrow$ target $\rightarrow$ pathway) currently reports the minimum edge weight, which is conservative but crude. Path-level uncertainty is deferred to v2 and listed as L8.

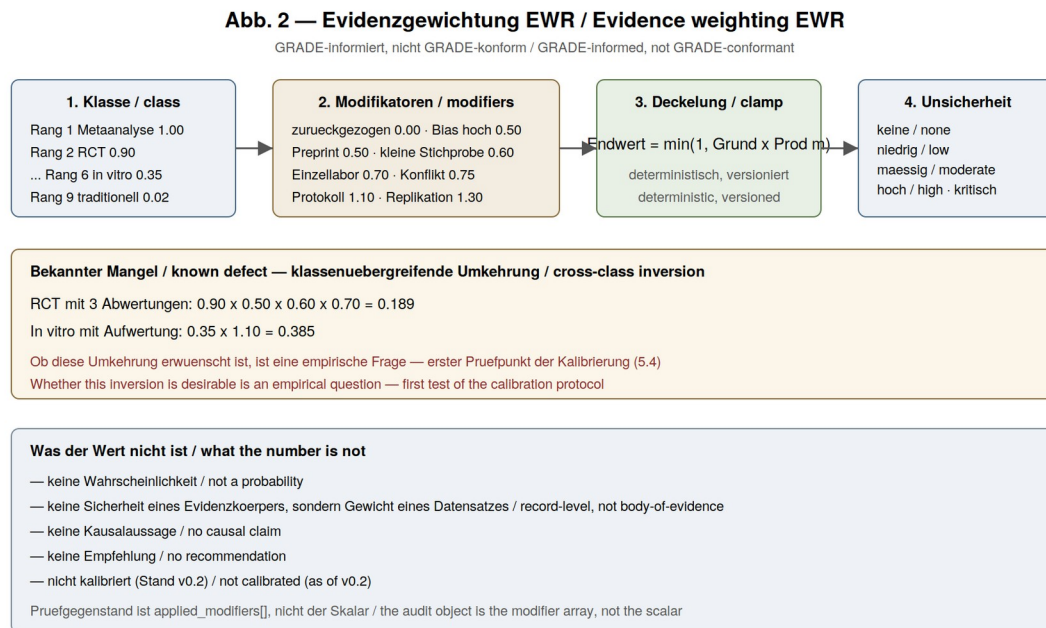


Figure 2 — Evidence weighting EWR, including the cross-class inversion of §4.7 and what the number is not.

5. Calibration protocol

This section is new in v0.3 and is the part that can refute the design. It is written so that it can be preregistered before data collection.

5.1 Object

Test whether EWR weights order evidence in the same way as trained human raters do, and whether the modifier structure carries information beyond the class rank alone.

5.2 Reference set

A stratified set of 300 relationship records drawn from the ingested corpus: 30 per evidence class, sampled at random within class, with a documented seed. Records that would identify a person are excluded by construction, since only population-level records enter the reference set.

5.3 Human rating

Three raters with methodological training, working independently, blinded to the machine weight. Each record is rated on two items: (a) a five-point certainty judgement, and (b) a pairwise preference within randomly assigned pairs from the same stratum. Inter-rater agreement is reported as Krippendorff's α with bootstrap confidence intervals. If $\alpha < 0.60$, the rating instrument is revised and the round repeated before any comparison with the machine — a machine cannot be calibrated against a reference that does not agree with itself.

5.4 Primary endpoint

Spearman correlation between EWR final weight and the mean human certainty rating across all 300 records.

- **Pre-specified success:** $\rho \geq 0.70$ with lower CI bound ≥ 0.55 .
- **Pre-specified failure:** $\rho < 0.50$, or systematic inversion in more than 15 % of cross-class pairs where raters and machine disagree in direction. On failure, the class-weight table is rebuilt from the rating data rather than patched.

5.5 Secondary endpoints

- Ablation: weights computed from class rank alone versus full modifier set. If the full model does not improve ρ by at least 0.05, the modifier apparatus is not carrying its complexity cost and is reduced.
- Alternative composition: log-additive penalties with a capped total downgrade, compared on the same reference set. The better model wins and the paper is amended, not the data.
- Retraction handling: a seeded set of 15 known-retracted records must be scored 0.00 with `critical` uncertainty in 15/15 cases. This is a pass/fail gate, not a correlation.

5.6 What calibration cannot establish

That the weights are correct in any absolute sense. Human raters are a convention, not a ground truth. A successful calibration licenses one statement only: the system orders records approximately as trained raters would. Anything beyond that would need outcome data, which this design does not collect and is not intended to collect.

6. Data quality and validation

6.1 Quality gate

Every ingested record passes four stages: (1) schema validation (Pydantic v2), (2) reference validation against foreign keys and controlled vocabularies, (3) duplicate detection (InChIKey / PMID / NCT-ID / DOI), (4) plausibility ($IC_{50} > 0$, $p \in [0,1]$, molecular weight > 0 , SMILES parseable via RDKit, LOINC code resolvable). Failures are quarantined with the failing rule recorded, and remain queryable. Deletion of a quarantined record is an explicit curator action with an audit entry.

6.2 Automated tests

51 tests: Evidence engine 20, quality gate 7, laboratory parsing 9, biomarker status 4, hypothesis integration 3, GDPR operations 7 (Art. 15 export, Art. 17 erasure, consent state machine), API 15. The counts overlap where a test covers two areas; the test identifiers are listed in the validation protocol.

What the tests do not cover, stated because the v0.1 claim “validated with 51 automated tests” was too broad: they establish that the implementation matches the specification. They establish nothing about whether the specification is right. Scoring correctness in the sense of §5 is untested. Connector behaviour against live APIs under rate limiting, schema drift in upstream sources, and long-run graph consistency are not covered and are listed as L9–L11.

6.3 Reproducibility

Deterministic scoring functions without randomness; versioned scoring configuration; configuration version stamped on every score; append-only audit log for scoring changes; container images pinned by digest. A recomputation of the full corpus under a given configuration version must reproduce every stored score bit-identically; this is itself a test (T-REP-01) and is run before any release.

Abb. 3 — Vierstufiges Qualitätsstor / Four-stage quality gate

Quarantaene statt Loeschung / quarantine instead of deletion

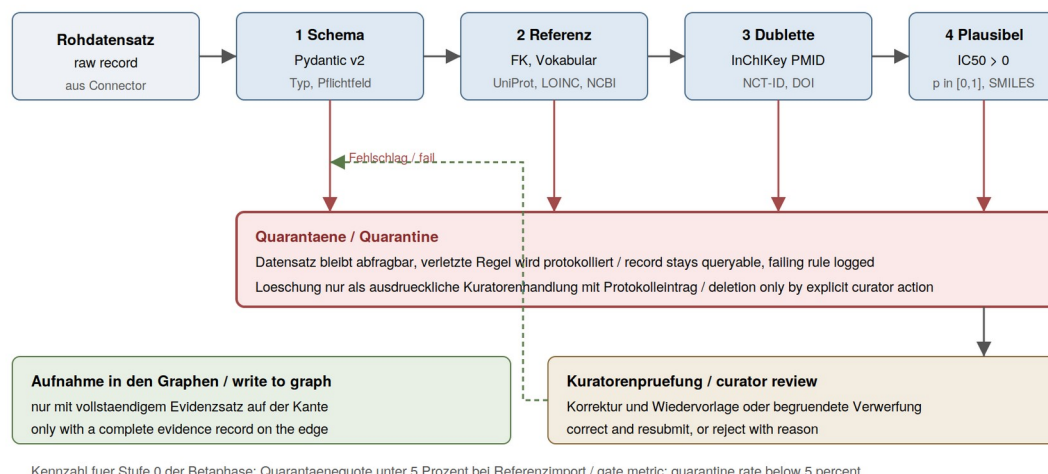


Figure 3 — Quality gate. Failed records are quarantined and remain queryable.

7. Regulatory positioning

This section replaces the single disclaimer of v0.1. It is a positioning, not legal advice; the author is not a lawyer, and the points marked [L] require review by qualified counsel before any release beyond a closed test group.

7.1 Medical device qualification (MDR/IVDR)

Qualification follows the intended purpose, not the technology, and under MDCG 2019-11 Rev.1 (June 2025) a multi-module product is assessed module by module: a medical module inside an otherwise non-medical platform brings only that module into the device regime, provided boundaries, interfaces and dependencies are documented [16]. That revision also introduces the term medical device AI and cross-references the European Health Data Space, which is why §3.5 defines module boundaries explicitly.

Applied to this system: M1–M3, M6 and M7 pursue a research purpose and do not provide information for the diagnosis or treatment of a disease. M5 (biomarker context) is the boundary. As specified — displaying a reading against the laboratory’s own reference range, with no interpretation, no risk score, no suggestion of further testing — M5 does not qualify. The following changes would flip that assessment:

Trigger	Effect
Interpreting a reading as indicating a condition	qualifies as MDSW; classification via Annex VIII Rule 11 [L]
Producing a risk score for a disease	qualifies; likely class IIa or higher [L]
Recommending further diagnostic testing	qualifies (the MDCG worked example on risk scores is directly analogous) [L]

Trigger	Effect
Combining biomarker readings into a derived index presented as health information	qualifies [L]
Marketing wording that promises detection, prevention or monitoring of disease	qualifies regardless of the technical design

The last row deserves the emphasis: the intended purpose is established by what is claimed, including in marketing material. A disclaimer does not cure a claim made elsewhere on the same page. The flyer and landing page in this bundle are written against that constraint.

7.2 Artificial intelligence regulation

Two questions, in order.

Is it an AI system? The scoring engine is a deterministic rule evaluation with fixed coefficients. It does not infer from inputs how to generate outputs, and it does not adapt. On that basis it falls outside the definition in Art. 3(1) of Regulation (EU) 2024/1689. This holds only for the current version; the NLP extraction deferred to v2 (L5) would change the answer for that module.

If it were, what would apply? Art. 2(6) exempts AI systems and models developed and put into service solely for scientific research and development. That exemption covers the current use but ends at the point where the system is placed on the market or put into service for non-research purposes — which is precisely the transition the beta programme in §8 approaches.

The timetable, as amended in 2026: general application from 2 August 2026, including the Art. 50 transparency obligations for AI-generated content and interaction with people. The obligations for stand-alone high-risk systems under Annex III were deferred by the Digital Omnibus from 2 August 2026 to 2 December 2027, and for AI embedded in regulated products under Annex I to 2 August 2028. Further prohibitions apply from 2 December 2026. [L] for the concrete effect on any future v2 NLP module, since a deferred obligation is not a repealed one.

7.3 Data protection

Biomarker readings are health data under Art. 9(1) GDPR. The lawful basis is explicit consent under Art. 9(2)(a), separately for each processing purpose, revocable, with the revocation path in the interface rather than in a policy document. A data protection impact assessment under Art. 35 exists in draft and requires review [L]. Data subject rights are implemented as API operations with tests (Art. 15 export, Art. 17 erasure, Art. 20 portability). Storage limitation is configured per data category. Pseudonymisation separates the identity table from the reading table; this reduces but does not remove the personal character of the data, and the paper does not claim anonymisation.

For an independent researcher without an institutional ethics committee, the ethics route must be chosen before the first external test round: either an independent ethics board opinion or a documented, written justification for why none is required, filed before the first participant is enrolled.

7.4 European Health Data Space

Regulation (EU) 2025/327 entered into force on 26 March 2025 and applies in stages: general application from 26 March 2027, with the secondary-use chapter — health data access bodies, data permits, HealthData@EU — applying from 26 March 2029, and specified data categories following in 2031. Two consequences for this design: any future route to population health data runs through a data access body and a data permit, not through bilateral agreement; and the interoperability requirements for electronic health record systems form the reference architecture for the laboratory import module, which is why LOINC coding is mandatory rather than optional in M4.

7.5 Summary of positioning

Instrument	Current status	Trigger for change
MDR/IVDR	not a medical device; M5 is the boundary	interpretation, risk scores, diagnostic suggestion, disease claims in marketing
AI Act	outside Art. 3(1); research exemption Art. 2(6) applies	v2 NLP module; placing on market beyond research
GDPR	Art. 9(2)(a) consent, DPIA in draft [L]	any processing purpose not covered by the consent text
EHDS	not in scope before 2027/2029	any use of health data from data holders
Austrian law	DSG, ECG §5 disclosure obligations for the public site	commercial operation, advertising

Abb. 4 — Regulatorischer Entscheidungspfad mit Kipppunkten / Regulatory decision path with triggers

Einordnung, keine Rechtsberatung — mit [R] markierte Punkte sind prüfungsbedürftig / positioning, not legal advice

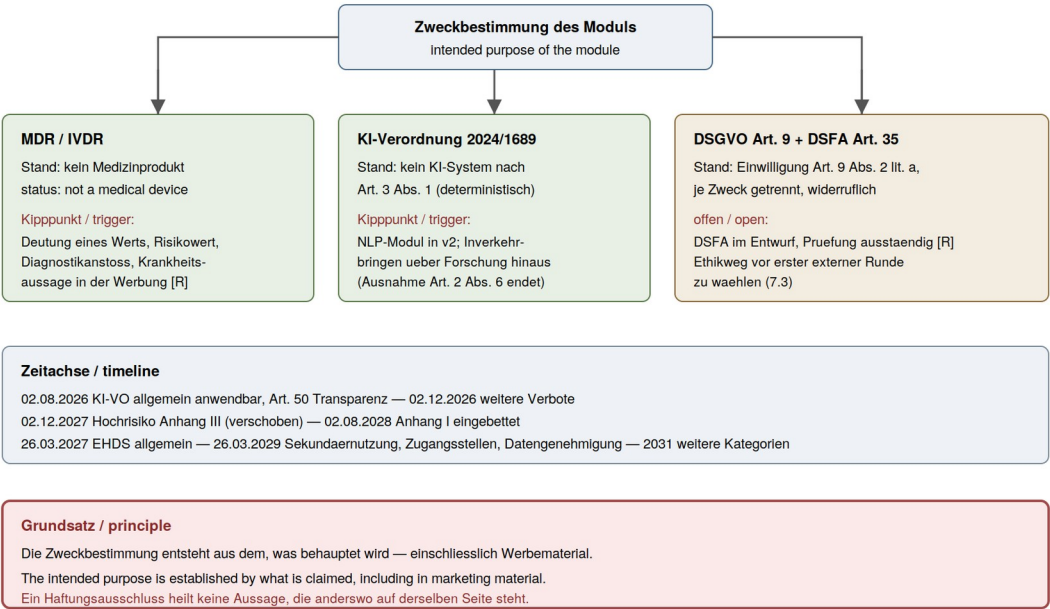


Figure 4 — Regulatory positioning with triggers and timeline. Points marked [L] require qualified review.

8. Evaluation and beta programme

8.1 Purpose

To test usability, comprehensibility of the uncertainty labels, and connector robustness — not to test the health effect of anything. No health outcome is measured, no intervention is administered, no advice is given.

8.2 Phases

Phase 0 — internal. Synthetic data only. Exit criterion: all 51 tests green, T-REP-01 reproducible, quarantine rate below 5 % on a reference import.

Phase 1 — closed alpha, 5–8 participants, no personal health data. Population-level features only. Endpoint: can participants state correctly, from the interface alone, what an uncertainty label means? Measured as a 10-item comprehension check; pre-specified threshold 80 % correct.

Phase 2 — closed beta, 15–25 participants, own laboratory reports, explicit consent. Not started before: DPIA reviewed [L], ethics route settled (§7.3), erasure path tested end to end by a third party, and the §5 calibration completed or the interface labelled “uncalibrated ordering” on every score display.

Phase 3 — open beta. Only after Phase 2 stopping criteria are met and the regulatory positioning has been reviewed [L].

8.3 Stopping rules

Any of the following halts the programme and triggers a written incident record: an uncorrected export or erasure defect; a participant reporting that they took a health decision on the basis of a display; a connector delivering data attributed to the wrong source; a score whose derivation cannot be reconstructed from the stored `applied_modifiers[]`.

8.4 Data minimisation

Participants upload only reports they choose to upload. No account merging, no third-party analytics, no external fonts or scripts on the public site. Feedback is collected in a form that does not require the participant to restate any health data.

9. Limitations

L1 — No empirical calibration. Weights are argued, not measured. §5 is a plan, not a result. Until it is executed, no score should be read as a probability or as a certainty estimate.

L2 — No causal claims. Associations only, for every score value.

L3 — Assay incomparability. IC50 values from different assay systems are not directly comparable. Assay metadata are annotated; no cross-assay normalisation is performed, and none is planned without a validated method.

L4 — Taxonomic instability. Historical publications use superseded plant names. NCBI Taxonomy canonicalisation does not resolve all historical synonyms; unresolved names are flagged rather than guessed.

L5 — NLP deferred. Automated extraction from full text is deferred to v2 because of the false-positive risk and because it would change the AI Act analysis (§7.2). v1 uses structured imports only.

L6 — Legal review pending. All points marked [L] in §7 require qualified review before any release beyond a closed group.

L7 — Modifier independence assumed. The multiplicative model treats modifiers as independent; they are not (§4.7). The alternative model is pre-specified in §5.5.

L8 — No path-level uncertainty. Multi-hop traversals report the weakest edge. This is conservative and imprecise.

L9 — Upstream schema drift. Connectors are tested against fixtures, not against live upstream changes. A silent upstream field change would currently surface as a quarantine spike rather than as a named error. One documented case exists: the classic ClinicalTrials.gov interface was retired on 25 June 2024, API v2 with a different base URL and pagination model is now binding, and code written against the old version breaks. The trial connector must be updated accordingly.

L10 — Source verification is partial. Of 55 registered sources, 18 are verified (32.7 %, as of 2026-09-04, across two rounds). The remaining 35 are carried over from v0.1 and must not be cited as verified. The second round produced four changes out of six checks — including a retired interface and a licence condition that does not fit a researcher without institutional affiliation (KEGG). Further false assumptions are to be expected in the unverified remainder.

L11 — Single-maintainer risk. One person builds, reviews and releases. There is no independent code review. This is stated because it bounds the credibility of every “validated” claim in §6.

L12 — Reference set not yet drawn. The 300-record reference set of §5.2 does not exist yet. The protocol is publishable; the result is not, and this version claims none.

10. Conclusion and route to v1.0

The system integrates natural products evidence and individual biomarker readings under one weighting and uncertainty semantics, keeps provenance on every edge, and refuses two specific transfers (extract—compound, animal—human) at write time. Version 0.3 withdraws the GRADE-adaptation claim, states what the numbers do not mean, and adds the calibration protocol that can refute the scoring design.

Route to version 1.0, in order and with gates:

1. Execute \$5 calibration; publish the result whichever way it falls.
2. Legal review of all [L] points; DPIA finalised.
3. Phase 1 comprehension study; revise labels if below threshold.
4. Independent review of the scoring code by a second person.
5. Source register verification completed to 100 % or the unverified remainder removed.
6. Journal submission at threshold B: primary target *Journal of Cheminformatics* (software article; requires full reproducibility and source code under an OSI licence; APC applies), alternative *Database* (Oxford) or *JAMIA Open*; the software component separately to *JOSS*, which charges no fee and reviews the software rather than a narrative.

Version 1.0 is reached when 1–5 are complete, not when the code is finished.

References

1. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926. doi:10.1136/bmj.39489.470347.AD
2. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*, version 6. Cochrane; 2019.
3. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi:10.1136/bmj.l4898
4. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC International Chemical Identifier. *J Cheminform*. 2015;7:23. doi:10.1186/s13321-015-0068-4
5. Weininger D. SMILES, a chemical language and information system. *J Chem Inf Comput Sci*. 1988;28(1):31–36. doi:10.1021/ci00057a005
6. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations. *Clin Chem*. 2003;49(4):624–633. doi:10.1373/49.4.624
7. Moreau L, Missier P (eds). *PROV-DM: The PROV Data Model*. W3C Recommendation; 2013.
8. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. doi:10.1038/sdata.2016.18
9. Rutz A, Sorokina M, Galgonek J, et al. The LOTUS initiative for open knowledge management in natural products research. *eLife*. 2022;11:e70780. doi:10.7554/eLife.70780
10. Zdrazil B, Felix E, Hunter F, et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res*. 2024;52(D1):D1180–D1192. doi:10.1093/nar/gkad1004
11. Zdrazil B, Leach AR, et al. Fifteen years of ChEMBL and its role in cheminformatics and drug discovery. *J Cheminform*. 2025. doi:10.1186/s13321-025-00963-z

12. Milacic M, Beavers D, Conley P, et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* 2024;52(D1):D672–D678. doi:10.1093/nar/gkad1025
 13. Buniello A, Suveges D, Cruz-Castillo C, et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* 2025;53(D1):D1467–D1475. doi:10.1093/nar/gkae1128
 14. UniProt Consortium. UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Res.* 2023;51(D1):D523–D531. doi:10.1093/nar/gkac1052
 15. U.S. Food and Drug Administration. FAERS — FDA Adverse Event Reporting System, public dashboard and quarterly data files. Accessed 2026-09-04.
 16. Medical Device Coordination Group. MDCG 2019-11 Rev.1 — Guidance on qualification and classification of software in Regulation (EU) 2017/745 (MDR) and Regulation (EU) 2017/746 (IVDR). European Commission; June 2025.
 17. Schünemann H, Brożek J, Guyatt G, Oxman A (eds). GRADE Handbook. GRADE Working Group; updated edition.
 18. Regulation (EU) 2016/679 (General Data Protection Regulation).
 19. Regulation (EU) 2017/745 on medical devices (MDR).
 20. Regulation (EU) 2017/746 on in vitro diagnostic medical devices (IVDR).
 21. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act), as amended 2026.
 22. Regulation (EU) 2025/327 on the European Health Data Space, OJ L, 2025/327, 5 March 2025.
 23. Rigden DJ, Fernández XM. The 2026 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 2026;54(D1). doi:10.1093/nar/gkaf1427
 24. Krippendorff K. Content Analysis: An Introduction to Its Methodology. 4th ed. SAGE; 2018. (α statistic used in §5.3)
 25. Schlesinger S, Schwingshackl L, Neuenschwander M, Barbaresko J. A critical reflection on the grading of the certainty of evidence in umbrella reviews. *Eur J Epidemiol.* 2019;34:889–890. doi:10.1007/s10654-019-00531-4
 26. Smith AM, Niemeyer KE, Katz DS, et al. Journal of Open Source Software (JOSS): design and first-year review. *PeerJ Comput Sci.* 2018;4:e147. doi:10.7717/peerj-cs.147
 27. Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz, DSGVO), Austria, as amended.
 28. Bundesgesetz über den elektronischen Geschäftsverkehr (E-Commerce-Gesetz, ECG), Austria, §5 disclosure obligations.
-

Appendix A — Figures

- **Figure 1** — System architecture with module boundaries (M1–M7) and the regulatory fault line at M5.
- **Figure 2** — Evidence weighting: class weight, modifier cascade, clamp, derived uncertainty.
- **Figure 3** — Four-stage quality gate with quarantine loop and curator action.

- **Figure 4** — Regulatory decision path with trigger points (MDR/IVDR, AI Act, GDPR, EHDS).

Appendix B — Verification status of the reference base

Of 28 references, 14 were re-verified against the publisher record on 2026-09-04 (DOI, volume, pages, year): [1], [4], [5], [9], [10], [11], [12], [13], [16], [21], [22], [23], [25], [26]. The remaining entries are standard works and legal texts carried forward; legal texts are cited by their official identifier and were checked for current amendment status where the amendment matters to the argument ([21], [22]). Reference [15] is a data resource, not a publication, and is cited by access date.

Footer: Thomas Schüller · Syntagma Forschung, Vienna · forschung@tschueller.com · ORCID 0009-0003-9799-6747 · CC BY 4.0 · v0.3 · 2026-09-04 · FORSCH-PROV ≥0427 (provisional) · slug biologic-fbs-paper