zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

An Open-Source Data Infrastructure for Alloy Synthesis Research: Stage 2 — Validated Multi-Instrument Characterization Pipelines

Iliya Radulov

October 2026

Technical report. Sample identifiers are anonymized throughout, except where results are validated against named public reference standards or published literature (cited where used).

Abstract

Stage 1 of this project established a structured database and desktop interface for alloy composition design, screening, and literature cross-referencing. Stage 2 turns that foundation into a working data layer for the laboratory's day-to-day research, through five components of comparable weight. (i) An *element-fraction table* converts every recorded composition into a fixed-width, one-column-per-element feature vector for machine-learning use, and is filled automatically at sample entry, with aimed, measured and reference compositions kept apart by an explicit type field. (ii) A *synthesis-route log* stores the screening module's suggested alternative routes, together with the physical reasoning behind each, as database records instead of transient screen output. (iii) A *multi-machine architecture* moves the database from a development container to a dedicated always-on server, verified row by row against the source, and is accompanied by a stage-scoped repository reorganization and a path-repair utility. (iv) Validated *characterization pipelines* for X-ray diffraction (XRD), vibrating-sample magnetometry (VSM) and scanning electron microscopy (SEM) share one three-layer pattern and one database schema; the XRD pipeline recovers the certified lattice parameter of NIST SRM 660b (LaB_6) to about 0.0001 Å. (v) A catalogue of *engineering lessons* records the silent failure modes found along the way. As in the Stage 1 report, the emphasis is on software architecture, validation methodology and these lessons rather than on the underlying materials-science results, which remain unpublished.

1 Introduction

Stage 1 (reported separately [1]) addressed the front end of the research workflow: composition entry, a multi-metric screening calculator (valence electron concentration, atomic size mismatch, mixing enthalpy, and a melt/boil-point-based synthesis-feasibility check),

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

and cross-referencing against public materials databases. That work produced a desktop application and a normalized schema, but left the laboratory's actual measurement data — diffraction patterns, magnetometry loops, and micrographs — as unstructured files on disk, parsed informally if at all.

Stage 2 closes that gap. Its scope, agreed at the outset, was threefold: (1) build and independently validate a correct, instrument-aware analysis pipeline for each characterization technique in use; (2) integrate each pipeline into the shared database with a schema that anticipates machine-learning consumption, not just archival storage; and (3) make the results usable day-to-day through interactive tools, not only batch scripts. A fourth, cross-cutting item — restructuring the repository and the database infrastructure to support the project's growth across multiple machines — emerged partway through and is reported alongside the three instrument pipelines.

Development proceeded one technique at a time rather than in parallel, with XRD first as a deliberate "learning integration" (the simplest single-table case) before VSM's more structurally complex, multi-table pipeline, and SEM last, since it carried an unresolved GUI-toolkit question from Stage 1. This sequencing is visible in the validation narratives below and is itself one of the project's working principles: establish a pattern on the simplest case, then reuse it with eyes open for where it needs to change.

Contributions.

1. Three independently validated, instrument-specific characterization pipelines (XRD, VSM, SEM), each checked against an external ground truth — a certified reference standard, a published measurement protocol, or an internal cross-sample physical consistency check — rather than only self-consistency.

2. A database schema extension (nine new tables — two for XRD and seven for VSM — plus a composition-type field) designed jointly for day-to-day record-keeping and for fixed-width, machine-learning-ready feature extraction, populated by a consistent three-layer pipeline pattern (standalone analysis module, database integration module, interactive operator tool) reused with modifications across all three techniques.

3. A documented, reproducible account of engineering findings across all three pipelines — silent failure modes (NaN propagation, duplicate-peak fitting, a type mismatch at a driver boundary, a metadata field's type assumption breaking across formats) and one deliberate human-in-the-loop design decision (SEM phase-fraction judgment) adopted after automated thresholding was shown, on real data, to fail in sample-dependent ways — each generalized into a lesson, in the same spirit as Stage 1's Section 5.

4. A cross-cutting data layer (an auto-computed element-fraction table and an automatic synthesis-route recommendation log) and a migration of the project's infrastructure to a shared, always-on, multi-machine architecture, both reported here since they were built alongside the three pipelines rather than as part of Stage 1 or Stage 3.

The rest of this paper is organized as follows. Section 2 positions this work relative to existing characterization-analysis tooling and the specific published references used for

validation. Section 3 gives the shared system-level pattern reused across all three pipelines. Sections 4–6 report each pipeline's architecture, validation, and engineering lessons in turn. Section 7 covers the cross-cutting data and infrastructure systems. Section 8 discusses scope and limitations, and Section 9 concludes.

2 Related Work and Positioning

Diffraction analysis. Full-profile refinement packages such as GSAS-II [8] are the established route to quantitative structural results, and open-source tools such as `powerxrd` cover quick peak fitting and crystallite-size estimates. The XRD module here has a narrower aim: reproducible single-peak parameters (position, width, area, fit quality) from routine laboratory scans, written to a database as soon as the file arrives. The underlying corrections are classical: the Scherrer relation for crystallite size [9, 10], the Rachinger separation of the Kα_2 component [11], and Nelson–Riley extrapolation for lattice-parameter error [12]. The module exists because the fitting defaults of the tool used as a cross-check (fixed fit windows, undocumented unit conventions) were found, on real data, to produce silently wrong results near closely spaced peaks (§4.3). That is a finding about one tool's defaults, not a claim that no adequate tool exists.

Magnetometry. Demagnetization effects on sheared loops are corrected here with two published cuboid results, Aharoni's closed-form magnetometric factor for a rectangular prism [2] and the Prozorov–Kogan effective factor [3]; §5.1 and §5.2 reconcile them explicitly instead of treating them as competing claims. The isothermal magnetic entropy change follows the discontinuous, zero-field-warm-up protocol and Maxwell-relation formulation of Law et al. [4], building on the general magnetocaloric framework of Pecharsky and Gschneidner [7]. That paper serves as the independent, citable check in §5.2.

Microscopy image analysis. General-purpose image-analysis environments such as ImageJ [15] and libraries such as scikit-image [16] provide the building blocks used here: Otsu thresholding [13], Canny edge detection [14] and watershed segmentation. This work adds the instrument-specific part around them, namely unified metadata parsing across three manufacturers' formats and a human-in-the-loop rule for phase-fraction judgment (§6).

Research-data infrastructure. Platforms such as NOMAD [17] and AiiDA [18] provide general, schema-based and provenance-tracking research-data management, and public databases such as the Materials Project [19] supply reference data (cross-referenced in Stage 1 [1]). This project occupies a different niche: a small, laboratory-local PostgreSQL database in which each instrument's analysis is validated against an external ground truth, and whose schema is designed jointly for lab-book use and for fixed-width machine-learning feature extraction. We make no claim that it replaces these platforms. It targets the layer below them, where instrument files become trustworthy numbers.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

3 System Overview

Each characterization technique is implemented as a three-layer stack, a pattern established with XRD and reused, with modifications, for VSM and SEM:

1. **Standalone analysis module** — pure functions operating on parsed instrument data, with no database connection and no GUI dependency. Produces structured results (peaks, segments, features) as plain data.
2. **Database integration module** — transforms a standalone module's output into rows matching the shared schema, applies instrument-specific duplicate-detection rules, and writes to PostgreSQL. Kept deliberately separate from the analysis module itself (analysis and persistence are different concerns), and separate again from a thin driver-compatibility layer (§4.3, §5.3) that sanitizes NumPy scalar types before they reach `psycopg2`.
3. **Interactive standalone tool** — a `customtkinter` desktop tool, launchable directly or pre-loaded from the main application's Data Viewer, exposing the technique's key adjustable parameters (peak-fit sensitivity, branch-detection sensitivity, segmentation threshold) with a live preview and an explicit "Save to DB" action that persists only what the operator has reviewed and accepted.

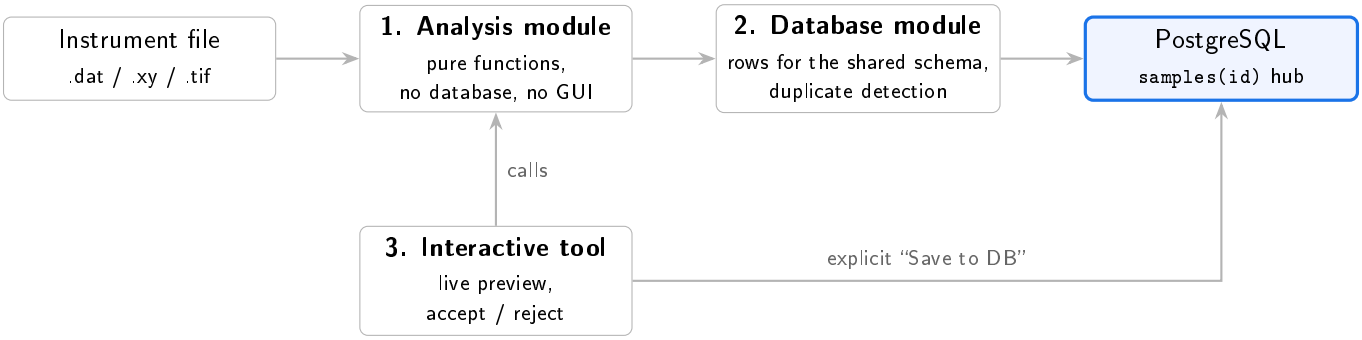


Figure 1: The three-layer pattern shared by the XRD, VSM and SEM pipelines. The analysis module (1) is independent of the database and the interface; the database module (2) turns its output into rows for the shared schema; the interactive tool (3) calls the analysis module and writes to the database only when the operator saves what they have reviewed and accepted. The tool's save step may also replace rows that the import (1→2) stored automatically.

This repeated pattern reflects a consistent design stance: fully automate what is reliably automatable, and make the remaining judgment calls fast and visible rather than either hiding them in a black box or leaving them as a manual offline step. The SEM phase-fraction tool (§6.1) is the clearest expression of this stance, but it is present throughout — XRD's per-peak accept/reject and VSM's per-segment accept/reject follow the same principle.

The underlying database is PostgreSQL, extended from Stage 1's `samples`-centered schema with technique-specific tables, each linked to `samples(id)` through a chain of foreign keys (Fig. 2). Stage 1 provided eight tables; Stage 2 adds nine more, two for XRD

and seven for VSM, for seventeen in total. Section 7 covers the schema's cross-cutting extensions (element-fraction table, synthesis-route logging) and the infrastructure that now serves it across three machines.

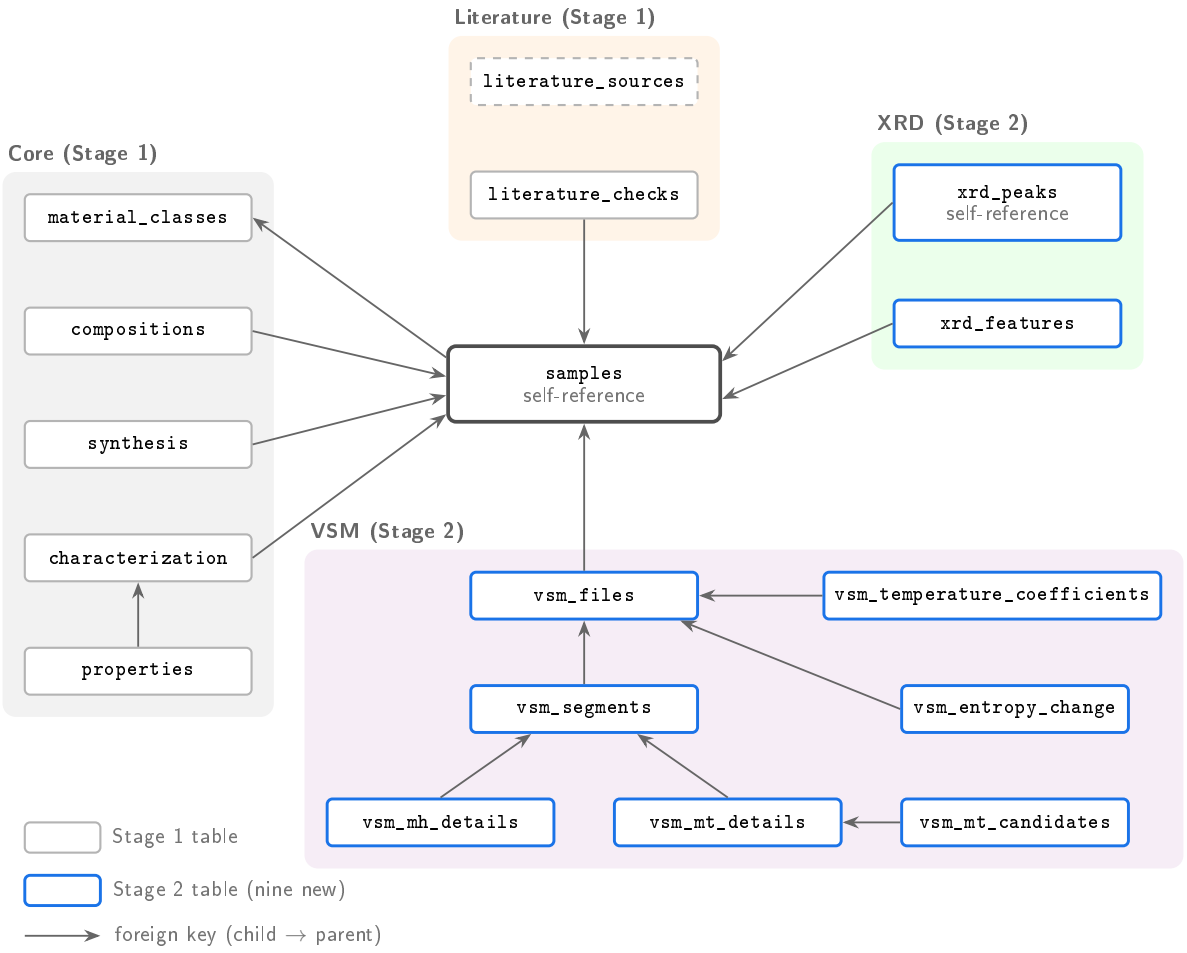


Figure 2: The seventeen tables of the shared schema. Each arrow is a foreign key, drawn from the referencing table to the table it references; `samples` is the hub, reached directly or through a short chain of keys from every table that holds measurement data. Blue-outlined tables are the nine added in Stage 2 (two for XRD, seven for VSM); the others come from Stage 1. `samples` and `xrd_peaks` each also reference themselves (`parent_sample_id` for sample lineage, `doublet_partner_peak_id` for paired XRD peaks). `literature_sources` has no foreign key to or from any other table; the SEM results are stored as rows of `properties` under `characterization`.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

4 XRD Pipeline

4.1 Architecture and Implementation

The XRD module (`xrd_analyzer`, developed through versions `v2` → `v3` → `dev1`) performs background subtraction, peak finding, and per-peak Gaussian fitting on raw two-column (2θ, intensity) pattern data. Its design evolved through several deliberate, physically motivated stages:

- **Neighbor-aware fit windowing.** Each peak's fit window and allowed center-shift are capped at roughly 90% of half the distance to its nearest neighboring peak, rather than a fixed angular window. This replaced an earlier fixed $\pm 1.5°$ window that allowed closely spaced peaks to "steal" each other's fits, producing duplicate fitted centers (§4.3).

- **Physical-unit peak separation.** The minimum-separation parameter passed to the peak finder is specified in degrees 2θ and converted internally to sample count using a step size measured directly from each file's own data, rather than assumed constant across instruments or acquisitions.

- **Wavelength handling.** A six-anode lookup table (Cu, Mo, Co, Cr, Fe, Ag) stores both Kα_1 and Kα_2 wavelengths from user-supplied reference values, with Kα_1 as the default — corrected from an earlier version that had used the Kα_1/Kα_2 weighted mean, a consistent $\sim$0.0013 Å source of error given that the laboratory's instruments have no Kα_2-stripping monochromator. A `wavelength_override` parameter additionally supports non-standard (e.g., synchrotron) sources, with an explicit guard preventing it from being combined with Kα_2 stripping, which is not physically meaningful for a monochromatic source.

- **Kα_2 handling.** Two complementary mechanisms: a default-on, whole-pattern Rachinger stripping correction (removing Kα_2's contribution to peak broadening even where the doublet is not visibly resolved, which otherwise corrupts Scherrer crystallite-size estimates), and an optional, constrained two-Gaussian doublet fit for cases where Kα_1/Kα_2 are already visibly resolved.

- **Fit-quality and derived quantities.** An R^2 metric computed from each peak's own local fit residuals; Scherrer crystallite size (from the module's own fitted center/FWHM by default, with an optional cross-check against PowderXRD's own re-fit); and d-spacing via Bragg's law.

- **Optional relative-intensity filtering.** A peak-ranking filter based on intensity relative to the pattern's strongest peak ($I/I_{\max}$), following the standard ICDD convention, provided as an explicit, non-default tool rather than applied automatically — it is useful for quick phase identification but actively discards some of the highest-precision, high-angle peaks, and is therefore inappropriate for precision lattice work (§4.2).

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

Results are transformed into two database tables by a dedicated builder module, deliberately separate from the analysis code itself: `xrd_peaks`, a long-format audit trail with one row per fitted peak (full fidelity, safe to recompute downstream features from without reparsing raw files), and `xrd_features`, a fixed-width table with one row per sample, carrying the top-five peaks by intensity as named columns (NULL-padded where fewer peaks exist) plus sample-level aggregates. The aggregates — mean crystallite size and mean R^2 — are computed only from peaks meeting a quality threshold ($R^2 \geq 0.9$ by default), since on real data roughly 10–15% of peaks that converge to a fit nonetheless do so poorly, and including them undamped would silently degrade the sample-level statistics that a downstream model would train on.

A separate, unvalidated utility exists alongside this pipeline for quick visual inspection during data import — a rough lattice-parameter read-off, not a substitute for the fitted-peak analysis above. It was corrected during the October 2026 review: its hand-typed reference reflection table was found to be 20–40° off in 2θ for every entry, and its lattice-parameter formula omitted the l^2/c^2 term entirely, so every indexed reflection with a nonzero l index was silently wrong. It now computes expected reflections from the correct reference cell and its space group's reflection conditions, matches a peak only when exactly one reflection falls within a tight angular tolerance, and fits both lattice constants together by least squares with reported standard errors — validated on a synthetic pattern with known lattice constants, recovering both to four significant figures. Because this tool's old output was wrong by a large, obvious margin (roughly double the correct lattice parameter), any value it produced and stored prior to this fix should be treated as unreliable and recomputed rather than trusted.

Figure 3 shows the interactive tool. The operator chooses the anode and whether to strip the Kα_2 component, sets the prominence and minimum peak distance, and reviews the fitted peaks in a table with one accept/reject checkbox per peak; only the accepted peaks are written to the database when "Save to DB" is used.

4.2 Validation

The pipeline was validated at increasing levels of rigor. Internally, neighbor-aware windowing was confirmed to eliminate duplicate fitted centers across multiple real patterns (zero duplicates where the earlier fixed-window version had produced them), and the physical-unit separation parameter was confirmed to reproduce identical peak counts to the previous behavior at the previous default, while correctly reducing peak count as the window was widened.

The strongest validation is external: the pipeline was run on a synchrotron diffraction pattern of **NIST Standard Reference Material 660b (LaB_6)** [5, 6], analysed with the wavelength supplied by the beamline, $\lambda = 0.738012$ Å. SRM 660b is a certified lattice-parameter standard with large, strain-free crystallites chosen specifically to minimize real peak broadening; its certified lattice parameter is $a = 4.15691(8)$ Å at 22.5 °C [6]. Fit quality across the pattern was high (R^2 improving from 0.91 to 0.998 relative to an earlier, uncorrected comparison run), and the lattice parameter computed from the pipeline's own fitted peak positions converged, by the twelfth peak in the pattern ($2\theta \approx 35.8°$), to **4.1568 Å, about 0.0001 Å (three parts in 10^5) from the certified value**. Low-angle peaks

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

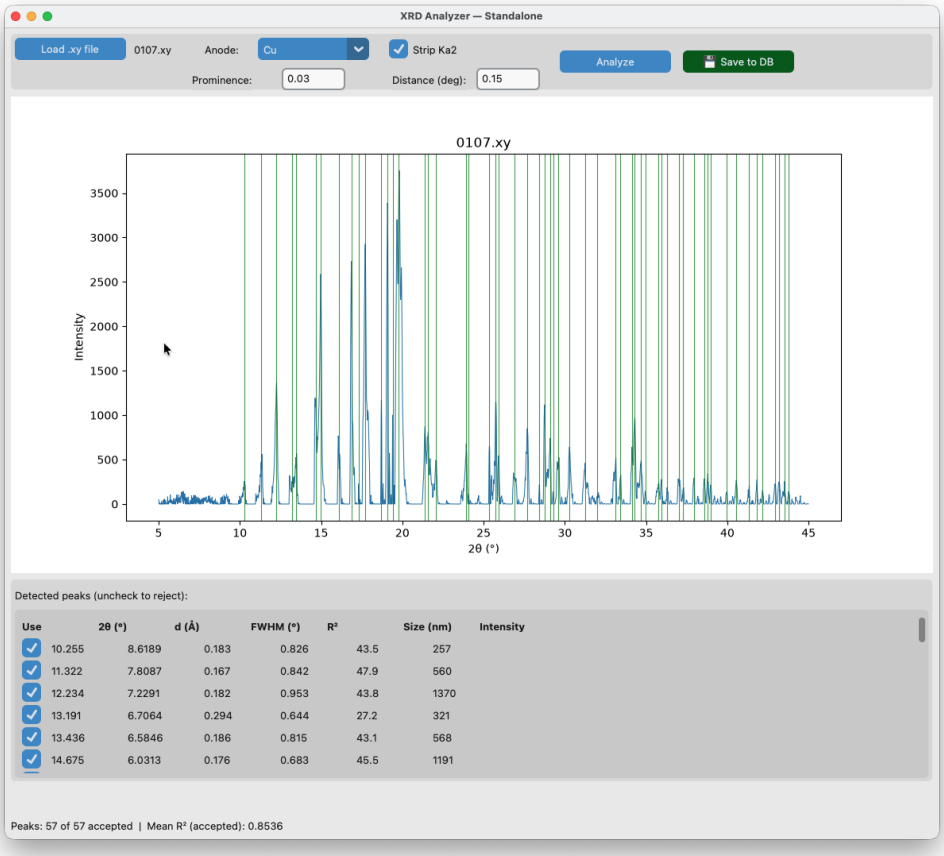


Figure 3: The standalone XRD analyzer on an example pattern (Cu anode, Kα_2 stripped). Detected peaks are marked on the pattern and listed below with position, d-spacing, FWHM, fit R^2 and crystallite size; each can be rejected individually. The status line reports how many peaks are accepted and their mean R^2.

showed larger disagreement than high-angle peaks, improving systematically with 2θ; this is the expected behavior predicted by Bragg's-law error propagation (sensitivity of d-spacing to angular error scales with $\cot\theta$) and is the physical basis of the established Nelson–Riley extrapolation technique — the pipeline's own data reproduced this trend without being built to. One limit of this check should be stated: synchrotron wavelengths are commonly calibrated against a reference powder, often LaB_6 itself. The agreement therefore validates the pipeline's peak positions and its d-spacing and lattice-parameter chain at the given wavelength; it is not an independent test of the measurement's absolute accuracy.

4.3 Engineering Lessons

Duplicate fitted peaks from fixed-window stealing. Defect. A fixed $\pm 1.5°$ fit window and center-shift bound allowed two real, closely spaced peaks to converge to the same fitted center on real data, silently producing duplicate rows with different intensities. **Fix.** Cap

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

each peak's window and allowed shift at a fraction of its distance to its nearest neighbor. **Generalizable lesson.** A fit window sized for the typical case will silently fail for the atypical one; sizing it relative to the actual local data (not a global constant) removes the failure mode rather than merely reducing its frequency.

A documented interface that does not match the implementation. Defect. PowderXRD's `Chart` class documents `K` and `lambdaKa` as constructor parameters; the actual code silently ignores any such arguments and requires them to be set as instance attributes afterward — and `lambdaKa` is in nanometers, not the Ångströms used throughout this project's own table. **Fix.** Confirmed from source rather than docstring, and set as attributes post-construction, with the unit conversion made explicit. **Generalizable lesson.** A third-party library's docstring is a claim, not a guarantee; for a value with physical consequences, verifying against the actual executed code is worth the extra step before trusting it.

Position matching alone is insufficient evidence of a Kα_2 doublet. Defect. An early doublet-detection criterion based solely on predicted-vs-observed peak position produced a false positive — a real peak pair in test data matched position prediction to 0.1° but had an independently fit amplitude ratio of ∼1.10, nearly double the physically fixed ∼0.5 Kα_2/Kα_1 ratio, and more plausibly reflected genuine crystallographic splitting. **Fix.** Added a second, independent amplitude-ratio check; both must pass. **Generalizable lesson.** Two physically independent signals pointing the same direction are meaningfully stronger evidence than one, especially where a single criterion can be satisfied by an unrelated physical cause.

Dead code surviving a partial edit. Defect. An earlier edit accidentally deleted a helper function's `def` line while leaving its body in place; Python treats code after a `return` statement as syntactically valid but unreachable, so the file imported and ran cleanly until the deleted function was actually needed (Rachinger stripping), at which point it failed at runtime with no connection to the edit that caused it. **Fix.** Root-caused via AST parsing rather than by guessing at the call site; a full regression check across all validated real files was re-run to confirm nothing else had silently broken. **Generalizable lesson.** A partial edit can leave code that is syntactically valid and behaviorally wrong in a way that only a full exercise of the affected call chain will surface — a pattern repeated from a different partial-deletion bug in Stage 1 (see [1], §5.2).

A driver-level type mismatch with a misleading error message. Defect. `psycopg2` has no default adapter for NumPy scalar types, which the XRD (and, as anticipated, VSM) analysis modules produce throughout; the resulting failure surfaced as `InvalidSchemaName: schema "np" does not exist` — the NumPy value's `repr()` leaking into the query, rather than any message naming the actual type-adaptation problem. **Fix.** A small, reusable sanitization step applied only at the database-insertion boundary, deliberately not built into the analysis modules themselves (producing correct values and satisfying one driver's type quirks are different responsibilities). **Generalizable lesson.** A database driver error message can describe the symptom's shape rather than its cause; building the fix as a boundary concern, once, meant the identical issue was already solved when VSM reached the same integration step. (This bug also surfaces in the Stage 1 paper's own account, §5.3, since the failing insert there was of a computed lattice parameter from this pipeline sharing Stage 1's persistence layer — the same defect, reported from each

paper's own side of the boundary.)

5 VSM Pipeline

5.1 Architecture and Implementation

The VSM pipeline handles files from three instrument platforms (PPMS/DynaCool, MPMS3, ACMS) in current laboratory use, each with a different column layout and, in two cases, a different convention for where real moment data actually lives. Instrument-type detection was found, on real data, to be unreliable by column count (the same software version produced different counts across files; different versions of the same software produced different counts for the same measurement type) — the pipeline instead checks for specific diagnostic column names present or absent regardless of count (e.g., `DC Moment Fixed Ctr (emu)` for MPMS3; `M-DC (emu)` together with `Calcoil' (emu)` for ACMS; neither, by elimination, for PPMS/DynaCool).

The pipeline is a chain of nine composable modules, which fall into three groups.

Preparing the file (modules 1–4).

1. *Type detection*, as described above.
2. *Mass extraction*: the header value is preferred, with a confidence-flagged filename-pattern fallback for files whose header was stripped.
3. *Segmentation* into MH (field-swept), MT (temperature-swept), idle and corrupted regions, using a rolling-window local range on field and temperature; a gap-bridging pass absorbs short idle gaps flanked by the same segment type.
4. *Self-centering quality flags*, from two signals (a blank-moment row and a center-position shift), since real data showed these are not fully redundant.

Extracting features (modules 5–7).

5. *MH features*: H_c and M_r from the second quadrant, i.e. the descending branch, since the practically relevant measurement is the demagnetization curve rather than a full closed loop. M_r is read at zero applied field, i.e. it is the apparent remanence of the open sample (§5.2).
6. *MT candidate features*: local extrema in M and dM/dT, deliberately left unclassified rather than labeled as a transition type, because Curie-, Néel- and blocking-type behavior look different and the correct reading is not knowable in advance across the alloy systems studied here.
7. *Temperature coefficients*: α for M_r and β for H_c (the permanent-magnet convention), linear in T and normalized to a 293 K reference.

Deriving figures of merit (modules 8–9).

8. *Isothermal magnetic entropy change* ΔS_M, from the Maxwell relation (validated in §5.2).
9. *Demagnetizing/shearing correction* producing $BH_{\max}$ for cuboid samples, the standard permanent-magnet figure of merit (below).

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

Demagnetizing correction. The correction uses the laboratory's established cuboid approximation (Prozorov–Kogan [3]),

$$N^{-1} = 1 + \frac{3}{4}\frac{c}{a}\left(1 + \frac{a}{b}\right), \tag{1}$$

with the field axis taken as the sample's third dimension, c — confirmed, not assumed: $c \to 0$ recovers $N \to 1$ (thin-plate limit) and $c \to \infty$ recovers $N \to 0$ (long-rod limit), the two physically unambiguous checks used to fix this convention during the October 2026 pre-publication review, after an earlier internal docstring had stated the field axis as b. The edge of the sample actually parallel to the applied field must be passed as the formula's c argument; §5.2 reports an open item regarding this formula's domain of applicability.

Database layer. A genuinely different design decision from XRD's flat row-building pattern: VSM's foreign-key chain (file → segment → detail/candidate rows) means child rows require a database-generated parent ID before they can be constructed, so the VSM database-builder module is written as a single orchestration function taking a live connection, rather than the pure, connection-free row-builder functions used for XRD. The import stores the automatic result for every segment; it does not compute $BH_{\max}$, which requires per-sample density and dimensions that the import does not receive. $BH_{\max}$ is computed in the interactive VSM tool (or by calling the pipeline with those inputs), and the tool's save step then replaces the automatic values with the operator's reviewed ones: accepted segments carry the re-analysed values, every other segment is stored without Hc/Mr/$BH_{\max}$ and flagged, and the temperature coefficients are refitted from the accepted segments only.

Stabilization policy. Whether simultaneous field-and-temperature variation during a measurement transition should be treated as a hard rejection or merely flagged was resolved as a deliberate design choice rather than a fixed rule: real adherence to recommended stabilization protocol varies across users and instruments, and the acceptable tolerance depends on analysis intent (acceptable near saturation for a rough M_s extraction, not acceptable for anything precision-sensitive) — so the segmenter reports the condition, and whether to discard or retain-with-warning is left as a downstream, record-creation-time decision.

Figure 4 shows the MH tab of the interactive tool. Segmentation parameters, the sample mass (read from the file or filename, or entered when absent) and the optional density and cuboid dimensions for the $BH_{\max}$ calculation are set at the top; each detected segment is listed on the left and can be accepted or rejected, with its coercivity and remanence shown below the plot. When density and dimensions are entered, $BH_{\max}$ is also computed and both demagnetizing factors are displayed, the Prozorov–Kogan value that is used and the Aharoni value for comparison, which is not saved, as in the figure; without them, $BH_{\max}$ is simply not computed. The remaining tabs cover temperature coefficients, entropy change and the MT candidates.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

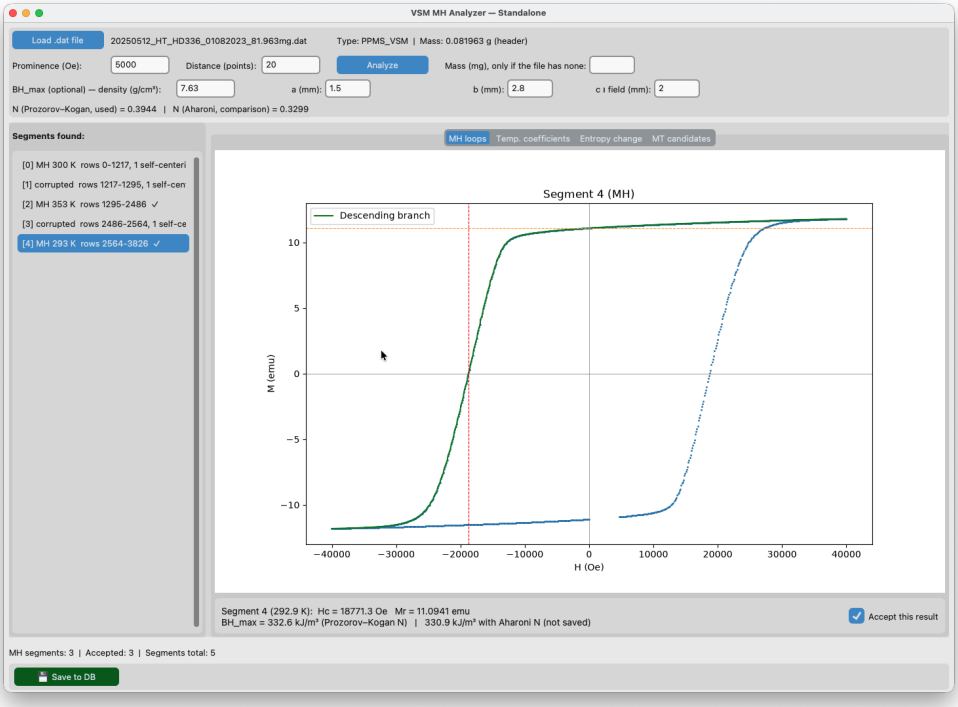


Figure 4: The standalone VSM analyzer, MH tab, on sample HD336 (hot-deformed at 825 °C), 293 K segment. The detected segment is plotted with its descending branch highlighted and the coercivity crossing marked. With density (7.63 g/cm^3) and cuboid dimensions ($1.5 \times 2.8 \times 2$ mm, the 2 mm edge along the field) entered, $BH_{\max}$ is computed with the Prozorov–Kogan factor (332.6 kJ/m^3) and the Aharoni value is shown for comparison (330.9 kJ/m^3).

5.2 Validation

Validation proceeded at three levels, from a physical consistency check, to a reimplemented calculation checked against a published protocol, to a derived mathematical result.

Level 1: cross-sample agreement. Independently measured, independently processed real samples produced nearly identical temperature coefficients ($\beta(H_c)$ of -0.607%/K and -0.611%/K, $R^2 = 0.996$ and 0.9999; $\alpha(M_r)$ of -0.123%/K and -0.110%/K, $R^2 = 0.997$ and 0.9995) — a genuine cross-validation against real physics rather than mere internal self-consistency, and incidentally a validation of the standard linear-in-T assumption over the measured range rather than an assumption adopted without checking.

Level 2: entropy change against a published protocol. The isothermal magnetic entropy change module was checked against real, previously trusted output from an earlier (now inaccessible) implementation, agreeing to within 1% at nearly every temperature point across two independent target field changes, using the real sample mass found in a header the user had forgotten still existed — a mass that itself matched the value back-calculated

from the validation to within 0.3%, before the header value was located. The measurement protocol and Maxwell-relation formulation follow Law et al. [4], confirming both the discontinuous, isothermal, zero-field-warm-up protocol and the decision to use only the up-sweep branch of each isotherm, appropriate given this material's genuine magnetic hysteresis.

Level 3: shearing correction and $BH_{\max}$. This level rests on a synthetic ground truth rather than on measured data. A synthetic test of the demagnetizing (shearing) correction recovered a known intrinsic H_c to within $\sim 10^{-13}$ Oe, confirming that the correction is numerically exact. The correction, $H_{\text{int}} = H_{\text{applied}} - NM$ in SI units, moves every point of the loop along the field axis by an amount proportional to that point's own magnetization. It therefore leaves H_c exactly unchanged, because $M = 0$ at the coercivity crossing. It does change the remanence. M_r is defined at zero *internal* field, which on the measured loop lies at $H_{\text{applied}} = NM_r > 0$, not at zero applied field. The value read at $H_{\text{applied}} = 0$, which is what the pipeline stores, is the apparent remanence of the open sample; it is lower than the intrinsic value by a fraction of about $\chi N/(1 + \chi N)$, where χ is the reversible (recoil) susceptibility near remanence. For HD336 ($N = 0.394$), the intrinsic remanence, read on the measured descending branch at $H_{\text{applied}} = 5181$ Oe, is 11.230 emu against the stored 11.094 emu: 1.2% higher ($J_r = 1.314$ T instead of 1.298 T), corresponding to $\chi \approx 0.03$. §5.3 records how this effect was first missed. $BH_{\max}$ is computed from the fully corrected curve, and it is the quantity most strongly affected by shearing (a maximum of the product $B \times H$ over the whole second quadrant, not a single crossing); it was therefore the main validation target. The resulting $BH_{\max}$ value, computed for a real hot-deformed sample (HD336, hot-deformed at 825 °C, measured at 293 K, density 7.63 g/cm^3), was 332.6 kJ/m^3 — within the range expected for a good sintered or hot-deformed NdFeB-grade magnet (commercial grades typically ∼320–430 kJ/m^3).

The same pipeline's H_c and M_r were also compared with an internal, unpublished reference table. Both measured H_c values (converted to kA/m) fell inside the table's corresponding ranges. M_r, after density correction with the density of 7.61 g/cm^3 specified for the commercial starting powder of these samples, came out roughly 4–7% *below* the table. (HD336, made from recycled powder whose heavy rare earths and cobalt raise the density, has 7.63 g/cm^3.) Part of this gap is expected from Level 3 above: the stored M_r is the apparent remanence of the open sample in the VSM, whereas the reference values were measured in closed circuit with a hysteresisgraph, where the sample sits between the pole pieces and shows no demagnetizing field. For HD336 this correction is 1.2%; for samples of similar shape it would account for about one percentage point of the gap. The remainder is plausibly because the measured specimens are not guaranteed to be the exact ones behind the table row, rather than a pipeline error, but this has not been independently confirmed. It is consistent in sign with a $BH_{\max}$ at the lower end of the expected range.

The correction was also reconciled with the laboratory's own spreadsheet practice. That practice performs the demagnetization subtraction in SI-consistent kA/m for both H and M, where the standard relation carries no 4π, whereas the independently derived CGS form ($H_i = H_{\text{applied}} - 4\pi NM$) does; the 4π in the laboratory's chain appears instead in the Oe-to-kA/m field conversion and in the polarisation conversion. The two are therefore the same physics in different unit conventions, not competing results. The demagnetizing-factor and

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

polarisation formulas reproduced two worked examples from the laboratory's own materials.

Sample geometry and axis assignment. The demagnetizing factor depends strongly on which sample edge lies along the field. For the sample behind the 332.6 kJ/m^3 figure (HD336, a $1.5 \times 2.8 \times 2$ mm piece), taking the 2 mm edge as the field axis gives $N = 0.394$, whereas taking the 1.5 mm edge gives $N = 0.509$ and the 2.8 mm edge gives $N = 0.290$. This spread would have produced a confidently wrong $BH_{\max}$ if the assignment had been guessed. Since $\frac{c}{a}\left(1 + \frac{a}{b}\right) = \frac{c}{a} + \frac{c}{b}$, the formula is symmetric in the two edges perpendicular to the field, and only the choice of the field edge matters. For this sample the 2 mm edge is the thickness of the pressed disc and was taken as the field axis, giving $N = 0.3944$; the two in-plane edges (1.5 and 2.8 mm) are interchangeable in the formula; this value is unchanged by the October 2026 docstring correction, which concerned the documentation of the convention rather than this number.

Open caveats on $BH_{\max}$. Two caveats remain open. First, the Prozorov–Kogan formula is strictly derived for diamagnetic ($\chi = -1$) samples. It was adopted because it reproduces the laboratory's established practice exactly, not because it was shown to be the more accurate choice for a ferromagnet. For this sample's geometry the more general Aharoni magnetometric factor [2] (implemented alongside, §5.1) gives a materially different value (0.330 versus 0.394). The resulting $BH_{\max}$, however, differs by less than one percent (330.9 versus 332.6 kJ/m^3 for the same loop, both from the interactive tool). Second, the field-axis assignment has been checked only for the sample above; it has not been re-confirmed against the corrected convention for every other sample the pipeline has processed. Neither caveat changes the pipeline's internal correctness, since the numerical exactness check holds for either factor, but the first bears on whether 332.6 kJ/m^3 is the right number and the second on whether other stored $BH_{\max}$ values are. Both are carried into §8. The final conversion step ($B \times H$ to kJ/m^3) was also derived from the exact SI identity ($\mathrm{T} \cdot \mathrm{A/m} = \mathrm{J/m^3}$) rather than by replicating the spreadsheet's own conversion step, and a direct cross-check against a trusted $BH_{\max}$ reference number from the laboratory's own workflow remains outstanding (§8).

5.3 Engineering Lessons

NaN propagation silently masking real physical variability. Defect. The segmenter's rolling-window range calculation used plain `max()`/`min()`, which propagate NaN across an entire window if even one blank value (e.g., an initial instrument settling row) falls inside it; since `NaN > threshold` silently evaluates to `False` rather than raising, this masked a genuine ∼60 K temperature jump between isotherms, merging three physically distinct MH loops into one segment. **Fix.** NaN-aware (`nanmax`/`nanmin`) range calculation, confirmed via full regression to change nothing on the files that happened not to have a NaN landing at a critical transition. **Generalizable lesson.** A NaN-tolerant comparison does not fail loudly — it fails by quietly answering "no" to a question it was never actually able to evaluate, which is a more dangerous failure mode than a crash.

A real decimal point defeating a filename-parsing pattern. Defect. The mass

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

extractor's filename fallback handled plain integers and underscore-as-decimal conventions, but a genuine decimal point in a real filename (e.g., an `85.304mg` pattern) fell through to the loosest matching pattern, which silently captured only the digits after the period, returning a mass roughly 280 times too small. **Fix.** A new, higher-priority pattern for literal decimal points, checked before the looser fallback, with full regression confirmed against every previously validated filename. **Generalizable lesson.** A fallback pattern ordered from strict to loose is only as safe as its coverage of real, observed formats — a new real file is often the actual test of an assumption that synthetic or historical test cases never exercised.

A structural import assumption invisible until deployed in the real directory layout. Defect. Two modules used bare sibling imports that resolve correctly only when Python's module search path happens to include the importing file's own directory — true in every prior sandbox test (always run from the same folder as the target file) but false in the real application, where the integration module imports these from a different directory than the modules themselves live in. **Fix.** Explicit `sys.path` insertion of each module's own directory at import time, verified by reproducing the exact real error in isolation, confirming a negative control still reproduced it with the fix removed, and then confirming the fix resolved it — not simply patched and assumed. **Generalizable lesson.** An import pattern that works in every test environment can still depend on an accident of how that environment happens to be invoked; the real deployment layout is the only test that actually exercises the assumption.

A symmetry argument that held for one crossing and not the other. Defect. The shearing module's documentation stated that the correction leaves both H_c and M_r unchanged, because the correction term is proportional to M, "which vanishes at both crossings". M vanishes at the H_c crossing only; at the remanence crossing it is the field that is zero, and after the correction the point of zero internal field is a different point of the loop. The synthetic test confirmed the H_c half of the claim, and the M_r half was never tested on its own, so stored M_r values are apparent rather than intrinsic remanence (§5.2). **Fix.** Found during the final review of this report; the stored values are now described as apparent remanence, and computing the intrinsic remanence is listed as an open item (§8). **Generalizable lesson.** A test confirms only what it actually compares. When a correction moves the points of a curve, a quantity defined by a condition on the curve must be found again by applying that condition after the correction — and each claimed invariance needs its own check, even when one argument seems to cover several.

Filename conventions. When the sample mass is not stored in the file header, it is read from the filename. The parser accepts the conventions found in our own data and in a test batch from another operator: a value followed by `mg`, with an underscore, point or comma as the decimal separator and an optional leading `m`, in upper or lower case. That test batch exposed a real defect: names with a point or comma decimal (for example `95.98mg`) were read silently as only the digits after the separator (98 mg), with errors of up to a factor of seventy. This is now fixed and covered by a test. Arbitrary naming cannot be anticipated, so we recommend a consistent, simple scheme (one decimal separator, one unit, one file extension); where the mass cannot be read, the parser reports none and the interactive tool accepts it as a manual entry.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

6 SEM Pipeline

6.1 Architecture and Implementation

Early in Stage 2, "SEM characterization" was recognized to conflate two genuinely different tasks: low-magnification phase, particle, and pore quality control — general, applicable to essentially every sample shortly after synthesis — and high-resolution grain-alignment and texture analysis, meaningful only for samples that have already undergone a specific process such as hot deformation, and closer in character to dedicated research on that process than to general Stage 2 characterization. This project explicitly scoped itself to the former for Stage 2, deferring the latter.

Three manufacturers' instruments are in routine use, each with an incompatible metadata convention, unified under one parser with per-format auto-detection:

- **Zeiss**: metadata is embedded in proprietary compressed binary TIFF tags (34118/34119) — the mechanism originally diagnosed during this project's earlier characterization work, predating the three-layer pipeline pattern used elsewhere in this paper and out of the Stage 1 paper's own scope (its Section 3 explicitly excludes raw-file quantitative extraction). A further real bug was found and fixed here: extracted values arrive as unit-embedded strings (e.g., `"25.00 K X"`, `"11.09 nm"`), which a naïve `float()` conversion silently fails on; proper regex-based extraction, including correct interpretation of the `K` magnification suffix, was cross-checked exactly against a historically validated pixel-size value.
- **JEOL**: the TIFF file itself carries no usable calibration metadata (confirmed directly — only generic tags, and the apparent resolution field is a print-DPI setting, not a real calibration value). Real metadata lives in a sidecar `.txt` file in a simple `$KEY value` format, directly providing magnification, accelerating voltage, working distance, signal mode, and — critically — a scale-bar length in both pixels and physical units for direct calibration.
- **Tescan**: a sidecar `.hdr` file in standard INI format, parsed with Python's built-in `configparser`, giving pixel size directly in meters with no conversion required. Tescan image data is natively 16-bit, unlike the 8-bit data from the other two instruments, which required a corresponding adaptation in the grain-detection module (§6.3).
- Files missing their required sidecar are handled by returning an explicit `format='unknown'` rather than a guessed result.

Grain detection uses watershed segmentation (Canny edge detection → dilation → distance transform → connected-component labeling → watershed), extracted from an earlier exploratory interactive tool and validated as a standalone module with the analysis math unchanged. Two further refinements were required before it worked reliably on real data: automatic detection and cropping of each image's databar footer (via a row-brightness-drop heuristic, robust to varying image sizes rather than a hardcoded row range), and adaptive, per-image Canny thresholds computed from each image's own gradient-magnitude percentiles, replacing fixed thresholds that failed outright on the majority of real, genuinely low-contrast images (§6.3).

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

Phase-fraction analysis is deliberately not fully automated. Automatic (Otsu) thresholding was tested against real images from all three formats and found to fail in two materially different, sample-dependent ways (§6.3), leading to an explicit design decision: automate what is reliably automatable — file loading, pixel calibration, footer cropping — and keep the actual phase-boundary judgment with a human operator, via an interactive tool offering Otsu's result as a starting suggestion with a live threshold slider. The resulting operator-confirmed fraction and threshold are the only values ever persisted; nothing from automatic thresholding is saved without explicit confirmation.

Figure 5 shows the interactive tool, opened from the main application's Data Viewer (left), which displays the selected micrograph with the metadata read from its file. In the grain tab (right), the edge percentiles, distance threshold, edge dilation and minimum grain area are adjusted with sliders, and the overlay, the grain count and the size and shape histograms update live; a second tab provides the threshold-based phase fraction. The tool reads the pixel calibration and footer crop from the file's own metadata, and its results are saved only on request, together with the parameters that produced them.

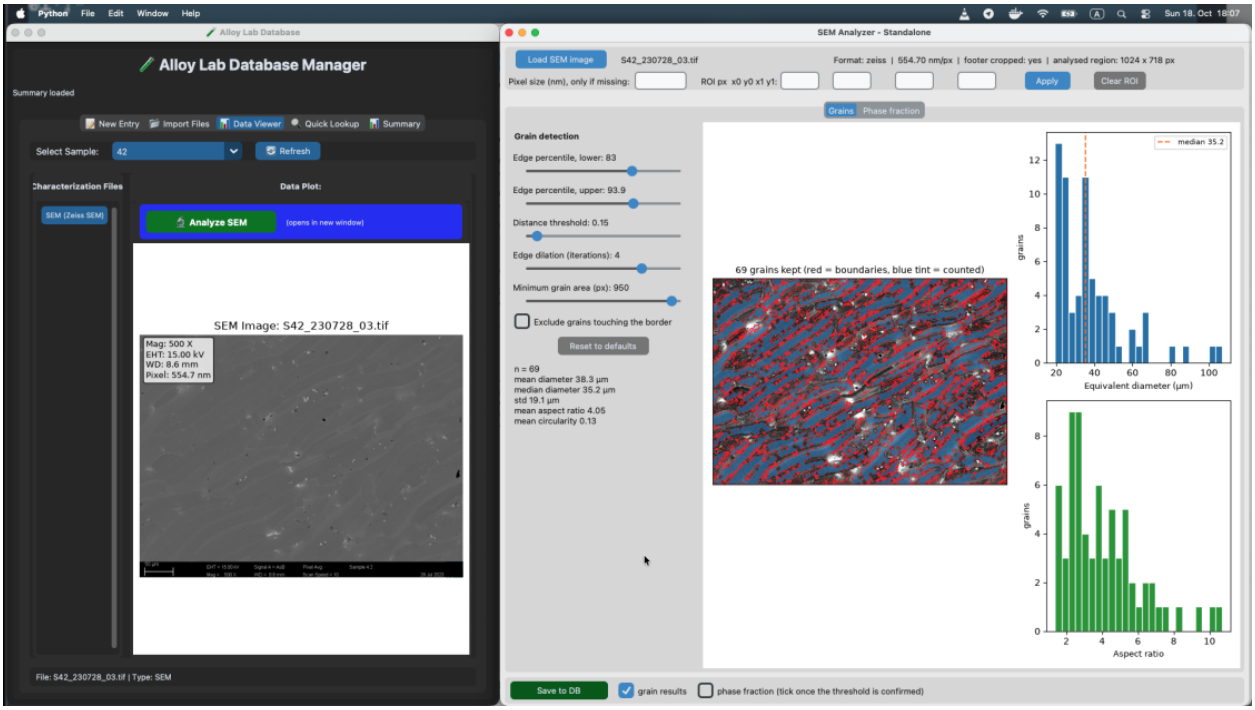


Figure 5: The SEM analyzer opened from the Data Viewer. Left: the main application showing the selected micrograph with the metadata read from the file (magnification, accelerating voltage, working distance, pixel size). Right: the standalone analyzer, grain tab, with the operator-chosen settings for this image (edge percentiles 83 and 93.9, distance threshold 0.15, dilation 4, minimum area 950 px, border grains not excluded). Red lines mark detected boundaries and the blue tint the grains that are counted; the histograms show equivalent diameter and aspect ratio.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

6.2 Validation

The grain-detection module was validated against sixteen real images of deliberately mixed quality. With fixed Canny thresholds, nine of twelve test images failed outright, each returning a single "grain" that was in fact the entire cropped frame (confirmed quantitatively: edge-pixel coverage near zero, reported grain area matching the full frame area). Adaptive, per-image thresholding resolved seven of these nine; the remaining two were confirmed to be images the operator had deliberately selected as lower-quality test cases, and the tool's correct behavior was to flag them as unreliable rather than force a confident answer — a result treated as a successful validation outcome, not a remaining defect. A further, physically grounded limit was identified directly rather than chased with more threshold tuning: on fresh fracture surfaces, the tool was found to be tracing a compositional intergranular phase rather than actual grain boundaries, because a fracture surface presents three-dimensional topographic facets that the algorithm's two-dimensional compositional-contrast assumption cannot resolve — correctly diagnosed as requiring a different technique (e.g., EBSD) rather than further algorithmic adjustment.

The phase-fraction tool's internal logic was validated by construction rather than against an external standard: threshold changes were confirmed to produce monotonic phase-fraction and area changes (stricter threshold $\rightarrow$ lower fraction), inverting the phase selection produced an exact complement ($13.91\% + 86.09\% = 100.00\%$ on one real test image), and the calibrated area readout matched a manual hand calculation exactly for a known pixel count and calibration value. The underlying design decision not to fully automate threshold selection was itself validated by a real, cross-technique finding: an apparent segmentation failure on a Tescan image, initially suspected to be a directional surface-texture artifact confusing the algorithm, was confirmed instead to be genuine physical structure via a Kerr microscopy image of the same feature class (Kerr microscopy specifically images magnetic domain contrast, not topography or composition) — the simple thresholding approach was not wrong, it was correctly detecting real contrast arising from a physical source other than the one intended for this particular measurement.

6.3 Engineering Lessons

A fixed detection threshold failing on the majority of real images. Defect. Fixed Canny edge-detection thresholds, adequate on the images used during initial tool development, failed outright on nine of twelve independently selected real test images, each genuinely low-contrast (gradient magnitude maxing around 100–120, well below the fixed threshold of 150). **Fix.** Per-image adaptive thresholds computed from each image's own gradient-magnitude distribution. **Generalizable lesson.** A threshold tuned against a small, self-selected set of development images is a statement about that set's contrast characteristics, not a general property of the instrument or the technique; real image diversity is the actual specification.

Automatic segmentation correctly detecting the wrong physical signal. Defect/non-defect. Automatic (Otsu) phase thresholding appeared to fail on a real Tescan image, attributed initially to directional surface texture confusing the algorithm. **Investigation.** A Kerr microscopy image of the same feature class confirmed the "texture"

was real physical structure, not an artifact — the algorithm was detecting genuine contrast, just not the contrast intended for this analysis. **Resulting decision.** Rather than attempt to engineer around sample- and technique-dependent failure modes that are not really algorithmic errors, keep phase-boundary judgment with the operator and automate only the genuinely reliable preprocessing steps. **Generalizable lesson.** A segmentation "failure" is not always a bug to fix — sometimes it is correct detection of the wrong thing for the purpose at hand, and the right fix is a change in where human judgment sits in the pipeline, not a better threshold.

A metadata field silently returning the wrong type across formats. Defect. The grain-analyzer's internal metadata parser was written against Zeiss's string-with-units convention (`"11.09 nm"`); the unified metadata module instead returns Tescan's pixel size as a plain float, and the Zeiss-oriented regex match silently failed against it, leaving Tescan grain-size measurements uncalibrated with no error raised. **Fix.** An explicit type check branching on string-vs-numeric input before parsing. **Generalizable lesson.** Unifying several formats behind one interface removes the need for calling code to know which format it received — but only if every downstream consumer is updated to the unified interface's actual output types, not just its field names.

Scope and limitations. The SEM analyzer is intended as a first-pass tool: it gives a quick, quantitative impression of the microstructure (approximate grain count, size and shape, and a threshold-based phase fraction) before any dedicated analysis. It is not meant to replace the specialised microscopy-analysis software and instrument-specific tools that the community already maintains, and for precise or publication-grade quantification we recommend those tools. Grain detection is semi-automatic: the edge-percentile, distance-threshold, dilation and minimum-area settings are chosen by the operator and are stored with every saved result, so each value can be traced back to the settings that produced it. Results depend on image contrast and resolution; grains smaller than roughly ten pixels across are not resolved reliably, which, for example, limits the analysis of nanoscale grains at lower magnification. Values should therefore be compared between samples only when acquired and analysed under identical conditions.

7 Cross-Cutting Systems

7.1 Element-Fraction Table

Stage 1's `compositions` table already stored composition data in long format (one row per sample–element pair, with weight-percent and an unpopulated atomic-percent column). A backfill script computed atomic percent from weight percent for existing samples using standard atomic weights spanning the full periodic table (not limited to elements currently in use, so it would not require extension as new elements appear in future samples), validated by confirming every sample's atomic percentages summed to exactly 100.000% and that the resulting compositions were physically sensible for the alloy systems in question.

A wide-format reshaping function — deliberately not a new physical database table, which would require a schema migration each time a new element appeared in any future sample and would otherwise be mostly empty cells per row — pivots the long-format data

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

into one row per sample and one column per element, zero-filling elements a given sample does not contain (a real zero, not a missing value), for direct consumption by downstream Python/ML tooling, with CSV export as a simple, language-agnostic handoff format.

This was later wired automatically into new-sample entry: a dedicated database method accepts the already-computed element components directly from the existing mass calculator (which already computes both weight and atomic percent per element) and writes them with an explicit `composition_type` field distinguishing aimed (nominal/target), measured (from instrumental analysis), and reference (literature) composition records — added as a schema extension (`CHECK` constraint on an enumerated text column) rather than a new table, preserving the single long-format source of truth.

Automatic entry covers only compositions that pass through the mass calculator. For compositions known by other means — a measured composition from instrumental analysis, or a literature reference — the layer is completed by an interactive standalone tool of the same kind as for the other techniques (§3). The operator types element and value rows (or pastes them, e.g. "Nd 26.7, Fe 72.3, B 1.0") in either weight or atomic percent and presses Evaluate. The tool returns both percentages, the wide-format row, and validation messages: unknown or duplicate elements and non-positive values are errors rather than being skipped, since dropping an element would silently change every other element's normalization, and a sum that is not about 100 is flagged. Nothing is written without an explicit save under a chosen `composition_type`, replacing existing rows of that type only after confirmation, in a single transaction. The calculation core is a set of pure functions checked against a hand calculation for $Nd_2Fe_{14}B$ (26.68, 72.32 and 1.00 wt%) and an exact weight–atomic round trip.

7.2 Synthesis-Route Auto-Save

The Stage-1 screening module already computed a synthesis-feasibility assessment and, where the target composition appeared difficult to produce by open melting, a list of suggested alternative routes (e.g., mechanical alloying, powder sintering, diffusion bonding) — displayed to the operator but never persisted. An auto-save step, triggered immediately after a new sample's composition is recorded, extracts this assessment and writes each suggested route as its own record in the existing (previously unused) synthesis table, with the full physical reasoning preserved in a notes field, so a route choice made months earlier remains self-explanatory without requiring the original screening session to be reconstructed. A first implementation attempt failed silently because the suggested-routes list is nested one level deeper in the screening result than the auto-save code assumed; corrected once traced to the actual result structure.

7.3 Infrastructure and Multi-Machine Architecture

The repository was reorganized into stage-scoped packages (`stage_one/`, `stage_two/`, `stage_three/` as a placeholder) with several explicit, discussed decisions: the Stage 1 schema-creation SQL remains a frozen, shared foundation, while the live database instance itself is treated as belonging to Stage 2 going forward, since every table added since Stage 1 already references the same `samples` table; shared credentials and configuration remain

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

at the true repository root rather than duplicated per stage, since secrets are a different category from stable stage-specific logic; and the proven, working application entry point remains at the root rather than being moved for the sake of directory tidiness alone. During execution, a bulk cleanup operation was found, via routine `git status` inspection before committing, to have affected a far larger set of files than intended — traced to an earlier "archive copy" that had actually been a move, leaving some files existing only in an external backup location rather than in both places as assumed. Because nothing had yet been committed, full recovery was possible via a straightforward restore, and the practice of checking the actual scope of a bulk change before committing was adopted explicitly going forward.

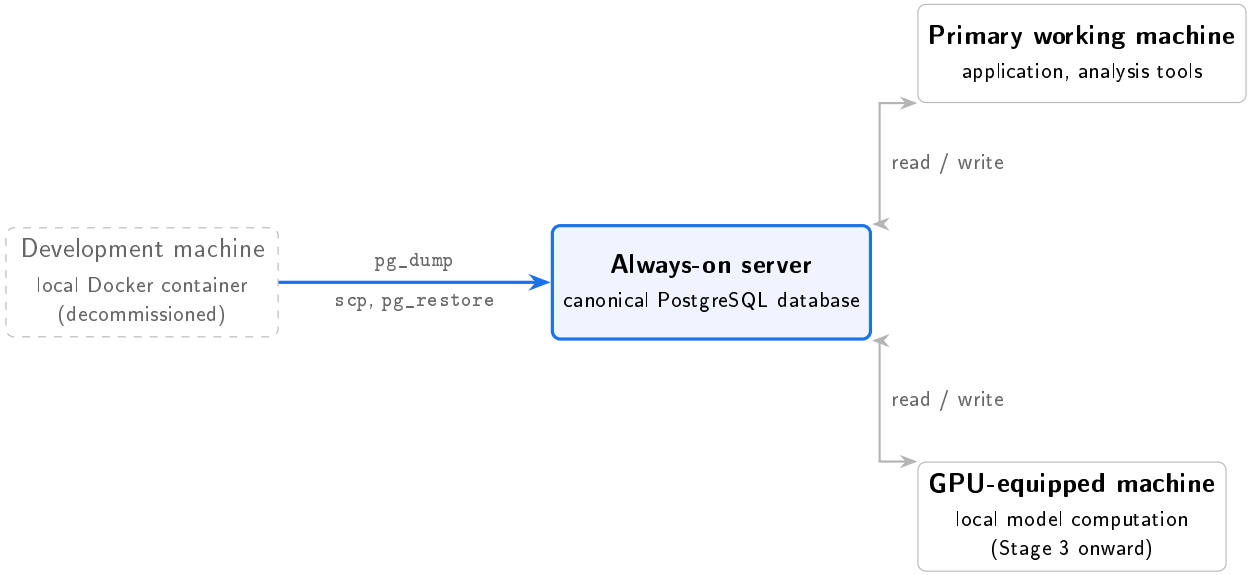


Figure 6: Multi-machine architecture after the database migration. The database moved from a local container on the development machine to a dedicated, always-on server (`pg_dump`, `scp`, `pg_restore`), verified row by row against the source; the server is now the single canonical database for the working machine and the GPU-equipped machine.

The project's database was subsequently migrated from a development machine's local Docker container to a dedicated, always-on home server, via a standard `pg_dump`/`scp`/`pg_restore` sequence. The migration was verified at the row level — fifty samples, thirty composition records, one hundred seventy XRD peaks, ninety-two characterization records, and four synthesis records, across the sixteen tables then present, with zero discrepancy against the source — and confirmed live by writing a new sample to the migrated database from a client machine before the original local database was decommissioned. This completed a three-machine architecture (a primary working machine, a GPU-equipped machine for the computation-heavy work of the later stages, such as local language models and model training, and the always-on server as the single canonical database), with a small accompanying utility (`path_repair.py`) built in response to a real incident where a folder rename broke stored file-path references, scanning all path-bearing columns and offering reporting, automatic, and interactive repair modes.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

8 Discussion and Limitations

Each pipeline in §§4–6 was checked against something outside itself rather than only against its own output. Each check, however, covers a particular part of its pipeline, and the limits of that coverage are the main limitations of this work.

What the external checks cover. The XRD check on SRM 660b exercises peak finding, peak fitting and the d-spacing and lattice-parameter chain at a known, monochromatic wavelength. It does not exercise the parts specific to the laboratory diffractometer, the Kα_1/Kα_2 wavelength table and the Rachinger stripping, which a synchrotron pattern does not need; a reference powder measured on the laboratory's own Cu-anode instrument would close that gap. Since the synchrotron wavelength is itself usually calibrated against a reference powder, the check is also not an independent test of absolute accuracy (§4.2). For VSM, the three levels of §5.2 check different things — consistency between two independently measured samples, agreement with an earlier implementation of the same published protocol, and the numerical exactness of the shearing correction — but no stored quantity has yet been compared with an independent measurement of the same specimen, such as a closed-circuit hysteresisgraph value. For SEM, the phase-fraction tool was validated by construction, and the grain detection by whether it succeeded or correctly flagged failure on sixteen images; grain sizes were not compared with an independent count of the same images.

Stored values that need attention. Three groups of values already in the database should not be used downstream as they stand.

- $BH_{\max}$ **(§5.2).** The Prozorov–Kogan demagnetizing factor used for the reported figures is formally derived for diamagnetic samples; for this project's sample geometry it differs from the more general Aharoni factor (0.394 vs. 0.330), although the resulting $BH_{\max}$ differs by less than one percent. The field axis was checked only for the sample behind the reported 332.6 kJ/m^3 and has not been re-confirmed for every other measured sample, and a direct cross-check of the final unit conversion against a trusted reference $BH_{\max}$ remains outstanding. Until these are resolved, reported $BH_{\max}$ values should be read as provisional.

- **Remanence (§5.2, §5.3).** Stored M_r values are apparent remanence, read at zero applied field — for HD336, 1.2% below the intrinsic value. For comparison with closed-circuit permanent-magnet data, the intrinsic remanence should be read at zero internal field on the same loop, which requires each sample's density and dimensions. The temperature coefficient $\alpha(M_r)$ is a normalized slope and is affected only through the temperature dependence of the recoil susceptibility; this is expected to be small but has not been checked.

- **XRD quick-look lattice parameters (§4.1).** Any `lattice_parameter_a` value stored before the October 2026 correction reflects the old, incorrect reference table and formula and should be recalculated or removed.

Scale and generality. At the time of the migration the database held fifty samples, thirty composition records and four synthesis records (§7.3). The schema is built for fixed-width feature extraction, but that readiness is a property of the design: a dataset large enough to train a model has not yet been collected, and the first real use of the feature tables will show whether the chosen columns are the right ones. The file handling covers the instruments, software versions and naming habits found in this laboratory and in one test batch from another operator (§5.3); files from other sources will need the same kind of checks. The engineering lessons come from a single project. They are reported with the evidence for each, as observations, not as a systematic study of failure modes in research software.

Human judgment in the loop. Keeping the operator in the loop — SEM thresholds, per-peak and per-segment acceptance — was a deliberate choice wherever automation was shown to fail on real data. Its benefit is that no automatic result is stored unreviewed. Its cost is throughput, and agreement between operators: the stored settings make every value traceable to the procedure that produced it, but they do not make two operators' choices agree, and that variability has not been measured.

9 Conclusion and Future Work

Stage 2 established, for each of the laboratory's three primary characterization techniques, a pipeline that is independently validated against an external ground truth rather than only internally self-consistent, integrated into a shared database schema designed jointly for day-to-day record-keeping and for eventual machine-learning feature extraction, and made usable through interactive tools that keep a human operator in the loop wherever automated judgment was shown, on real data, to be unreliable. The resulting database — seventeen tables, spanning composition, synthesis routing, and three independently validated measurement modalities, now hosted on dedicated infrastructure shared across the project's working machines — is the data layer on which the later stages build.

The immediate next steps for Stage 2 itself are the open items of §8: the intrinsic remanence, the demagnetizing-factor and field-axis checks behind $BH_{\max}$, and a reference measurement on the laboratory's own diffractometer. Stage 3 (reported separately) builds advisory tools on this data layer, starting with a CALPHAD-based phase-stability advisor. Machine-learning models trained on the fixed-width feature tables are left to later work, as is recording each recommendation together with the measured outcome of the sample made from it, so that suggestion and measurement are linked in the same database.

References

[1] Radulov, I.A. (2026) An Open-Source Data Infrastructure for Alloy Synthesis Research: Stage 1 -Composition Screening and Literature Cross-Referencing. Zenodo. Available at: https://doi.org/10.5281/zenodo.23158894.

[2] Aharoni, A. (1998). Demagnetizing factors for rectangular ferromagnetic prisms. *Journal of Applied Physics*, 83(6), 3432–3434.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416

[3] Prozorov, R., & Kogan, V. G. (2018). Effective demagnetizing factors of diamagnetic samples of various shapes. *Physical Review Applied*, 10(1), 014030.

[4] Law, J. Y., Franco, V., Moreno-Ramírez, L. M., Conde, A., Karpenkov, D. Y., Radulov, I., Skokov, K. P., & Gutfleisch, O. (2018). A quantitative criterion for determining the order of magnetic phase transitions using the magnetocaloric effect. *Nature Communications*, 9, 2680.

[5] NIST Standard Reference Material 660b (Lanthanum Hexaboride Powder, Line Position and Line Shape Standard for Powder Diffraction) — certificate of analysis, National Institute of Standards and Technology, Gaithersburg, MD.

[6] Black, D. R., Windover, D., Henins, A., Filliben, J. J., & Cline, J. P. (2011). Certification of Standard Reference Material 660b. *Powder Diffraction*, 26(2), 155–158. doi:10.1154/1.3591064.

[7] Pecharsky, V. K., & Gschneidner, K. A., Jr. (1999). Magnetocaloric effect and magnetic refrigeration. *Journal of Magnetism and Magnetic Materials*, 200, 44–56.

[8] Toby, B. H., & Von Dreele, R. B. (2013). GSAS-II: the genesis of a modern open-source all purpose crystallography software package. *Journal of Applied Crystallography*, 46, 544–549.

[9] Scherrer, P. (1918). Bestimmung der Größe und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 98–100.

[10] Langford, J. I., & Wilson, A. J. C. (1978). Scherrer after sixty years: a survey and some new results in the determination of crystallite size. *Journal of Applied Crystallography*, 11, 102–113.

[11] Rachinger, W. A. (1948). A correction for the $\alpha_1\alpha_2$ doublet in the measurement of widths of X-ray diffraction lines. *Journal of Scientific Instruments*, 25, 254–255.

[12] Nelson, J. B., & Riley, D. P. (1945). An experimental investigation of extrapolation methods in the derivation of accurate unit-cell dimensions of crystals. *Proceedings of the Physical Society*, 57, 160–177.

[13] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.

[14] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679–698.

[15] Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9, 671–675.

[16] van der Walt, S., et al. (2014). scikit-image: image processing in Python. *PeerJ*, 2, e453.

[17] Scheidgen, M., et al. (2023). NOMAD: A distributed web-based platform for managing materials science research data. *Journal of Open Source Software*, 8(90), 5388.

[18] Huber, S. P., et al. (2020). AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific Data*, 7, 300.

[19] Jain, A., et al. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1, 011002.

zenodo:23499416 [cond-mat.mtrl-sci] 20 Oct 2026 DOI: 10.5281/zenodo.23499416