WHITE PAPER 05 / EVIDENCE INTEGRITY

The Native File as the Acquisition GroundTruth™

How native numerical arrays, axes, scale, channels, acquisition metadata, file identity, and provenance establish what the microscope actually recorded—and why every normalized representation, display, measurement, and claim must remain accountable to that source.

DR. BRIAN T. BENNETT
Co-Founder and CSO

PAPER 05 OF 10
AUGUST 25, 2026

NATIVE EVIDENCE / DETERMINISTIC MEASUREMENT / GOVERNED INTERPRETATION / DEFENSIBLE PUBLICATION

CONTENTS

Paper structure

The native file is authoritative for the acquisition record. The complete experiment determines how that record may be interpreted.

All numbered citations link to the corresponding reference. Linked titles in the reference list open the cited source.

ABSTRACT

Paper 4 established the material boundary of fluorescence evidence. Target definition, antibody identity, labeling architecture, specimen preparation, optical compatibility, controls, and limitations determine what fluorescence can be presented to the microscope. Once that material system reaches the instrument, the authority changes. The microscope produces an acquisition record, and the preserved native acquisition dataset becomes the authoritative persistent computational record of the instrument output.

That record is more than a displayed image. It may contain one or more numerical arrays; X, Y, Z, channel, time, position, or other axes; physical calibration; bit depth and data type; channel descriptions; instrument and detector settings; acquisition timing; series relationships; reconstruction or processing state; and vendor-specific metadata. These elements determine the exact plane, channel, volume, timepoint, field of view, or acquisition position under analysis—and therefore which measurements are scientifically eligible.

The native file is therefore the Acquisition GroundTruth™. The term is deliberately bounded. It does not mean absolute biological truth, perfect instrument performance, complete metadata, or freedom from acquisition and reconstruction artifacts. It means that the preserved native acquisition evidence and context—not a screenshot, projection, presentation image, or language-model description—remain authoritative for what the instrument recorded.

Microscopes from ZEISS, Leica, Nikon, Olympus, and other manufacturers encode numerical arrays, dimensions, channels, physical calibration, instrument state, and metadata differently. Standards and software such as the Open Microscopy Environment and Bio-Formats improve access and interoperability across these heterogeneous records, but translation into a common representation does not erase the scientific authority of the source. The reader and version, mappings performed, warnings or missing information, and relationship to the original acquisition must remain preserved.[5–11]

The governing rule is strict: ingest without erasing origin. The source object must remain immutable; its identity and hash must remain recorded; arrays and metadata must be decoded with explicit mappings; missing or ambiguous state must remain visible; normalized data must be treated as derived computational representations; and every downstream representation, correction, measurement, and claim must remain traceable to the native acquisition record.

> Central thesis: The native file is the Acquisition GroundTruth™ because it is the authoritative computational record of what the instrument acquired. It is not absolute biological truth. Its arrays, axes, scale, channels, metadata, acquisition state, limitations, and provenance must remain connected so that downstream transformation cannot silently become new evidence.

1. Paper 4 hands competent material to the instrument

Paper 4 ended at a precise boundary. A reagent and labeling system are scientifically competent only within a defined target, specimen, labeling strategy, preparation, optical environment, control state, and intended measurement or claim.[1] Those conditions determine what fluorescence is capable of reaching the detector.

The instrument cannot repair an upstream material failure. It cannot retrospectively distinguish intended target-associated fluorescence from an unrecognized cross-reactive signal, restore an epitope altered during preparation, or recover spatial information degraded by refractive-index mismatch. It records the optical signal presented to it under the acquisition conditions that were used.

Once acquisition begins, a second failure becomes possible: the instrument may create a rich numerical and metadata record, but the downstream workflow may preserve only a partial representation. A screenshot may discard bit depth and scale. An exported TIFF may omit vendor-specific acquisition fields. A maximum-intensity projection may collapse Z. A rendered RGB image may replace separate quantitative channels with display colors. A cropped figure may preserve appearance while losing its relationship to the larger field, series, or acquisition.

Paper 5 therefore inherits Paper 4’s question exactly:

> Once a biologically and materially competent sample reaches the microscope, what record must remain authoritative for what the instrument actually acquired?

The answer is the native acquisition record.

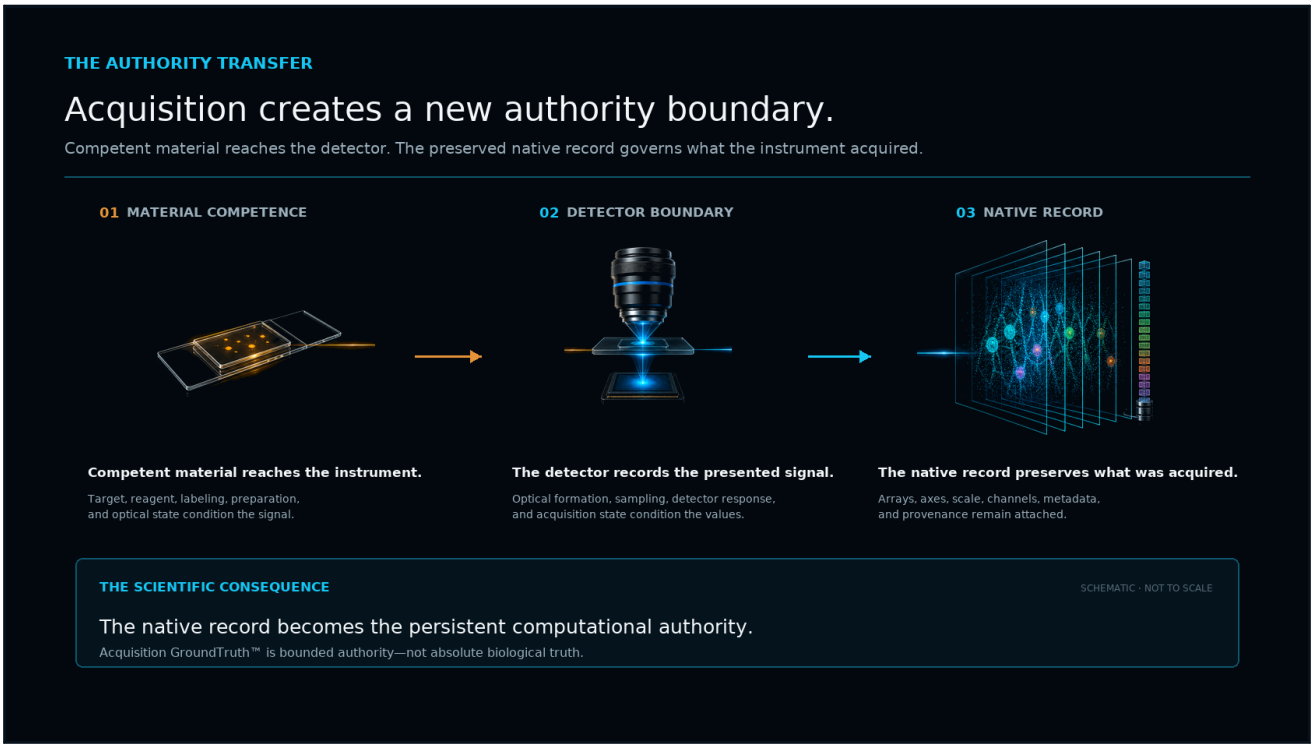

Figure 1. The authority transfer. Competent material and optical conditions cross the detector boundary; the native acquisition record becomes the persistent computational authority. Material competence and acquisition authority remain distinct, and neither is absolute biological truth.

2. “GroundTruth™” is a bounded term

In machine learning, ground truth often refers to a reference label or annotation used to train or evaluate a model. In microscopy, that usage can become dangerous if it suggests that an image is an unmediated representation of biology.

Every microscopy acquisition is conditioned. The specimen and label were already shaped upstream. The optical system forms an image through a point-spread function. Sampling, detector response, noise, dynamic range, exposure, gain, scanning, reconstruction, and processing affect the recorded values. Some modalities write detector intensities directly. Others write reconstructed images or coordinate estimates produced through instrument-specific computation.

BENNETT therefore uses Acquisition GroundTruth™ in a narrower sense:

- Acquisition identifies the boundary of authority: what the instrument recorded or produced as its native acquisition output.
- GroundTruth™ identifies the preserved reference to which downstream computational representations remain accountable.

- The term does not declare that the record is complete, artifact-free, biologically true, or competent for every measurement.

The native file may contain evidence of saturation, inadequate sampling, missing calibration, incomplete metadata, reconstruction, or other limitations. Those limitations do not reduce the need to preserve the file. They are part of the state that must remain visible.

The governing distinction is:

> The native file is authoritative for the acquisition record. The complete experiment determines how that record may be interpreted.

3. The native file is a scientific object, not merely a filename

A native microscopy dataset may be one file, several companion files, a directory structure, a database-backed acquisition, or a collection of linked objects. The scientifically relevant object is the complete native dataset required to resolve the acquisition—not merely the file selected in a dialog box.

That object carries several connected layers. Its recorded evidence may be stored intensity values, reconstructed values, label-free measurements, localization coordinates, or other modality-specific numerical evidence. Its acquisition structure establishes dimensions and axes; series, fields, positions, scenes, tiles, or resolutions; physical calibration; and the channel paths through which signal was recorded. Its instrument state can preserve the microscope, objective, detector, timing, illumination, acquisition sequence, environmental information, and any processing or reconstruction already applied. Its provenance connects those elements to source identity, creation time, file relationships, acquisition software, import method, reader and version, warnings, and integrity hashes.

The absence of a field is also scientific state. If physical scale is missing, the workflow must not silently assume one. If axis order is ambiguous, the workflow must not infer a convenient interpretation and proceed as though it were recorded fact. If a dataset requires companion files, ingest is incomplete until that relationship is resolved or explicitly declared unresolved.

4. Arrays are the measurement substrate

A displayed microscopy image is a mapping from numerical values to visible brightness and color. The numerical array, not the screen rendering, is the substrate for deterministic measurement.

This distinction becomes visible whenever display settings change. Black and white points may be adjusted. Gamma may alter apparent contrast. A lookup table may convert one channel to

green and another to magenta. A Z stack may be projected. A time series may be reduced to one frame. A three-dimensional volume may be rendered from a chosen angle with selected opacity and threshold settings.

Those operations can be scientifically useful. They can make structure visible, support review, or communicate a result. They do not rewrite the recorded numerical values unless a new transformed dataset is explicitly created.

The governed acquisition model must therefore preserve three separable states:

- Source array state: the decoded numerical evidence associated with the native record.
- Analytical state: the explicitly transformed or selected data used by a deterministic method.
- Presentation state: the display mapping used for human review or communication.

A presentation image may help a scientist decide what to inspect. It cannot silently become the source of an intensity, distance, volume, temporal, or spatial-relationship measurement when the native numerical data are available.

5. Axes determine what an index means

A number in an array has no complete scientific meaning until its axes are resolved.

The same data dimensions can be interpreted incorrectly if axis order is wrong. A sequence may represent Z rather than time. Several series may represent positions rather than repeated measurements. Channels may be interleaved. Tiles may require spatial assembly. A multidimensional dataset may contain more than one scene, pyramid level, or resolution representation.

Axis governance therefore preserves identity, order, length, physical unit and increment where available, origin or coordinate mapping where relevant, relationships among series, scenes, positions, tiles, and pyramid levels, and any ambiguity, reader warning, or user-confirmed mapping.

This is not a software detail. Axis identity determines eligibility. A two-dimensional plane cannot support a true volumetric measurement. A Z series cannot be treated as time. A maximum-intensity projection cannot substitute for its underlying stack. A pyramid level intended for rapid viewing cannot automatically replace the full-resolution source for quantitative analysis.

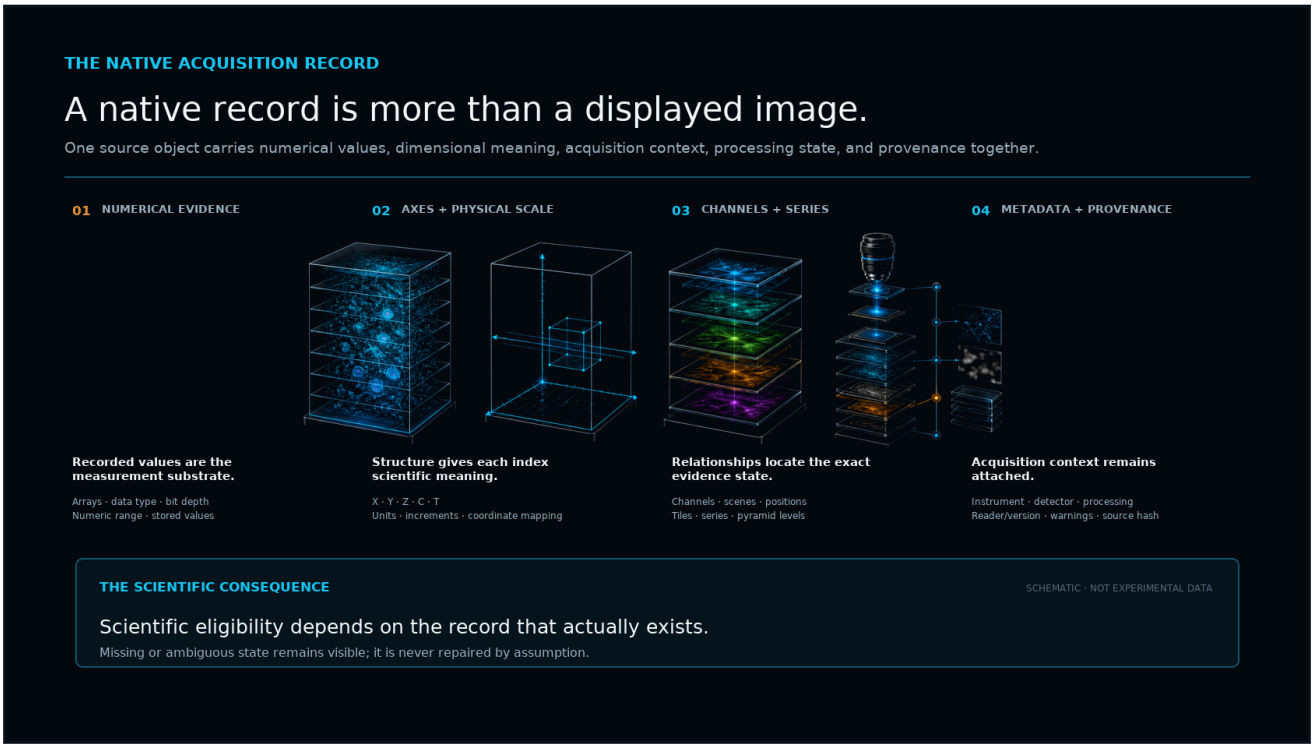

Figure 2. The anatomy of the native acquisition record. One source object resolves into numerical arrays, axes, physical calibration, channels, series relationships, acquisition metadata, processing state, and provenance. Missing or ambiguous fields remain visible rather than filled by assumption.

6. Physical scale turns array distance into measurement

Pixel and voxel indices are not physical units. Distance, area, volume, velocity, density, and many spatial comparisons require a valid mapping from array coordinates to physical scale.

That mapping may include X and Y pixel size, Z spacing, time interval, stage position, wavelength, and modality-specific calibration. The units must remain attached. A value of 0.1 is not scientifically interpretable as a spatial increment unless the record establishes whether it means micrometers, nanometers, millimeters, or another unit.

Calibration also has a scope. A nominal objective and camera combination does not guarantee that every dataset carries correct physical scale. Reconstructed modalities may change output sampling. Cropping, binning, resampling, stitching, or export may alter array geometry. Stage coordinates and image coordinates may occupy different frames of reference.

The acquisition record must distinguish recorded physical calibration from calibration derived through a declared transformation, calibration supplied or corrected from authoritative evidence, nominal instrument information that has not been validated for the current record, and calibration that is missing or contradictory.

When authoritative evidence supports a calibration correction, the corrected value becomes a declared derivative state that preserves the original recorded value, the revised value, units, supporting evidence, and affected scope. The native source is not silently rewritten.

7. Channels are acquisition paths—not biological identities by themselves

A native file may identify channels through numbers, detector paths, wavelengths, hardware names, or user-entered labels. A channel's display color is not its biological identity.

Paper 4 established the required upstream chain:

> target → primary reagent → conjugate or secondary reagent → fluorophore → optical channel → staining order → controls → measured relationship

Paper 5 preserves the acquisition side of that mapping. The native record should retain the channel structure and instrument metadata actually written by the acquisition system. The complete experiment supplies the biological and material mapping that explains what each channel was intended to represent.

Those two records must meet without being collapsed. A channel called "green" does not prove that it contains a particular target. A user-entered channel name may be informative but incorrect. Excitation and emission information can support a mapping but may not uniquely establish reagent identity. If the channel-to-target relationship is missing or contradictory, the workflow must require supporting evidence or restrict downstream interpretation.

Multicolor claims require this discipline. Apparent overlap between display colors is not evidence of direct molecular interaction. Before any relationship is measured, the channel identities, dimensional state, registration, controls, and measurement definition must be eligible.

8. Metadata are evidence—but metadata can be partial

The Open Microscopy Environment established a common data model for biological imaging, and Bio-Formats provides readers that translate many proprietary file structures into standardized pixel and metadata representations.[5,6] Bio-Formats distinguishes core metadata needed to interpret the basic pixel structure, original format-specific metadata, and OME metadata translated into the common model.[6]

That distinction is important because metadata translation is not automatically complete. Vendor formats differ. Fields can be absent, inconsistently named, stored in companion

objects, unsupported by a reader, or mapped imperfectly. The Bio-Formats documentation explicitly describes metadata conversion as useful but not complete or error-free.[6]

Community standards reinforce why metadata matter. REMBI separates the semantic question of what information is needed to understand and reuse biological images from the syntactic question of how a data model represents it.[7] The 4DN-BINA-OME specifications use tiered requirements that scale with experimental complexity and extend the OME model for hardware, acquisition, and quality-control metadata.[8] QUAREP-LiMi addresses instrument and image quality-control practices needed for reproducibility.[9]

These standards should be used without overstating what they establish. A standard representation improves access and comparability; it does not certify the correctness of the source metadata, the performance of the instrument, or the suitability of the acquisition for a particular claim.

The governed rule is:

> Preserve what the native record says, preserve how it was translated, and preserve what remains unknown.

9. Native does not always mean raw—and raw does not mean correct

Microscopy vendors and modalities use “raw,” “native,” “processed,” and “reconstructed” differently. A native file is the format produced by the acquisition system, but its numerical content may already include instrument-side correction, reconstruction, averaging, compression, stitching, deconvolution, spectral unmixing, or localization fitting.

Single-molecule localization illustrates the boundary. Camera frames and a reconstructed localization table are different evidence objects. Structured illumination data and a reconstructed SIM image are different evidence states. A confocal acquisition and a deconvolved volume are related but not interchangeable. A whole-slide image may include multiple pyramid levels designed for different viewing scales.

The acquisition framework must therefore record the evidence state rather than rely on the word “native” as a quality claim. For each supported dataset, the governed record should identify, where possible, what numerical object is present; whether it is detector-level, corrected, reconstructed, rendered, or derived; which processing steps and source objects are recorded; which parameters and software versions are available; whether the transformation is reversible or reproducible; and which measurements are eligible from that state.

Raw data can still be saturated, undersampled, noisy, poorly calibrated, or biologically compromised. Native acquisition status preserves authority for what was recorded; it does not waive quality control.

10. Exports and displays are derivatives with defined jobs

Native microscopy data may arrive in proprietary formats such as Leica .lif, ZEISS .czi or .lsm, Nikon .nd2, and Olympus/Evident .oir, .oib, .oif, or .vsi, among others. These native acquisition datasets are distinct from exchange, analytical, or presentation representations such as OME-TIFF, OME-Zarr, TIFF, PNG, JPEG, CSV, rendered volumes, projections, and publication figures.

None of these derivative representations is inherently invalid. Each can serve a legitimate scientific, computational, or communication purpose. The problem arises when a derivative loses its declared relationship to the native source or silently inherits authority that it no longer contains.

A multichannel Z stack illustrates the distinction. The native dataset preserves the recorded planes, channels, calibration, and acquisition context. A selected plane or subvolume defines a bounded analytical scope. A projection creates a useful overview while collapsing declared dimensional information. A publication figure communicates a selected view. Each representation can perform its assigned job, but none may silently replace the authority or context of its parent.

The following boundaries should remain explicit:

Representation	Legitimate role	Authority that must not be assumed
Native acquisition dataset	Authoritative computational record of the instrument output within its recorded state	Absolute biological truth or universal measurement eligibility
Lossless normalized representation	Cross-format computation with declared mappings and provenance	Replacement of the immutable source object
Selected plane or subvolume	Bounded analysis or review scope	Representation of unselected dimensions
Projection	Overview or declared projection-based analysis	Preservation of axial separation or volumetric structure
Display rendering	Human review and communication	Quantitative intensity or color authority
Screenshot or presentation figure	Communication of a selected view	Recovery of native arrays, metadata, scale, or full context
Segmentation mask or measurement table	Derived analytical evidence under a defined method	Independent source evidence

Every derivative needs a parent. Its provenance should identify the source dataset, selected series and axes, transformation, parameters, software or method version, scientist actions where relevant, and integrity state.

11. Normalization is a governed translation—not a replacement

Cross-vendor operation requires harmonization. A deterministic system cannot execute reliably if every proprietary format is treated through ad hoc assumptions. Bio-Formats, the OME data model, OME-TIFF, and OME-NGFF demonstrate the value of common representations for heterogeneous biological imaging data.[5,6,10]

Cross-vendor computation therefore requires normalization, but normalization must remain subordinate to source preservation.

A governed normalization contract preserves the immutable native dataset and its cryptographic identity; the structure, reader, library, and version required for decoding; the resolved series, axes, dimensions, units, scale, data type, bit depth, numeric range, channels, and acquisition metadata; accessible original vendor metadata; the identity of the normalized representation; all warnings, unsupported fields, ambiguities, and user-confirmed mappings; and machine-readable lineage from each normalized element back to its source.

Normalization must not silently apply background subtraction, contrast stretching, gamma, thresholding, channel reassignment, Z projection, resampling, denoising, deconvolution, segmentation, or biological interpretation. Those are separate operations with separate states.

OME-NGFF addresses scalable access to large multidimensional bioimaging data and proposes common metadata across storage vessels such as Zarr.[10] That capability is valuable for computation and sharing. It does not change the authority rule: an interoperable representation is scientifically strongest when its exact relationship to the native source is preserved.

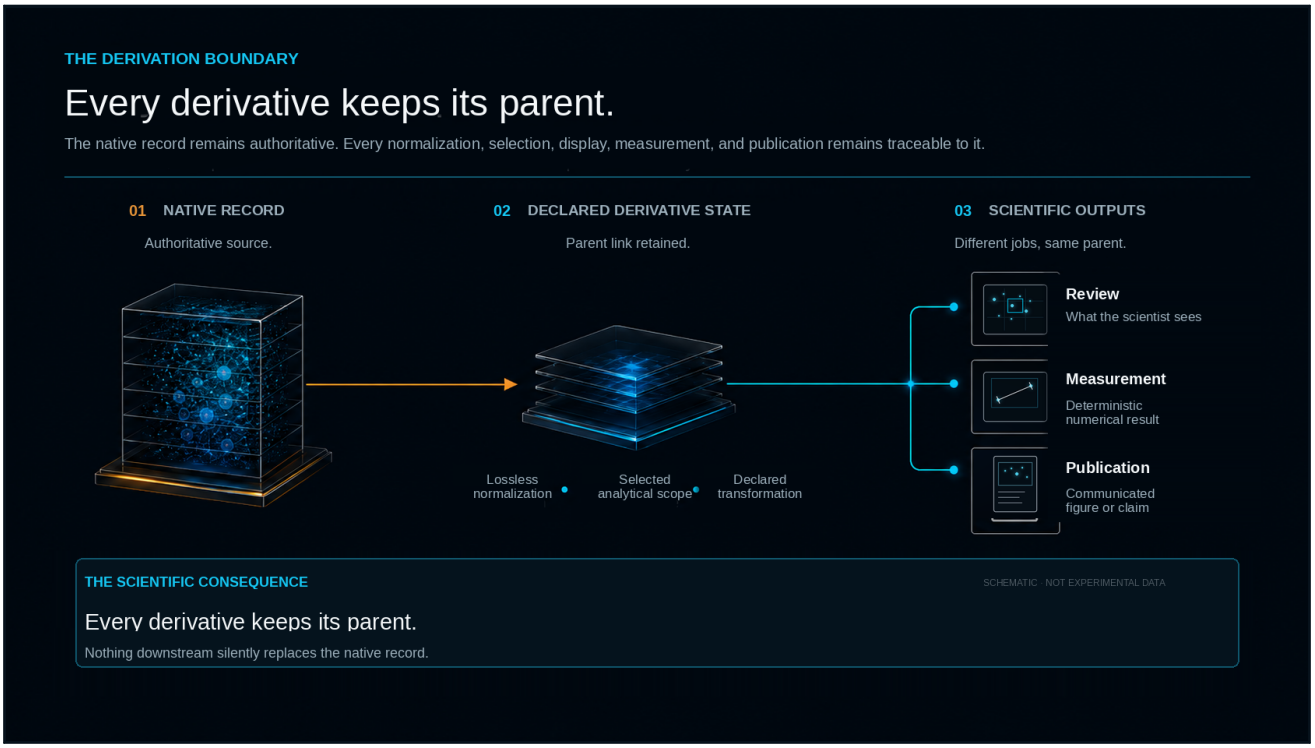

Figure 3. The derivation boundary. The native acquisition record remains authoritative while normalized representations, analytical selections, review displays, measurements, and publication objects remain declared descendants with parent lineage preserved.

12. File identity and provenance make the route auditable

Filename, folder location, and upload time are not sufficient identifiers. Files can be renamed, copied, exported, or replaced while retaining similar names.

A governed acquisition record should preserve cryptographic identity for every source file or required dataset component; source size and structural inventory; acquisition or creation timestamps without treating filesystem copy time as equivalent; original source context where appropriate; available instrument and acquisition-software information; ingest time and responsible user or system; reader and version; normalized-object identity; warnings and validation status; and parent-child relationships for every derivative.

A hash establishes identity, not scientific validity. Two matching hashes show that two binary objects are identical. They do not prove that the acquisition was well designed. Conversely, a different hash does not by itself establish a scientifically meaningful difference; the change must be characterized.

The purpose of identity control is to make later questions answerable. Which exact acquisition was measured? Did the source change? Which reader decoded it? Which series and channels

were selected? Was the physical scale recorded, corrected, or assumed? Which output came from which state?

13. The minimum BENNETT acquisition record

Before downstream measurement, the governed acquisition record must represent the following state or explicitly declare it missing:

Domain	Minimum state	Scientific question protected
Source identity	Dataset components, filenames, sizes, hashes, source relationship	Which exact acquisition object is authoritative?
Format and reader	Detected format, reader, reader version, import status, warnings	How was the source decoded?
Numerical structure	Data type, bit depth, numeric range, dimensions, array shape	What values were recorded and how are they organized?
Axes and series	Axis identity/order, series, scene, position, tile, pyramid level	What does each index represent?
Physical calibration	X/Y scale, Z increment, time interval, units, coordinate mapping	Which physical measurements are eligible?
Channels	Channel structure, instrument fields, wavelength/detector context, upstream target mapping state	What acquisition path produced each signal?
Acquisition context	Instrument, objective, detector, exposure/dwell, gain, illumination and modality fields where available	Under what recorded conditions was the evidence acquired?
Processing state	Raw/corrected/reconstructed/stitched/deconvolved/localized/rendered state and available parameters	What transformations already condition the numerical evidence?
Quality and limitations	Saturation, missing metadata, unsupported fields, ambiguity, calibration or reconstruction concerns	What must be reviewed or blocked?
Normalization lineage	Source-to-normalized mappings, software/version, normalized hash	Can computation be traced back to the native source?
Authoritative corrections	Confirmed mappings or corrections, source value, revised value, units, supporting evidence, affected scope	Which declared correction changed the governed acquisition state?

This record does not need every possible metadata field before any work can begin. Requirements should be proportional to the modality, complexity, intended analysis, and claim, consistent with tiered community metadata approaches.[7,8] But a missing prerequisite must remain missing. It cannot be converted into an assumed fact merely to keep the workflow moving.

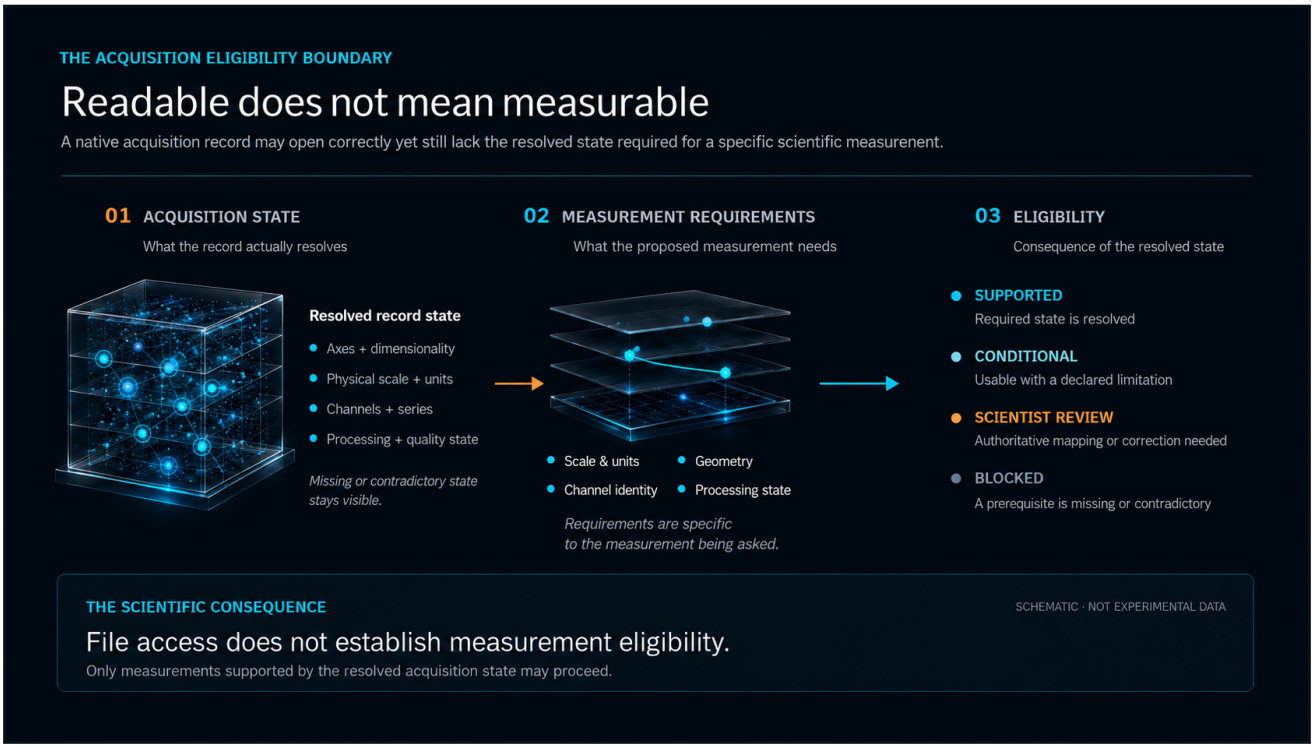

Figure 4. The acquisition eligibility boundary. A readable file does not make every measurement eligible. The resolved acquisition state determines whether a proposed measurement is supported, conditional, requires scientist review, or is blocked.

14. What this framework does—and does not—establish

This framework can establish which native dataset is authoritative, whether required companion objects are present, which arrays and acquisition structures were decoded, which metadata and processing state were recovered, which reader performed the translation, what remains missing or unresolved, how derivatives relate to the source, and which analyses are eligible from the resolved acquisition state.

It does not make the acquisition absolute biological truth, make native synonymous with raw, certify raw data as high quality, guarantee that vendor metadata or standardized conversion are complete, make a screenshot or projection numerically equivalent to the source, permit calibration to be inferred from appearance, make a channel name proof of biological identity, authorize every available analysis merely because an array can be accessed, or allow an LLM to supply missing values or become the source of measurement.

The native file establishes acquisition authority. Scientific eligibility still depends on the complete experiment, material competence, modality, calibration, quality state, analytical method, and intended claim.

15. The native record hands evidence to governed practice

Papers 1–5 now establish the evidence that a governed scientific system must preserve:

- Paper 1 establishes why observation creates scientific responsibility.
- Paper 2 establishes the complete experiment as the connected evidence-bearing object.
- Paper 3 establishes that observations and claims must be tested against the competence of the evidence environment.
- Paper 4 establishes the competence and provenance of the material presented to the microscope.
- Paper 5 establishes the authority and provenance of the record produced by the microscope.

The next question is not whether the file contains data. It is how expert scientific practice becomes reproducible operation around those data.

Paper 5 therefore hands Paper 6 this question:

> Once native acquisition evidence and its boundaries are preserved, how can expert microscopy practice be converted into explicit, reproducible states, gates, decisions, and validation rules?

Paper 6 answers by moving from expert practice to governed scientific intelligence.

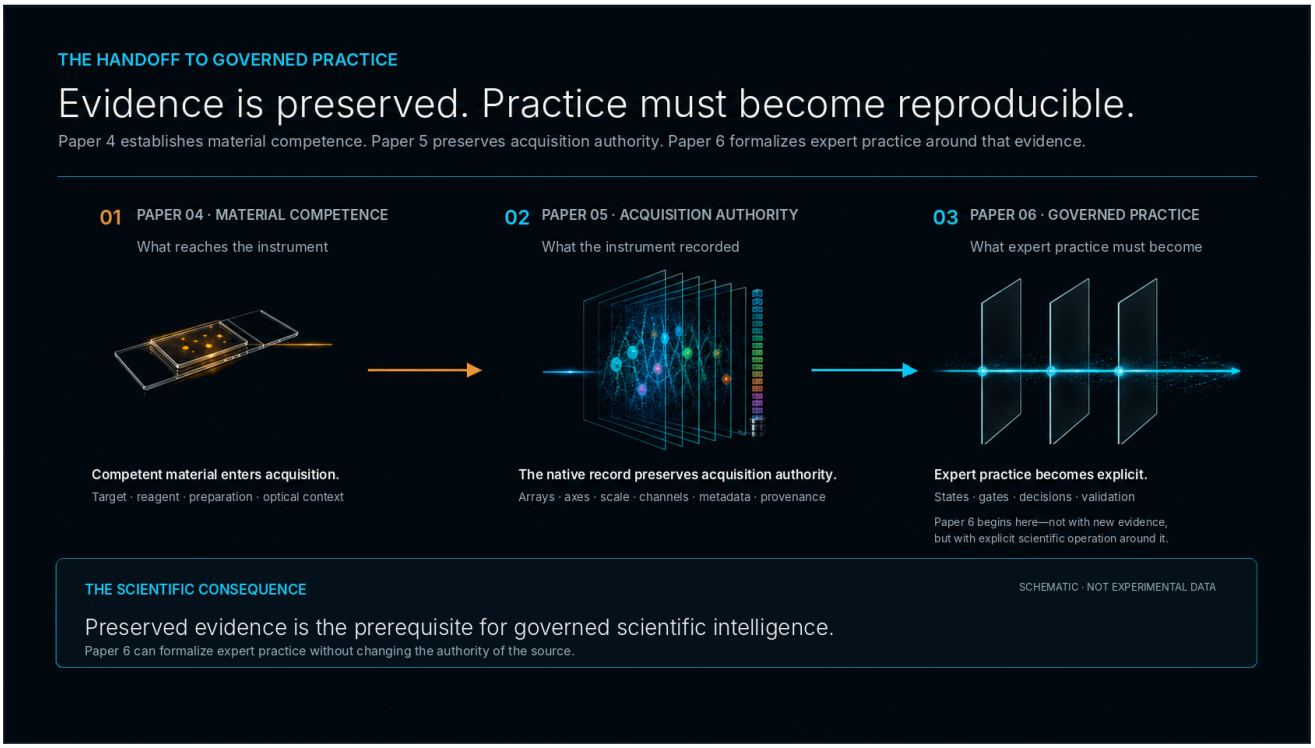

Figure 5. The handoff to governed practice. Paper 4 establishes material competence; Paper 5 preserves acquisition authority; Paper 6 begins with explicit scientific states, gates, decisions, and validation around preserved evidence.

Conclusion

The microscope does not produce a generic picture. It produces a structured acquisition record whose numerical arrays, dimensions, physical calibration, channels, series, metadata, processing state, and provenance define what downstream computation has received.

That record is the Acquisition GroundTruth™. The phrase does not elevate an image to biological truth. It establishes a controlled authority boundary. The native acquisition object is the reference for what the instrument recorded; the complete experiment provides the scientific context that determines what the record can support.

This distinction prevents several common substitutions. Display is not measurement. Color is not channel identity. Array index is not physical distance. A projection is not the underlying volume. A normalized representation is not an unrecorded replacement for the source. A successful file read is not proof of complete metadata. A native file is not automatically raw, artifact-free, calibrated, or eligible for every analysis.

Open standards make heterogeneous microscopy evidence more accessible and reusable, but accessibility does not supersede source authority. Interoperable representations must be used without erasing provenance, uncertainty, or format-specific state that remain scientifically consequential.[5–11]

The governing rule is simple:

> Preserve the native source. Resolve what it contains. Declare what remains unknown. Record every transformation. Measure from eligible numerical evidence. Keep every downstream representation accountable to its origin.

Paper 4 established the material presented to the instrument. Paper 5 establishes the record produced by the instrument. Paper 6 can now ask how expert practice becomes an executable, governed scientific system around that evidence.

References

1. Bennett BT. Antibodies You Can Trust. Data You Can Defend. BENNETT White Paper 04. Canonical locked v1.1, 2026.

2. Bennett BT. Resolution Stepdown. BENNETT White Paper 03. Canonical locked v1.1, 2026.

3. Bennett BT. The Complete Experiment. BENNETT White Paper 02. Canonical locked v1.3, 2026.

4. Bennett BT. From 4Pi Microscopy to Governed Scientific Intelligence. BENNETT White Paper 01. Canonical locked v1.3, 2026.

5. Goldberg IG, Allan C, Burel JM, et al. The Open Microscopy Environment Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biology. 2005;6:R47. doi:10.1186/gb-2005-6-5-r47.

6. Open Microscopy Environment. Bio-Formats documentation: data and metadata access, core/original/OME metadata, supported formats, and translation limitations. Official technical documentation. Accessed August 2026.

7. Sarkans U, Chiu W, Collinson L, et al. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. Nature Methods. 2021;18:1418–1422. doi:10.1038/s41592-021-01166-8.

8. Hammer M, Huisman M, Rigano A, et al. Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model. Nature Methods. 2021;18:1428–1440. doi:10.1038/s41592-021-01327-9.

9. Boehm U, Nelson G, Brown CM, et al. QUAREP-LiMi: a community endeavor to advance quality assessment and reproducibility in light microscopy. Nature Methods. 2021;18:1423–1426. doi:10.1038/s41592-021-01162-y.

10. Moore J, Allan C, Besson S, et al. OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. Nature Methods. 2021;18:1496–1498. doi:10.1038/s41592-021-01326-w.

11. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3:160018. doi:10.1038/sdata.2016.18.