

# An Auditable AI Assistant for Scientific Metadata Curation within ISAAC

Building an Evidence-Grounded Metadata Assistant and Record Verifier for Scientific AI

Krish Verma<sup>1,2</sup> · Angel Garcia<sup>1</sup> · Dean Skoien<sup>1</sup>

<sup>1</sup> SLAC National Accelerator Laboratory · <sup>2</sup> San José State University, Biomedical Engineering

LCLS Summer Internship Program · 2026



San José State University

CORE PRINCIPLE

AI-Assisted Metadata Must Be Grounded Before It Becomes a Scientific Record.



AI Assists



Rules Verify



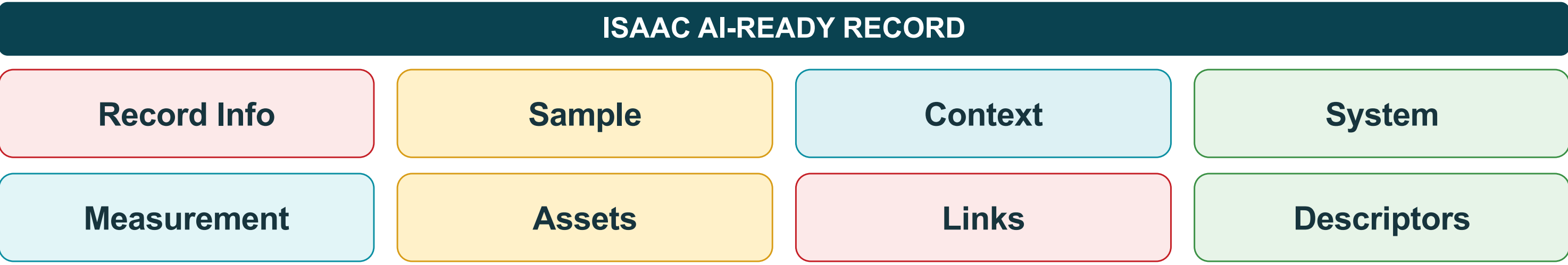
Scientists Decide

Every AI-assisted field remains traceable, attributable, and reviewable.

## 1 ISAAC and the Problem We Are Solving

ISAAC is a DOE multi-lab effort to turn heterogeneous catalysis data into interoperable, provenance-rich records for trustworthy AI-assisted discovery.

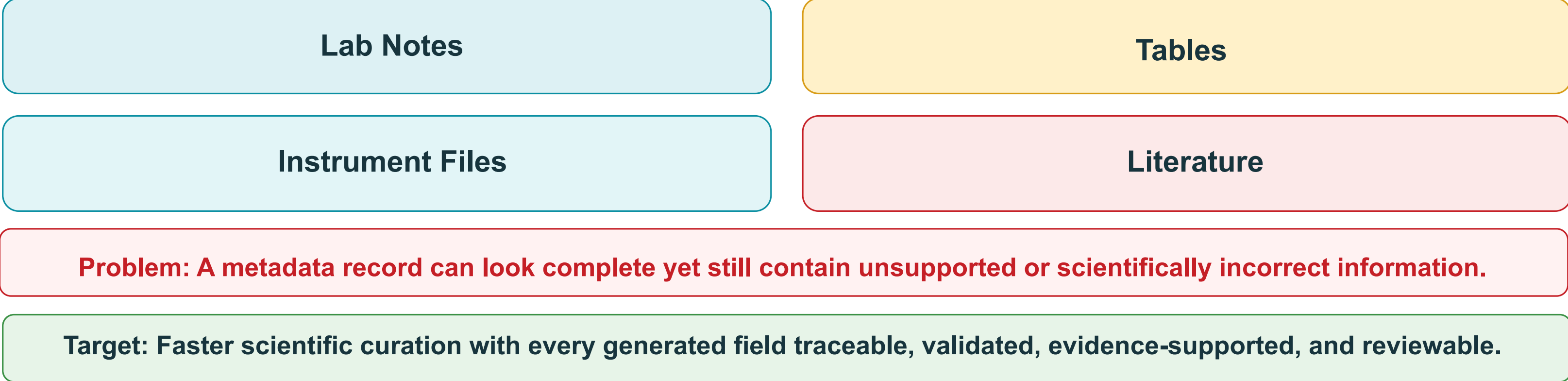
**My Focus:** I built and tested the metadata curation, schema validation, and evidence-verification layer that checks whether AI-assisted metadata is structurally valid and scientifically supported.



Each record connects measurements with experimental context, descriptors, provenance, and linked source assets.

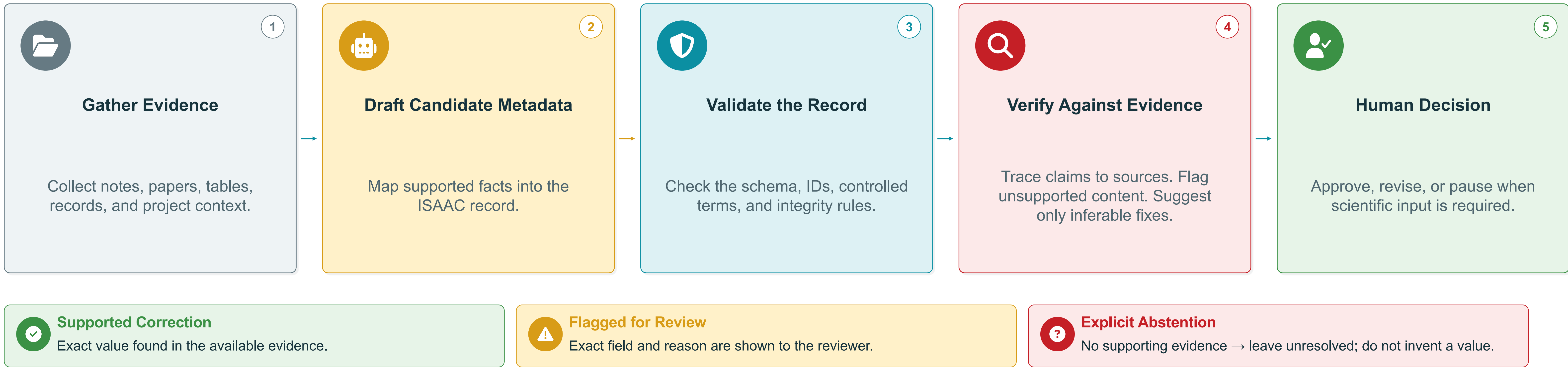
### The Problem / Target

Scientific evidence is fragmented across:



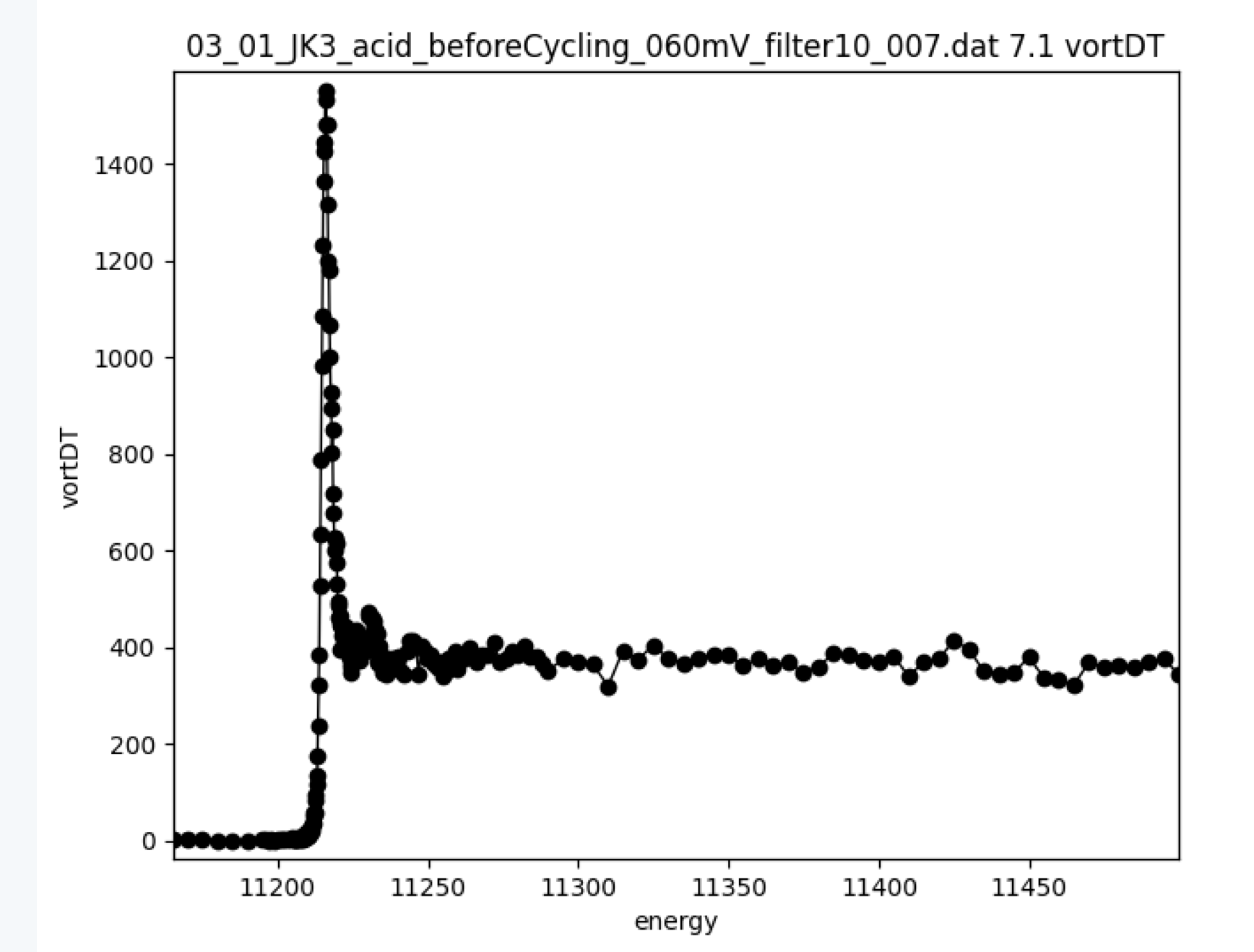
## 2 A Guarded Curation Workflow

Validation checks the record against rules; verification checks its claims against evidence.

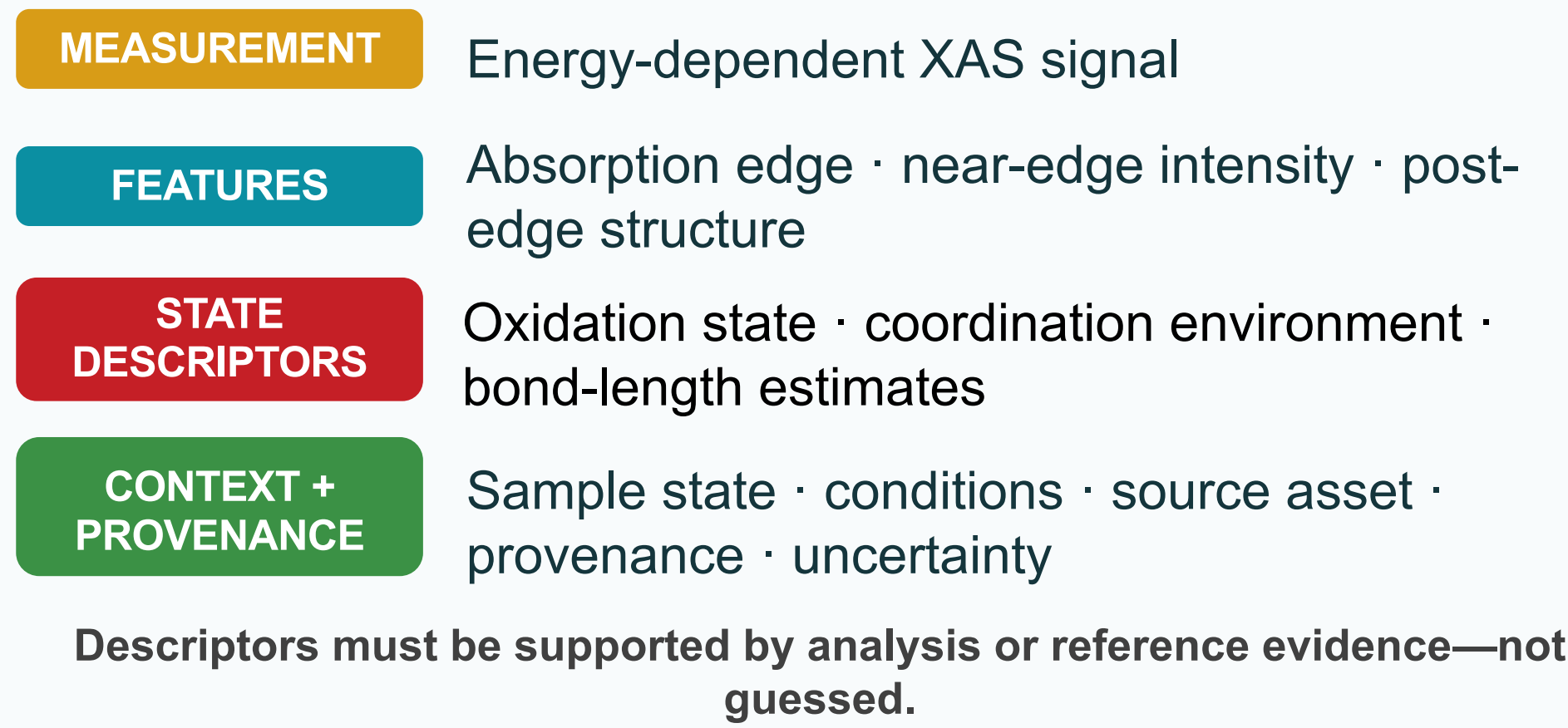


## 3 From Real Beamline Data to an Auditable ISAAC Record

Actual SSRL BL15-2 XAS data linked to experimental context, descriptors, and provenance.

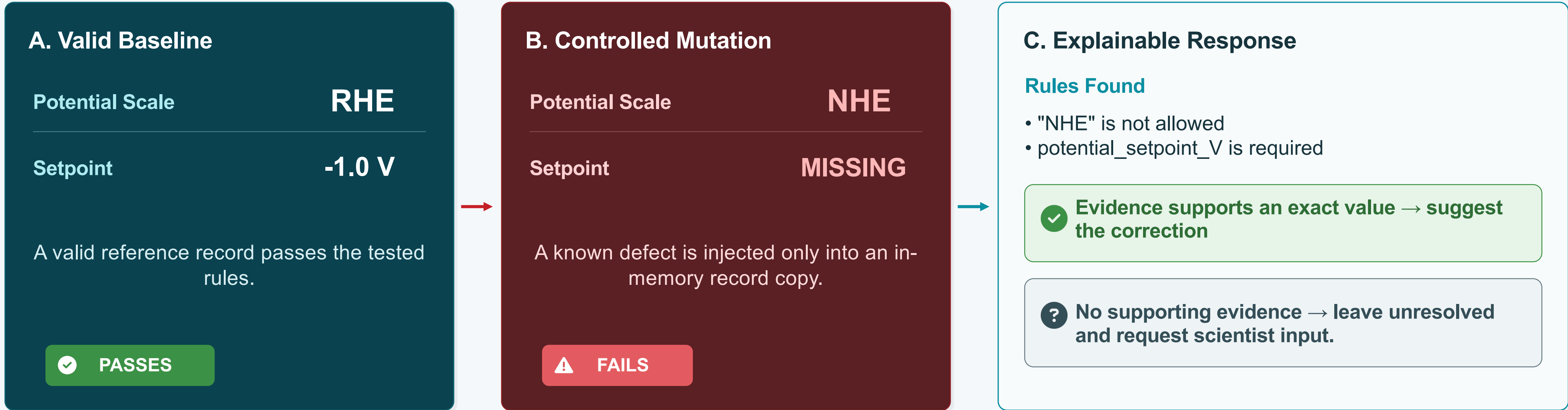


### HOW ISAAC STRUCTURES THE MEASUREMENT

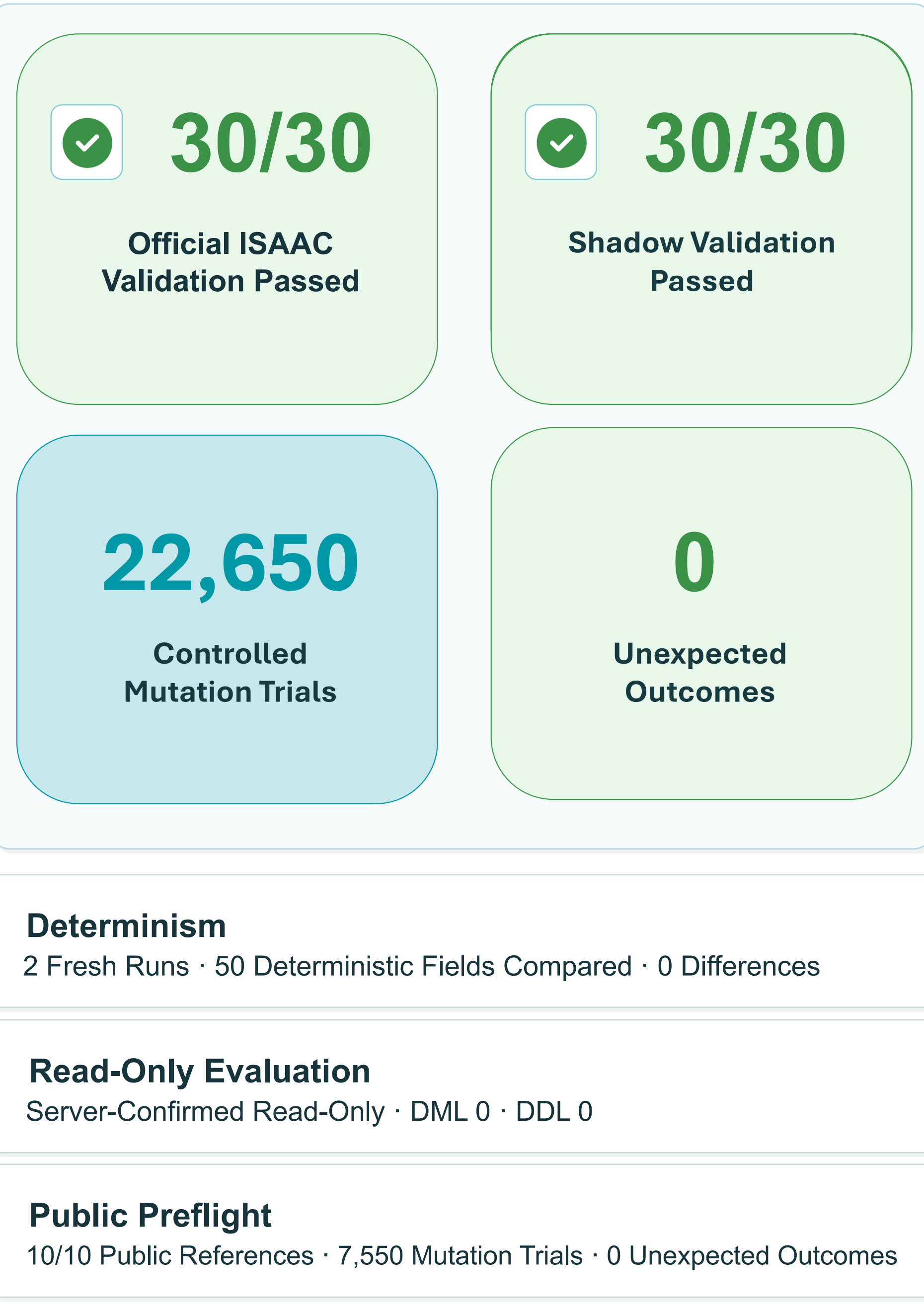


### Controlled Verification Test

Starting from a valid record copy, inject one known metadata defect and verify the expected response.



## 4 Evaluation



5

## Future Work

### Scientist-Friendly Data Capture

Make it easy for scientists to capture experimental context and metadata directly in ISAAC during their workflow.

### Cross-Facility State Descriptors

Standardize context-rich, uncertainty-aware state descriptors across experiments, techniques, and facilities.

### Physics-Informed Reasoning

Connect verified descriptors with mechanistic priors and simulations to support explanations of observed scientific behavior.

### Next-Experiment Recommendation

Use accumulated evidence and uncertainty to recommend the most informative next human-run experiment.

AI ASSISTS. RULES VERIFY. SCIENTISTS DECIDE.

AI-assisted metadata is validated against schema, verified against evidence, and reviewed before submission.

krish.verma2808@gmail.com