ZT-WP-002 | ZERO-TRANSIT RESEARCH WORKING PAPER

The Natural Fluctuation Reference Library: Reversible Baselines for Multi-Domain Anomaly Research

Sean Roberts
Zero-Transit Research Program, Stella Nova Education Foundation
8 August 2026

Research status: Foundational concept / methods architecture. This working paper defines testable research structure and does not report a validated long-range detection effect, operational observatory, government partnership, or deployed capability.

ABSTRACT

An anomaly is meaningful only relative to a defensible baseline. This paper specifies the Natural Fluctuation Reference Library (NFRL) as the comparative foundation of the Zero-Transit research program. The NFRL is designed to preserve known natural, environmental, instrumental, and anthropogenic events as versioned reference layers rather than subtracting them irreversibly from source observations. The architecture combines authoritative public feeds, local sensor context, provenance, timing uncertainty, and reproducible analytical transformations. The central design goal is epistemic: an unresolved residual should emerge because known explanations have been tested and documented, not because preprocessing silently removed inconvenient data. The paper defines baseline families, ingestion requirements, event localization, layer versioning, false-discovery controls, and criteria for promoting a recurring residual into a provisional pattern class.

Keywords: *natural baselines, reference library, provenance, space weather, seismology, anomaly detection, reproducibility*

1. Why a Reference Library Is Foundational

High-dimensional measurement systems can find patterns in almost any sufficiently large dataset. The scientific problem is therefore not simply detecting unusual values; it is determining whether unusual values remain unusual after the environment, instrument, and analysis history are considered. The NFRL addresses this problem by treating baseline knowledge as a first-class research object.

Instead of subtracting a natural model from raw data and keeping only the residual, Zero-Transit preserves the original observation and stores each baseline comparison as a named, versioned layer. An analyst can later remove, replace, or update a baseline without changing the source record. This design supports both falsification and retrospective discovery: an event initially unexplained may later match a newly characterized natural process.

2. Baseline Families

The initial taxonomy spans geophysical, atmospheric, ionospheric, solar and space-weather, cosmic/particle, oceanic and hydrologic, local environmental, and human/infrastructure context. The categories are deliberately broad because the observatory concept crosses conventional instrument boundaries. Each family can contain authoritative external records, local measurements, derived indices, and quality flags.

Two public sources already represented in the Zero-Transit application illustrate the approach. The U.S. Geological Survey provides programmatic GeoJSON earthquake feeds with event metadata and geometry suitable for time-location comparison. NOAA Space Weather Prediction Center publishes a one-minute planetary K-index JSON product for geomagnetic context. These feeds do not constitute Zero-Transit observations; they are independent reference layers against which local or synthetic observatory data can be compared.

3. Event-Centric Contextualization

The NFRL should support two complementary query modes. A time-centric query asks what known events were active around a candidate timestamp. A location-centric query asks what environmental processes were relevant near a sensor or inferred region. The combination allows the system to construct an event context packet without deleting any observation.

Context packets should record the external source, retrieval time, provider timestamp, geographic uncertainty, transformation steps, and the relationship between the source record and the observatory interval. If a feed is revised later, the new version should be stored without overwriting the original evidence used in an earlier analysis.

4. Provenance and Immutable Source Records

FAIR data principles provide a useful high-level target for reusable scientific data, especially when observations must be found, interpreted, and integrated by both people and machines (Wilkinson et al. 2016). For Zero-Transit, FAIRness is necessary but not sufficient. The system must additionally demonstrate derivation: what source produced a layer, which code transformed it, and who approved a classification.

PROV-O offers a standards-based vocabulary for representing entities, activities, agents, and derivation relationships (Lebo, Sahoo, and McGuinness 2013). The NFRL can map a raw external record to a normalized event entity, then to a spatial or temporal comparison activity, then to the resulting context layer. The key rule is that the lineage is append-only from the perspective of scientific evidence.

5. Pattern Classes Without Premature Explanation

Recurring residuals should initially receive neutral identifiers rather than explanatory names. For example, a pattern class might be defined by a reproducible temporal morphology, cross-channel coincidence, duration range, and spatial recurrence. The label should not contain “vacuum,” “quantum,” “propagation,” or another causal term unless an independent evidentiary pathway justifies that interpretation.

This naming discipline reduces confirmation pressure and supports blind comparison. A future analyst can test whether a new interval matches a class definition without knowing the favored mechanism. Pattern promotion requires recurrence, data-quality sufficiency, independence from known baseline families, and documented review.

6. Multiplicity and Discovery Control

Every additional sensor, time lag, frequency band, and contextual variable expands the number of possible comparisons. Without explicit control, the observatory could generate a steady stream of statistically impressive but accidental relationships. False-discovery-rate methods provide one established way to manage multiplicity when many hypotheses are tested (Benjamini and Hochberg 1995).

Zero-Transit should distinguish exploratory scanning from confirmatory testing. Exploratory results can be broad but must be labeled provisional. Confirmation should use held-out periods, predeclared windows, or independent sites. Repeated tuning on the same event cannot count as independent validation.

7. Operational Pathway

The NFRL begins before any custom detector exists. Phase one is taxonomy and ingestion: define schemas, import authoritative sources, and preserve source hashes. Phase two adds local environmental reference instruments and calibration records. Phase three builds automatic context packets around synthetic or known events. Only after the baseline machinery works should blind anomaly trials begin.

As the program expands geographically, baseline behavior must remain site-specific where necessary. A mountain observatory, urban laboratory, coastal site, and orbital platform experience different environmental regimes. Global architecture therefore requires shared semantics without pretending that one global background model explains every location.

8. Conclusion

The NFRL is intended to make the Zero-Transit program harder to fool. Its value is measured not by how many anomalies it preserves, but by how many plausible explanations it can test transparently while retaining the ability to revisit every decision. If the program ultimately finds no residual beyond known causes, the library still produces a reusable multi-domain environmental and instrumental reference system. If a residual survives, the NFRL provides the evidentiary record needed to show what was tested first.

References

U.S. Geological Survey. 2026. “Earthquake Hazards Program: GeoJSON Summary Feed and Format Documentation.” U.S. Geological Survey.

NOAA Space Weather Prediction Center. 2026. “Planetary K-index, One-Minute JSON Product and K-index Documentation.” National Oceanic and Atmospheric Administration.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” Scientific Data 3:160018. DOI: 10.1038/sdata.2016.18.

Lebo, Timothy, Satya Sahoo, and Deborah McGuinness, eds. 2013. PROV-O: The PROV Ontology. W3C Recommendation, 30 April 2013.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” Journal of the Royal Statistical Society: Series B 57(1):289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Beaulieu-Jones, Brett K., and Casey S. Greene. 2017. “Reproducibility of Computational Workflows Is Automated Using Continuous Analysis.” Nature Biotechnology 35(4):342-346. DOI: 10.1038/nbt.3780.

Suggested citation: Roberts, Sean. 2026. “The Natural Fluctuation Reference Library: Reversible Baselines for Multi-Domain Anomaly Research.” ZT-WP-002, Zero-Transit Research Working Paper. Stella Nova Education Foundation.