

EVIDENCE SCOPE AND DATA PROVENANCE

A Summary With a Gap Answers a Different Question

Scope, provenance and the limits of proportionate disclosure in computational-load evidence.

RESEARCH AS AT

7 May 2026

Three findings on how far a summary can be reduced before it answers a different question. Sources no later than this cutoff.

ABSTRACT

Evidence about a computational facility need not expose every aspect of it, but it must preserve enough context to answer its stated question. This paper establishes three limits on how far a summary can be reduced. An average computed over partial coverage is not an average of the period it appears to describe: the same figure of 14 MW is consistent with a true hourly average of 13 MW or 18 MW, depending on one missing quarter-hour. Provenance describes derivation, not truth - it establishes neither the accuracy of the instruments nor the coverage of the interval, and the word verified collapses two separate uncertainties into one. And withholding material as confidential leaves the proposition it would have addressed unresolved; it does not establish it. Proportionality is a question about which context is load-bearing, not about file size.

01

Specify the purpose

02

Preserve context

03

Limit disclosure

Conceptual sequence. Independent evidence does not replace the recipient's decision.

Edition and research provenance

Research as at 7 May 2026. The evidence set is the dated CANUS submission to the AEMC and fixed provenance, audit and timestamping standards; nothing later is used. Findings 1 and 3 are deductive. Company-authored research, not peer-reviewed. Not legal advice.

Research scope: Three limits on proportionate disclosure, with worked arithmetic. No implemented platform, recipient right, physical validation or empirical benefit is reported.

1. Principal findings

Finding 1 - A partial-coverage average is not an average of the period it appears to describe. Where a summary reports a mean over an interval it did not fully cover, the figure is arithmetically correct and the period attached to it is wrong. Section 4 shows one missing quarter-hour moving the true hourly answer across a 5 MW range without changing the reported number at all.

Finding 2 - Provenance describes derivation, not truth. A record of who produced an information object and how it was transformed supports review of the derivation. It does not establish that the instruments were accurate, that the source file covered the intended interval, or that the number suits a question it was not computed for. Document history and physical measurement support are distinct, and a package that reports both as “verified”; has reported neither.

Finding 3 - Confidentiality cannot substitute for support. Where material is withheld as confidential, the proposition it would have addressed remains unresolved. Labelling the omission does not change its effect on the conclusion. The available responses are authorised restricted access, a narrower conclusion, or leaving the question open - not treating the withheld support as supplied.

2. Proportionality is about load-bearing context, not size

Data minimisation does not mean producing the smallest file. A one-line assertion is compact and unusable. Assurance does not mean obtaining every operational record either: more information raises review burden and exposure without necessarily resolving the uncertainty that matters.

The workable starting point is a purpose statement. Which question must be answered, for which boundary and period, and with what consequence if the conclusion is wrong? Each information category requested should carry an explanation connecting it to that purpose. A request without one is a candidate for removal, narrower access, or a less revealing substitute.

Substitution has limits, and they are the subject of Finding 1. A summary is inadequate where it hides a material interval, combines unlike boundaries, or drops the qualification its conclusion depends on. In the other direction, a reviewer should not demand workload content merely because it exists, where the question concerns an electrical quantity assessable without it.

3. What a disclosure decision actually turns on

Decision question	Context that is load-bearing	Disclosure needing separate justification
What happened at the connection point?	Quantity, boundary, period, material limitations	Unrelated workload or customer content
Is a summary representative of its stated interval?	Coverage, exclusions, aggregation description	Detailed records irrelevant to that interval
Does net behaviour describe underlying load?	Relevant generation and storage distinctions	Internal business information unrelated to attribution
Can a reviewer understand the reported number?	Source identification, calculation description, qualifications	Unrestricted access to every source system
Does the information suit a different question?	Comparison of the new purpose with the original scope	Assumed entitlement to all underlying material

Neither total disclosure nor complete secrecy is a default. Additional information needs a reason, and so does each omission.

4. The arithmetic of Finding 1

Take a request for the average net import of a facility over one hour. The source file holds three consecutive fifteen-minute averages: 12 MW, 12 MW, 18 MW. The fourth is missing. Equal interval lengths, consistent units, no other complications.

The mean of the available values is 14 MW. Reporting that as the hour's average changes the question: it is the average of the observed forty-five minutes. If the missing interval averaged 10 MW, the complete hour averages 13 MW. If it averaged 30 MW, the hour averages 18 MW. With no constraint on the missing value beyond non-negativity, the hourly average is bounded below at 10.5 MW and unbounded above - while the reported figure stays at 14 MW throughout.

Summary option	What it communicates	Assessment
Hourly average: 14 MW	Implies complete-hour coverage	Unsupported by the stated inputs
Observed-period average: 14 MW; 45 of 60 minutes available	Preserves the gap and the denominator	Accurate description of a limited calculation
Full customer and workload export	Reveals additional operational content	Does not recover the missing electrical interval

The third row is the point of contact with Finding 3. Volume is not support. An export containing every customer record in the facility leaves the hourly question exactly as unresolved as the one-line summary does, while exposing material the question never required.

A proportionate package can describe the observation period, the missing interval, the available-period calculation and the resulting limitation without identifying whose jobs were running. That is enough to understand why the hourly question remains open. It is not enough to conclude that any electrical limit was met. If the missing observation later becomes available, the new calculation should state its expanded coverage rather than recast the earlier one as complete.

5. Where a summary fails

The analyst belongs in the failure model: a conscientious editor can remove load-bearing context while trying to simplify a report.

Failure mode	Risk to the decision	Review question
Missing interval omitted from the summary	Partial coverage appears complete	Does the denominator match the stated period?
Boundary relabelled during aggregation	A site result becomes a subsystem claim	Is the physical subject unchanged?
Confidentiality used in place of support	An unexamined proposition appears established	What can be concluded from what is available?
Excessive information requested	Avoidable disclosure and review burden	Which decision depends on each category?
Source material altered without explanation	The summary cannot be reconciled with its inputs	Is the basis of the calculation intelligible?

W3C PROV-DM supplies the vocabulary for the second and fifth rows: concepts for information objects, the activities producing or transforming them, the responsible agents, and the derivation and revision relations between them. [2, sections 2.1, 5.2, 5.3] It describes how a number came to exist. Finding 2 is the limit on what that description carries.

NIST SP 800-53 revision 5 addresses protection of audit information against unauthorised change and its retention for retrieval. [3, controls AU-9 and AU-11] RFC 3161 describes timestamping as support for evidence that data existed before a stated time. [4, section 1] Neither establishes public availability, nor the truth of the recorded event.

6. What follows

The three findings describe a tension rather than a defect: the same reduction that removes unnecessary exposure can remove the context a conclusion rests on, and both effects are invisible in the finished summary. A reader cannot tell from a well-formed number whether its denominator matches its stated period.

Two paths lead from there, and this paper does not choose between them.

The first: the reader asks. A sophisticated recipient specifies coverage, boundary and qualification, and rejects a package that omits them.

The second: that capacity is conditional on the reader being present when the requirement was written, and on the summary reaching them directly. Where evidence is produced for one purpose and later read for another – by a financier, an insurer, a purchaser – the conditions in Finding 1 and Finding 3 are not met by default, and the failure is silent.

Testable claim: given matched packages of identical underlying material – extensive source records, a purpose-specific summary retaining load-bearing context, and an over-reduced summary – qualified reviewers will differ materially in the conclusions they accept, and the over-reduced group will express higher confidence in unsupported answers than the extensive group. The proposal for proportionate summaries is weakened if it conceals material gaps, if it requires reconstruction merely to establish its own scope, or if material withheld as irrelevant repeatedly proves necessary. That study has not been performed. It is stated so the claim can be tested rather than asserted.

7. Limits

The dated CANUS submission to the Australian Energy Market Commission argues for evidence appropriate to a defined electrical boundary without unnecessary disclosure of customer, workload, model or internal operational information, and distinguishes net connection-point behaviour from underlying consumption where generation and storage coexist with load; section 9 recommends proportionate requirements. [1, pages 5–7] It supports the policy concern this paper works from. It does not document an implemented data-provenance system.

The unit of analysis is an evidence package for a specified electrical assessment, not a continuing commercial record system. Nothing here creates a right to obtain confidential information, binds a decision-maker, or alters an operator's existing obligations. The figures in section 4 are an explanatory case with stated assumptions, not observations from a facility. The cited standards are background for the distinctions drawn and not evidence that any particular package complies with them. No implemented platform, recipient right, physical validation or empirical benefit is reported. Findings 1 and 3 are deductive; Finding 2 is a statement about what a provenance record contains.

References

1. CANUS Technologies. Submission to the Australian Energy Market Commission: Improving the NEM Access Standards - Package 2 (ERC0394). Document dated 7 May 2026. Eight pages; page 1, sections 6-7, pages 5-6, and section 9, page 7.
2. W3C. PROV-DM: The PROV Data Model. W3C Recommendation, 30 April 2013. Sections 2.1, 5.2 and 5.3. Fixed-version official URL: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>.
3. NIST. Security and Privacy Controls for Information Systems and Organizations. SP 800-53 revision 5, September 2020; cited edition incorporates updates dated 10 December 2020. Controls AU-9 and AU-11, printed pages 74-75 and 77. Official edition URL: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf>. Later catalogue releases are not used.
4. IETF. Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). RFC 3161, August 2001. Section 1, printed pages 1-2. Fixed RFC URL: <https://www.rfc-editor.org/rfc/rfc3161.html>.