# Accelerating Commercial Data Integration through Fit-for-Purpose Certification

*A Framework for Scalable, Trust-Based Earth Data Certification*

**Zorana Jelenak**

UCAR/CPAESS

*We don’t accelerate transition into operation by doing less validation; we accelerate it by doing the right validation at the right time.*

## 1. Executive Summary

**The transition of commercial Earth Observation data into government operations is too slow.** Innovative sensors launch faster than the system can absorb them, and calibration and validation (Cal/Val) is one of the critical gates every dataset must pass through on the way to operations. When that gate lacks a common standard and its results are not portable across the community, effort is inevitably repeated and the bottleneck compounds every other source of delay. Meanwhile, AI and machine learning systems increasingly consume Earth observation data with no common assurance of quality, characterization of uncertainty, or fitness for use. Streamlining this gate, without weakening it, is one of the most direct levers available to accelerate the entire transition pipeline.

**This paper shows how to fix that without compromising quality.** Speed and rigor are not in conflict. The framework proposed here accelerates transition by structuring validation around applications, running stages in parallel, and reusing completed work across missions, not by cutting corners or lowering the bar. Every quality gate remains enforced. Every dataset earns its certification through measured performance and demonstrated methodological rigor. Critically, this framework certifies the Cal/Val process itself, not just the data products it generates: the methods, tools, independence, and reproducibility of the validation are scored alongside the empirical results.

The core problem is structural. Commercial data providers want to begin earning returns as soon as they launch. Government agencies need high-confidence data to support operational decisions. Other buyers (international agencies, research institutions, the private sector) need to understand what the data can and cannot do before they commit resources to it. Yet the dialogue about data quality and readiness starts far too late. The process of understanding what a vendor has, what it is suitable for, and what additional Cal/Val work is needed should begin much earlier, so that users are prepared, vendors have performed as much validation as they can, and both sides share a common language for data maturity.

Today, that early engagement rarely happens in a structured way. Without a shared framework that defines what adequate Cal/Val looks like for a given application, the level of investment in validation naturally varies across the industry. This is not a reflection of vendor intent but of a system that provides no standardized expectations, no clear return on quality investment, and no common language for communicating data maturity. Even agencies and organizations that prioritize data quality often lack the tools to articulate precisely what they need. The result is a structural disconnect: vendors cannot effectively signal quality, buyers cannot consistently specify it, and the transition to operations stalls.

This white paper presents a certification methodology that addresses all sides of this problem. Unlike current approaches that rely primarily on verifying a few mission-imposed instrument performance thresholds, this methodology certifies the Cal/Val process itself, not just the outputs it produces. It evaluates how rigorously, independently, and reproducibly validation was performed, and then measures data quality against metrics that matter for specific applications. The result is not a single pass/fail verdict but a quantitative, application-specific maturity score that any stakeholder can interpret and act on.

The methodology stands on its own. Any organization that adopts its two-stage architecture, four-tier scoring, and application-based thresholds can produce high-confidence certification results, whether or not a centralized body exists. That said, the paper also proposes the establishment of a Data Certification Center: a neutral, nonprofit-hosted body that provides the authoritative implementation currently missing from the ecosystem. International Cal/Val standards exist, but no entity today issues certifications against them, maintains a shared registry of results, or accredits the organizations that perform validation. The center fills that gap. It issues data certifications based on published, quantitative criteria. It accredits independent validators (government labs, universities, and commercial firms alike) so that certified Cal/Val can be performed by a growing pool of qualified organizations, not just a single bottleneck. And it maintains the standards registry and coordinates shared tests so that completed work is portable and recognized across all users. The center does not mandate participation or dictate how organizations conduct their internal Cal/Val.

What it provides is a trusted vehicle that makes the standards actionable: organizations that engage earn recognized certifications, and the market rewards that investment. By lowering the barrier for new validators to enter the ecosystem, the center also encourages innovation in Cal/Val methods and tools. The institutional model follows proven precedents: UL (Underwriters Laboratories) independently certifies product safety, and LEED (Leadership in Energy and Environmental Design) certifies building sustainability. Both are voluntary, market-driven systems where the certification body sets standards, accredits testing, and maintains a registry, but does not regulate or guarantee outcomes. Certification is a rating of demonstrated quality, not a warranty of performance in every operational scenario; liability remains with the data provider and the operational user, just as product liability remains with the manufacturer, not with UL. Figure 1 illustrates how such a center connects government agencies, independent labs, academia, and commercial providers into a single federated framework, giving all parties confidence in data validity regardless of who performs the validation.

Together, methodology and center offer a single, transparent alternative to today's fragmented, siloed Cal/Val practices. They give vendors clear, published standards to invest against. They give any buyer (government, commercial, or academic) a certified maturity score that can inform evaluation and procurement decisions. And they structure the validation process so that the right work happens at the right time, getting data into the field faster while every quality gate remains enforced. Higher certification tiers justify higher pricing, creating a market mechanism that finally rewards investment in data quality rather than data volume alone.

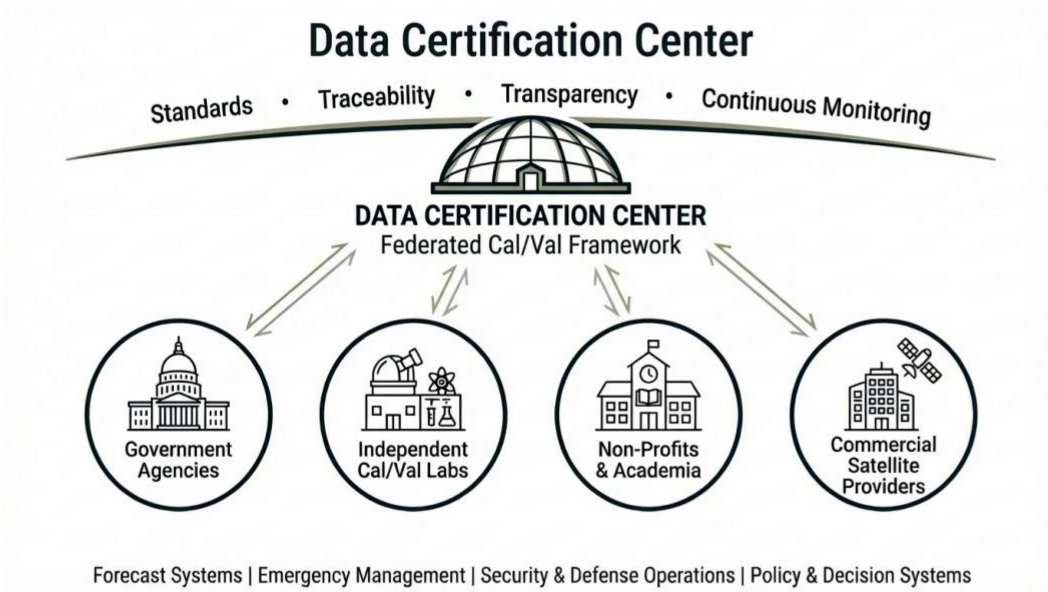


*Figure 1. Data Certification Center: a federated Cal/Val framework connecting government agencies, independent labs, academia, and commercial providers.*

## 2. The Crisis of Trust and Quality in Earth Observation

### 2.1 The Absence of a Trustworthy Market Signal

No neutral mechanism exists for commercial data vendors to clearly signal their data quality to any buyer. International Cal/Val standards exist, but no effective system enforces and translates them into certified, verifiable processes. This creates a significant gap in procedural governance, leading to confusion and mistrust for data providers, government agencies, and any organization that depends on Earth observation data for decisions. Pricing does not incentivize investment in improving data quality and maturity, because the market lacks a quality signal to reward it.

### 2.2 The Compliance Trap

Many procurement contracts rely on narrow, single-number instrument-level metrics that miss systematic errors tied to geography, orbit, or environmental conditions. Without maturity standards, the market rewards rapid launch cycles and data volume over rigorous uncertainty characterization. Additionally, integrating data based on volume rather than verified quality risks poisoning AI/ML pipelines with sensor noise and systemic biases, leading to untrustworthy automated decisions.

### 2.3 Fragmented Methodologies and Duplicated Effort

Even where international standards exist, without authoritative implementation, current siloed approaches breed mistrust, duplication, and delayed transition. Calibration efforts are often tailored to simplified mission-specific requirements, focusing on instrument capability rather than application-based fitness-for-purpose. Without a shared standard that all parties trust, validation work is inevitably repeated, forcing each organization to perform its own internal review. This leads to redundant spending and sequential delays that keep useful data out of the hands of users for months or years.

### 2.4 The Need for a Neutral Arbiter

When vendors perform their own Cal/Val, results may lack the transparency and independence required for high-stakes applications. When a data feed drifts or breaks, it is often unclear who owns the response: the vendor, the agency user, or the processing lab. Small, innovative commercial firms often cannot navigate opaque validation requirements, preventing access to the very innovation the community seeks to buy. Robust, automated, and shared Cal/Val infrastructure is now as critical as building the satellites themselves.

### 2.5 Current Market State

The commercial Earth observation market presents a distinct set of realities that any certification framework must accommodate. Innovation cycles are fast: commercial providers move quickly from concept to deployment and must demonstrate value rapidly. Operational markets are difficult to access: government and operational agencies require high confidence in data quality and reliability. Measurement capability alone is not enough: operational users need validated products, well-characterized performance, and clear guidance for use. Validation expectations are often unclear: commercial providers may not know what level of evidence is required for operational adoption. Operational users need trust before they adopt: independent validation and transparent methodologies are essential for building that trust. And integration with operational systems takes time: forecast models, decision systems, and workflows must adapt to new observations.

## 3. The Proposed Solution: A Unified Certification Framework

The proposed solution rests on four principles:

- **Unified Cal/Val Framework:** Replacing fragmented processes with a single, streamlined path from raw measurement to operational certification.
- **Open Standards:** Anyone should be capable of performing Cal/Val using published standards, fostering subject matter expertise and competition.
- **Neutral Authority:** A unified, neutral implementation authority scores products based on measured performance and demonstrated adherence to standardized procedures.
- **Market Signal:** A certified maturity score utilized by both buyer and vendor to set a price, creating a critical market mechanism that rewards quality investment.

A natural concern with any new certification system is that it may introduce another layer of bureaucracy. The framework is explicitly designed to prevent this, as described in Section 7.

# CERTIFICATION METHODOLOGY

*Two-Stage Architecture and Four-Tier Maturity Model*

This section describes the certification methodology in full. It is designed as a self-contained reference that can be extracted and applied independently of the broader framework context described in the surrounding sections.

A distinguishing feature of this methodology is what it certifies. Current validation practices typically verify whether data meets a set of mission-imposed instrument performance thresholds, for example, whether wind speed accuracy falls within 2 m/s. This approach is necessary but insufficient: a single threshold number cannot capture systematic errors tied to geography, orbit, or environmental conditions, nor can it reveal whether the validation process itself was rigorous, independent, or reproducible. The methodology proposed here takes a fundamentally different approach: it certifies both the Cal/Val process and the data product. Rather than asking only whether the data passed a threshold, it asks whether the validation itself was conducted with sufficient rigor, independence, and reproducibility. It evaluates how Cal/Val was performed (what tools, reference datasets, and levels of independence were used) and it measures data quality against application-specific performance metrics rather than a single universal threshold. The result is a multidimensional maturity assessment that tells any user not just whether data passed a test, but how much confidence they should place in it for their specific application.

This does not mean more work. Government laboratories, universities, and vendors already perform most of these evaluations: they assess methodologies, run comparisons against reference datasets, test stability across environments, and document results. The problem is that this work is performed in silos, repeated by multiple agencies for the same dataset, and the results are not portable. The methodology proposed here does not add new science to the Cal/Val process. It organizes the work that is already being done into a standardized structure so that each evaluation is performed once, documented to a common standard, and recognized across all users. The net effect is less total effort, not more, because duplication is eliminated and every completed test carries forward.

## 4.1 Two-Stage Architecture: Speed through Modularity

The framework separates scientific evidence from operational fitness. This separation is the key to speed: it allows the government to verify the technical integrity of data rapidly while progressively proving its value for complex missions (Figure 2).

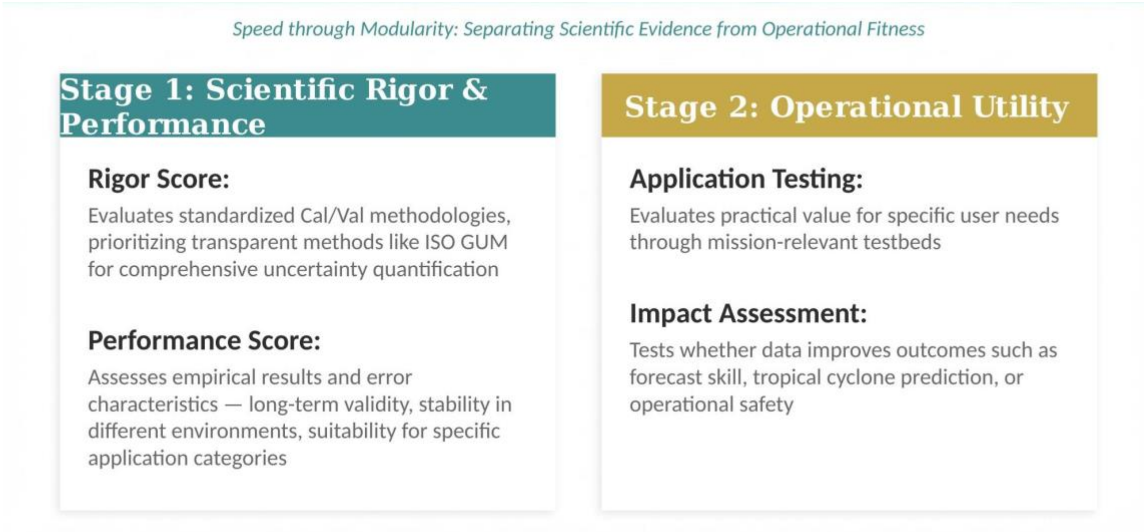


*Figure 2. Two-Stage Architecture: separating scientific evidence from operational fitness.*

### 4.1.1 Stage 1: Scientific Rigor and Performance

Stage 1 establishes scientific trust through two complementary scores:

- **Rigor Score:** Evaluates the standardized Cal/Val methodologies and tools used to characterize errors, prioritizing transparent methods like ISO GUM for comprehensive uncertainty quantification. It scales from foundational instrument health checks to deep-dive efforts like cross-sensor calibration and stability analysis.
- **Performance Score:** Assesses the empirical results and error characteristics derived from those tools, identifying long-term validity, stability in different environments, and suitability for specific application categories based on observed error bounds.

The Rigor Score captures the quality of the validation process itself: were the methods sound, independent, and reproducible? The Performance Score captures the quality of the data as measured by that process. Both are necessary: rigorous methods applied to poor data, or good data validated with weak methods, both result in low confidence. This dual evaluation is the core innovation: certification depends on how validation was done, not only on what it found.

### 4.1.2 Stage 2: Operational Utility

Stage 2 evaluates the practical value of the dataset for specific user needs. This involves testing data within mission-relevant testbeds to determine whether it improves outcomes such as forecast skill, tropical cyclone prediction, data assimilation impact, or operational safety. Stage 2 is application-specific: different applications define different operational tests.

Critically, Stage 2 applications validate the same data product levels (L0/L1, L2, L3+) that Stage 1 evaluates scientifically, but from the perspective of operational use. This overlap is what makes the two-stage architecture robust: a wind product, for example, undergoes rigorous uncertainty analysis in Stage 1 and is independently tested in a hurricane forecast model in Stage 2. The iterative loop between calibration, validation, application testing, and provider feedback (Figure 3) shows how both stages reinforce each other across all data levels, and why application-based validation is essential rather than optional.

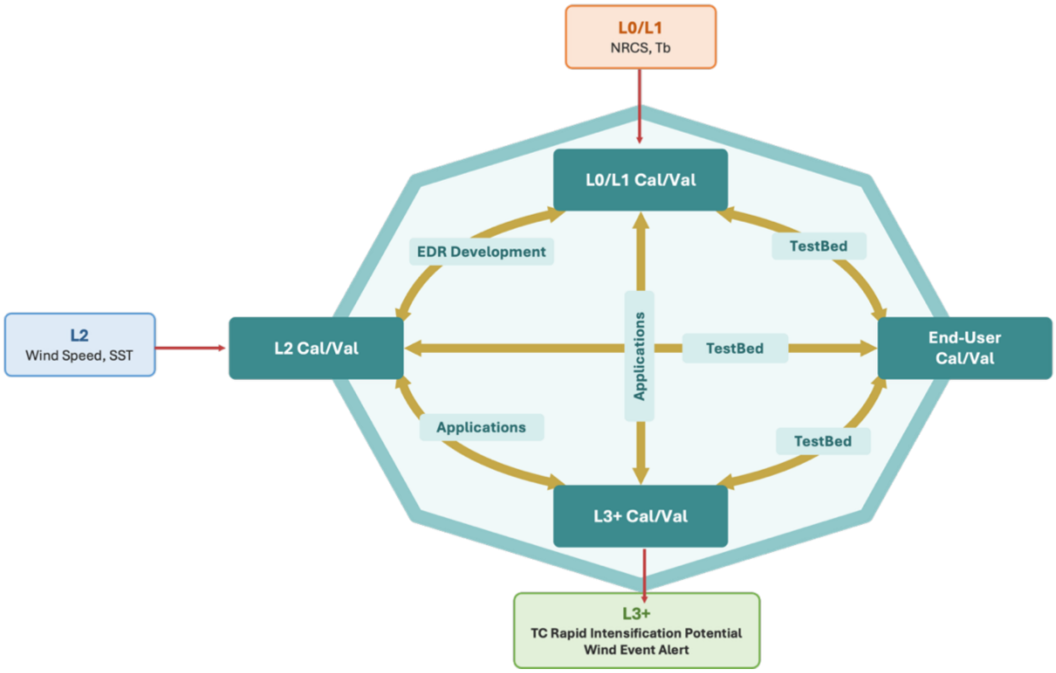


*Figure 3. Cal/Val and product development loop: iterative relationship between calibration, validation, application testing, and provider feedback across data levels.*

### 4.1.3 Four Certification Tiers

The two stages produce quantitative scores that map directly to one of four certification tiers representing increasing maturity and operational readiness (Figure 4):

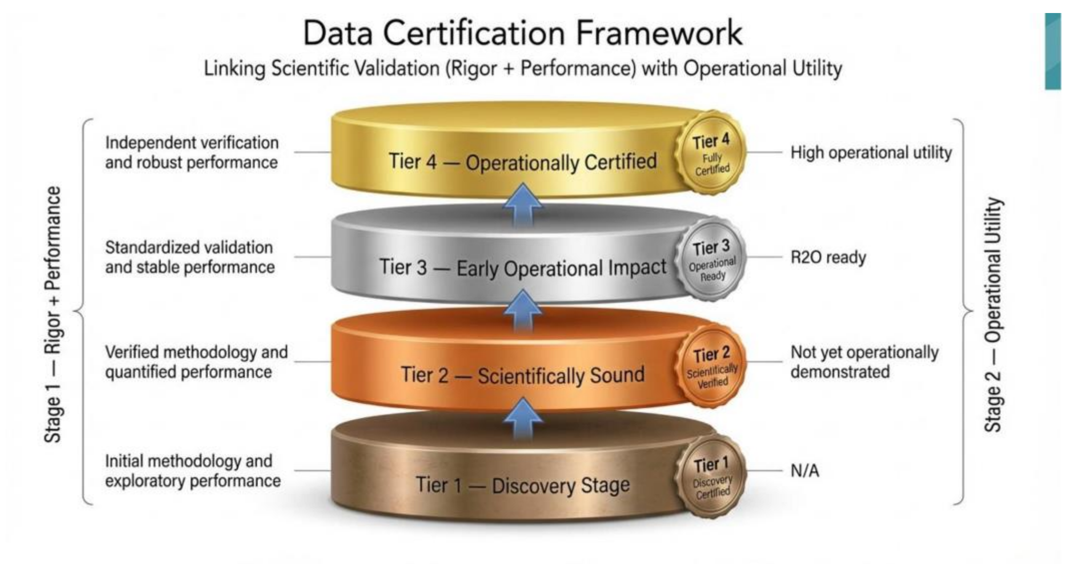


*Figure 4. Data Certification Framework: four tiers linking scientific validation with operational utility. (AI Generated image)*

| Tier | Name | Score Range | Meaning |
|---|---|---|---|
| 1 | **Discovery Stage** | < 0.25 | Initial methodology established; data available for exploratory use |

| | | | |
|---|---|---|---|
| 2 | **Scientifically Sound** | 0.25 – 0.50 | Performance quantified against standards; suitable for research applications |
| 3 | **Early Operational Impact** | 0.50 – 0.75 | Validation stable and reproducible; data ready for research-to-operations transition |
| 4 | **Operationally Certified** | > 0.75 | Independent verification complete; full operational deployment authorized |

*Note: The score ranges and tier names shown above are illustrative examples of how certification tiers can be defined. Actual thresholds, tier boundaries, and naming conventions would be established by subject-matter experts during the implementation phase, tailored to the specific measurement domain and application requirements.*

### 4.1.4 Progressive Tier Certification

Tier certifications are milestones reached during validation, not just at the end. As validation progresses through the iterative loop between Stage 1 and Stage 2, scores improve and tier thresholds are crossed. Tiers 1 through 3 can be awarded during validation, enabling data to be used for appropriate applications before the full process is complete. Tier 4, Operationally Certified, signals readiness for T2O. This progressive model means data does not sit idle waiting for full certification; it enters use as soon as it reaches the maturity threshold required for a given application (Figure 5).

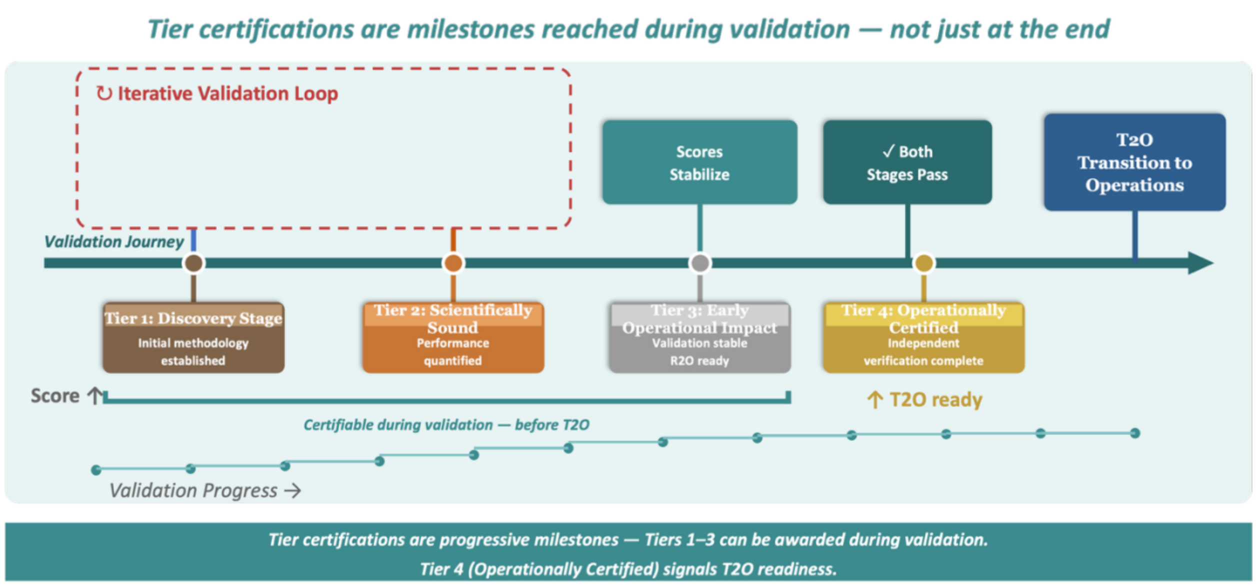


*Figure 5. Tier certifications are earned progressively during validation, not just at the end.*

### 4.1.5 Iterative Validation Loop and Parallel Execution

Stage 1 and Stage 2 are not sequential gates. Once data passes foundational integrity checks in Stage 1, Stage 2 operational verification can begin in parallel. The two stages form an iterative validation loop: findings from Stage 2 application testing feed back into Stage 1 scientific assessment, and vice versa, until both stages converge. The transition to operations (T2O) begins only when both stages pass.

This parallel architecture is the primary mechanism for acceleration. Not all applications require every level of scientific validation. A situational awareness path for qualitative imagery can move to Stage 2 after initial foundational certification, while a data assimilation path for high-stakes numerical weather prediction continues toward the highest Stage 1 tiers in parallel. Shared tests carry forward across applications: if forecasting has already completed a set of rigor and performance tests, data assimilation adds only the incremental tests it requires, rather than repeating work already done (Figure 6).

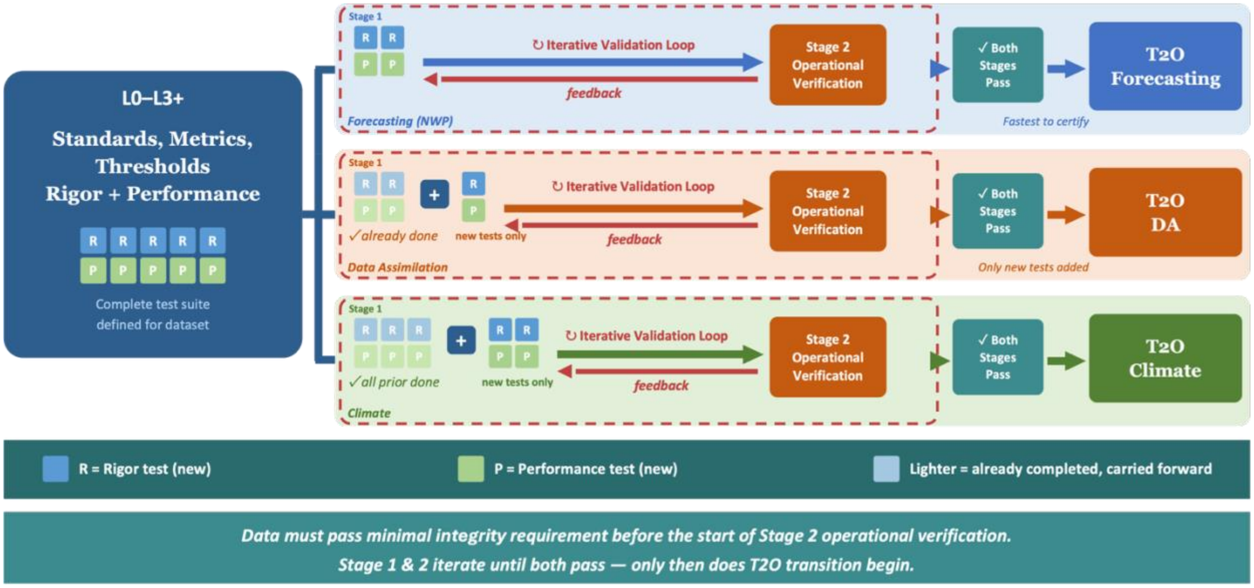


*Figure 6. Parallel certification framework showing iterative validation loop, shared test reuse, and application-specific paths.*

## 4.2 The Rigor Scorecard: Five Dimensions of Methodological Maturity

The rigor assessment is structured as a multi-dimensional scorecard. Rather than a single pass/fail determination, maturity is defined by the combination of categories achieved across five dimensions, each with four levels of increasing rigor (Figure 7).

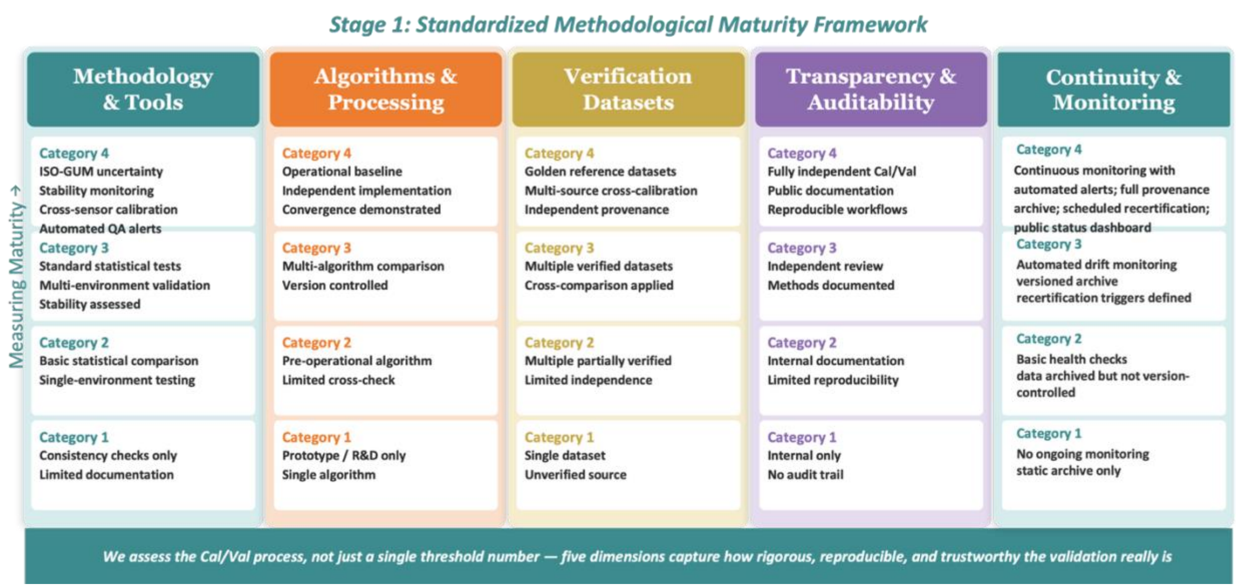


*Figure 7. Rigor Master Scorecard: five dimensions of methodological maturity, each with four categories of increasing rigor.*

The scorecard assesses the Cal/Val process, not just a single threshold number. Five dimensions capture how rigorous, reproducible, and trustworthy the validation really is, and whether quality is maintained over time. A key principle is that verification datasets are themselves subject to maturity assessment: high-tier Cal/Val cannot rely on low-confidence reference data.

### 4.2.1 Performance Thresholds

Each rigor category defines how many performance metrics are tested. Higher rigor categories employ more metrics. Each metric has its own quantitative thresholds across four performance categories, removing subjective judgment from the assessment. Thresholds can be relaxed per application: forecasting may accept a lower category for geographic uniformity while climate monitoring requires the highest. This application-specific flexibility allows faster certification for uses that need fewer tests at lower thresholds (Figure 8).

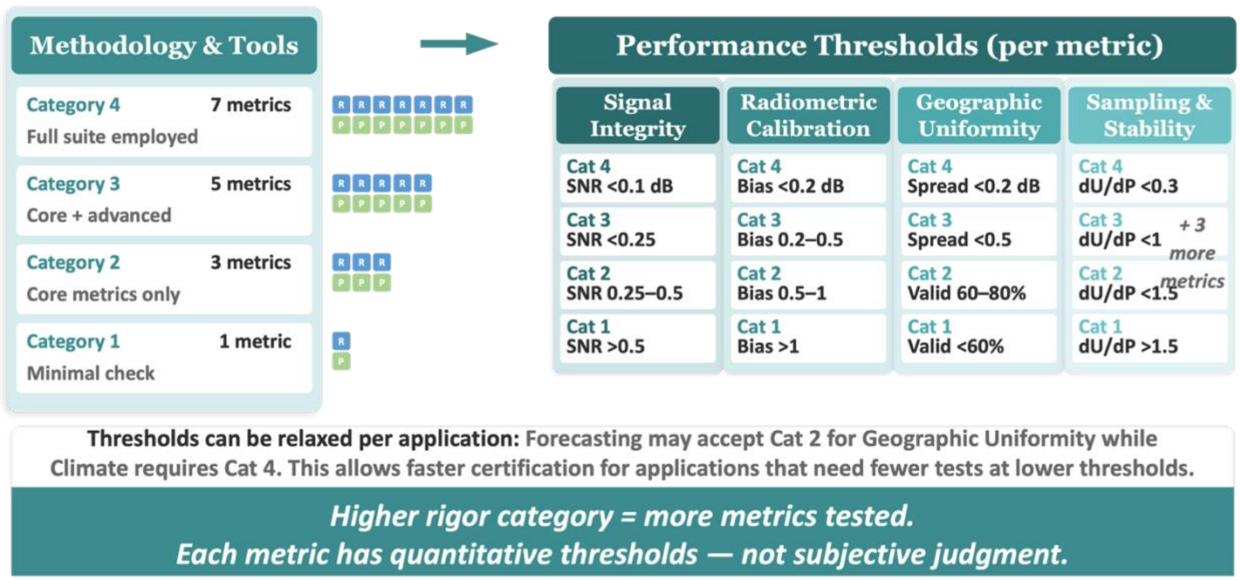


*Figure 8. Performance threshold categories defined for each rigor test.*

## 4.3 From Scores to Maturity: Quantitative Indices

The framework produces two quantitative indices that combine into a single certification score:

**Validation Rigor Index (VRI):** A weighted combination of scores across the rigor sub-dimensions (fleet representativeness, algorithm maturity, dataset diversity, transparency, technical rigor, and continuity of monitoring). Each dimension is weighted by importance. VRI ranges from 0 to 1.0.

Fleet representativeness applies when a program operates multiple satellites or sensor units. Validating one unit does not guarantee the performance of another: manufacturing variation, orbital differences, and aging effects mean each unit in a constellation may behave differently. This dimension captures whether Cal/Val has been performed across the full fleet or only on a single representative unit. When only one satellite exists in a program, this dimension carries less weight; when a constellation of dozens exists, it becomes critical. The framework does not assume that certifying one sensor certifies all.

To illustrate why VRI increases with rigor: a dataset validated with only consistency checks, a single algorithm, and one unverified reference dataset scores low across all dimensions. A dataset validated with ISO-GUM uncertainty quantification, multiple independent algorithm implementations, golden reference datasets, and fully public documentation scores high. The table below shows how representative maturity levels map to rigor categories across the five dimensions:

| Dimension | Low Maturity | | | High Maturity |
|---|---|---|---|---|

| | Cat 1 | Cat 2 | Cat 3 | Cat 4 |
|---|---|---|---|---|
| **Methodology & Tools** | Consistency checks | Basic statistical tests | Multi-environment; stability assessed | ISO-GUM; cross-sensor; automated QA |
| **Algorithms** | Single prototype | Pre-operational | Multi-algorithm comparison | Independent implementation; convergence shown |
| **Verification Data** | Single unverified source | Partially verified | Multiple verified; cross-compared | Golden reference; independent provenance |
| **Transparency** | Internal only | Limited documentation | Independent review | Fully public; reproducible workflows |
| Continuity & Monitoring | No monitoring; static archive | Basic health checks; archived | Automated drift alerts; versioned | Continuous monitoring; full provenance; public dashboard |
| **Resulting VRI** | **Low** | ↑ | ↑ | **High** |

Overall dataset maturity is determined by the aggregate scorecard, not by any single dimension. A dataset may be high-tier in methodology but limited by verification data, or transparent in reporting but early in algorithm maturity. Critically, required maturity is application-dependent: qualitative imagery may require fewer categories at lower levels, while data assimilation requires high categories across multiple dimensions.

**Measured Performance Index (MPI):** A weighted combination of measured performance in each rigor dimension. MPI ranges from 0 to 1.0. Where VRI captures the quality of the process, MPI captures the quality of the results.

**Composite Score = VRI × MPI.** Performance is adjusted by methodological credibility. This single, quantitative, instrument-agnostic number determines the maturity tier. No subjective judgment is required (Figure 9).

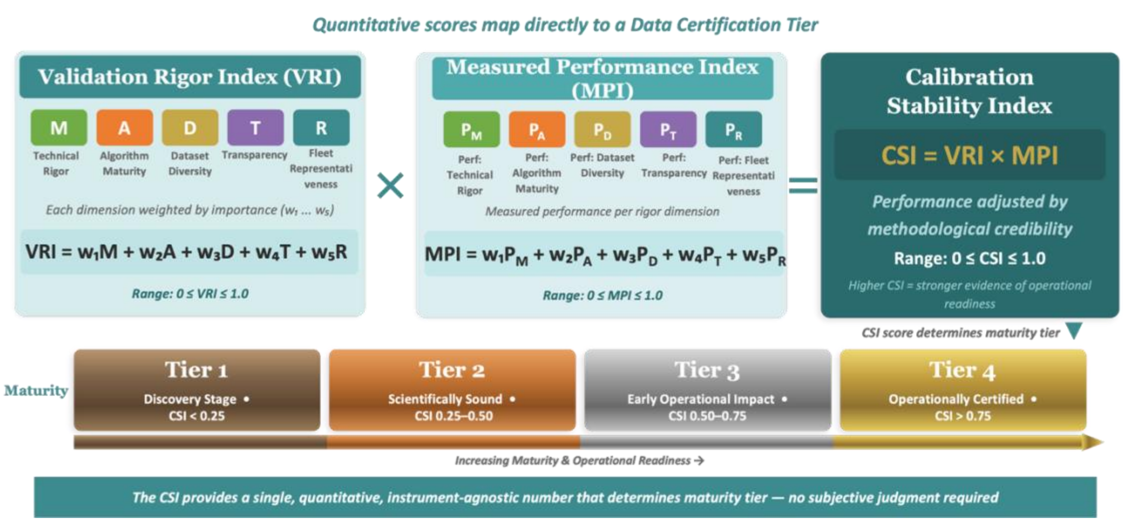


*Figure 9. From scores to maturity level: VRI, MPI, and composite score mapping to certification tiers.*

### 4.4 Data Levels and Tiered Execution

The framework is organized around standard satellite data levels: Level 0 (raw measurements), Level 1 (calibrated observables), Level 2 (derived geophysical variables), and Level 3 (application-oriented products). Each data level can support multiple tiers of calibration and validation. Tiers are executed within each level, as required by the intended application, allowing Cal/Val effort to scale both vertically across levels and horizontally across tiers.

This level-aware structure exists because every processing step in the chain introduces decisions that affect downstream data quality. Algorithm selection, interpolation methods, quality control flagging, and ancillary data sources are all points where errors can be introduced, amplified, or masked. If a systematic bias is embedded during L1 calibration, no amount of L2 or L3 validation will identify the root cause; it will manifest as unexplained error in the final product. Validating only the end product, as many current approaches do, can hide compensating errors: two processing artifacts may cancel in a global mean bias statistic while regional errors persist undetected. By executing certification tiers within each data level, the framework traces quality through the full processing chain rather than treating the data as a black box.

## 5. Parallel Execution: Speed through Structure

The most critical feature of the framework is that while Stage 1 provides the necessary evidence for Stage 2, the entire scientific deep dive does not need to be finished before operational testing begins.

- **Minimum Viable Rigor:** Data must pass the foundational integrity portion of Stage 1 to enter Stage 2 application tests, ensuring testbeds are not poisoned by uncalibrated or fundamentally broken data.
- **Shared Test Reuse:** Tests already completed for one application carry forward to the next. A forecasting certification that required two rigor and two performance tests means data assimilation only needs to add its incremental requirements, not repeat the full suite.
- **Application-Specific Paths:** A situational awareness path can move to Stage 2 after initial foundational certification, while a climate monitoring path continues toward the highest Stage 1 tiers in parallel.

Because the framework is tiered, an issue discovered in a high-tier deep-dive does not necessarily invalidate the data for lower-tier uses. If a bias is found that makes data unsuitable for data assimilation, that same data may still be certified for situational awareness while the fix is implemented. This functional redundancy, the ability of data to remain certified for some applications while issues are resolved for others, is a key advantage over binary pass/fail systems.

## 6. The Strategic Shift

| Current Approach | Proposed Approach | Mission Impact |
|---|---|---|
| Siloed & Instrument-Specific | Standardized, Application-Based Certification | Risk Management: Matches data to the right mission risk level |
| Fixed, Dataset-Wide Maturity Gates | Per-Application Maturity Gates | Agility: Gets data into the field in weeks, not months |
| Redundant Reviews | Centralized Certification Registry | Cost Savings: Eliminates paying for the same Cal/Val twice |
| Restricted Participation | Inclusive, Role-Based Participation | Adoption: All stakeholders contribute through accepted, transparent processes |

# 7. Preventing Bureaucracy by Design

A natural concern with any new certification framework is that it risks becoming another bureaucratic layer that slows down the very process it was designed to accelerate. This concern is valid, and the framework addresses it directly. Every design choice is made to facilitate adoption, not to police compliance. The framework is built on six principles that together ensure it remains a tool of enablement rather than a source of friction.

## 7.1 Open Standards, Not Gatekeeping

The framework does not restrict who can perform Cal/Val. Anyone, whether a government lab, a university research group, or a commercial vendor, can perform calibration and validation using the same published standards. The framework standardizes how evaluation is done; it does not create a monopoly on who can do it. This openness is essential: it grows the pool of qualified Cal/Val practitioners, fosters competition in validation services, and ensures that subject matter expertise, rather than institutional affiliation, determines who contributes.

## 7.2 Quantitative Scores, Not Committee Opinions

Certification decisions are not made by committee vote or subjective expert judgment. The framework produces quantitative maturity scores that are formula-driven and reproducible. The Validation Rigor Index and Measured Performance Index combine into a composite score that maps directly to a certification tier. Because the formulas and weights are published, vendors can self-assess against the same criteria before submitting, eliminating surprises and reducing the back-and-forth that characterizes many current review processes.

## 7.3 Published Thresholds, No Ambiguity

Each certification tier has clear, measurable requirements that are defined in advance and available to all participants. There is no ambiguity about what is required to reach a given tier. Vendors know exactly what to aim for, and buyers (whether government agencies, international partners, or commercial users) know exactly what each tier guarantees. This transparency eliminates the frustration of opaque requirements and ensures that the path to certification is predictable and plannable.

## 7.4 Incentive Structure, Not Compliance Burden

The framework aligns economic incentives with quality investment. Higher maturity scores justify higher pricing, because they represent a higher level of verified confidence in the data. Vendors invest in quality not because a regulator demands it, but because the market rewards it. This is a fundamental distinction: the framework creates a market mechanism rather than a compliance obligation. A vendor at Tier 2 is not penalized; they simply access a different segment of the market than a vendor at Tier 4. The incentive to move up is built into the commercial structure, not imposed from above.

## 7.5 Neutral Nonprofit Host, Not a Regulatory Body

The certification center is modeled after organizations like UL (product safety certification) or LEED (building sustainability certification): an industry-serving body that sets standards, maintains a registry, and facilitates trust, but does not regulate or enforce. It is built to serve the ecosystem, not to police it. Hosting the center within a neutral nonprofit (i.e UCAR) ensures independence from both commercial competition and government procurement authority, maintaining the credibility of the trust signal on which the entire framework depends.

## 7.6 Government Retains Oversight Without Doing All the Work

A critical advantage of this framework is that it shifts the burden of execution without shifting the burden of authority. Under the current model, government agencies often face a difficult choice: either accept vendor-provided validation at face value, or dedicate scarce internal resources to repeat the entire analysis themselves. The framework offers a third path. Commercial providers and independent laboratories perform Cal/Val under open, published standards. Government validates the results and retains independent assessment authority, rather than repeating the analysis from scratch. The result is that government oversight is maintained with a fraction of the effort, freeing expert resources to focus on the highest-stakes evaluations and policy decisions rather than routine verification.

This is a win for all sides. Vendors gain a clear, predictable path to certification and market access. Government agencies gain a trusted, auditable quality signal without having to resource every validation internally. And the broader community of data users (international partners, research institutions, the commercial sector) gains a common language for understanding what a dataset is ready for and what it is not, regardless of who the purchaser is.

## 7.7 Built-In Custodianship across the Processing Chain

In many current workflows, custodianship of data quality is fragmented: a vendor owns instrument calibration at Level 0/L1, a government laboratory may produce Level 2 retrievals, another group generates Level 3 application products, and the operational end user receives the final output with limited visibility into what happened upstream. When an issue is discovered at one level, there is no systematic mechanism to trace its impact on downstream products or to notify affected users.

The framework addresses this by embedding custodianship into the certification process itself. Because certification is executed within each data level, every level has a documented owner, a recorded methodology, and a transparency score. The Rigor Scorecard's Transparency dimension, from internal-only documentation at Category 1 to fully public, reproducible workflows at Category 4, is in effect a measure of custodianship maturity. When an L1 recalibration occurs, the certification registry traces which L2 and L3 certifications depend on it and flags them for re-evaluation. This chain-of-custody visibility ensures that quality is not just assessed at the end of the pipeline but maintained and traceable throughout it.

## 7.8 Certification as Rating, Not Warranty

A legitimate concern with any certification system is whether issuing a certification creates liability for the certifying body. If data certified at Tier 4 subsequently causes an issue in an operational forecast, is the certification center responsible? The answer, grounded in established institutional precedent, is no.

The certification center is modeled after organizations like UL (Underwriters Laboratories) and LEED (Leadership in Energy and Environmental Design). UL independently certifies that products meet published safety standards; LEED certifies that buildings meet published sustainability criteria. Both have operated for decades under a clear principle: certification is a rating of demonstrated conformance to standards at the time of assessment, not a warranty of future performance or a guarantee of outcomes. When a UL-listed product fails, product liability rests with the manufacturer, not with UL. When a LEED-certified building underperforms its energy targets, the USGBC (which administers LEED) is not liable for the shortfall. This distinction has been tested in court: a $100 million class-action lawsuit against USGBC alleging that LEED certifications constituted false advertising was dismissed by a federal judge, affirming that a voluntary rating system based on published criteria does not create the same obligations as a product warranty.

The Data Certification Center operates on the same principle. A maturity score certifies that the Cal/Val process was conducted according to published standards and that measured performance reached a quantified level at the time of assessment. It does not guarantee that the data will perform correctly in every future operational context, under every environmental condition, or within every downstream application.

Liability for data quality remains with the data provider; liability for operational decisions made using the data remains with the operational user. The center's role is to provide a transparent, reproducible, and independently verified assessment that both parties can rely on to make informed decisions. This is the same relationship that exists between UL and the manufacturers whose products it certifies: UL provides the trusted assessment, but the manufacturer owns the product and its performance.

Because certification reflects the state of data quality at the time of assessment, the framework also addresses what happens after certification is issued. Sensor performance can drift, processing algorithms can change, and environmental conditions can shift. A certification that ignores post-assessment stewardship would be incomplete. The rigor scorecard addresses this directly through its Continuity and Monitoring dimension (Section 4.2), which evaluates whether the dataset is actively monitored for drift and anomalies, whether it is archived with version control and provenance tracking, and whether recertification triggers are defined. Datasets with robust ongoing stewardship score higher on this dimension, and the market signal reflects that difference. This approach avoids imposing a bureaucratic expiration date on certifications. Instead, it ensures that the maturity score itself captures whether the data provider has invested in the infrastructure needed to maintain quality over time.

## 8. Alignment with the Scientific Executive Order

The framework directly addresses the principles of the Gold Standard Science Executive Order:

- **Reproducible:** All Cal/Val results include documented methods, datasets, and tools so any qualified lab can replicate the outcome.
- **Transparent:** Open visibility into who performed validation, what tier was achieved, and what metrics were used.
- **Communicative of Error and Uncertainty:** Each maturity level includes quantified uncertainty estimates and performance metrics communicated to users.
- **Collaborative and Interdisciplinary:** Government, academia, commercial vendors, and independent labs work under shared standards.
- **Skeptical of Findings:** Independent validation and third-party audits reduce over-reliance on self-reported claims.
- **Structured for Falsifiability:** Tiers and metrics test whether data meets operational claims. Failure is as informative as success.
- **Subject to Peer Review:** Cal/Val results are auditable by independent experts; labs must demonstrate independence.
- **Without Conflicts of Interest:** Transparent disclosure of who performed validation; independence rules enforced for third-party labs.

## 9. Implementation Roadmap: The Data Certification Center

### 9.1 Mission and Structure

To bridge the gap between agency-level expertise and national efficiency, the framework proposes the establishment of a Data Certification Center. This center provides the authoritative implementation that international Cal/Val standards currently lack: it issues data certifications, accredits independent validators, and maintains the registry and trust signals that allow the maturity framework to scale across the federal enterprise and commercial market alike.

- **A Federated, Collaborative Hub:** The center integrates expertise from government laboratories (NOAA, NASA, USSF), academia, and commercial partners under one trusted structure.

- **Ending the Double-Pay Cycle:** A central certification registry means a Stage 1 scientific integrity certification is performed once and utilized by all, eliminating redundant spending.
- **Utilizing Government SMEs:** Leading experts in Stage 1 validation reside within government labs. The center provides the framework for these internal SMEs to contribute to a central registry.
- **Standardizing the How:** The center codifies ad-hoc practices into authoritative implementations, providing commercial vendors with a clear roadmap of tools and methodologies required to meet government needs.
- **Accrediting Independent Validators:** The center certifies organizations (government laboratories, universities, and commercial firms) as qualified independent validators. This grows the pool of accredited Cal/Val providers, removes the single-point bottleneck of government-only validation, and creates a competitive market for validation services that accelerates the overall process.
- **Encouraging Innovation:** By publishing open standards and accrediting new entrants, the center lowers the barrier for organizations to develop novel Cal/Val methods, tools, and services. Innovation in validation technology is rewarded rather than constrained, because any organization that meets the accreditation criteria can compete to deliver certified results.

## 9.2 Host Authority

To maintain neutrality, scientific rigor, and freedom from conflicts of interest, the host organization is proposed to be the University Corporation for Atmospheric Research (UCAR). As a nonprofit consortium of over 120 member universities, UCAR provides the necessary distance from commercial competition to maintain a neutral trust signal. Its existing collaborations with NOAA, NASA, and the USSF make it the logical host to define maturity scorecards and accredit external labs.

An existing international body such as the Committee on Earth Observation Satellites (CEOS) might appear to be a natural candidate, given its role in coordinating Cal/Val best practices across space agencies through its Working Group on Calibration and Validation (WGCV). However, the certification center must be equally accessible to commercial industry and to government, and CEOS is structured as an agency-to-agency coordination body. Its membership consists of national space agencies and affiliated organizations; commercial data providers do not have a direct seat at the table. A certification framework that aims to create a market signal for commercial data quality must offer vendors the same standing as government participants: the ability to submit data for certification, access published standards, and receive scores under the same transparent process. CEOS governance, which operates by consensus among member agencies on multi-year planning cycles, is not designed for this kind of vendor-facing, market-responsive function. Equally important, CEOS member agencies are themselves major data buyers, which creates a structural conflict of interest if the same organizations that procure data also control the certification scores. The framework should build on CEOS/WGCV best practices, intercomparison protocols, and reference datasets wherever possible, and CEOS endorsement of the standards would strengthen their credibility. But the operational host must be an independent entity that serves all participants, government and commercial alike, on equal terms.

## 9.3 Phase 1: SME-Led Technical Sprints

The first implementation step is to convene Subject Matter Experts from across agencies to define the initial Stage 1 and Stage 2 metrics for the certification registry. SME groups will identify standardized tools and reference datasets required to characterize errors for specific sensors, define distribution metrics to distinguish between global and localized acquisition needs, and establish the initial entries in the certification registry.

## 10. Call to Action

***Establish a Data Certification Center that accelerates the path to operations and gives every stakeholder a trusted seat at the table.***

The framework proposed in this paper is not theoretical. It is built on proven Cal/Val practices already in use, brought together into a coherent, trusted structure. It preserves scientific rigor while increasing agility, enables earlier and lower-risk use of new data, supports parallel rather than serial execution, improves communication of data readiness and limitations, and provides a defensible basis for operational and procurement decisions.

By defining Cal/Val maturity as a structured, transparent assessment of both the validation process and its results, rather than a binary label on data products alone, the framework aligns Cal/Val effort with real-world application needs and accelerates responsible transition to operations. The time to act is now.