

The Value of Data

Defining the Blueprint for Civil Engineering Benchmarks

Value of Data (VD) Task Force | ASCE DSA Committee

Dr. Dong Zhao

The Data Fragmentation Dilemma

Civil Engineering produces massive volumes of high-fidelity data—from geotechnical sensor streams to structural digital twins. Yet, this data remains trapped, creating a remarkably data-poor ecosystem despite high collection rates.

CURRENT REALITY

Ad-hoc sharing

Project-specific

Unstandardized formats



MISSED OPPORTUNITY

AI/ML Training

Algorithmic Validation

Global Predictive Modeling



From Research Byproduct to Strategic Asset

We must shift the culture from treating data as a temporary exhaust to a rigorously evaluated, reusable asset that compounds in utility over time.

	Current Paradigm	Benchmark Paradigm
Focus	Data as an exhaust/byproduct	Data as a primary, FAIR strategic asset
Delivery	Scattered files, PDFs, localized ad-hoc hosting	Machine-accessible APIs, DOIs, structured databases
Value	Single-use analysis, isolated project validation	Compounding community utility, global algorithmic training

Focus: Data Benchmark Establishment for Civil Engineering

While “data sharing” is a prerequisite, it’s a process, not an end goal. The task force’s focus is on “data benchmark establishment,” a concrete, high-value outcome that directly addresses a critical need in the field.



Benchmarks Drive Innovation

High-quality benchmark datasets are essential for validating new computational models, comparing algorithm performance, and training AI/ML systems. They accelerate progress by providing a common ground for the entire research community.



Benchmarks Demonstrate Value

A successful benchmark dataset is the ultimate demonstration of the value of data. It creates a tangible asset that practitioners and researchers will actively use, making the case for data sharing more powerful.



A Goal-Oriented Approach

Framing the TF's mission around establishing benchmarks gives our efforts a clear, aspirational target. We can then promote data sharing as the necessary pathway to achieving this critical goal.

Task Force Work Plan: A Multi-Pronged Approach



Work Area 1: Systematic Review

The Foundation

Provides an objective, evidence-based understanding of the current academic landscape of data-centric research.



Work Area 2: ASCE Member Survey

The Pulse Check

Quantifies the attitudes, perceived barriers, and practical needs of the wider professional community.



Work Area 3: Awardee Interviews

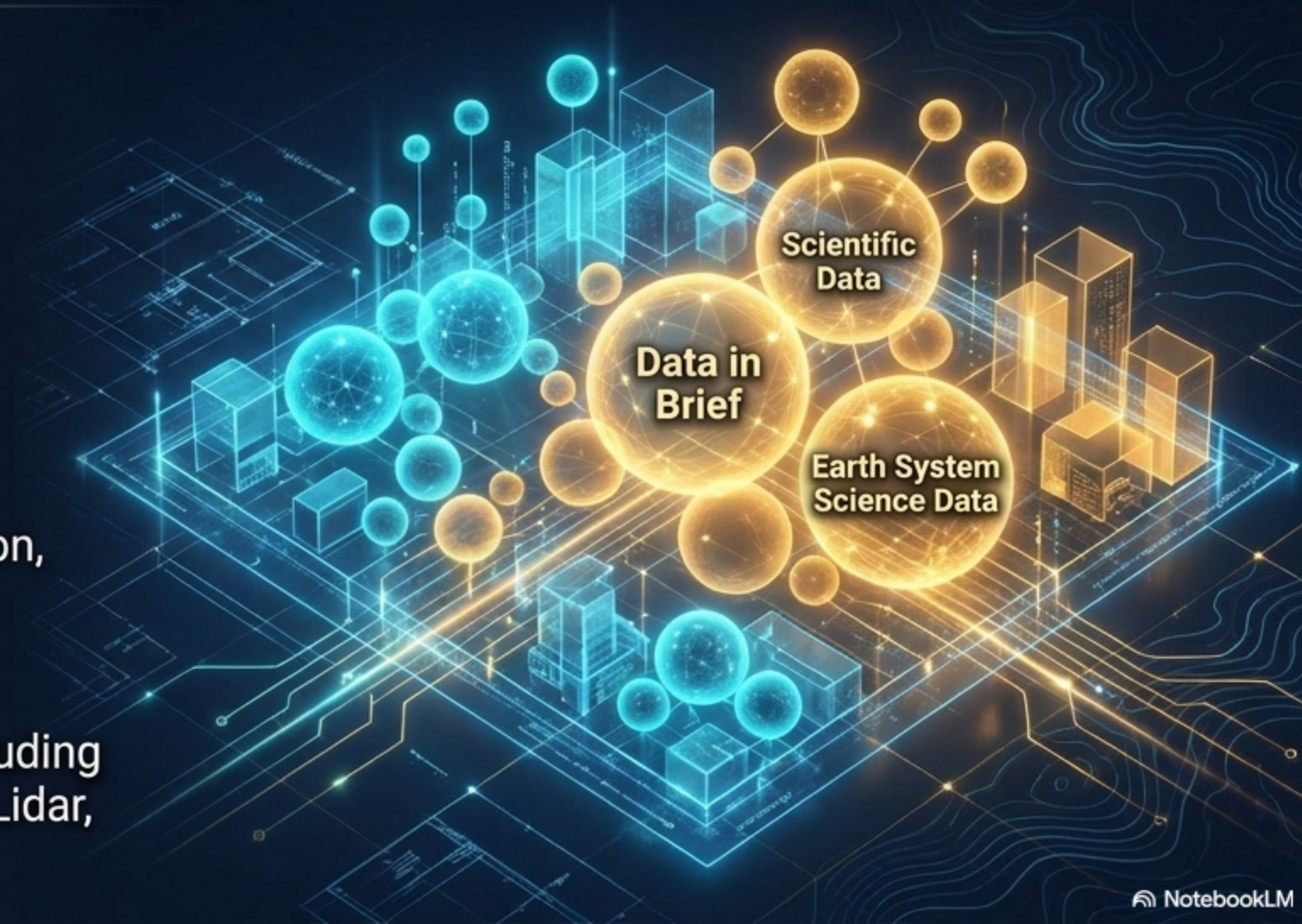
Deep Qualitative Insight

Captures nuanced, real-world experiences and success stories to understand how and why data sharing works.

Work Area 1: Mapping the Publication Landscape

Establishing a queryable, evidence-based database of current data-sharing practices across the profession.

- **Outlets:** Heavy concentration in open-access data journals (e.g., Data in Brief currently leads with nearly 2,900 related papers specifically mentioning civil/construction domains, followed by Scientific Data and Earth System Science Data) rather than traditional domain publications.
- **Domains:** Categorizing across Construction, Transportation, Structure, Environment, Geotechnical, and Water Resources.
- **Modalities:** Tracking specific formats including structured tabular data, multimodal sets, Lidar, imagery, and sensor time-series.



Benchmark Matrix

	Imag	Video	Photogrammetry	Meter	Lidar	Sensing	Text	Survey	Audio	2D/3D Graphs	Tabular
1. Occupational Health and Safety	☒	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
2. Robotics and Automation	☐	☐	☐	☒	☐	☐	☐	☐	☐	☐	☐
3. BIM, Reality Capture, and Digital Twin	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
4. Visualization, Simulation, and Extended Reality	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5. Smart, Sustainable and Resilient Cities	☐	☒	☐	☐	☐	☐	☐	☐	☐	☐	☐
6. Human-Machine Interaction	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
7. Energy and Building Systems	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
8. Project Planning and Management	☐	☐	☐	☐	☐	☐	☒	☐	☐	☐	☐
9. Education and Workforce Development	☐	☐	☐	☐	☐	☐	☐	☒	☐	☐	☐
10. Applied Data Science, AI (LLM) & Sensing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Work Area 1: Progress on Journal Data Mapping

Quantifying the landscape of data papers across key open-access journals.

- **Data in Brief (Elsevier):** ~2,900 papers (Leads the field)
- **China Scientific Data:** 446 papers
- **Scientific Data (Nature):** 326 papers
- **Earth System Science Data:** 248 papers
- **Data (MDPI):** 194 papers



Work Area 2: ASCE Member Survey

General Introduction

This work area gauges the profession's attitudes, capabilities, and needs regarding data sharing, including identifying critical data types, understanding barriers, and exploring incentives for open data practices.

Sample Survey Questions

Background Information:

- Primary research domain (e.g., Structural, Geotechnical)
- Primary research interest and focus
- Research approaches used (e.g., Modeling, Data-Driven Methods)

Data Needs:

- Criticality of data to research activities
- Most critical data types (e.g., Field Data, Simulation Data)
- Importance of data characteristics (e.g., Quality, Metadata)
- Primary challenges in obtaining data

Benefits of Data:

- Primary benefits of increased data sharing
- Impact of shared data access on research outcomes
- Valuable data types for practice

Attitudes and Barriers:

- Willingness to share data
- Frequency of past data sharing
- Barriers to sharing (e.g., Legal, Professional, Resource)

Agreement on Approaches:

- Preferences for data-sharing governance and timing
- Conditions affecting willingness to share (e.g., Credit, Risk)

Work Area 3: Lessons from the Vanguard

Moving beyond statistics to capture the nuanced, real-world experiences behind successful dataset design, governance, and long-term sustainability within the AEC/FM community.

Status Update

- Interview protocol established and refined for the AEC/FM sector.
- IRB exemptions and ethical compliance fully secured.
- Remote Zoom interview deployment activated, ensuring de-identified data extraction.



Work Area 3: Interview Themes & Questions

Moving beyond statistics to capture nuanced, real-world experiences and lessons learned.



- **Dataset Creation & Design:** How was the dataset conceived, designed, and structured?



- **Governance & Ethics:** What governance models were used? How were ethical considerations managed?



- **Sustainability & Maintenance:** How is the dataset maintained and funded long-term?



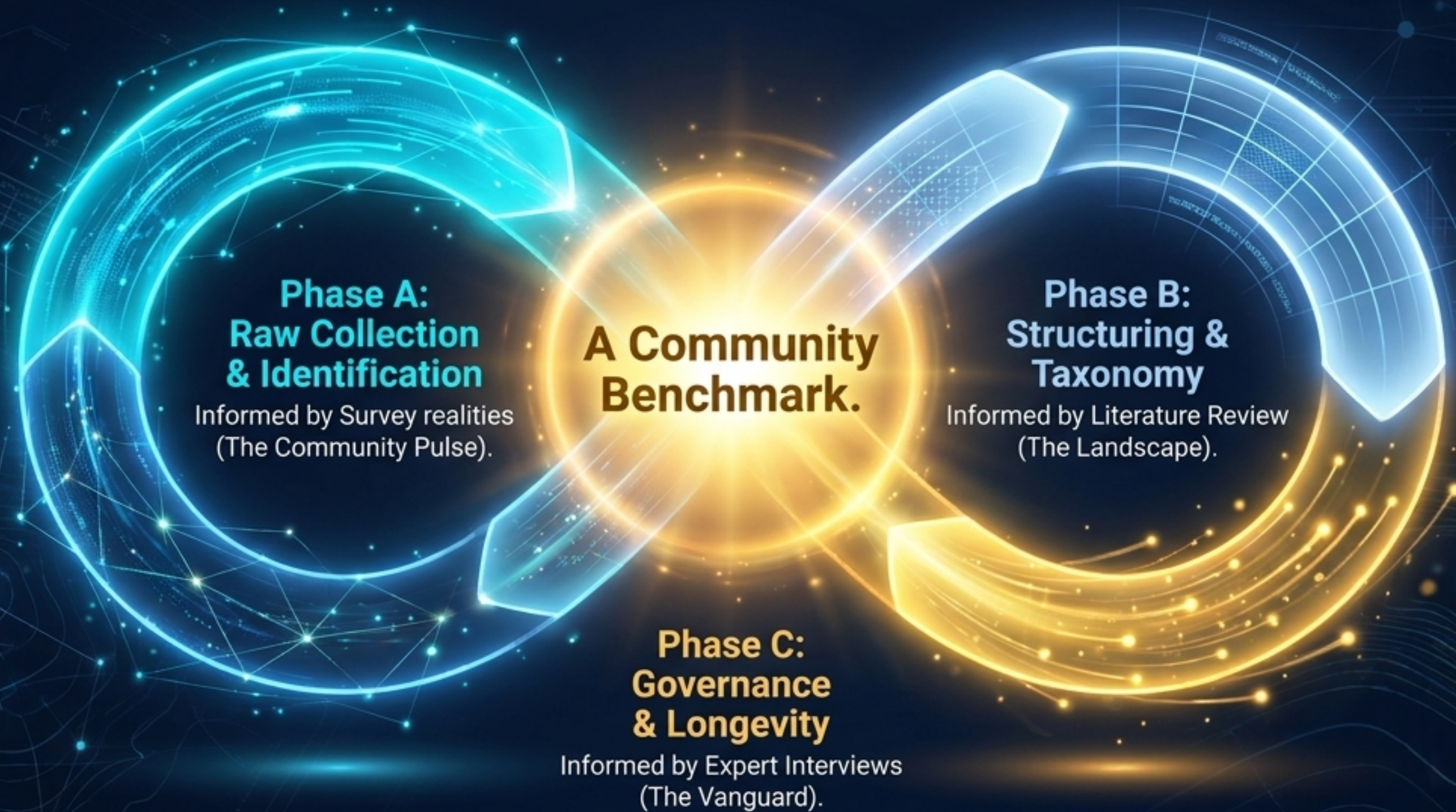
- **Value & Impact:** What is the realized value of the dataset for the community?



- **Lessons Learned:** What were the key challenges and success factors?

Convergence: Engineering a Standard

How the insights from our three Work Areas map directly to the lifecycle of future ASCE benchmarks.



A Joint Effort: Participating Institutions



University of
New Haven



STEVENS
INSTITUTE OF TECHNOLOGY

Build the Benchmark With Us

Transitioning from siloed files to global benchmarks requires community consensus.

- Participate in the upcoming ASCE Member Survey.
- Share your data stories and use cases.
- Join the DSA Committee effort.

