sustainability

MDPI

Article

Assessing the Usability of Existing ITS Data for Autonomous-Driving AI: A Task-Specific Framework for Sustainable Data Reuse

Young-Min Yoon

ITS Performance Evaluation Center, Construction Test & Certification Department, Korea Institute of Civil Engineering and Building Technology, Ilsanseo-gu, Goyang-si 10223, Republic of Korea; ymyoon@kict.re.kr

Abstract

This study assesses whether existing intelligent transportation system (ITS) data can be reused for autonomous-driving artificial intelligence (AI) based on task-specific information requirements and the recoverability of missing information. A task-specific framework was applied to five heterogeneous ITS data assets—roadside closed-circuit television (CCTV), camera-based and radar-based automatic incident detection system (AIDS) data, radar-based smart intersection system (SIS) data, and vehicle detection system (VDS) performance-evaluation data—across four data-use tasks: 2D road-user detection training, detection/measurement output validation, event/structured traffic-state retrieval, and road-referenced trajectory reconstruction. For each asset–task pair, mandatory information requirements were identified and classified as explicitly available, directly derivable or linkable, recoverable through feasible transformation, not reliably recoverable, or unconfirmed; these information-level states were then translated into task-specific usability. The results showed that usability varied substantially by task. Camera-based AIDS provided the broadest direct usability, whereas roadside CCTV showed greater reuse potential through transformation. Radar-based AIDS, radar-based SIS, and VDS performance-evaluation data were directly usable for validation and structured traffic-state retrieval, while visual-training and trajectory-reconstruction usability was more limited. Performance-evaluation records showed particular value for validation because they preserved target outputs, validated reference information, and their correspondence. These findings demonstrate that the reuse value of existing ITS data depends on the information actually retained and its recoverability rather than on sensing modality alone. By identifying reusable data and selective future-retention needs, the framework supports more resource-efficient management of digital ITS assets and links task-specific data reuse to sustainable transportation.

Keywords: intelligent transportation systems (ITS); autonomous driving; artificial intelligence; sustainable transportation; sustainable data reuse; data usability; recoverability; performance-evaluation data; data retention; digital transportation infrastructure

Check for updates

Academic Editor: Heng Li

Received: 18 September 2026
Revised: 7 October 2026
Accepted: 16 October 2026
Published: 22 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Artificial intelligence (AI) has become central to the development of automated and autonomous driving, where model performance increasingly depends on the availability of data that are suitable for specific development and validation tasks. Accordingly, recent research on data readiness for AI has emphasized that data should be assessed not only in terms of availability and quality, but also in terms of suitability, accessibility, and fitness for the intended application [1,2].

In parallel with these advances, extensive intelligent transportation system (ITS) infrastructure has already been deployed across road networks worldwide. Roadside closed-circuit television (CCTV) cameras, vehicle detection systems (VDSs), automatic incident detection systems, smart-intersection systems, and other roadside facilities continuously generate traffic-related information, which may be retained in different forms depending on operational and evaluation needs. These data may include video, traffic volumes and speeds, detected objects, event records, structured traffic states, and performance-evaluation records. Unlike purpose-built autonomous-driving datasets, however, existing ITS data are generally collected and retained to support traffic monitoring, detection, control, incident response, system operation, or performance evaluation. Their potential value for autonomous-driving AI is therefore different from that of datasets designed specifically for AI development.

The value of roadside information for autonomous-driving research is well established. Purpose-built datasets such as Rope3D, DAIR-V2X, V2X-Seq, and TUMTraf-V2X demonstrate that roadside and vehicle–infrastructure sensing can support perception, cooperative detection, tracking, forecasting, and trajectory-oriented analysis [3–6]. These datasets, however, are intentionally constructed around their target tasks and commonly retain annotations, calibration information, spatial references, persistent identities, sequential observations, or other information required for downstream use. By contrast, existing operational ITS data are not generally retained according to these downstream AI information requirements.

The existence of a camera, radar sensor, or structured traffic record does not by itself establish that the information required for a new AI task has been retained. A camera-based system may not preserve reusable raw video, and repeated detections or measurements do not necessarily provide the identity continuity and spatial reference required for trajectory reconstruction. Conversely, information that is not explicitly stored may still be recoverable from retained data through annotation, linkage, calibration, semantic conversion, or other processing. The potential usability of existing ITS data therefore depends not only on sensing modality or nominal system capability, but also on what information was actually retained and whether missing information can be recovered with sufficient reliability.

Evidence from operational traffic-data reuse supports this view. Yu et al. reconstructed approximately five million vehicle trajectories from archived records obtained from more than 3000 existing traffic cameras in two Chinese metropolitan areas using spatial–temporal vehicle re-identification, path inference, and trajectory recovery [7]. This example shows that data collected for traffic monitoring can support substantially different downstream uses when sufficient source information has been retained. It also illustrates the practical boundary of reuse: the value of repurposing depends on what remains in the original records and what can still be reconstructed from them. Existing data should therefore be neither assumed to be AI-ready nor dismissed simply because they were not collected for autonomous-driving purposes.

From a sustainability perspective, the reuse of existing ITS data can also contribute to more resource-efficient management of digital transportation infrastructure. When retained data can support new AI-related purposes, their reuse may extend the useful value of existing information assets and reduce the need for redundant data collection. Conversely, identifying information that cannot be recovered retrospectively can support more selective future data-retention strategies rather than indiscriminate storage of all available raw data. In this sense, assessing the task-specific reuse potential of existing ITS data is relevant not only to AI-data readiness but also to the sustainable management of transportation data resources.

General AI-readiness research and recent task-centric data-quality studies provide important foundations for this problem [1,2,8]. However, the assessment of an existing ITS data asset for a particular autonomous-driving AI task involves an additional question: whether the task-specific information required for that use is actually retained, directly obtainable, recoverable through transformation, not reliably recoverable, or simply unconfirmed from the available evidence. These situations have different implications for reuse. Treating all missing or unavailable information as a single category can underestimate the value of retained data, while assuming that technically inferable information is recoverable can overestimate its practical usability.

These differences become more consequential when heterogeneous ITS assets are assessed together. Operational video, event-detection records, structured intersection data, and performance-evaluation records may each preserve different subsets of information, even when they originate from the same roadside environment. A data asset that is inadequate for visual training may still be useful for validation, while structured traffic records that are well suited to retrieval may not support road-referenced trajectory reconstruction. The usability of an existing ITS data asset is therefore better understood as task specific rather than as a single fixed property of the dataset or sensing system.

Against this background, this study develops a task-specific framework for assessing the usability of existing ITS data for autonomous-driving AI. The study considers four representative data-use tasks: 2D road-user detection training, detection/measurement output validation, event/structured traffic-state retrieval, and road-referenced trajectory reconstruction. These tasks were selected to represent different functional uses and different information requirements rather than different levels of AI capability.

Accordingly, the study does not attempt to evaluate the full set of onboard autonomous-driving functions or to provide a general survey of all ITS data sources applicable to machine learning. Instead, it focuses on four explicitly defined data-use functions that support autonomous-driving AI development, validation, and downstream analysis, allowing the same retained ITS asset to be evaluated against different task-specific information requirements.

The empirical assessment uses five types of roadside ITS data assets, including operational data and data retained through performance-evaluation activities.

The framework first defines the mandatory information required for each task and compares those requirements with the information actually retained in each data asset. When required information is not directly available, its recoverability is assessed before task-specific usability is determined. The approach therefore distinguishes information that is explicitly available or directly obtainable, information that can be recovered through feasible transformation, information that cannot be reliably recovered retrospectively, and information whose status cannot be confirmed from the evidence examined. The resulting usability assessment further distinguishes data that are directly usable, usable after transformation, currently unsuitable, or unconfirmed for a given task. Information gaps are subsequently linked either to data-transformation requirements or, where retrospective recovery is not feasible, to future data-retention needs. The empirical scope is therefore explicitly bounded by the five examined ITS data assets—roadside CCTV, camera-based AIDS, radar-based AIDS, radar-based SIS, and VDS performance-evaluation data—and the four selected data-use tasks defined in this study.

The study addresses the following research questions:

RQ1. What mandatory information requirements determine the usability of the five examined ITS data assets for the four selected data-use tasks: 2D road-user detection training, detection/measurement output validation, event/structured traffic-state retrieval, and road-referenced trajectory reconstruction?

RQ2. To what extent are these task-specific information requirements retained, directly obtainable, recoverable through transformation, not reliably recoverable, or unconfirmed across heterogeneous existing ITS data assets?

RQ3. How do these information-level differences affect the task-specific usability of operational and performance-evaluation ITS data?

RQ4. What transformation and future data-retention requirements arise from the information gaps identified for different autonomous-driving AI tasks?

This study contributes an integrated assessment framework that links task-specific information requirements with the information actually retained in heterogeneous ITS data assets, the recoverability of missing information, and the resulting task-specific usability. The framework preserves the information-level basis for each usability judgment rather than reducing the assessment to a single overall AI-readiness score. It also distinguishes uncertainty from confirmed irrecoverability and links the resulting gaps to either transformation of existing data or future retention of information that cannot be reconstructed retrospectively. By applying this approach to both operational and performance-evaluation ITS data, the study demonstrates how the usability of existing roadside data changes according to the intended task and the information actually retained.

2. Related Work

2.1. Data Quality and Readiness for Artificial Intelligence

Conventional data-quality research has expanded toward the broader concept of data readiness for artificial intelligence (AI), which considers not only whether data are available but also whether they are sufficiently reliable, accessible, representative, and suitable for the intended application. Hiniduma et al. organized data-readiness metrics across structured and unstructured data and showed that AI readiness involves multiple dimensions beyond conventional measures such as completeness or correctness [1]. Related readiness frameworks have similarly incorporated broader considerations such as quality, bias, fairness, privacy, and fitness for use [2].

This broader view is also reflected in international standardization. The ISO/IEC 5259 series provides a common foundation for data quality in analytics and machine learning. ISO/IEC 5259-1 establishes terminology and conceptual foundations, while ISO/IEC 5259-2 defines measurable data-quality characteristics for assessing whether data are sufficiently trustworthy and fit for purpose in analytics and machine-learning applications [9,10]. These developments reinforce the view that data readiness should not be treated as a single intrinsic property of a dataset. Its practical meaning depends on both the characteristics of the available data and the requirements of the intended use.

This perspective provides an important basis for assessing transportation data, but general AI-readiness frameworks are not designed around the information structures of roadside ITS data. Autonomous-driving applications may depend on information such as object-level semantics, spatial referencing, temporal continuity, reference–target relationships, traffic-event semantics, or road-referenced trajectories, depending on the intended task. The relevance of these information elements cannot be determined from general data-quality characteristics alone. Assessing existing ITS data for autonomous-driving AI therefore requires a domain-specific and task-aware interpretation of data usability.

2.2. Data Requirements for Autonomous-Driving AI Development

Autonomous-driving research has generated a large body of purpose-built datasets for perception, tracking, prediction, validation, and related AI tasks. Unlike general-purpose datasets, these resources typically incorporate annotations and contextual information selected for specific driving tasks. Recent task-centric data-quality research has similarly

emphasized that the relevance of data characteristics depends on the autonomous-driving task being considered [8]. This task-oriented view suggests that the information required from a dataset should be defined in relation to its intended use rather than by applying a single set of data characteristics across all autonomous-driving applications.

For perception-oriented training, visual datasets commonly retain object classes and image-domain localization information such as bounding boxes, together with the underlying visual input [3,4,6]. Standardized annotation formats such as ASAM OpenLABEL likewise represent object classes and image-space locations, reflecting their role in object-detection data [11]. Sequential tasks impose different requirements. Tracking and trajectory-oriented datasets preserve observations across time together with persistent object identities and successive positions, while road-referenced trajectory reconstruction additionally requires a consistent spatial reference [5,6,12,13]. V2X-Seq, for example, combines sequential observations with trajectories and map information for cooperative tracking and forecasting [5]. These data structures illustrate why repeated detections or measurements alone are not equivalent to a reconstructable road-referenced trajectory.

Validation-oriented use requires a different information structure. The output being evaluated must be associated with sufficiently reliable reference information, and the relationship between the target output and the corresponding reference event or measurement must be identifiable [14–18]. Measurement semantics are also necessary to interpret what is being compared. The reference information should therefore not be assumed to constitute perfect ground truth merely because it is available; its role depends on how it was obtained, validated, and linked to the target output [14,18].

Event- or traffic-state-oriented retrieval has another set of requirements. Such use depends on locating relevant observations in time and associating them with machine-readable event or traffic-state semantics so that records can be searched, selected, or organized according to the intended analytical purpose [11,19,20]. This differs from trajectory reconstruction, which requires continuity of object identity and position across successive observations; data may therefore support retrieval of structured traffic states or events without containing the information needed to reconstruct continuous road-referenced trajectories.

These task-specific requirements are particularly relevant to data that were not originally collected for autonomous-driving development, because the information retained for their original purpose may differ from that required for a new AI task. Their potential usability therefore depends not only on the type of data available, but also on the specific information retained within the data asset.

2.3. Roadside and Infrastructure-Based Data for Autonomous Driving

Roadside sensing has been widely investigated as a source of information for autonomous-driving applications. Compared with vehicle-mounted sensing, elevated roadside sensors can provide a broader and more persistent view of traffic scenes and may reduce some occlusion and sensing-range limitations. Rope3D was developed specifically for roadside monocular 3D perception and contains approximately 50,000 images and more than 1.5 million annotated 3D objects collected under heterogeneous camera configurations, viewpoints, and environmental conditions [3]. The dataset illustrates how roadside camera data can support perception-oriented autonomous-driving research under diverse viewing and environmental conditions.

Vehicle–infrastructure cooperative datasets have extended this concept further. DAIR-V2X provides large-scale camera and LiDAR data from real-world vehicle and infrastructure viewpoints together with 3D annotations [4]. The study also identifies temporal asynchrony between vehicle and infrastructure sensors as a challenge for cooperative perception,

illustrating that infrastructure-based data use depends not only on sensing capability but also on temporal alignment and data integration.

Subsequent datasets have incorporated richer sequential and contextual information. V2X-Seq extends vehicle–infrastructure data toward sequential perception and forecasting by providing trajectories, maps, and traffic-light information [5]. TUMTraf-V2X integrates roadside and onboard sensing and includes labeled images and point clouds together with track identities, GPS and IMU information, and diverse traffic scenarios [6]. These datasets show that roadside and cooperative data can support a range of autonomous-driving research tasks, from perception to tracking and trajectory-oriented analysis.

The studies above establish the usefulness of roadside and infrastructure-based data for autonomous-driving research, but they also highlight a key difference from the operational ITS data examined in this study. Rope3D, DAIR-V2X, V2X-Seq, and TUMTraf-V2X were developed specifically to support autonomous-driving research, and their annotations, calibration information, spatial references, temporal structures, and other data characteristics were intentionally organized around their intended tasks [3–6]. Existing ITS data, by contrast, are generally retained according to the requirements of their original traffic-management, system-operation, or performance-evaluation purposes. The demonstrated value of purpose-built roadside datasets therefore cannot be directly generalized to existing ITS data. Whether an existing ITS asset can support a new AI task depends on what information was actually retained and, where required information is missing, whether it can be reliably recovered from that asset.

2.4. Repurposing Existing Operational Traffic Data

Beyond purpose-built autonomous-driving datasets, previous studies have examined whether data collected for traffic monitoring or other operational purposes can support new uses. Yu et al. provide a notable example by reconstructing approximately five million vehicle trajectories from archived records obtained from more than 3000 existing traffic cameras [7]. Their approach combined spatial–temporal vehicle re-identification, path inference, and trajectory recovery, demonstrating that operational traffic data can support substantially different downstream uses when sufficient source information remains available.

More broadly, data-reuse research has examined how existing datasets can be adapted to purposes beyond those for which they were originally collected. Dataset harmonization and reuse frameworks have shown that data from heterogeneous sources can be transformed, linked, or reorganized to support analyses beyond their original applications [19,21]. In the transportation domain, data-reuse approaches have also examined how information generated for one mobility or infrastructure purpose can support other ITS applications [22]. These studies suggest that the potential value of existing data should not be judged solely by their original collection purpose or current data structure.

However, repurposing also has limits. Some information that is not explicitly represented in an existing dataset can be derived or reconstructed from retained source data through annotation, semantic conversion, data linkage, calibration, or spatial and temporal reconstruction. Other information may depend on source data, identifiers, reference relationships, calibration parameters, or spatial information that were not retained and therefore cannot be reconstructed reliably after collection. A third situation arises when the available evidence is insufficient to determine whether the required information exists or can be recovered. These cases are materially different for secondary use, even though each may initially appear simply as “missing information”.

The reuse of operational ITS data therefore involves an intermediate question between retained information and final task usability: whether information required for a new task can be recovered with sufficient reliability from what was actually retained. Existing studies

demonstrate the feasibility and value of data reuse, transformation, and harmonization, but they have not generally provided a systematic basis for distinguishing recoverable information gaps from gaps that cannot be resolved retrospectively, particularly across heterogeneous ITS data assets and different autonomous-driving AI tasks.

2.5. Research Gap and Positioning of the Present Study

The literature reviewed above establishes several relevant foundations. General AI-readiness research has shown that data suitability depends on the intended application rather than on data quality alone [1,2]. Autonomous-driving datasets and task-centric research further demonstrate that different tasks require different combinations of semantic, spatial, temporal, and reference information [3–6,8]. Roadside and vehicle–infrastructure datasets confirm the value of infrastructure-based sensing for autonomous-driving research [3–6], while studies on data reuse show that information collected for one purpose can sometimes be transformed, harmonized, or linked to support a different use [19,21,22].

Across these research streams, an intermediate problem remains insufficiently addressed. Existing ITS data assets are typically retained for traffic management, system operation, or performance evaluation rather than for autonomous-driving AI. As a result, the information available in a retained asset may only partially correspond to the information required for a new AI task. Assessing such data therefore requires more than identifying the sensor type, nominal system capability, or general data-quality characteristics. It requires determining which task-specific information elements are actually retained and, when required information is not explicitly available, whether it can be recovered with sufficient reliability from the retained data and associated evidence.

Information that appears to be missing can have different implications for reuse. Some information may be obtained directly from retained records or recovered through feasible transformation, whereas other information may depend on source data or contextual information that is no longer available and therefore cannot be reconstructed reliably. In other cases, the available evidence may be insufficient to determine whether the required information exists or is recoverable. Treating these situations as a single category of missing data can either underestimate the reuse potential of existing ITS assets or overestimate their readiness for a new AI task.

The present study addresses this gap by integrating task-specific information requirements, retained-information assessment, recoverability, and usability within a common evaluation framework for existing ITS data assets. Rather than assigning a single overall AI-readiness score to each asset, the framework evaluates usability separately for selected autonomous-driving AI tasks and preserves the information-level basis for each judgment. It further distinguishes information gaps that can be addressed through feasible transformation of existing data from those that require additional information to be retained in future ITS data assets. The framework is applied to heterogeneous operational and performance-evaluation ITS data assets to examine how their usability varies across AI tasks and to identify the transformation or future data-retention requirements associated with the observed information gaps.

3. Methodology

3.1. Research Framework

This study developed a task-specific framework for assessing whether data retained from existing intelligent transportation systems (ITS) can support selected AI tasks for autonomous-driving systems. The framework is based on the premise that the usability of existing ITS data cannot be determined solely from sensor modality, system capability, or data format. Different AI tasks require different types and levels of information, and infor-

mation that is not explicitly retained may, in some cases, still be derived or reconstructed from the available data. The assessment therefore considers both the information retained in an existing data asset and the recoverability of information required for a specific AI task.

Figure 1 presents the overall assessment framework, which consists of six steps. First, the existing ITS data assets available for assessment are identified (Step 1). These assets may include operational data as well as data retained through performance-evaluation activities. Second, the intended autonomous-driving AI task is specified (Step 2). Four tasks are considered in this study: 2D road-user detection training (T1), detection/measurement output validation (T2), event/structured traffic-state retrieval (T3A), and road-referenced trajectory reconstruction (T3B). These tasks were selected to represent distinct AI uses with different information requirements rather than different levels of AI capability.

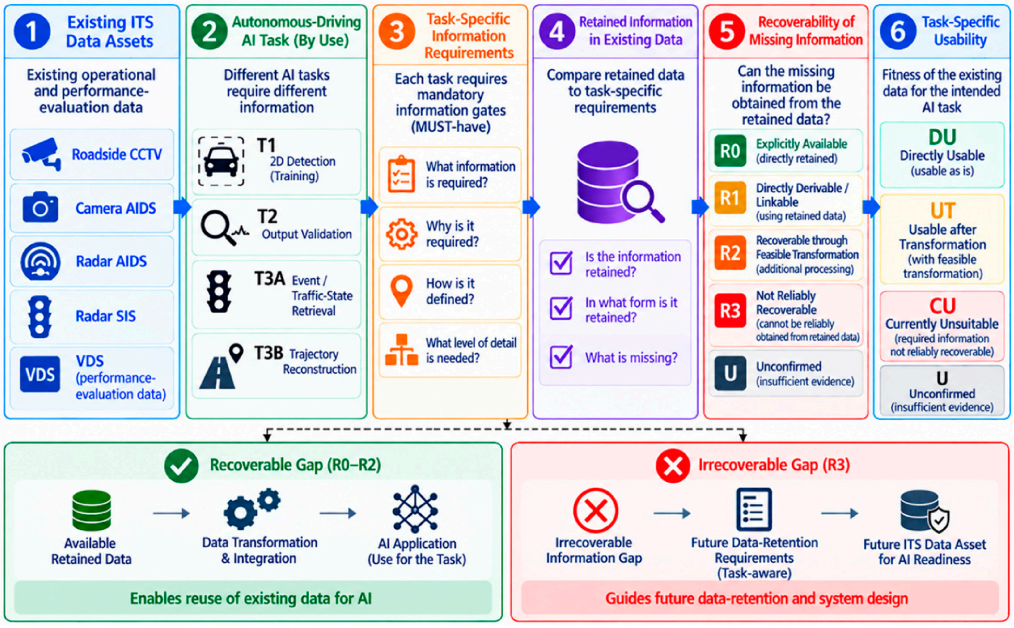


Figure 1. Task-Specific Framework for Assessing the Usability of Existing ITS Data for Autonomous-Driving AI.

For each task, the information required to perform the intended use is then defined (Step 3). Information designated as mandatory is treated as a set of information gates that must be satisfied when determining task-specific usability. The information actually retained in each ITS data asset is subsequently compared with these requirements (Step 4). This assessment is based on the retained evidence available for examination rather than on the information that the original sensor or system may technically have been capable of generating. This distinction prevents system capability from being interpreted as evidence of data availability when the corresponding raw data, processed outputs, metadata, or reference information were not retained or could not be confirmed.

When mandatory information is not explicitly available, its recoverability from the retained data is assessed (Step 5). Five recoverability statuses are used: R0 for information that is explicitly available; R1 for information that is directly derivable or linkable; R2 for information that is recoverable through feasible transformation; R3 for information that is not reliably recoverable from the retained information; and U when availability or recoverability cannot be confirmed from the evidence examined. These statuses characterize the availability and recoverability of required information and are not intended to constitute an ordinal scale of data quality or system performance.

The recoverability statuses of the mandatory information gates are then used to determine the task-specific usability of each data asset (Step 6). Four usability states are distinguished: directly usable (DU), usable after transformation (UT), currently unsuitable (CU), and unconfirmed (U). Accordingly, the resulting usability state represents the fitness of a particular retained data asset for a specified AI task, rather than a general quality rating of the ITS equipment or dataset. The detailed task definitions, mandatory information requirements, recoverability criteria, and usability decision rules are described in Sections 3.3–3.5.

The framework further distinguishes recoverable information pathways (R0–R2) from retrospectively irrecoverable information gaps (R3). Information assigned R0–R2 can support the intended task either directly or through derivation, linkage, or transformation, thereby providing a pathway for the reuse of existing ITS data. By contrast, an R3 status indicates that the required information cannot be reliably recovered from the retained data. Such an information gap constrains current reuse for the corresponding task, while also identifying information that should be considered in future data-retention and system-design strategies if the same AI use is anticipated. The framework thus supports both the assessment of existing ITS data for current AI applications and the identification of task-aware data-retention requirements for future ITS deployments.

3.2. Empirical ITS Data Assets

The empirical assessment was conducted using five types of data assets obtained from existing roadside ITS systems and associated performance-evaluation activities: roadside closed-circuit television (CCTV), camera-based automatic incident detection system (AIDS), radar-based AIDS, radar-based smart intersection system (SIS), and vehicle detection system (VDS) performance-evaluation data. These assets provided contrasts in sensing modality, data structure, and retention practice within the empirical material available for this study. The analysis focused on the information actually retained and available for examination in each asset, rather than on the full technical capabilities of the corresponding equipment. To improve the traceability of the subsequent recoverability assessment, Supplementary Table S1 summarizes the characteristics of the retained holdings actually examined, including the data period, approximate data volume, recording or sampling basis, file or record format, key retained information or fields, and assessment-relevant metadata. Because the five assets include continuous video, event-based records, cycle-based traffic records, and performance-evaluation measurements, the applicable recording or sampling basis is reported according to the structure of each asset rather than by imposing a common sampling-frequency measure. Representative examples of the retained data are provided in Supplementary Figure S1.

Roadside CCTV represented a primarily visual data source. The retained asset consisted of recorded video from fixed roadside cameras together with associated temporal information, providing observable road users, traffic movements, and surrounding roadway context. Such imagery can preserve substantial visual information; however, information required for AI use, such as object annotations, explicit lane references, or road-coordinate mappings, is not necessarily stored with the video itself. The CCTV asset was therefore assessed on the basis of the video and associated information that could be confirmed from the available material, without assuming the availability of additional calibration or metadata solely from the presence of the camera system.

The camera-based AIDS and radar-based AIDS assets represented systems designed to detect traffic events and generate associated detection or measurement outputs. Although the two systems employ different sensing modalities, the assessment did not infer data availability from their sensing principles. Instead, the retained records and outputs

available for each system were examined to determine whether task-relevant information, including detection results, event-related information, traffic-state variables, and supporting metadata, was explicitly retained or could be obtained from the available records. This approach allowed differences in retained information to be evaluated independently of the nominal sensing capability of the underlying systems. For camera-based AIDS, the examined performance-evaluation holdings also included video retained for evaluation and verification purposes. Such video was distinguished from routine operational retention and was not taken to imply that continuous video is normally retained by the system.

The radar-based SIS asset provided structured information associated with intersection traffic monitoring. The retained data included system-generated traffic-state information that could be examined for task-specific use. As with the AIDS assets, the analysis was restricted to information that was explicitly available or recoverable from the retained records. The presence of radar sensing was not taken as evidence that raw radar measurements, object-level trajectories, or other intermediate sensing products had been retained unless such information could be verified from the available data.

The VDS asset differed from the other system-derived assets because the material examined in this study consisted specifically of performance-evaluation data, rather than a general archive of operational VDS data. These records were produced and retained in the context of ITS performance evaluation and contained structured VDS outputs together with information used to assess the corresponding detection or measurement results. The term VDS performance-evaluation data is therefore used throughout this study to distinguish this asset from the native sensing stream or routine operational data that a VDS may generate. In particular, visual evidence used during performance evaluation was treated as evaluation evidence and was not interpreted as the native raw-data stream of the evaluated VDS.

This distinction was especially relevant for camera-based sensing. The fact that a system employs a camera does not establish that continuous raw video was retained as part of the system data available for subsequent reuse. Similarly, performance-evaluation CCTV or other auxiliary visual evidence was treated as evaluation evidence and was not attributed to the target ITS system unless it originated from that system and formed part of the examined data holdings. These distinctions were maintained throughout the assessment to avoid attributing information to a data asset on the basis of sensing capability or auxiliary evaluation evidence rather than confirmed data retention.

Table 1 summarizes the five empirical data assets, including their sensing or system type, original operational purpose, principal retained information, and the role of each asset in the subsequent task-specific assessment. Because the assets differ substantially in both data modality and retention structure, they were not compared using a single measure of data richness or completeness. Instead, each asset was evaluated against the information requirements of the selected AI tasks defined in the following sections.

Table 1. Overview of the Existing ITS Data Assets Used in the Empirical Assessment.

ITS Data Asset	Primary Purpose	Data Source	Retained Information Examined
Roadside CCTV	Traffic monitoring	Roadside camera	Recorded video with temporal road-scene information
Camera-based AIDS	Incident detection and traffic monitoring	Roadside camera	Event/detection outputs, visual evidence, and performance-evaluation records

Table 1. *Cont.*

ITS Data Asset	Primary Purpose	Data Source	Retained Information Examined
Radar-based AIDS	Incident detection and traffic monitoring	Roadside radar	Event/detection outputs, traffic information, and performance-evaluation records
Radar-based SIS	Intersection traffic monitoring and operational efficiency	Roadside radar	Vehicle/traffic-state outputs, traffic information, and performance-evaluation records
VDS performance-evaluation data	Traffic measurement and performance evaluation	Camera/radar VDS and laser reference equipment	VDS measurements, validated reference measurements, correspondence information, and performance-analysis results

Note: AIDS, automatic incident detection system; SIS, smart intersection system; VDS, vehicle detection system. Retained information refers only to data actually preserved and examined in this study and does not represent the full sensing or processing capability of each ITS system type. Detailed characteristics of the examined holdings are provided in Supplementary Table S1, with representative retained-data examples shown in Supplementary Figure S1.

3.3. Autonomous-Driving AI Tasks

To operationalize task-specific usability, four representative AI tasks were defined to reflect distinct ways in which existing ITS data may be reused for autonomous-driving applications. The tasks were selected to span different functional purposes and information structures rather than to represent a hierarchy of AI sophistication. In particular, the selected tasks cover model training, performance validation, structured traffic-state retrieval, and trajectory reconstruction. They represent data-use functions supporting autonomous-driving AI development, validation, and downstream analysis rather than onboard driving functions themselves.

The usefulness of an ITS data asset depends on the information required by the intended task, not simply on the apparent richness of the underlying sensor data.

The first task, T1: 2D road-user detection training, represents the use of retained ITS data as training material for image- or frame-based road-user detection. For this purpose, the essential information is not limited to the presence of visual data. The training sample must also support class interpretation and two-dimensional localization of the target object within the image domain. Accordingly, the usability of an existing data asset for T1 depends on whether visual input, class semantics, and 2D object localization are available or can be reliably recovered from the retained data.

This task therefore captures a common reuse scenario in which raw or recorded visual content may have potential value but still requires task-specific semantic preparation before it becomes suitable for AI training.

The second task, T2: detection/measurement output validation, addresses the use of existing ITS data to evaluate the correctness of a system-generated detection or measurement result. Unlike T1, this task does not primarily require rich raw sensory input. Instead, it requires a clearly defined target output, a validated reference event or measurement, a reliable correspondence between the target result and the reference, and sufficient measurement semantics to interpret the comparison. In this study, a validated reference denotes an event, state, or measurement that was verified and adopted as the reference within the relevant evaluation procedure; it is not assumed to constitute perfect ground truth. Validation-oriented reuse depends strongly on traceability and reference–target correspondence, even when the underlying data are less visually rich.

The third task, T3A: event/structured traffic-state retrieval, represents the extraction or reuse of temporally localized, machine-readable traffic information from retained ITS

data. This task includes event-oriented information when such semantics are available, but also accommodates structured traffic-state information such as detections, counts, speeds, occupancies, or related system outputs. T3A was defined broadly enough to accommodate the different output structures of the empirical ITS assets while maintaining a common requirement that the retained information be interpretable in a structured form. For VDS performance-evaluation data, T3A is limited to structured traffic-state retrieval rather than semantic incident retrieval, reflecting the type of information actually retained in that asset.

The fourth task, T3B: road-referenced trajectory reconstruction, represents a more demanding form of reuse in which successive observations must be organized into the movement history of a road user within a road- or world-referenced spatial framework. This requires persistent identity and temporal continuity across successive observations, together with sufficient positional information to reconstruct movement within a consistent road- or world-referenced spatial framework. These characteristics distinguish trajectory reconstruction from the simple availability of repeated detections or distance measurements. A data asset may contain temporally ordered or object-related records without providing sufficient information to establish persistent identity or reconstruct positions in a common road-referenced coordinate system.

Accordingly, timestamps or chronologically ordered records alone were not treated as a temporal sequence of successive observations of the same road user. Thus, the presence of repeated measurements alone was not treated as evidence of trajectory usability.

Taken together, the four tasks were intended to create a controlled contrast in information demand. T1 emphasizes visual content together with semantic labeling and image-space localization; T2 emphasizes validated comparison and reference–target traceability; T3A emphasizes temporally indexed, machine-readable traffic semantics; and T3B emphasizes identity continuity, sequential positioning, and spatial referencing. This structure enables the same retained ITS data asset to be assessed differently depending on the intended AI use. The task-specific information requirements used for these four tasks are defined in Section 3.4.

3.4. Task-Specific Information Requirements

For each AI task defined in Section 3.3, a set of task-specific information requirements was established before assessing the retained information of the individual ITS data assets. The purpose of this step was to define the minimum information needed for the intended use independently of whether a particular asset happened to contain that information. This separation between task requirements and observed data availability was necessary to avoid deriving the assessment criteria retrospectively from the characteristics of the available datasets.

The requirements were organized into mandatory and additional information. Mandatory information represents the functional minimum needed to perform the basic form of the intended task with interpretable and verifiable inputs or outputs, rather than the full set of information that may be required for a specific implementation or performance level. Each mandatory item therefore acts as an information gate in the subsequent usability assessment. If a mandatory item is not explicitly retained, its recoverability must be evaluated before the data asset can be considered usable for the task. Additional information may improve data quality, robustness, interpretability, or the range of downstream analyses, but its absence alone does not determine the usability state. This distinction allows the framework to evaluate task fitness without requiring every potentially useful attribute to be present. The usability states defined in this study represent information-level task suitability rather than comprehensive data readiness. Satisfying the mandatory information gates indicates that the information structurally required for the basic task is available or

recoverable, but does not by itself establish that the data meet all quality requirements for practical AI development or validation. Factors such as spatial or temporal resolution, occlusion, coverage, sample diversity, class balance, measurement accuracy, and environmental representativeness may further determine downstream fitness for use and should therefore be evaluated separately when a specific implementation is considered.

For T1 (2D road-user detection training), the mandatory requirements are visual input, class semantics, and 2D object localization. Visual input provides the observable road-user content from which the model learns, whereas class semantics define the target category associated with an object. Two-dimensional localization specifies the position or spatial extent of that object within the image domain, for example through a bounding box or an equivalent representation. The availability of video or image data alone is therefore insufficient for direct use in supervised 2D detection training if the semantic class and object localization required to construct training labels are absent. Such missing information may nevertheless be recoverable through annotation or other feasible transformation, which is evaluated separately in the recoverability step.

For T2 (detection/measurement output validation), four mandatory requirements are defined: target output, validated reference event or measurement, reference–target correspondence, and measurement semantics. The target output identifies the detection or measurement being evaluated, while the validated reference event or measurement provides the corresponding reference against which that output can be assessed. Reference–target correspondence establishes that the two records refer to the same observation or evaluation instance, for example through temporal, spatial, identifier-based, or other reliable linkage. Measurement semantics specify what the recorded values represent, including the relevant variable, unit, category, or interpretation needed for comparison. These requirements reflect the fact that validation depends not simply on the coexistence of target and reference records, but on their traceable and interpretable correspondence. As noted in Section 3.3, the term validated reference refers to information verified and adopted as the reference within the relevant evaluation procedure and does not imply error-free ground truth.

For T3A (event/structured traffic-state retrieval), the mandatory requirements are temporal localization and machine-readable event or traffic-state semantics. Temporal localization is required to associate the retained information with a defined time or interval, whereas machine-readable semantics are required to determine what traffic event, state, or measurement the record represents. Depending on the source system, such information may describe an event or may consist of structured traffic variables such as detections, counts, speeds, or occupancies. The requirement is therefore defined in terms of interpretable structured traffic information rather than a single event taxonomy. This formulation allows heterogeneous ITS outputs to be assessed under a common task definition while preserving the distinction between semantic incident information and structured traffic-state information. Accordingly, an asset was evaluated only for the form of structured retrieval supported by its retained information; in particular, VDS performance-evaluation data were assessed for structured traffic-state retrieval and were not assumed to contain semantic incident information.

For T3B (road-referenced trajectory reconstruction), the mandatory requirements are persistent identity, temporal sequence, sequential position, and metric road/world reference. Persistent identity is needed to associate successive observations with the same road user, and temporal sequence establishes their ordering over time. Accordingly, temporal sequence refers to the ordering of successive observations of the same road user, rather than merely the presence of timestamps or chronologically ordered records.

Sequential position provides the spatial observations required to reconstruct movement, while metric road/world reference relates those positions to a consistent roadway or real-world coordinate framework. All four elements are required because isolated detections, repeated measurements, or relative distance records do not by themselves constitute a road-referenced trajectory. In particular, repeated observations without reliable identity continuity or without a reconstructable metric spatial reference were not treated as satisfying the trajectory requirement.

The mandatory requirements were applied consistently across all empirical assets for a given task. Thus, the criteria were not relaxed or redefined according to sensor modality, original system purpose, or apparent data richness. This common-gate approach enables heterogeneous ITS data assets to be compared against the same functional requirements while allowing their recoverability pathways to differ. For example, information not explicitly available in one asset may be directly derived, linked, or recovered through feasible transformation, whereas the same information may already be explicitly retained in another asset. These differences are captured by the recoverability assessment rather than by changing the task requirements themselves.

Table 2 summarizes the mandatory and additional information considered for each task. The mandatory items form the decision basis for the usability classification, whereas additional information is used to characterize information that may enhance the practical value or reliability of the data without acting as a mandatory gate. The recoverability categories applied to these requirements and the rule used to derive the final task-specific usability states are described in Section 3.5.

Table 2. Task-Specific Information Requirements for the Selected Autonomous-Driving AI Tasks.

Task	Intended AI Use	Mandatory Information Requirements	Supporting Basis
T1. 2D road-user detection training	Training object detectors using roadside data	Visual input; class semantics; 2D object localization	Roadside perception datasets and annotation standards [3,4,6,11]
T2. Detection/ measurement output validation	Comparing ITS-generated outputs with validated reference measurements	Target output; validated reference; reference–target correspondence; measurement semantics	Reference-based ADAS/ADS evaluation and sensor-validation studies [14–18]
T3A. Event/structured traffic-state retrieval	Retrieving events or structured traffic states from retained ITS data	Temporal localization; machine-readable event or traffic-state semantics	Scenario retrieval and structured annotation approaches [11,19,20]
T3B. Road-referenced trajectory reconstruction	Reconstructing continuous object movement in road/world coordinates	Persistent identity; temporal sequence; sequential position; metric road/world reference	Sequential V2X datasets, trajectory studies, and scenario standards [5,6,12,13]

Note: Only information required to perform the basic function of each task is listed as mandatory. Additional information, such as environmental conditions, detailed road geometry, sensor calibration, object attributes, or velocity and heading, may be necessary for specific implementations, performance levels, or analysis objectives, but its absence alone does not preclude the basic task defined here.

3.5. Recoverability Assessment and Task-Specific Usability Classification

After the task-specific information requirements were defined, each empirical data asset was assessed at the level of individual mandatory information items. The assessment was conducted in two stages. First, each required information item was assigned a recoverability status (R0–R3 or U) according to the evidence retained in the asset and the feasibility of obtaining the required information from that evidence. Second, the recoverability sta-

tuses of all mandatory information gates for a given asset–task pair were combined using a deterministic decision rule to assign the final task-specific usability state (DU, UT, CU, or U). Figure 2 summarizes this two-stage procedure.

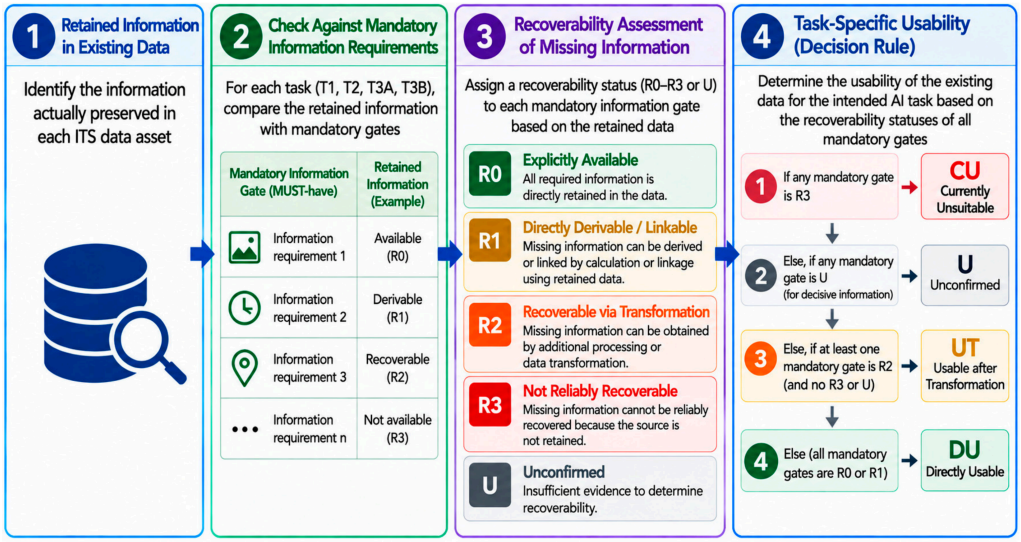


Figure 2. Decision Logic for Recoverability Assessment and Task-Specific Usability Classification.

For each asset, the retained files, structured records, metadata, and available evaluation evidence were first inventoried. Each mandatory information requirement was then assessed against the information retained in the target data asset, with evaluation evidence used where applicable to verify the interpretation or correspondence of that information. When a direct match was unavailable, direct derivability or linkage from the retained information was examined first, followed by the feasibility of transformation-based recovery. If the available evidence was insufficient to support any of these judgments, the requirement was assigned U rather than being treated as absent.

The assessment was performed by the sole author; therefore, inter-rater agreement could not be assessed. To improve consistency and traceability, the same predefined task-specific information requirements and recoverability criteria were systematically applied across all five ITS data assets. The judgments were based on the retained files, structured records, metadata, and available evaluation evidence for each asset, and cases with insufficient evidence were assigned U rather than resolved by assumption. The supporting retained-data characteristics and evidence are further documented in Supplementary Table S1 and Figure S1.

Five recoverability statuses were used. R0 (explicitly available) was assigned when the required information was directly retained and identifiable in the available data. R1 (directly derivable or linkable) was assigned when the information was not necessarily stored as a standalone field but could be obtained through direct calculation, deterministic derivation, or reliable linkage using retained information, without reconstructing missing source content. R2 (recoverable through transformation) was assigned when the required information was not directly available or derivable but could be reconstructed from the retained data through a feasible additional processing step, such as annotation, semantic harmonization, calibration-based reconstruction, or coordinate transformation. R3 (not reliably recoverable) was assigned when the retained data did not contain sufficient source information to reconstruct the required item with defensible reliability. U (unconfirmed) was assigned when the evidence available for this study was insufficient to determine whether the required information was retained or recoverable.

The distinction between R1 and R2 was based on whether obtaining the required information involved direct derivation or linkage from retained values, or instead required a substantive transformation that introduced an additional reconstruction step. R2 was therefore assigned only when an identifiable and feasible transformation pathway could be supported by the retained data. The theoretical possibility that information might be inferred was not, by itself, considered sufficient for an R2 assignment.

A transformation was considered feasible when the required information could be produced from the retained evidence through an identifiable processing procedure without assuming unavailable source observations. Where recoverability depended on auxiliary information that was not confirmed in the retained holdings, the requirement was assigned U rather than R2. An R2 assignment therefore indicates that a defensible recovery pathway was identified from the retained evidence; it does not imply that every transformation was implemented or that the recovered information was independently validated for all downstream applications.

Conversely, R3 was used when a required source element was absent from the retained data such that retrospective reconstruction could not be supported reliably. The U status was kept separate from R3 because lack of evidence in the material examined does not demonstrate that the information was absent from the original system or is irrecoverable from evidence that may exist outside the material examined.

The recoverability assessment was applied to the mandatory information gates defined for T1, T2, T3A, and T3B. Table 3 reports the resulting information-level recoverability matrix. Each cell therefore describes the status of a specific information requirement for a particular retained data asset and should not be interpreted as a general performance rating of the corresponding ITS equipment. The R0–R3 categories are also not treated as numerical scores or as an ordinal scale to be averaged across requirements. Their role is to identify how each required information item is available, derivable, transformable, or unavailable for the intended task.

The information-level results in Table 3 were then aggregated into the task-specific usability states reported in Table 4. Only mandatory information gates were used in this classification. Here, U denotes an unconfirmed status for a mandatory information gate that prevents a conclusive usability classification unless an R3 gate already determines the asset–task pair to be currently unsuitable. The decision rule was applied hierarchically. If any mandatory gate was assigned R3, the asset–task pair was classified as currently unsuitable (CU) because at least one functionally required information item could not be reliably recovered. R3 takes precedence over U because confirmation of an irrecoverable mandatory requirement is sufficient to rule out current task usability regardless of uncertainty in other mandatory gates. If no R3 was present but at least one mandatory gate was unconfirmed (U), the pair was classified as unconfirmed (U) because the available evidence was insufficient to support a more definitive usability judgment. If all mandatory gates were confirmed and at least one was assigned R2, with no R3 or U, the pair was classified as usable after transformation (UT). Finally, if all mandatory gates were R0 or R1, the pair was classified as directly usable (DU).

Accordingly, the classification rule can be expressed as follows:

$$Usability(a,t) = \begin{cases} \text{CU}, & \text{if any mandatory gate is R3,} \\ \text{U}, & \text{if no R3 exists and a mandatory gate is U,} \\ \text{UT}, & \text{if neither R3 nor U exists and at least one mandatory gate is R2,} \\ \text{DU}, & \text{if all mandatory gates are R0 or R1,} \end{cases}$$

where a denotes an ITS data asset and t denotes an AI task.

Table 3. Recoverability of Mandatory Information Requirements across Existing ITS Data Assets.

AI Task	Mandatory Information Requirement	Roadside CCTV	Camera-Based AIDS	Radar-Based AIDS	Radar-Based SIS	VDS Performance-Evaluation Data
T1. 2D road-user detection training	Visual input	R0	R0	U	U	R3
	Class semantics	R2	R0	R0	R0	R3
	2D object localization	R2	R0	U	U	R3
T2. Detection/measurement output validation	Target output	R3	R0	R0	R0	R0
	Validated reference event or measurement	R3	R0	R0	R0	R0
	Reference–target correspondence	R3	R1	R1	R1	R0
	Measurement semantics	R3	R0	R0	R0	R0
T3A. Event/structured traffic-state retrieval	Temporal localization	R0	R0	R0	R0	R0
	Machine-readable event or traffic-state semantics	R2	R0	R0	R0	R0
T3B. Road-referenced trajectory reconstruction	Persistent identity	R2	U	U	U	R3
	Temporal sequence	R0	U	U	U	R3
	Sequential position	R2	U	U	R3	R3
	Metric road/world reference	U	U	U	R3	R3

Note: R0 = explicitly available; R1 = directly derivable or linkable; R2 = recoverable through feasible transformation; R3 = not reliably recoverable from the retained information; U = unconfirmed based on the evidence examined. For roadside CCTV, metric road/world reference was classified as U because camera-to-road calibration or sufficient evidence for its reliable reconstruction was not confirmed in the retained holdings. For camera-based AIDS, trajectory-related requirements were classified as U where the retained evidence was insufficient to confirm continuous object-level information; evaluation-management CCTV was not treated as equivalent to the native sensing stream of the target system. For VDS data, T3A refers to structured traffic-state retrieval rather than semantic incident retrieval. Visual evidence retained for performance evaluation was not treated as equivalent to the native sensing stream of the target ITS system.

Table 4. Task-Specific Usability of Existing ITS Data Assets for Autonomous-Driving AI Tasks.

ITS Data Asset	T1. 2D Road-User Detection Training	T2. Detection/Measurement Output Validation	T3A. Event/Structured Traffic-State Retrieval	T3B. Road-Referenced Trajectory Reconstruction
Roadside CCTV	UT	CU	UT	U
Camera-based AIDS	DU	DU	DU	U
Radar-based AIDS	U	DU	DU	U
Radar-based SIS	U	DU	DU	CU
VDS performance-evaluation data	CU	DU	DU †	CU

Note: DU = directly usable; UT = usable after transformation; CU = currently unsuitable; U = unconfirmed. † For VDS performance-evaluation data, T3A usability is limited to structured traffic-state retrieval rather than semantic incident retrieval.

This mandatory-gate rule intentionally avoids compensatory scoring. In other words, the presence of several readily available information items cannot offset the absence of another item that is functionally required for the task. Similarly, an R2 requirement does

not make a data asset unusable if the missing information can be recovered through a feasible transformation; instead, the transformation becomes a prerequisite for task use. This logic preserves the task-specific meaning of the assessment and prevents information richness in one dimension from masking a critical information gap in another.

For mandatory information items assigned R2, the required transformation was additionally characterized according to its functional role, including annotation, semantic harmonization, and spatial/temporal reconstruction; calibration or coordinate transformation was treated as a supporting procedure where required. These transformation pathways provide an explicit bridge between the recoverability assessment and practical data preparation by indicating the additional processing required before the corresponding data asset can support the intended AI task.

By contrast, R3 identifies an information-retention gap that cannot be addressed retrospectively using the retained asset and therefore motivates future data retention rather than transformation of the existing asset. The corresponding transformation and retention requirements are presented in Section 4.3.

Figure 2 summarizes the decision logic used in the assessment. The resulting information-level recoverability matrix and task-specific usability classifications are reported in Tables 3 and 4, respectively, in Section 4. This separation preserves traceability from the final usability classification back to the specific information requirement and retained evidence on which the judgment was based.

4. Results and Discussion

4.1. Information-Level Recoverability Across ITS Data Assets

Table 3 presents the recoverability of the mandatory information requirements for the four AI tasks across the five ITS data assets. The classifications varied across both assets and information requirements. Some requirements were explicitly available or directly linkable (R0–R1), whereas others required transformation (R2), remained unconfirmed (U), or could not be reliably recovered (R3). The resulting patterns differed by task-specific information requirement and could not be explained by sensor modality alone.

For T1, roadside CCTV retained the required visual input (R0), while class semantics and 2D object localization were not explicitly available but could be recovered through feasible transformation (R2). Camera-based AIDS contained all three mandatory information elements as R0. In contrast, the evidence examined for radar-based AIDS and radar-based SIS did not confirm visual input or 2D object localization, although class semantics were explicitly available in both assets. Visual input and 2D object localization were therefore classified as U rather than being treated as absent. VDS performance-evaluation data showed a different limitation. The three mandatory T1 requirements were classified as R3 because the retained VDS performance-evaluation data did not contain the source information needed to reconstruct a reusable image-domain training sample, including native visual input and corresponding object-level class and localization information. CCTV footage used during performance evaluation served as auxiliary evaluation evidence and was therefore not treated as part of the VDS native data stream.

For T2, most structured system and performance-evaluation assets retained the mandatory validation information. Camera-based AIDS, radar-based AIDS, and radar-based SIS explicitly retained the target output, validated reference event or measurement, and measurement semantics (R0). Reference–target correspondence was directly derivable or linkable (R1) for these three assets. VDS performance-evaluation data retained all four mandatory T2 information elements as R0, including explicit reference–target correspondence. For roadside CCTV, none of the four mandatory T2 information requirements could be reliably recovered from the retained video, and all were therefore classified as R3.

For T3A, temporal localization was explicitly available (R0) in all five assets. Machine-readable event or traffic-state semantics were also explicitly available in both AIDS assets, radar-based SIS, and VDS performance-evaluation data. Roadside CCTV differed in that such semantics were not explicitly retained but could be generated from the recorded visual information through feasible transformation, resulting in an R2 classification. For VDS performance-evaluation data, this result refers specifically to structured traffic-state retrieval rather than semantic incident retrieval, consistent with the type of information retained in the evaluation records.

The largest limitations and uncertainties were observed for T3B. For roadside CCTV, temporal sequence was explicitly available (R0), while persistent identity and sequential position were classified as R2 because they could be recovered from the retained continuous video through additional tracking-based processing. Metric road/world reference, however, was classified as U because the retained holdings did not contain confirmed camera-to-road calibration information or sufficient evidence to establish that such calibration could be reconstructed with adequate reliability. For camera-based AIDS, all four trajectory-related requirements were classified as U. Although event records and event snapshots were retained, the submitted native system video was not available for re-examination, and evaluation-management CCTV was not treated as part of the native system data. The available evidence was therefore insufficient to confirm persistent identity, temporal sequence, sequential position, or metric road/world reference. For radar-based AIDS, all four trajectory-related requirements likewise remained unconfirmed. For radar-based SIS, persistent identity and temporal sequence remained unconfirmed (U), while sequential position and metric road/world reference were classified as R3. VDS performance-evaluation data showed the strongest trajectory-related limitation, with all four mandatory T3B requirements classified as R3.

The information-level results show a general contrast between information aligned with the original purpose of the retained asset and information required for a different AI use. Structured system outputs and evaluation records generally preserved the information needed for validation and structured traffic-state retrieval, whereas visual training information and road-referenced trajectory information more often required transformation, additional evidence, or information that could not be reliably recovered from the retained data. These information-level differences provide the basis for the task-specific usability classifications presented in Section 4.2.

4.2. Task-Specific Usability of Existing ITS Data Assets

Table 4 presents the task-specific usability classifications derived from the information-level recoverability results in Table 3, and Figure 3 provides an integrated view of the resulting asset–task patterns.

The classifications varied substantially across tasks, including within the same data asset. No asset was therefore characterized by a single usability level across all four tasks; instead, its usability depended on whether the mandatory information required for the intended task was retained, recoverable, or unconfirmed.

For T1, roadside CCTV was classified as usable after transformation (UT). Although visual input was explicitly available, the class semantics and 2D object localization required for detection training had to be generated from the retained imagery through annotation or equivalent transformation. Camera-based AIDS was directly usable (DU) because all mandatory T1 information was explicitly available in the retained asset. By contrast, radar-based AIDS and radar-based SIS were classified as unconfirmed (U), reflecting unresolved availability of mandatory visual or image-domain information in the evidence examined. VDS performance-evaluation data were classified as currently unsuitable (CU) because the

retained VDS performance-evaluation data did not contain the source information needed to reconstruct the image-domain training information required for T1. CCTV footage used during performance evaluation served as auxiliary evaluation evidence and was not treated as part of the VDS native data stream.

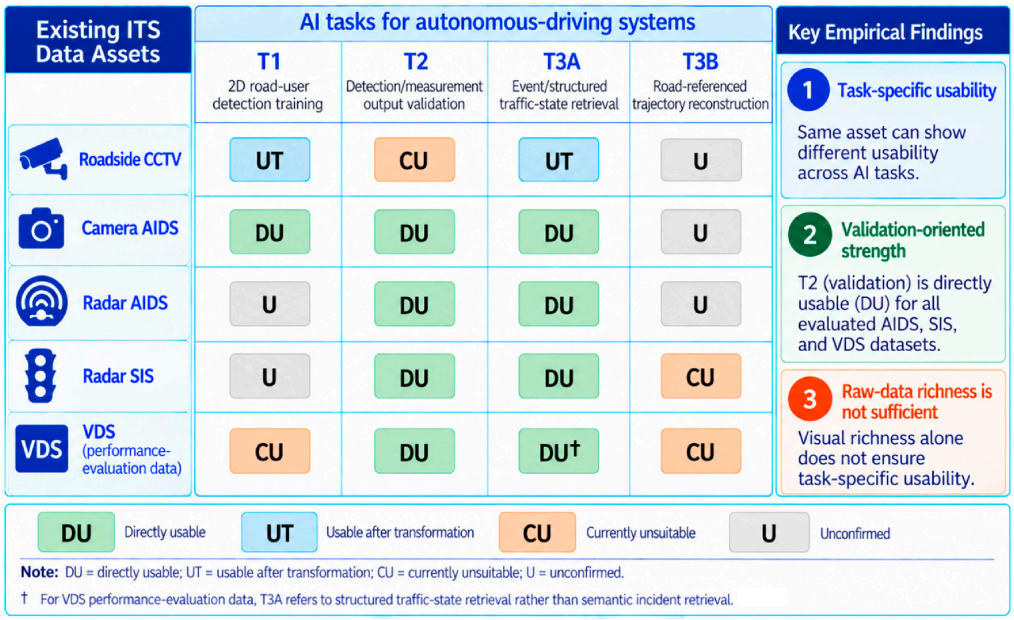


Figure 3. Task-Specific Usability Matrix and Key Empirical Findings across Existing ITS Data Assets.

T2 produced the broadest pattern of direct usability among the system-generated and performance-evaluation assets. Camera-based AIDS, radar-based AIDS, and radar-based SIS were classified as DU because their mandatory validation information was either explicitly retained or directly linkable. VDS performance-evaluation data were also classified as DU, with the target output, validated reference event or measurement, reference–target correspondence, and measurement semantics retained within the evaluation records. In contrast, roadside CCTV was classified as CU because the mandatory target and reference information required for detection/measurement output validation could not be reliably recovered from the retained video alone.

For T3A, all structured system-derived assets were classified as DU. Camera-based AIDS, radar-based AIDS, radar-based SIS, and VDS performance-evaluation data retained the temporal localization and machine-readable event or traffic-state semantics required for this task. For the VDS asset, this classification refers specifically to structured traffic-state retrieval rather than semantic incident retrieval. Roadside CCTV was classified as UT because temporal information was available, whereas machine-readable traffic semantics had to be generated from the retained imagery through transformation.

T3B showed the most constrained usability pattern. Roadside CCTV was classified as U because, although persistent identity and sequential position were recoverable through transformation and temporal sequence was explicitly available, the metric road/world reference required for road-referenced trajectory reconstruction could not be confirmed from the retained evidence. Because a reliable camera-to-road calibration could not be established from the retained evidence, this study did not specify a spatial-error tolerance for T3B; such a threshold would require an explicit calibration and validation procedure outside the scope of the present assessment.

Camera-based AIDS was also classified as U because the available retained data did not confirm any of the four mandatory trajectory-related information requirements. Radar-based AIDS was likewise classified as U because the mandatory trajectory-related

information remained unconfirmed. Radar-based SIS was classified as CU because at least one mandatory trajectory requirement was assigned R3. VDS performance-evaluation data were also classified as CU, with all four mandatory T3B requirements assigned R3. In particular, repeated detections, measurements, or temporally ordered records were not considered sufficient for road-referenced trajectory reconstruction when persistent identity, sequential position, or metric spatial reference could not be reliably established.

The usability matrix shows that no data asset had a uniform usability state across the four tasks. Camera-based AIDS provided the broadest direct usability, whereas roadside CCTV showed greater potential through transformation. The radar-based assets and VDS performance-evaluation data were directly usable for T2 and T3A but were limited for visual training or road-referenced trajectory reconstruction. Overall, task-specific usability could not be inferred from sensing modality or the apparent richness of an ITS data asset alone.

4.3. Transformation and Future Data-Retention Requirements

The recoverability and usability results identified three distinct types of requirements for extending the use of existing ITS data. The first concerns information that was not explicitly available but remained recoverable from the retained asset (R2). In these cases, additional processing is required before the data can support the intended task. The second concerns mandatory information whose availability or recoverability remained unconfirmed (U). Such cases require further verification before a transformation or future-retention pathway can be assigned. The third concerns mandatory information that was not reliably recoverable from the retained asset (R3). Such cases cannot be addressed by transforming the existing data alone and instead indicate information that would need to be retained in future data collection or system operation. Table 5 summarizes these transformation, verification, and retention requirements.

The R2 cases were associated with three main transformation pathways. Annotation (TR1) was required when information needed for AI use could be generated from retained visual content but was not explicitly encoded in the original data. This was particularly relevant to roadside CCTV for T1, where class semantics and 2D object localization could be added through annotation. Semantic harmonization (TR2) was applicable when retained information had to be converted or mapped into the semantic representation required by the target task. Spatial/temporal reconstruction (TR3) was required when retained observations had to be transformed into a consistent spatial or temporal representation. For roadside CCTV, tracking-based reconstruction could support persistent identity and sequential position; however, T3B usability remained unconfirmed because the metric road/world reference required for the task could not be established from the retained evidence.

These transformation pathways apply to information that remains recoverable from the existing asset and therefore correspond to R2 cases. Where the availability or recoverability of mandatory information remained unconfirmed (U), further verification was required before either a transformation or future-retention pathway could be assigned. By contrast, R3 cases require future data retention because the source information needed for reliable retrospective recovery was not retained. In such cases, the requirement is not an additional transformation of the existing asset but the retention of the missing information in future data collection or system operation. Examples include preserving visual or image-domain source information when future visual AI training is anticipated, and retaining persistent object identity, sequential position, or metric road/world reference when road-referenced trajectory reconstruction is an intended future use.

Table 5. Transformation and Data-Retention Requirements for Task-Specific Reuse of Existing ITS Data.

ITS Data Asset/AI Task	Key Gap	Pathway	Reuse Implication
Roadside CCTV/T1	Class semantics and 2D localization	TR1	Task-ready training data can be generated from retained video
Roadside CCTV/T3A	Machine-readable semantics	TR1, with TR2 where needed	Structured event/traffic-state data can be derived
Roadside CCTV/T3B	Trajectory reconstruction cannot be confirmed until a reliable metric road/world reference is established	Verification	Metric road/world reference remains unconfirmed
Camera AIDS/T3B; Radar AIDS/T1, T3B; Radar SIS/T1	Required information remains unconfirmed	Verification	Transformation pathway requires further evidence
Radar SIS/T3B	Sequential position and metric road/world reference not recoverable	Future data retention	Current data remain unsuitable; future retention can enable reuse
VDS/T1, T3B	Visual/object or trajectory-supporting information not recoverable	Future data retention	Current data remain unsuitable; future retention can enable reuse

Note: TR1 = annotation; TR2 = semantic harmonization; Verification indicates that the availability or recoverability of the required information must first be confirmed before a transformation pathway can be assigned. Future data retention indicates that the required information cannot be reliably recovered from the retained asset and therefore needs to be retained in future data collection or system operation.

The transformation requirements also differed across the evaluated assets and tasks. Roadside CCTV showed the greatest dependence on transformation: its retained imagery could support T1 and T3A after additional processing, while T3B remained unconfirmed because the required metric spatial reference could not be verified from the retained evidence. In contrast, the structured AIDS, SIS, and VDS performance-evaluation assets generally required little or no transformation for T2 and T3A because the relevant system outputs, semantics, and temporal or reference information were already retained or directly linkable. Their limitations became more pronounced for tasks requiring information outside the structure of the retained records, particularly visual training information and road-referenced trajectory information.

These results distinguish between information gaps that can be addressed through additional processing of existing data and those that require changes in future data retention. The former can be addressed through task-specific transformation pathways, whereas the latter require changes in what information is retained if the corresponding AI use is to be supported in the future. This distinction provides a direct link between the recoverability assessment of existing ITS data and the identification of task-specific data-preparation and retention requirements.

4.4. Discussion and Research Contributions

The results show that the usability of the existing ITS data examined in this study was not a fixed property of a sensor, system, or dataset. The same retained asset exhibited different usability states depending on the intended task. Roadside CCTV, for example, was usable after transformation for T1 and T3A, was currently unsuitable for T2, and

remained unconfirmed for T3B because the required metric road/world reference could not be established from the retained evidence. VDS performance-evaluation data showed the opposite pattern in several respects, being directly usable for T2 and T3A while remaining unsuitable for T1 and T3B. These differences support the use-specific perspective introduced in Sections 1 and 2: the relevant question is not whether an existing ITS dataset is generally "AI-ready", but whether the information required for a particular AI task is retained or can be recovered with sufficient reliability.

Accordingly, the DU and UT states should be interpreted as indicating information-level suitability for the basic task defined in this study, rather than comprehensive AI-data quality or readiness. For example, retained CCTV may contain the visual content and recoverable annotations required for T1 while still varying in image resolution, occlusion, viewpoint coverage, environmental diversity, class distribution, or other quality characteristics that can affect practical model development and performance. These broader quality dimensions therefore remain relevant even when the mandatory information gates are satisfied and should be evaluated separately for a specific downstream implementation.

This finding is consistent with prior research showing that AI data readiness depends on the intended application and that autonomous-driving datasets are typically organized around task-specific information needs [1,8]. Purpose-built roadside and vehicle–infrastructure datasets such as Rope3D, DAIR-V2X, V2X-Seq, and TUMTraf-V2X retain annotations, calibration information, spatial references, persistent identities, or sequential observations because these elements are required for their intended perception, tracking, and forecasting tasks [3–6]. The existing ITS assets examined in this study differ in an important respect: they were originally retained for traffic management, system operation, or performance evaluation rather than for autonomous-driving AI. Consequently, sensing modality alone did not determine downstream usability. The presence of a camera did not necessarily imply that reusable visual training data had been retained, and repeated or temporally ordered radar-system records did not by themselves provide a road-referenced trajectory. The empirical results therefore reinforce the distinction between what a system is technically capable of sensing and what information is actually preserved in an accessible data asset.

The results also show why recoverability needs to be considered separately from current data availability. Previous research on data readiness and dataset reuse has established the importance of fitness for purpose and has demonstrated that existing data can sometimes be transformed into new analytical resources [1,19,21,22]. However, treating missing information as simply absent may underestimate the reuse potential of retained data, while assuming that technically inferable information is recoverable may overestimate it. The R0–R3/U scheme addresses this intermediate problem by distinguishing information that is explicitly retained or directly obtainable, information that can be recovered through a feasible transformation, information that cannot be reliably recovered retrospectively, and information whose status cannot be confirmed from the available evidence. This distinction was consequential in the empirical results. Missing class semantics and 2D localization in retained CCTV imagery could be recovered through annotation, whereas missing target–reference information required for T2 could not be reconstructed from the video alone. Similarly, unresolved trajectory-related information in several radar-based assets was retained as U rather than being interpreted as proof of nonexistence.

The T2 results reveal another source of value that is easily overlooked in existing ITS data: performance-evaluation provenance. For T2, camera-based AIDS, radar-based AIDS, radar-based SIS, and VDS performance-evaluation data were all directly usable because the retained information preserved system outputs together with validated reference information and sufficient correspondence for comparison. In the VDS case, this value

did not depend on retaining the native sensing stream. Rather, it arose from the traceable relationship among the evaluated system output, the corresponding reference information, and the evidence retained during performance evaluation. This result complements prior work on dataset reuse and data provenance [22,23] by showing that evaluation records themselves can constitute useful validation-oriented data assets when their reference relationships and semantics are preserved. At the same time, the result underlines the need to distinguish supporting evaluation evidence from the native data produced by the evaluated ITS equipment. Accordingly, the stronger T2 usability observed for several assets should not be interpreted as an inherent advantage of the corresponding sensing modality or system type. It partly reflects the provenance and retention context of the examined holdings, particularly the preservation of validated reference information and reference–target correspondence during performance evaluation.

The transformation results also show that UT should not be interpreted as a measure of transformation effort. Annotation, semantic harmonization, and spatial/temporal reconstruction provide different pathways for converting retained information into task-usable forms. Generating class labels and 2D object locations from retained video clearly satisfied the R2 criterion because the necessary source information was present in the retained imagery. By contrast, trajectory-related recovery required more stringent evidence: where metric road/world referencing depended on unconfirmed calibration information, the corresponding requirement was retained as U rather than classified as R2. UT therefore identifies the existence of a feasible recovery pathway supported by the retained evidence, rather than implying that all technically conceivable transformations are usable in practice. This distinction is important when the framework is used to identify candidate data assets for practical reuse.

The distinction between R2 and R3 further connects current data reuse with future data-retention decisions. Where mandatory information remained recoverable, existing assets could potentially be reused through additional processing without changing the original sensing system. Where source information had not been retained and could not be reconstructed reliably, however, retrospective processing could not resolve the gap.

These cases indicate what information may need to be preserved if a similar AI task is anticipated in future ITS operation or performance evaluation. Such retention does not imply that all raw data should be stored indefinitely. Rather, the results support a task-aware approach in which selected metadata, reference relationships, calibration parameters, persistent identifiers, spatial information, or source data are retained when they are necessary for an intended downstream use. This provides a practical connection between reuse of existing ITS assets and more selective future data-retention strategies.

These findings also have implications for the sustainable management of digital transportation infrastructure. Reusing suitable existing ITS data can extend the functional value of previously collected information, while task-aware retention can help avoid unnecessary duplication of data collection and the indiscriminate accumulation of data with limited downstream value. Although the present study does not quantify environmental or economic benefits, the framework provides a structured basis for linking data reuse and selective retention to more resource-efficient long-term ITS data management.

The contribution of the proposed framework lies in how these elements are combined rather than in task-specific assessment alone. The proposed framework links task-specific mandatory information requirements with the information actually retained in heterogeneous existing ITS assets, evaluates the recoverability of missing information before assigning usability, distinguishes uncertainty from confirmed irrecoverability, and connects the resulting gaps to transformation or future-retention requirements.

The transferability of the framework lies primarily in its task-and-information-requirement structure rather than in the empirical usability ratings assigned to the five examined assets. For another ITS dataset or system, the same assessment logic can be applied by first defining the intended data-use task and its mandatory information requirements, and then evaluating whether those requirements are retained or recoverable. The mandatory requirements therefore represent the functional minimum needed for the basic task, whereas additional information may improve data quality, robustness, interpretability, or performance for a specific implementation. The empirical DU/UT/CU/U classifications reported in this study should not, however, be transferred directly to other installations or datasets without examining their actual retained information.

The empirical application further shows that conventional operational and performance-evaluation data can provide different forms of value for autonomous-driving AI, including model training, output validation, and structured traffic-state retrieval, while also revealing substantial information limitations for road-referenced trajectory reconstruction. Rather than reducing these differences to a single AI-readiness score, the assessment preserves the information-level basis for each task-specific usability judgment. Several limitations should be considered when interpreting these findings. The empirical assessment was based on five ITS data-asset types and four selected AI data-use tasks, and the resulting usability classifications therefore should not be interpreted as universal ratings for all implementations of CCTV, AIDS, SIS, or VDS systems. In addition, recoverability was assessed from the retained data and evaluation evidence available for this study; information that could not be confirmed was intentionally classified as U rather than inferred from nominal system capability. Finally, the framework identifies whether a feasible transformation pathway exists, but the identified transformations were not necessarily implemented or independently validated. The assessment also did not quantitatively evaluate broader data-quality dimensions such as spatial or temporal resolution, occlusion, coverage, representativeness, sample or class diversity, or measurement accuracy, nor did it quantify transformation cost, processing effort, post-transformation data quality, or downstream AI performance. Future work could extend the assessment to additional ITS assets and autonomous-driving tasks while incorporating quantitative measures of transformation effort and post-transformation data quality.

5. Conclusions

This study examined whether existing ITS data collected for traffic management, system operation, and performance evaluation can be reused for autonomous-driving AI. Rather than treating AI usability as an inherent property of a data asset or sensing system, the study assessed whether the information required for a specific task was actually retained and, when required information was not directly available, whether it could be recovered with sufficient reliability. The proposed framework was applied to five heterogeneous ITS data assets across four data-use tasks: 2D road-user detection training, detection/measurement output validation, event/structured traffic-state retrieval, and road-referenced trajectory reconstruction.

The empirical results showed that no data asset had a uniform usability state across all four tasks. Camera-based AIDS provided the broadest direct usability, whereas roadside CCTV showed greater potential through transformation. Radar-based AIDS, radar-based SIS, and VDS performance-evaluation data were directly usable for validation and structured traffic-state retrieval, whereas their usability for visual training and road-referenced trajectory reconstruction was either unconfirmed or currently unsuitable depending on the retained information. The results also showed that sensing modality alone was insufficient to determine AI usability. What mattered was the information actually retained for the

intended task, including semantic, temporal, spatial, identity, and reference information. In particular, the performance-evaluation records examined in this study showed distinctive value for validation-oriented use because they preserved target outputs together with validated reference information and their correspondence, even though they did not necessarily contain rich native sensor data.

A further finding was that missing information should not be treated as a single condition. Some information absent from the retained records could be recovered through feasible annotation, semantic harmonization, or spatial and temporal reconstruction, whereas other information could not be reliably reconstructed retrospectively because the necessary source information had not been retained. In other cases, the available evidence was insufficient to confirm recoverability. Distinguishing these conditions made it possible to separate data that were directly usable from those requiring transformation, those currently unsuitable for a particular task, and those whose usability remained unconfirmed. The contribution of the framework therefore lies in integrating task-specific information requirements, retained-information assessment, recoverability, and usability while preserving the information-level basis for each judgment rather than reducing heterogeneous ITS data to a single AI-readiness score.

These findings have practical implications for both the reuse of existing ITS data and the future management of ITS data assets. Existing ITS data should not be regarded as unsuitable solely because they were not originally collected for autonomous-driving AI. Where the required information is already retained or can be reliably recovered, reuse and transformation of existing data may provide an alternative source of data for AI development, validation, or downstream analysis. Conversely, when task-required information cannot be reconstructed retrospectively, the identified gaps can inform future ITS data-retention practices. This does not imply indiscriminate retention of all raw data. Rather, future ITS deployment and data-management strategies could selectively preserve task-relevant information that is difficult or impossible to reconstruct later, such as visual or image-domain source information, persistent object identities, sequential position information, and calibration or metric road/world reference where appropriate.

This reuse-and-retention perspective also provides a basis for considering the sustainable management of digital transportation infrastructure. Extending the useful value of existing ITS data, avoiding unnecessary duplication of data collection where suitable information already exists, and selectively retaining information according to anticipated uses may support more resource-efficient data management. The framework does not quantify reductions in data-acquisition cost, storage demand, or environmental impacts, and such benefits should therefore not be inferred directly from the present results. Nevertheless, the findings provide a basis for considering data reuse and selective retention as part of the long-term management of digital ITS infrastructure.

The study has several limitations. The empirical assessment was based on five heterogeneous ITS data assets and was intended to examine the applicability and discriminatory capability of the framework rather than to provide a statistically representative assessment of all ITS systems. Some information states remained unconfirmed because the available evidence was insufficient, and the transformations associated with recoverable information were assessed in terms of feasibility rather than implemented, independently validated, and quantitatively evaluated for all cases. The study also did not measure transformation effort, cost, post-transformation data quality, or downstream AI performance. Future research could therefore extend the assessment to additional ITS assets, systems, and autonomous-driving tasks; implement the identified transformation pathways and evaluate post-transformation data quality and downstream AI utility; and quantify the effort and cost associated with transforming existing data. Broader cross-system or cross-country ap-

plications could further examine the generalizability of the framework and the implications of different ITS data-retention practices.

Overall, the findings show that existing ITS data should not be judged as universally AI-ready or AI-unusable simply because they were collected for another purpose. Their value for autonomous-driving AI depends on the information required by the intended task, what was actually retained, and whether missing information can still be recovered with sufficient reliability. This task-specific distinction provides a basis for deciding what existing ITS data can be reused, what can be made usable through transformation, and what information should be considered for future retention when retrospective recovery is no longer possible.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su182110720/s1, Figure S1: Representative examples of the retained ITS data examined in this study; Table S1: Characteristics of the retained ITS data examined in the empirical assessment.

Funding: This research was supported by the Korea Institute of Civil Engineering and Building Technology (KICT) under the "2026 ITS Basic Performance Evaluation" project (Project No. 20260027-001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because they consist of operational and performance-evaluation data subject to institutional and operational restrictions. The characteristics of the data and the information relevant to the assessment are described in the article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADAS	Advanced Driver-Assistance Systems
ADS	Automated Driving Systems
AI	Artificial Intelligence
AIDS	Automatic Incident Detection System
CCTV	Closed-Circuit Television
CU	Currently Unsuitable
DU	Directly Usable
ITS	Intelligent Transportation Systems
SIS	Smart Intersection System
UT	Usable after Transformation
V2X	Vehicle-to-Everything
VDS	Vehicle Detection System

References

1. Hiniduma, K.; Byna, S.; Bez, J.L. Data Readiness for AI: A 360-Degree Survey. *ACM Comput. Surv.* **2025**, *57*, 219. [CrossRef]
2. Hiniduma, K.; Ryan, D.; Byna, S.; Bez, J.L.; Madduri, R.K. AIDRIN 2.0: A Framework to Assess Data Readiness for AI. *arXiv* **2025**, arXiv:2505.18213. [CrossRef]
3. Ye, X.; Shu, M.; Li, H.; Shi, Y.; Li, Y.; Wang, G.; Tan, X.; Ding, E. Rope3D: The Roadside Perception Dataset for Autonomous Driving and Monocular 3D Object Detection Task. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 21–24 June 2022; pp. 21341–21350. [CrossRef]
4. Yu, H.; Luo, Y.; Shu, M.; Huo, Y.; Yang, Z.; Shi, Y.; Guo, Z.; Li, H.; Hu, X.; Yuan, J.; et al. DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 21–24 June 2022; pp. 21361–21370. [CrossRef]

5. Yu, H.; Yang, W.; Ruan, H.; Yang, Z.; Tang, Y.; Gao, X.; Hao, X.; Shi, Y.; Pan, Y.; Sun, N.; et al. V2X-Seq: A Large-Scale Sequential Dataset for Vehicle-Infrastructure Cooperative Perception and Forecasting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–22 June 2023; pp. 5486–5495. [CrossRef]
6. Zimmer, W.; Wardana, G.A.; Sritharan, S.; Zhou, X.; Song, R.; Knoll, A.C. TUMTraf V2X Cooperative Perception Dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 22668–22677. [CrossRef]
7. Yu, F.; Yan, H.; Chen, R.; Zhang, G.; Liu, Y.; Chen, M.; Li, Y. City-scale Vehicle Trajectory Data from Traffic Camera Videos. *Sci. Data* **2023**, *10*, 711. [CrossRef] [PubMed]
8. Zhou, Y.; Chen, H.; Sha, K. A Novel Multilayer Task-Centric and Data Quality Framework for Autonomous Driving. *IEEE Internet Comput.* **2026**, *30*, 67–75. [CrossRef]
9. *ISO/IEC 5259-1:2024*; Artificial Intelligence—Data Quality for Analytics and Machine Learning (ML)—Part 1: Overview, Terminology, and Examples. International Organization for Standardization (ISO): Geneva, Switzerland; International Electrotechnical Commission: Geneva, Switzerland, 2024.
10. *ISO/IEC 5259-2:2024*; Artificial Intelligence—Data Quality for Analytics and Machine Learning (ML)—Part 2: Data Quality Measures. International Organization for Standardization (ISO): Geneva, Switzerland; International Electrotechnical Commission: Geneva, Switzerland, 2024.
11. ASAM e.V. *ASAM OpenLABEL 1.0.0*; Association for Standardization of Automation and Measuring Systems: Höhenkirchen, Germany, 2021.
12. Punzo, V.; Borzacchiello, M.T.; Ciuffo, B. On the Assessment of Vehicle Trajectory Data Accuracy and Application to the Next Generation SIMulation (NGSIM) Program Data. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 1243–1262. [CrossRef]
13. ASAM e.V. *ASAM OpenSCENARIO XML 1.4.0*; Association for Standardization of Automation and Measuring Systems: Höhenkirchen, Germany, 2026.
14. Haas, S.; Solmaz, S.; Reckenzaun, J.; Genser, S. ViF-GTAD: A New Automotive Dataset with Ground Truth for ADAS/AD Development, Testing, and Validation. *Int. J. Robot. Res.* **2023**, *42*, 614–630. [CrossRef]
15. Rövid, A.; Vincze, Z.; Pálinkás, T.; Kocsis, M.; Serrano, V.; Szalay, Z. Reference Platform for ADAS Camera System Evaluation. *Sensors* **2025**, *25*, 1690. [CrossRef] [PubMed]
16. Rehrl, K.; Gröchenig, S. Evaluating Localization Accuracy of Automated Driving Systems. *Sensors* **2021**, *21*, 5855. [CrossRef] [PubMed]
17. Zhao, C.; Ding, D.; Shi, Y.; Ji, Y.; Du, Y. Graph Matching-Based Spatiotemporal Calibration of Roadside Sensors in Cooperative Vehicle-Infrastructure Systems. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 9281–9295. [CrossRef]
18. Holder, M.; Elster, L.; Winner, H. Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation. *Energies* **2022**, *15*, 989. [CrossRef]
19. Li, Q.; Peng, Z.M.; Feng, L.; Liu, Z.; Duan, C.; Mo, W.; Zhou, B. ScenarioNet: Open-Source Platform for Large-Scale Traffic Scenario Simulation and Modeling. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 3894–3920. [CrossRef]
20. Davidson, C.; Ramanan, D.; Peri, N. RefAV: Towards Planning-Centric Scenario Mining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, CO, USA, 3–7 June 2026; pp. 21537–21548.
21. Koesten, L.; Vougiouklis, P.; Simperl, E.; Groth, P. Dataset Reuse: Toward Translating Principles to Practice. *Patterns* **2020**, *1*, 100136. [CrossRef] [PubMed]
22. Miranda, C.; da Silva, A.S.; da Costa, J.P.J.; Santos, G.A.; da Silva, D.A.; de Freitas, E.P.; Vinel, A. A Virtual Infrastructure Model Based on Data Reuse to Support Intelligent Transportation System Applications. *IEEE Access* **2025**, *13*, 40607–40620. [CrossRef]
23. Longpre, S.; Mahari, R.; Chen, A.; Obeng-Marnu, N.; Sileo, D.; Brannon, W.; Muennighoff, N.; Khazam, N.; Kabbara, J.; Perisetla, K.; et al. A Large-Scale Audit of Dataset Licensing and Attribution in AI. *Nat. Mach. Intell.* **2024**, *6*, 975–987. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.