

Real and synthetic scenarios generated for the development, training, virtual testing and validation of CCAM systems



D2.2 Storylines definition and technical and interoperability requirements

Document Type	Document, Report
Document Number	D2.2
Primary Author(s)	Erwan REVERT IRT SystemX
Document Version / Status	1.0 Final
Distribution Level	PU (public)

Project Acronym	SYNERGIES
Project Website	www.synergies-ccam.eu
Project Coordinator	IDIADA AUTOMOTIVE TECHNOLOGY SA
Grant Agreement Number	101146542
Date of latest version of Annex I against which the assessment will be made	2024-05-27



**Funded by
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

CONTRIBUTORS

Name	Organization	Name	Organization
Erwan Revert	IRT SystemX	Lucas Dulewicz	VUFO
Marc El Zeenny	IRT SystemX	Konrad Reisinger	VUFO
Michael Schuldes	RWTH IKA	Elena Daskalaki	ICCS
Sven Tarlowski	RWTH IKA	Anastasia Bolovinou	ICCS
Onur Yildirim	TU Eindhoven	Moisés Rial Martinez	CTAG
Joaquin Vanschoren	TU Eindhoven	Mohsen Alirezaei	Siemens NL
Mykola Pechenizkiy	TU Eindhoven	Edwin van Hassel	Siemens NL
Bahram Zonooz	TU Eindhoven	Konstantinos Gkentsidis	Siemens BE
Sara Garcia	MOSAIC	Thibault Griffon	STELLANTIS
Burcu Kolbay	MOSAIC	Stanislas Di Piazza	STELLANTIS
Erwin de Gelder	TNO	Patrick Irvine	UoW
Fredrik Warg	RISE	Jerein Jeyachandran	UoW
Ludwig Friedmann	BMW	Xizhe Zhang (Jason)	UoW

FORMAL REVIEWERS

Name	Organization	Date
Jose Diaz Mendoza	TNO	19/05/2025
Marc Perez Quintana	IDIADA	29/05/2025

DOCUMENT HISTORY

Revision	Date	Author Organization	Description
0.1	2024-11-29	Erwan REVERT IRT SystemX	Preliminary release of the document containing input from tasks 2.3 and 2.4 for milestone 3.
1.0	2025-05-23	Erwan REVERT IRT SystemX	Final revision, containing contributions from tasks 2.2, 2.3 and 2.4.

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	8
2	INTRODUCTION	9
2.1	Project	9
2.2	Purpose of the deliverable	9
2.3	Relation with other work packages	9
3	KEY CONCEPTS AND TERMINOLOGY	10
3.1	Platform overview	10
3.2	Scenarios	10
3.3	Data	11
3.4	Terms and definitions	13
4	STORYLINES	16
4.1	Storyline's glossary	16
4.1.1	Classes	18
4.1.2	Relations	21
4.1.3	Graph	24
4.2	Storyline's definition	25
4.3	High-level reference architecture	34
5	CONSTRAINTS	35
5.1	Interoperability	35
5.1.1	Semantic interoperability requirements	35
5.1.2	Pragmatic interoperability requirements	36
5.1.3	Technical interoperability requirements	37
5.1.4	Data interoperability requirements	40
5.2	Data and Scenario Traceability	42
5.2.1	Lineage and Processing of Data and Scenarios	42
5.2.2	Data and Scenario Source Integration	43
5.2.3	Change Management System	44
5.2.4	Data Traceability KPIs	45
5.3	Data and Scenario Trustworthiness	46
5.3.1	Data and Scenario Completeness Assurance	46
5.3.2	Data and Scenario Accuracy Assurance	47
5.3.3	Relevance Assurance	49
5.3.4	Timeliness Assurance	50
5.3.5	Consistency Assurance	51

5.3.6	Data trustworthiness KPIs	52
5.4	Cybersecurity	53
5.4.1	Data Protection Framework	53
5.4.2	GDPR Compliance Framework	56
5.4.3	Access Control and Authentication	57
5.4.4	Cybersecurity KPIs	58
6	TECHNICAL REQUIREMENTS	60
6.1	Metadata format	60
6.1.1	Common metadata properties	61
6.1.2	Raw data metadata	63
6.1.3	Scenario specific metadata	64
6.1.4	Quality descriptors	66
6.2	Scenario source data format	67
6.3	Scenario format	70
6.4	Ontologies	73
7	CONCLUSION	74
8	REFERENCES	75
	ABBREVIATIONS	76

LIST OF FIGURES

Figure 1: SYNERGIES preliminary high-level platform architecture	10
Figure 2: Relationship of functional scenario, abstract scenario, logical scenario, and concrete scenario (source: ISO 34501:2022 [2]).....	11
Figure 3: Examples of raw data from different data sources.....	12
Figure 4: Conversion of various raw data examples into scenario source data	13
Figure 5: Storylines definition to map user needs into User types, System components and Data flows.	16
Figure 6: Essential SYNERGIES Platform components.....	21
Figure 7: User Roles within the SYNERGIES storylines. Source: D2.1's user types and roles.	22
Figure 8: Graph with Classes and Relations used to define the SYNERGIES Storylines.....	24
Figure 9: High-level reference architecture of the SYNERGIES Platform derived from the storylines.....	34

LIST OF TABLES

Table 1: Terms and definitions.....	15
Table 2: Terms and definitions already defined in other documents.....	18
Table 3: Top-level classes.....	18
Table 4: Derived classes.....	20
Table 5: Components of the SYNERGIES Platform.....	20
Table 6: Relations between User Roles classes (i.e., user actions).....	22
Table 7: Relations between other classes.....	23
Table 8: Relations related with SUNRISE and external SCDBs.....	24
Table 9: Shared vocabulary for high-level platform concepts and technical terms requirements.	36
Table 10: Shared scenario descriptors requirements.....	36
Table 11: Explicit assets scope requirements.....	37
Table 12: Availability of metadata requirements.....	37
Table 13: Resources unique identification requirements.....	38
Table 14: Availability of interfaces documentations within the platform requirements.....	38
Table 15: Version management of assets requirements.....	39
Table 16: Availability of an Application Programming Interface (API) requirements.....	39
Table 17: Tooling governance requirements.....	39
Table 18: Scenario Source Data Model(s) requirements.....	40
Table 19: Scenario Source Data File requirements.....	41
Table 20: Scenario format requirements.....	41
Table 21: Road network format requirements.....	42
Table 22: Scenario Extraction Methodology Documentation requirements.....	42
Table 23: Data and Scenario Lineage Recording requirements.....	42
Table 24: Scenario Selection and Modification requirements.....	43
Table 25: Source Credibility Assessment requirements.....	43
Table 26: Provider Credential Management requirements.....	44
Table 27: Modification tracking requirements.....	44
Table 28: Change Rationale Documentation requirements.....	44
Table 29: Version Rollback Capabilities requirements.....	45
Table 30: Data and Scenario Modification Traceability requirements.....	45
Table 31: Data Traceability KPIs requirements.....	46
Table 32: Data and Scenario Coverage requirements.....	46
Table 33: Comprehensive Metadata requirements.....	47

Table 34: Edge Case Inclusion requirements	47
Table 35: Scenario Plausibility Checks.....	47
Table 36: Data and Scenario Source Verification requirements.....	48
Table 37: Error Detection and Correction requirements.....	48
Table 38: Validation by Experts requirements	49
Table 39: Scenario Relevance Filtering requirements.....	49
Table 40: Dynamic Relevance Scoring requirements.....	49
Table 41: Periodic Reassessment requirements.....	50
Table 42: Use Case Alignment requirements.....	50
Table 43: Data Freshness Checks requirements.....	50
Table 44: Update Automation requirements.....	51
Table 45: Semantic Consistency requirements.....	51
Table 46: Temporal Alignment requirements.....	52
Table 47: Data/Scenario Integration Protocols requirements.....	52
Table 48: Data trustworthiness KPIs requirements.....	53
Table 49: End-to-End Encryption Implementation requirements.....	54
Table 50: Secure Storage and Transmission requirements.....	54
Table 51: Security Audit Procedures requirements.....	55
Table 52: Incident Response Protocol requirements.....	56
Table 53: Personal Data Processing requirements.....	56
Table 54: Data Subject Rights Management requirements.....	57
Table 55: Role-Based Access Control requirements.....	58
Table 56: Multi-Factor Authentication requirements.....	58
Table 57: Cybersecurity KPIs requirements.....	59
Table 58: Metadata typology	60
Table 59: Common metadata properties.....	63
Table 60: Raw data metadata properties.....	64
Table 61: Scenario metadata requirements.....	65
Table 62: Quality descriptors requirements.....	66
Table 63: Scenarios source data formats requirements.....	70
Table 64: Scenario format requirements.....	72
Table 65: Ontologies requirements	73

1 EXECUTIVE SUMMARY

This deliverable, D2.2 "Storylines definition and technical and interoperability requirements," formalizes the foundational elements necessary for the design and development of the SYNERGIES Platform—a European initiative to support scenario-based development, testing, and validation of Connected, Cooperative, and Automated Mobility (CCAM) systems.

Building on Tasks T2.2, T2.3, and T2.4, the document presents a comprehensive definition of user storylines, which capture stakeholder interactions with the platform across different roles, such as data providers, tool users, and scenario consumers. These storylines establish how users engage with core components like the Scenario Dataspace and Marketplace, and inform the high-level reference architecture of the platform.

In parallel, the deliverable specifies the **technical and non-functional requirements** that will guide implementation. These are structured around four key pillars:

- **Interoperability**, covering semantic, pragmatic, technical, and data-level compatibility.
- **Traceability**, ensuring the full lifecycle of data and scenarios is transparent and auditable.
- **Trustworthiness**, establishing quality criteria for data and scenario completeness, accuracy, and relevance.
- **Cybersecurity**, including GDPR compliance, access control, and data protection mechanisms.

Together, these requirements ensure that the platform will be robust, scalable, and aligned with European regulatory and industrial expectations. The document also defines common metadata models, file formats, versioning policies, and ontological frameworks that support federated access to heterogeneous scenario databases and tools.

In *Section 3: Key concepts and terminology*, we provide an overview of the SYNERGIES platform, outline the envisioned scenarios, specify the data types involved, and define relevant terms and definitions.

Section 4: Storylines, describes how various stakeholders will interact with the SYNERGIES platform. These storylines serve as use-cases to guide the requirements specification process, ultimately driving the technical development of the platform to ensure it meets the stakeholders' needs effectively.

Section 5: Constraints, focuses on the general requirements that constrain the design of the platform according to four key aspects: interoperability, traceability, data trustworthiness, and cybersecurity.

Finally, *Section 6: Technical requirements* presents a detailed specification of the specific technical requirements the platform must fulfill. These requirements, derived from the storylines and constraints discussed in the previous sections, form the basis for the platform's architecture and implementation.

Keywords: Key concepts, Glossary, Storylines, Interoperability, Traceability, Data trustworthiness, Cybersecurity, Technical requirements

2 INTRODUCTION

2.1 Project

SYNERGIES confronts pivotal challenges within the CCAM community, such as the absence of interoperable scenario databases, time-consuming and expensive development cycles, and regulatory ambiguities. It achieves this by implementing the Safety Assurance Framework developed in HEADSTART and SUNRISE. SYNERGIES furnishes stakeholders with interoperable, federated scenario databases, incorporating data from Safety Pool Scenario Database™, AD-Scene, StreetWise, VV Methods, L3Pilot, Hi-Drive, and more. This facilitates standardized processes, streamlines development cycles, and ensures regulatory compliance.

To accomplish this, SYNERGIES will culminate in a European platform designed to enhance the development, training, virtual testing, and validation of CCAM systems. The SYNERGIES Platform comprises a Scenario Dataspace, aligned with Europe's approach to data sharing and competitiveness, and a Marketplace, ensuring continual updates and Dataspace scalability.

Furthermore, SYNERGIES encourages the inclusion of new initiatives into the scenario dataspace by offering the requisite tools and guidance, from data processing and scenario identification to scenario database governance. This presents a unique opportunity to amplify investments in research and development, consolidating Europe's leadership in CCAM development, all while prioritizing safety and data protection.

Source: Grant agreement [1]

2.2 Purpose of the deliverable

The purpose of this deliverable is to define user storylines based on stakeholder needs and to establish the technical and interoperability requirements for the SYNERGIES Platform. By specifying the necessary features, components and data flows, this deliverable creates a coherent functional and technical specification that aligns stakeholder expectations with the project's technical capabilities. Additionally, this document addresses non-functional constraints such as interoperability, data traceability, trustworthiness, and cybersecurity.

2.3 Relation with other work packages

This deliverable is part of the work package 2 "Requirements definitions" of the SYNERGIES project. It incorporates requirements formalized in deliverable 2.1 "Stakeholder high-level requirements" and includes storylines definitions and technical and interoperability requirements defined in tasks 2.2 "Storylines definitions and technical requirements", 2.3 "Interoperability requirements" and 2.4 "Data traceability, trustworthiness and cyber-security requirements".

It will serve as a guideline for tools developed in work packages 3 "Data analysis, collection, and generation", 4 "Heterogeneous AI data processing tools and data interoperability", and 5 "Scenario generation methodology and process". It will also act as a specification for work package 7 "SYNERGIES platform establishment, testing and delivery".

3 KEY CONCEPTS AND TERMINOLOGY

3.1 Platform overview

The SYNERGIES Platform consists of a (i) Scenario Dataspace, aligned with the European approach on data sharing and competitiveness, and (ii) Marketplace [...].

Source: grant agreement [1]

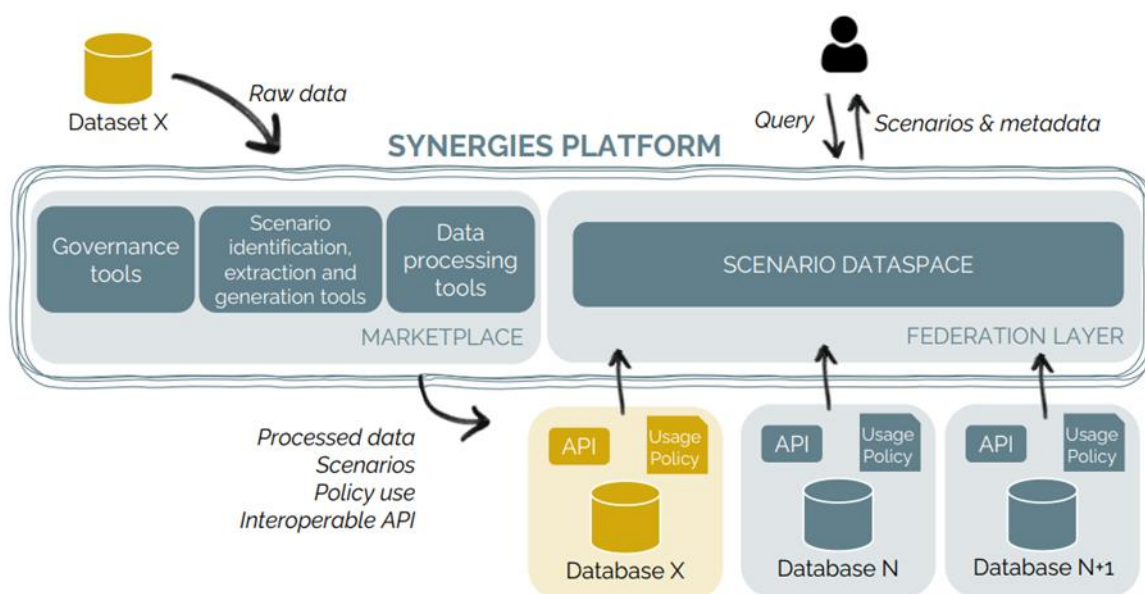


Figure 1: SYNERGIES preliminary high-level platform architecture

The requirements specified in this document are based on the assumptions outlined in this preliminary high-level architecture.

3.2 Scenarios

In the SYNERGIES project, the focus is on generating and providing scenarios from various data sources. A scenario, in this context, is understood as a structured representation formed by a specific set of pre-determined conditions and variables representing actual or theoretical states, situations, or interactions that may occur in the real world¹.

In the Autonomous Driving Systems (ADS) domain, a scenario involves placing the vehicle (ego) in situations characterised by specific events and actions within a defined environment, distinct from test or simulation parameters.

Moreover, scenarios may originate from a diverse set of sources, including synthetic data, theoretical frameworks, expert knowledge, real-world accident reports, driving data and standards set by organisations such as EURO NCAP.

¹ Cf. Scenario definition in the glossary of chapter 3.4.

While multiple representations of scenarios exist (keyframes/sequence of scenes; initial state with events and triggers; temporal data; text), the goal of SYNERGIES is to generate scenario files in standardized formats. The purpose of a scenario, which is representing states, situations, and interactions, remains unchanged despite differences in its structure.

Furthermore, in the SYNERGIES project, it is important for a scenario to encompass different levels of abstraction, regardless of the method employed in its creation. The literature provides insights into these abstraction levels along with detailed definitions.

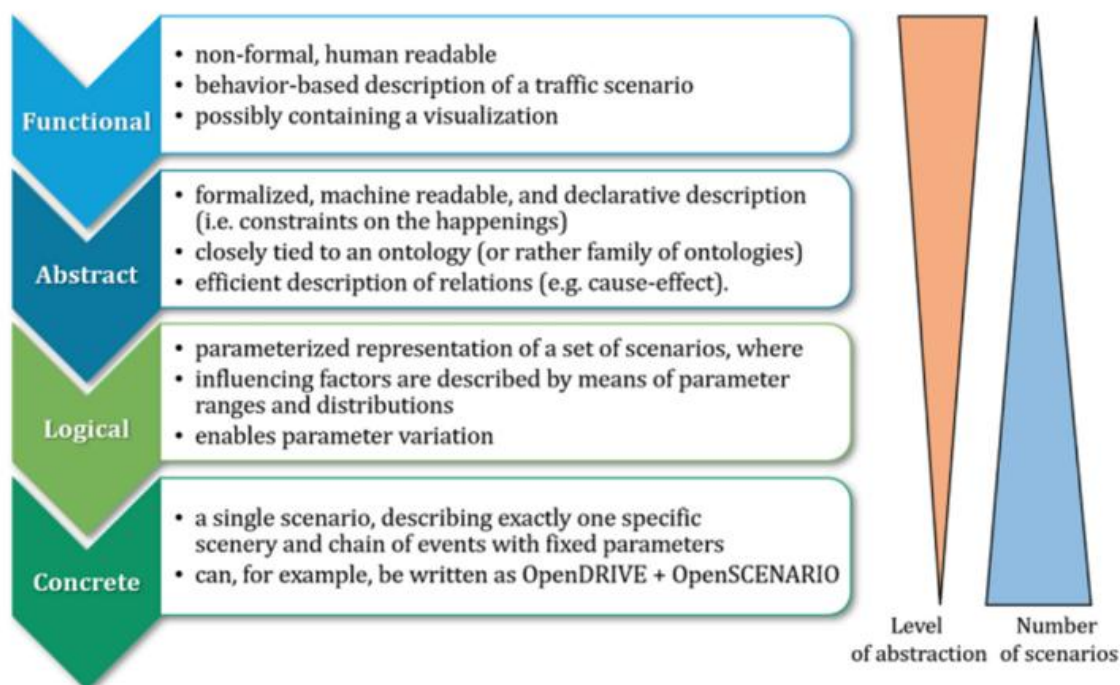


Figure 2: Relationship of functional scenario, abstract scenario, logical scenario, and concrete scenario (source: ISO 34501:2022 [2])

Therefore, to ensure clarity when discussing scenarios, it is crucial to introduce a prefix that specifies the type of scenario being referred to: functional, abstract, logical, and concrete scenario. SYNERGIES will utilize these various types of scenarios, with their definitions provided in the 3.4. Terms and definitions section.

3.3 Data

Raw data can come in various forms, such as video files captured by drones, datasets collected using instrumented vehicles or stationary sensors, or synthetic data containing timestamped packets, all with a binary payload. The figure below illustrates examples of the raw data that can be expected from different data sources.

**Drone observing traffic from above with a single camera**

Ex: Raw data is a single video file

**Instrumented vehicle**

Ex: Raw data is a folder containing:

- Per source-stream payload files
(e.g. video files, binary dumps of each sensor stream)
- Per source-stream index files (start/end offset of each packet/frame in payload)
- An index file containing lookup table: timestamp; source; packetID/frameID

**Synthetic raw data**

Ex: Raw data is a single file which contains timestamped packets, each containing binary payload (e.g. the binary content of a JPG file with its associated source and timestamp): timestamp; source; payload

Figure 3: Examples of raw data from different data sources

The nature of raw data depends on the sensors used for collection and the tools employed to timestamp and aggregate multiple sensor streams before flushing them to storage. This raw data, typically stored in non-standardized formats, lacks interoperability. To achieve interoperability, we must transform the raw data into usable information. This involves decoding payloads, synchronizing streams, fusing data, and performing tasks like object recognition and tracking. By converting the raw data to a standardized format, the processed information becomes algorithm-friendly (e.g. for data enrichment or scenario generation) and interoperable.

However, even after these processing steps, significant challenges remain. Differences in data collection constraints persist, alongside variations in intended use, semantics, modelling approaches for physical entities, and methods for structuring and serializing data. To address these challenges, **this project defines SSD (Scenario-Source Data)** as the input data used for scenario identification and extraction.

For seamless integration, data from heterogeneous sources must be utilized within a single tool-chain. To achieve this, we introduce the "S-CDF" (Scenario-Compatible Data Format) interface, where all diverse sources are converted into a unified format for interoperability.

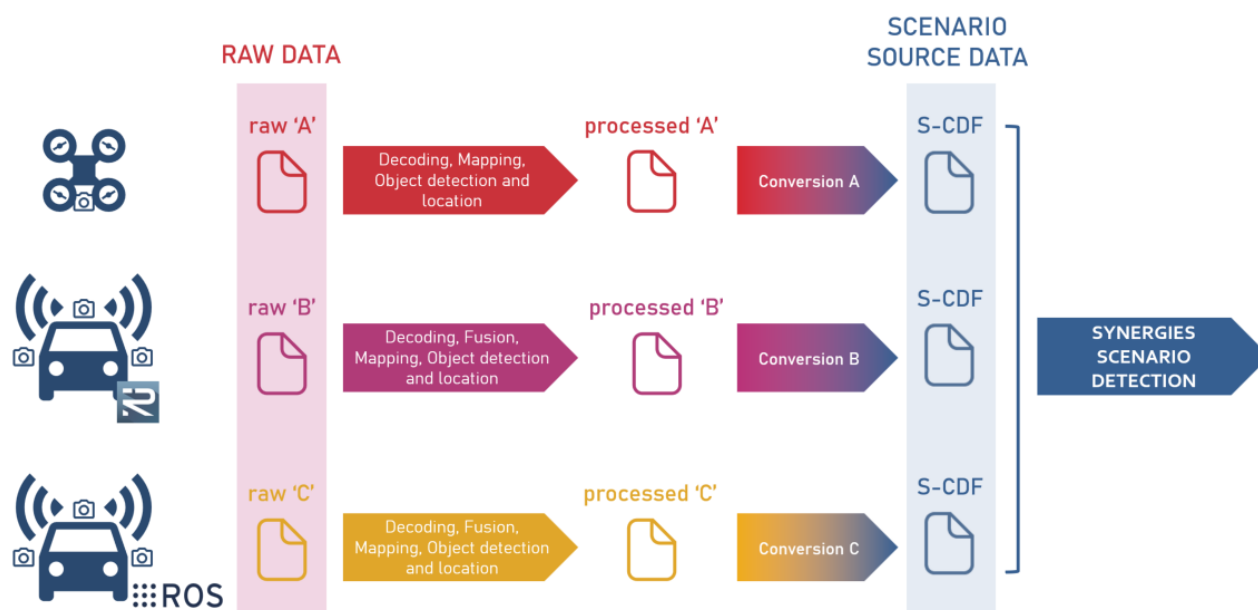


Figure 4: Conversion of various raw data examples into scenario source data

3.4 Terms and definitions

Term	Definition
Abstract scenario	<p>Formalized, declarative description derived from a functional scenario. The semantics of the description will be closely tied to an ontology, drastically increasing the precision of the employed terminology.</p> <p><i>Source: ISO 34501:2022 [2]</i></p>
Concrete scenario	<p>Scenario depicted with explicit parameters values, describing physical attributes. Parameter values can consist of default values, randomly chosen values or advisedly chosen values.</p> <p><i>Source: ISO 34501:2022 [2]</i></p>
Completeness	<p>"Completeness" is a measure of the extent of which something X captures the required details of something else Y. Note that completeness can only be defined for a certain purpose.</p>
Coverage	<p>"Coverage" is the extent that something X is addressed by something else Y.</p>
Data accuracy	<p>Degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.</p> <p><i>Source: ISO/IEC 25012, 2008a [3]</i></p>

Descriptor	A descriptor is a feature or attribute used to characterize an object or set of data. Descriptors are used to capture essential information about something in a way that can be used for further analysis, comparison or classification. They provide a representation of key properties.
Edge case	Within an edge case scenario one or more values achieve the system boundary. <i>Source: ISO 34505:2024, 6.3.3.2.3</i>
Functional scenario	Scenario described in natural language on a conceptional level, in general without specific physical values. <i>Source: ISO 34501:2022 [2]</i>
Logical scenario	Scenario described with the inclusion of parameters, including the set of allowable parameter values and, optionally, the (possibly joint) distribution of the parameters.
Operational design domain (ODD)	Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics. <i>Source: SAE J3016:2021 [4]</i>
Raw data	Raw data is information that has not been further processed after it has been collected (e.g., by sensors).
Representativeness	The "representativeness" of X with respect to Y expresses the extent to which the occurrences within X and the characteristics of these occurrences reflect the occurrences and their characteristics in Y. Note that the "relevant" characteristics depend on the purpose.
Scenario	A scenario is a structured depiction, specified by a set of pre-determined conditions and possibly variables, representing actual or theoretical states, situations, or interactions. In the context of vehicles and driving, this encompasses the developing view of both world-fixed (static) elements, such as road layout and road furniture, and world-changing (dynamic) elements, such as weather and lighting, vehicles, objects, people, and traffic light states. This description is independent of whether the environment is simulated, real, or any combination thereof.
Scenario catalogue	A collection of scenarios.
Scenario database	A data store in which scenarios are collected and characterized by a uniform set of parameters. It can serve as the source of a scenario catalogue.

Scenario metadata	Scenario metadata is information that is tagged to a scenario to enrich it with top-level properties that go beyond the content of the scenario itself.
Scenario relevance	Scenario relevance refers to the degree to which a given scenario is applicable and meaningful in the context of a specific system's intended operational domain, objectives, and evaluation criteria. A scenario is considered relevant if it reflects conditions, events, or behaviors that are likely to impact system performance, safety, or decision-making in real-world deployments.
Scenario source data (SSD)	In the context of SYNERGIES, "scenario source data" is raw data that has been processed and harmonized to the common file format (S-CDF) and modelled to be used for scenario extraction.
Test scenario	<p>Scenario intended for testing and assessment of ADS(s)/subject vehicle(s).</p> <p>Test scenarios may include additional items for the purpose of assessing the ADS performance or behaviour in addition to the scenario content.</p> <p>The additional items include, but are not limited to, sampling event, check for flagging SuT(s) error, relevant ODD data, success criteria, HMI event(s) that may trigger actions.</p> <p><i>Source: ISO 34501:2022 [2]</i></p>

Table 1: Terms and definitions

Note: [The taxonomy developed within the FAME project](#) [5] provides definitions for terms relevant in the CCAM domain.

4 STORYLINES

A "storyline" is a narrative that describes how a system works under its interaction with a user to achieve specific goals, expressing the system's functions based on user needs, data flows, interfaces, etc.

In task T2.2 of SYNERGIES, storylines have been designed to help identify User types, System components, and Data flows, which altogether help to map user needs into a reference architecture (which is to be developed in T7.1 of SYNERGIES).

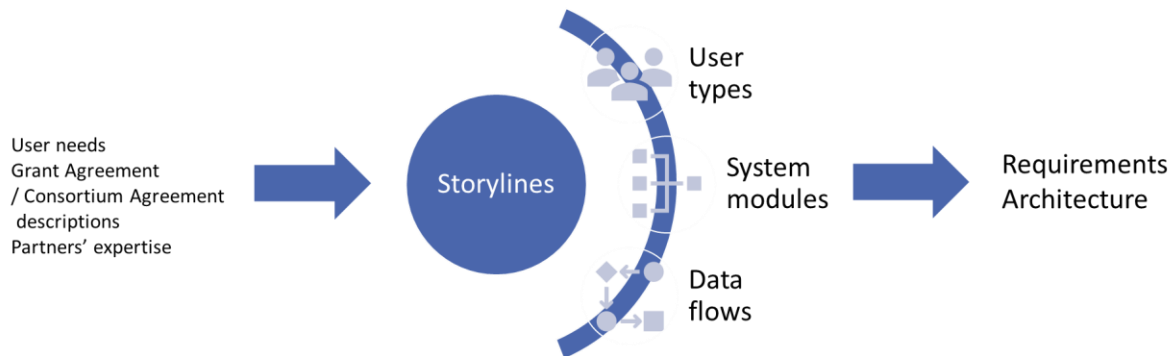


Figure 5: Storylines definition to map user needs into User types, System components and Data flows.

The format description of storylines can be simple natural language text, so storylines can be interpreted by human readers. However, for the sake of clarity and to avoid misunderstandings, the important terms used in these natural language descriptions have been carefully selected, debated and defined in the task T2.2. glossary, which is summarized as follows.

4.1 Storyline's glossary

The glossary includes a list of classes (specified by a term or word and an associated definition or definitions) needed to build the storylines as simple sentences, as unambiguously as possible, with linked concepts in semantic relations.

The following groups of classes are defined:

- **Top level classes:** the highest-level classes from which all other classes are derived.
- **Derived level classes:** basic elements that define the SYNERGIES ecosystem, defined with `subClassOf` relation with respect to parent classes.
- **Components of the SYNERGIES Platform:** special classes that define components or sub-systems (as `isPartOf` relations) of the SYNERGIES Platform.

The following groups of relations are defined:

- **Relations between User Role classes:** relations between different User Roles, understood as possible actions or activities that the users of SYNERGIES can perform.
- **Relations between other classes:** relations between users, components and other classes that build the fabric of the storylines of SYNERGIES.
- **Relations related to SUNRISE:** special relations that link to the SUNRISE project.

This set of classes and relations have been defined to provide a formal description that simplifies building storylines, and help reducing confusion and potential misunderstandings. It is not a formal ontology, although the type of relations used have been chosen to simplify the construction of an ontology. Such ontology may include definitions for relations like *hasAccessTo*, or other semantics that are not explicitly defined in this work.

The classes defined in the glossary in section 3.4 and at the Consortium Agreement (CA) document prevail. The following table inspects which classes are relevant to the storylines where defined:

Term	Definition & source
Scenario	<i>See definition in section 3.4.</i>
Scenario Database	<i>See definition in section 3.4.</i>
Raw Data	<i>See definition in section 3.4.</i>
Scenario Source Data	<i>See definition in section 3.4.</i>
Federation Layer	<p>The software structure, that enables the use of multiple scenario databases from a single point of access. Further details about the Federation Layer are given in the Grant Agreement. It is expected that this software structure is developed under the SUNRISE Project.</p> <p><i>Source: Consortium Agreement [6]</i></p>
SYNERGIES Database	<p>A comprehensive, structured and openly accessible for all Parties compilation of driving and environmental scenarios specifically designed for use in the testing, research and/or homologation of CCAM systems, generated during the Project execution and possibly updated afterwards. The SYNERGIES Database may include scenarios generated by the Parties outside of the Project. The SYNERGIES Database is a specific implementation within the broader SYNERGIES Scenario Dataspace. Further details about the SYNERGIES Database are given in Grant Agreement.</p> <p><i>Source: Consortium Agreement [6]</i></p>
Background Database	<p>A Scenario Database (SCDB) external to the SYNERGIES Project and that exists outside of the SYNERGIES Platform. It may be reachable via the SYNERGIES Scenario Dataspace.</p> <p><i>NOTE: Equivalent to "Background SCDB" and "Other SCDB".</i></p>
SYNERGIES Marketplace	<p>An openly accessible for all Parties platform part of SYNERGIES Platform that provides existing ("Background Tools") and newly developed cutting-edge tools (e.g. "SYNERGIES Tools") for data processing, scenario generation, and database governance, fostering seamless scalability and enabling the integration of new initiatives.</p>

<i>Source: Consortium Agreement [6]</i>	
SYNERGIES Platform	<p>The European platform designed to facilitate the development, training, virtual testing, and validation of CCAM systems. The SYNERGIES Platform consists of a (i) SYNERGIES Scenario Dataspace, aligned with the European approach on data sharing and competitiveness, and (ii) SYNERGIES Marketplace, to ensure continuous updates and scalability of the SYNERGIES Scenario Dataspace.</p> <p><i>Source: Consortium Agreement [6]</i></p>
SYNERGIES Scenario Dataspace	<p>A federated and interoperable repository that integrates existing ("Background Database") and newly developed scenario databases (e.g. "SYNERGIES Database"). This SYNERGIES Scenario Dataspace enables access to a wide array of scenarios across different regions, countries, and conditions, supporting the evaluation and development of CCAM systems.</p> <p><i>Source: Consortium Agreement [6]</i></p>
SYNERGIES Tools	<p>The software and methodologies generated in the Project and provided through the SYNERGIES Marketplace to support data processing, scenario identification, extraction, generation, and governance.</p> <p><i>Source: Consortium Agreement [6]</i></p> <p><i>NOTE: Equivalent to the term "Tools" defined in this section.</i></p>

Table 2: Terms and definitions already defined in other documents.

4.1.1 Classes

Top level classes are those classes that do not have any parent in the taxonomy and serve as a basis for the hierarchical definition of derived classes.

Class	Definition
Asset	An Asset is a Thing that is managed by SYNERGIES Platform as item of work.
User Role	A User Role defines the type of Roles a User can have to work with the SYNERGIES Platform.
SYNERGIES Platform	<i>See definition in Table 2.</i>
Scenario Database	<i>See definition in section 3.4.</i>

Table 3: Top-level classes.

Derived classes are those that inherit from either top-level classes or other derived classes. Child classes inherit the properties and behaviours of their parent classes. Derived classes are linked to their parent classes with a subClassOf relationship.

Class	Definition
Tool	Tool subClassOf Asset. <i>See CA's definition in Table 2.</i>
Data	Data subClassOf Asset. <ul style="list-style-type: none"> - A Data is an Asset that represents facts, figures, or information that can be stored in digital form that can be transmitted or processed.
Scenario	Scenario subClassOf Asset. <i>See definition in section 3.4.</i>
Raw Data	Raw Data subClassOf Data. <i>See definition in section 3.4.</i>
Scenario Source Data	Scenario Source Data subClassOf Data. <i>See definition in section 3.4.</i>
Administrator	Administrator subClassOf User Role. <ul style="list-style-type: none"> - An Administrator is a type of User Role that manages the SYNERGIES Platform.
Consumer	Consumer subClassOf User Role. <ul style="list-style-type: none"> - A Consumer is a type of User Role that uses Assets from SYNERGIES Platform.
Provider	Provider subClassOf User Role.. <ul style="list-style-type: none"> - A Provider is a type of User Role that provides Assets to SYNERGIES Platform.
Tool User	Tool User subClassOf Consumer. <ul style="list-style-type: none"> - A Tool User is a type of Consumer that accesses SYNERGIES Platform to find Tools and use them.
Scenario User	Scenario User subClassOf Consumer. <ul style="list-style-type: none"> - A Scenario User is a type of Consumer that accesses SYNERGIES Platform to find Scenarios and use them.
Data Provider	Data Provider subClassOf Provider. <ul style="list-style-type: none"> - A Data Provider is a type of Provider that accesses SYNERGIES Platform to register Data into the Data Registry.
Tool Provider	Tool Provider subClassOf Provider.

	<ul style="list-style-type: none"> - A Tool Provider is a type of Provider that accesses SYNERGIES Platform to register Tools into the SYNERGIES Marketplace.
Scenario Provider	<p>Scenario Provider subClassOf Provider.</p> <ul style="list-style-type: none"> - A Scenario Provider is a type of Provider that accesses SYNERGIES Platform to register Scenarios into Scenario Databases.
Scenario Database Provider	<p>Scenario Database Provider subClassOf Provider.</p> <ul style="list-style-type: none"> - A Scenario Database Provider is a type of Provider that accesses SYNERGIES Platform to register a Scenario Database into the Scenario Dataspace.

Table 4: Derived classes.

Components of the SYNERGIES Platform are specified with a relation of type `isPartOf`:

Relation	Comment
SYNERGIES Marketplace isPartOf SYNERGIES Platform	<i>See CA's definition in Table 2.</i>
Data Registry isPartOf SYNERGIES Platform	The Data Registry is a component of the SYNERGIES Platform that contains a registry of Data, in the form of reference metadata associated to each Data (i.e., the Data Registry does not contain the Data, but just administrative references and other descriptive metadata). The reference metadata might be a URI that enables a download or other forms of access mechanisms (e.g., Data Space connector or other end points subject to authentication or license/contract definition).
Scenario Dataspace isPartOf SYNERGIES Platform	<i>See CA's definition in Table 2.</i>
Knowledge Base isPartOf SYNERGIES Platform	The Knowledge Base is a component of the SYNERGIES Platform that contains information and documentation about the SYNERGIES Platform and its components and Assets
SYNERGIES Database isPartOf SYNERGIES Platform	<i>See CA's definition in Table 2.</i>

Table 5: Components of the SYNERGIES Platform.

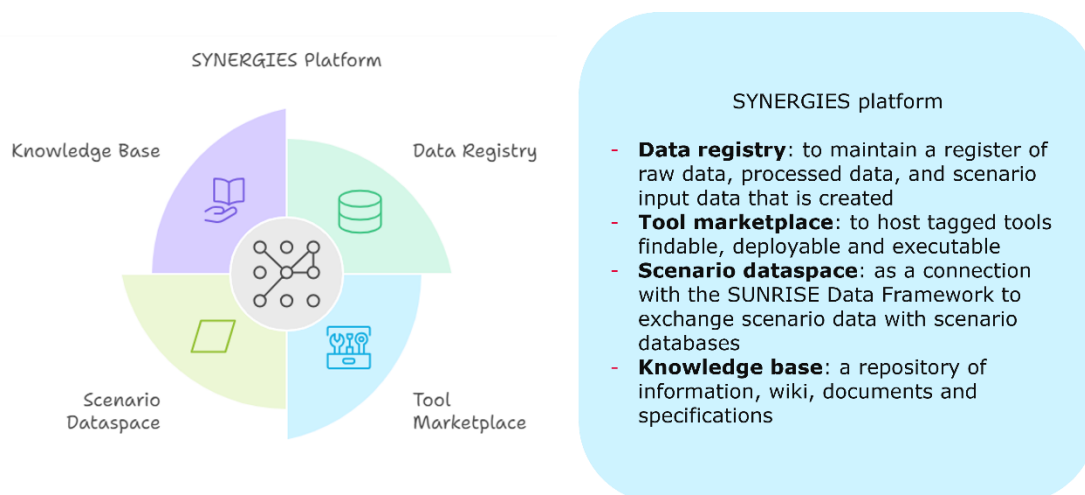


Figure 6: Essential SYNERGIES Platform components.

NOTE: Knowledge Base responds to D2.1's high-level requirements on "documentation and feedback" requirements.

4.1.2 Relations

Relations are defined as interdependencies between classes, describing specific actions or activities that certain users or components may perform on components of the SYNERGIES Platform. Relations constitute the fabric of storylines.

Relation	Description
User Role hasAccessTo SYNERGIES Marketplace	User Role can log in into the SYNERGIES Platform and get access to the component SYNERGIES Marketplace.
User Role searchesToolsAt SYNERGIES Marketplace	User Role can perform searches on the catalogue of Tools of the SYNERGIES Marketplace to find the pointers to Tools of interest.
User Role hasAccessTo Data Registry	User Role can log in into the SYNERGIES Platform and get access to the component Data Registry.
User Role searchesDataAt Data Registry	User Role can perform searches on the catalogue of Data of the Data Registry to find the pointers to Data of interest.
Tool Provider provides Tool	Tool Provider generates Tool and provides it to the SYNERGIES Marketplace.
Tool User uses Tool	Tool User uses Tool to process Data. <i>NOTE: using may imply downloading, executing, deploying or other forms of utilisation.</i>
Scenario Provider provides Scenario	Scenario Provider generates Scenario by using Tools on Data, and provides it to Scenario Databases via the Scenario Dataspace.

Scenario Provider getAccessTo Scenario Dataspace	Scenario Provider gets access to the Scenario Dataspace to perform actions on Scenario Databases.
Scenario Provider registersScenarioInto Scenario Database	Scenario Provider registers Scenario into Scenario Databases via the Scenario Dataspace. <i>NOTE: potentially multiple Scenarios uploaded to multiple Scenario Databases.</i>
Scenario User searchesScenarioAt Scenario Dataspace	Scenario User searches scenarios at the Scenario Dataspace (using harmonized queries).
Scenario User uses Scenario	Scenario User uses Scenario for developing or testing CCAM functions or other Verification & Validation activities. <i>NOTE: "uses" means "downloads and uses", as there is no way Scenarios can be used on the platform.</i>
Data Provider provides Data	Data Provider provides Data to the SYNERGIES Platform by recording or generating Data.
Data Provider registersDataInto Data Registry	Data Provider provides Data to the SYNERGIES Platform via registering the Data metadata into the Data Registry.
Scenario Database Provider re-gistersSCDBInto Scenario Dataspace	Scenario Database Provider registers a Scenario Database (SCDB) into the Scenario Dataspace.

Table 6: Relations between User Roles classes (i.e., user actions).

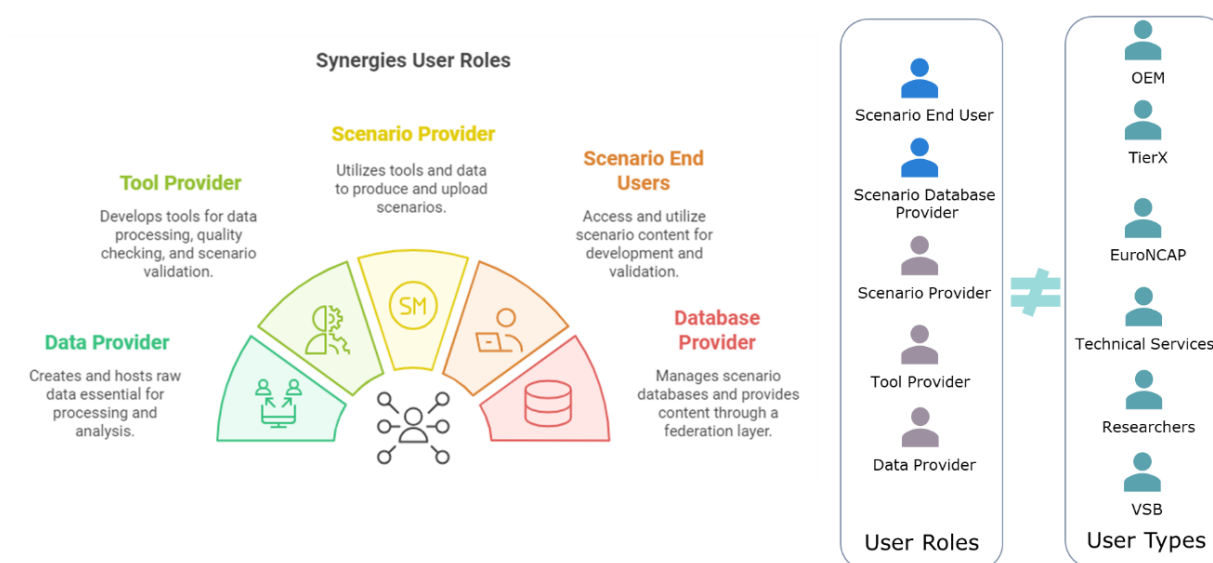


Figure 7: User Roles within the SYNERGIES storylines. Source: D2.1's user types and roles.

Apart from user actions, dataflows within the SYNERGIES Platform are also defined by means of the relations between certain components.

Relation	Description
Tool processes Data	Tools can process Data with a particular purpose, using input Data and producing output Data.
SYNERGIES Marketplace containsPointersTo Tool	The SYNERGIES Marketplace contains pointers to Tools, in the form of a catalogue with descriptive information of Tools, so Tools are findable, and usable (e.g., via downloading, or execution of services).
Data Registry containsPointer-sTo Data	The Data Registry contains pointers to Data, in the form of a catalogue with descriptive information of Data, so Data are findable, and usable (e.g., via downloading, streaming, or getting in contact with the Data Provider).
Scenario Database contains Scenario	Scenario Database contains Scenarios, so the Scenarios are findable and usable by User Roles.
Scenario Dataspace providesAccessTo Scenario Database	Scenario Dataspace manages user access to Scenario Databases, and manages queries to them.

Table 7: Relations between other classes.

There are some specific relations that refer to the components existing in the SUNRISE Data Framework, which shall be considered in SYNERGIES as an instance or implementation of the Scenario Dataspace concept.

Relation	Description
SUNRISE Data Framework isA Scenario Dataspace	The SUNRISE Data Framework is a platform that works as a Scenario Dataspace (i.e., it provides access to Scenario Databases).
SYNERGIES Database isA Scenario Database	SYNERGIES Database is a Scenario Database accessible via the SYNERGIES Dataspace.
ADSCENE isA Scenario Database	ADSCENE is a Scenario Database accessible via the SUNRISE Data Framework.
SafetyPool isA Scenario Database	SafetyPool is a Scenario Database accessible via the SUNRISE Data Framework.
Scenario.Center isA Scenario Database	Scenario.Center is a Scenario Database accessible via the SUNRISE Data Framework.
Scenius isA Scenario Database	Scenius is a Scenario Database accessible via the SUNRISE Data Framework.

4.2 Storyline's definition

With the defined classes and relations, it is possible to describe the storylines as single elements with as less ambiguity as possible.

ID	SL-1
Storyline description	A Data Provider creates and prepares Data.
RDF from Glossary	Data Provider provides Data.
Extensions / details of the process	A Data Provider creates (e.g., records or generates) Data, and prepares the Data in the specified standardized format defined in the Knowledge Base. Tools from the SYNERGIES Marketplace might be needed.
Storyline category	Data creation and registration.
Priority	Core*.
Involved Users	Data Provider.
Involved components	Local computing; SYNERGIES Marketplace; Knowledge Base.

NOTE: Priority levels are *Core* – for those storylines of higher priority for the SYNERGIES project, which are directly related to objectives or main functionalities of the platform. *User/System* – for those storylines of lower priority, not strictly related to functionalities but rather to technical details of the management of the platform being a software platform.

ID	SL-2
Storyline description	A Data Provider accesses the Data Registry and searches for available Data.
RDF from Glossary	User Role hasAccessTo Data Registry; User Role searches-DataAt Data Registry.
Extensions / details of the process	A Data Provider gets access to the Data Registry and explores and searches for existing or available Data in order to identify gaps or missing Data that the Data Provider might provide or update.
Storyline category	Data creation and registration.
Priority	Core.
Involved Users	Data Provider.
Involved components	Data registry.

ID	SL-3
Storyline description	A Data Provider registers Data at the Data Registry.
RDF from Glossary	Data Provider provides Data.
Extensions / details of the process	A Data Provider prepares Data (e.g., Raw Data, or Scenario Source Data), and adds a registry entry into the Data Registry, so that the Data is enlisted and findable by other users. The process requires the approval of the SYNERGIES Administrator.
Storyline category	Data creation and registration.
Priority	Core.
Involved Users	Data Provider; Administrator.
Involved components	Data registry.

ID	SL-4
Storyline description	A Tool Provider prepares (develop or adapt) and provides a Tool to the SYNERGIES Marketplace.
RDF from Glossary	User Role hasAccessTo SYNERGIES Marketplace; Tool Provider provides Tool.
Extensions / details of the process	Depending on the nature of the Tool (e.g., open code project, private application, cloud service), the Tool can be tested in different ways to check compliance with SYNERGIES defined standards and interfaces. After approval from the Administrator the Tool is registered into the SYNERGIES Marketplace.
Storyline category	Tool provision and marketplace.
Priority	Core.
Involved Users	Tool Provider; Administrator.
Involved components	SYNERGIES Marketplace.

ID	SL-5
Storyline description	A Tool User gets access and searches Tools at the SYNERGIES Marketplace.

RDF from Glossary	User Role hasAccessTo SYNERGIES Marketplace; User Role searchesToolAt SYNERGIES Marketplace.
Extensions / details of the process	A Tool User gets access to the SYNERGIES Marketplace and explores and searches for specific Tools using search capabilities, filtering or keywords. Information about the Tool is presented to the Tool User, such as technical documentation, and also terms of use, costs, etc.
Storyline category	Tool provision and marketplace.
Priority	Core.
Involved Users	Tool User.
Involved components	SYNERGIES Marketplace.

ID	SL-6
Storyline description	A Tool User uses a Tool from the Marketplace.
RDF from Glossary	Tool User uses Tool.
Extensions / details of the process	Depending on the type of Tool User and Tool (e.g., open code project, private application, cloud service), the Tool has to be downloaded, installed, or accessed in different ways. The Tool User then uses the Tool to process Data in different possible ways (e.g., Data might be Raw Data, or Scenario Source Data).
Storyline category	Tool provision and marketplace.
Priority	Core.
Involved Users	Tool User.
Involved components	SYNERGIES Marketplace.

ID	SL-7
Storyline description	A Scenario Database Provider registers a Scenario Database into the Scenario Dataspace.
RDF from Glossary	Scenario Database Provider registers SCDB Into Scenario Dataspace.
Extensions / details of the process	A Scenario Database Provider manages a Scenario Database (SCDB) and joins the Scenario Dataspace by following the onboarding process of the Scenario Dataspace, which includes

	approval from the SYNERGIES Administrator, and provision of an interface between the SCDB and the Scenario Dataspace.
Storyline category	Scenario databases.
Priority	Core.
Involved Users	Scenario Database Provider; Administrator.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-8
Storyline description	A Scenario Provider provides Scenario(s).
RDF from Glossary	Scenario Provider provides Scenario.
Extensions / details of the process	A Scenario Provider provides new Scenario files. Note that the scenarios might have been created (from knowledge), extracted (generated from data processing), acquired or other forms of provision by the Scenario Provider.
Storyline category	Scenario creation and upload.
Priority	Core.
Involved Users	Scenario Provider.
Involved components	Local computing.

ID	SL-9
Storyline description	A Scenario Provider registers Scenario(s) into the Scenario Dataspace.
RDF from Glossary	Scenario Provider getAccessTo Scenario Dataspace; Scenario Provider registersScenarioInto Scenario Database.
Extensions / details of the process	<p>A Scenario Provider gets access to the SYNERGIES Scenario Dataspace, where access to different Scenario Databases is possible. Then the Scenario Provider selects which Scenario Database to use and submits or uploads the Scenario(s).</p> <p><i>NOTE: This functionality will be available for the SYNERGIES Database. However, for other Background or Scenario Databases, write access from the SYNERGIES platform is not guaranteed. This means that the ability to write to or upload scenarios into these databases may not be supported. Whether this</i></p>

	<i>functionality is available depends on the governance policies of each individual Scenario Database.</i>
Storyline category	Scenario creation and upload.
Priority	Core.
Involved Users	Scenario Provider.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-10
Storyline description	A Scenario Provider updates existing Scenario (e.g., adjusting to new regulation, or new/different performance metric, or new/different metadata).
RDF from Glossary	Scenario Provider getAccessTo Scenario Dataspace; Scenario Provider registersScenarioInto Scenario Database.
Extensions / details of the process	<p>A Scenario Provider gets access to the SYNERGIES Scenario Dataspace, which provides access to scenario databases like the SYNERGIES Database (e.g., to upload Scenarios and the Metadata that contains the origin, creation or extraction process). The Scenario Provider requests the update of existing Scenarios to the SYNERGIES Database.</p> <p><i>NOTE: This functionality will be available for the SYNERGIES Database. However, for other Background or Scenario Databases, write access from the SYNERGIES platform is not guaranteed. This means that the ability to write to or upload scenarios into these databases may not be supported. Whether this functionality is available depends on the governance policies of each individual Scenario Database.</i></p>
Storyline category	Scenario creation and upload.
Priority	Core.
Involved Users	Scenario Provider.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-11
Storyline description	A Scenario Database Provider accepts or rejects update of one of its Scenarios.

RDF from Glossary	Scenario Provider getAccessTo Scenario Dataspace; Scenario Provider registersScenarioInto Scenario Database.
Extensions / details of the process	<p>The Scenario Database Provider receives and examines requests to update Scenarios at the Scenario Database. This process might be automated by granting permissions to Scenario Providers.</p> <p><i>NOTE: This functionality will be available for the SYNERGIES Database. However, for other Background or Scenario Databases, write access from the SYNERGIES platform is not guaranteed. This means that the ability to write to or upload scenarios into these databases may not be supported. Whether this functionality is available depends on the governance policies of each individual Scenario Database.</i></p>
Storyline category	Scenario creation and upload.
Priority	Core.
Involved Users	Scenario Database Provider.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-12
Storyline description	A User requests access to a Scenario Database.
RDF from Glossary	N/A.
Extensions / details of the process	A Scenario User or a Scenario Provider requests access to a specific Scenario Database, to get credentials to search for or upload Scenarios. The Scenario Database Provider accepts or rejects the request.
Storyline category	Scenario Database querying.
Priority	Core.
Involved Users	Scenario User; Scenario Provider; Scenario Database Provider.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-13
Storyline description	A User queries a Scenario Database to search for Scenarios.
RDF from Glossary	Scenario User searchesScenarioAt Scenario Dataspace.

Extensions / details of the process	A Scenario User or a Scenario Provider uses the Scenario Dataspace to create a query to find Scenarios at the Scenario Databases for which this user has credentials.
Storyline category	Scenario Database querying.
Priority	Core.
Involved Users	Scenario User; Scenario Provider.
Involved components	Scenario Dataspace; Scenario Database.

ID	SL-14
Storyline description	A User registers in and gets access to the SYNERGIES Platform.
RDF from Glossary	User Role hasAccessTo SYNERGIES Marketplace; User Role hasAccessTo Data Registry; Scenario Provider getAccessTo Scenario Dataspace.
Extensions / details of the process	User registers into SYNERGIES Platform to get access to the different components and services.
Storyline category	User registration and access.
Priority	User/System.
Involved Users	All Users.
Involved components	SYNERGIES Platform.

ID	SL-15
Storyline description	The Administrator reviews and approves requests of Tool Providers to register Tools into the SYNERGIES Marketplace.
RDF from Glossary	Administrator manages SYNERGIES Platform.
Extensions / details of the process	The Administrator accesses the management application at the SYNERGIES Platform to approve requests to register Tools to the SYNERGIES Marketplace.
Storyline category	Management.
Priority	User/System.
Involved Users	Administrator.

Involved components	SYNERGIES Platform.
----------------------------	---------------------

ID	SL-16
Storyline description	The Administrator reviews and approves requests of Data Providers to register Data into the Data Registry.
RDF from Glossary	Administrator manages SYNERGIES Platform.
Extensions / details of the process	The Administrator accesses the management application at the SYNERGIES Platform to approve requests to register Data into the Data Registry.
Storyline category	Management.
Priority	User/System.
Involved Users	Administrator.
Involved components	SYNERGIES Platform..

ID	SL-17
Storyline description	A User sends requirement or feedback to a Data Provider.
RDF from Glossary	User Role hasAccessTo Data Registry.
Extensions / details of the process	A User reviews or analyzes Data from the Data Registry, and requests a requirement or provides feedback to the Data Provider.
Storyline category	Data creation and registration.
Priority	User/System.
Involved Users	All Users, Data Provider.
Involved components	Data Registry.

ID	SL-18
Storyline description	A User explores, searches and accesses shared taxonomy and concepts documentation at the Knowledge Base.
RDF from Glossary	Knowledge Base isPartOf SYNERGIES Platform.

Extensions / details of the process	A User gets access to the Knowledge base to search for shared taxonomies and concepts documentation.
Storyline category	Semantic and technical interoperability.
Priority	User/System.
Involved Users	All Users.
Involved components	Knowledge Base.

ID	SL-19
Storyline description	A User explores scenario descriptors and related ontologies.
RDF from Glossary	Knowledge Base isPartOf SYNERGIES Platform.
Extensions / details of the process	A User gets access to the Knowledge base to search for scenario descriptors and related ontologies.
Storyline category	Semantic and technical interoperability.
Priority	User/System.
Involved Users	All Users.
Involved components	Knowledge Base.

ID	SL-20
Storyline description	A Tool Provider accesses the API and related documentation.
RDF from Glossary	Knowledge Base isPartOf SYNERGIES Platform.
Extensions / details of the process	A Tool Provider gets access to the Knowledge base to search for documentation about the APIs of SYNERGIES Platform components.
Storyline category	Semantic and technical interoperability.
Priority	User/System.
Involved Users	Tool Provider.
Involved components	Knowledge Base.

ID	SL-21
Storyline description	A User looks up to the history of updates made in a Scenario Database.
RDF from Glossary	Scenario User searchesScenarioAt Scenario Dataspace.
Extensions / details of the process	<p>A Scenario User or a Scenario Provider use the Scenario Dataspace to create a query to find updates made at the Scenario Databases for which this user has credentials.</p> <p><i>NOTE: this functionality might exist or not depending on the Scenario Database.</i></p>
Storyline category	Semantic and technical interoperability.
Priority	User/System.
Involved Users	Scenarios Users, Tool Provider.
Involved components	Scenario Dataspace; Scenario Database.

4.3 High-level reference architecture

From the Storylines, and from the definition of the different SYNERGIES Platform components and User Roles, it is possible to draft a high-level reference architecture of the SYNERGIES Platform.

This section contains a draft of such architecture, aiming to serve as a basis for the definition work in other tasks and work packages (e.g., T7.1).

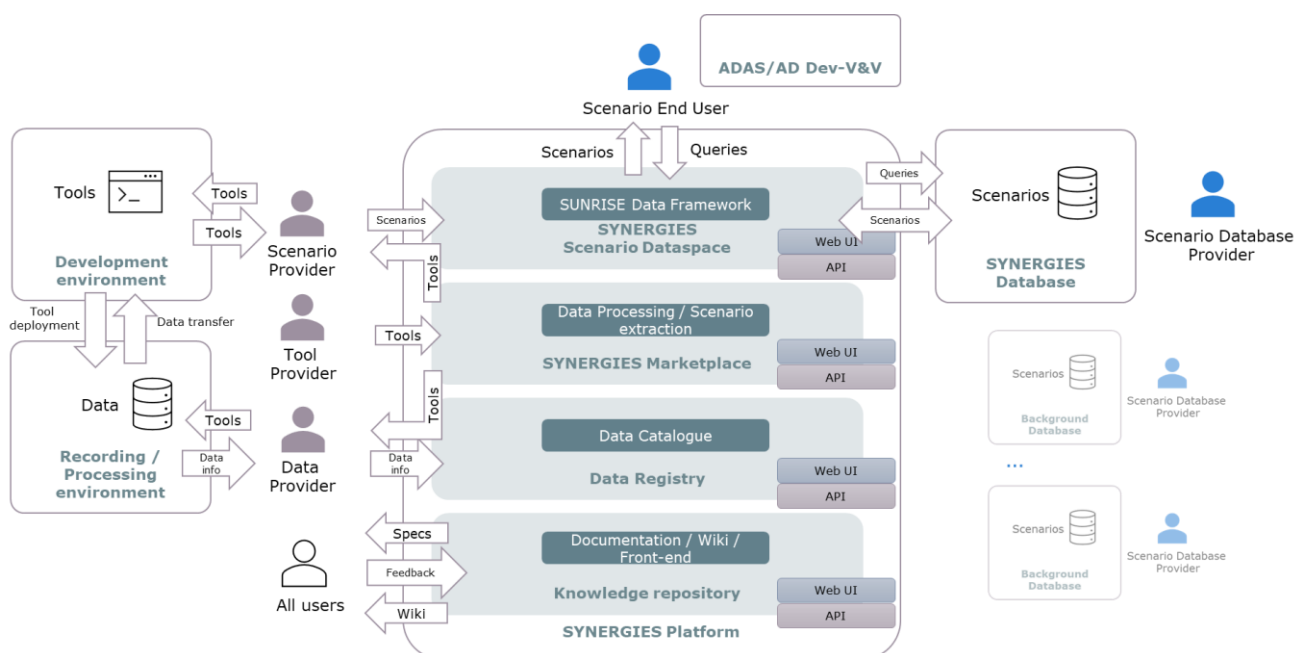


Figure 9: High-level reference architecture of the SYNERGIES Platform derived from the storylines.

5 CONSTRAINTS

5.1 Interoperability

Interoperability is the ability of different systems to work together and exchange information flawlessly and efficiently, enabling them to function as a unit despite their differences in technology and protocols.

In this document, interoperability requirements are organized into four main categories:

- **Semantic interoperability** gathers requirements referring to the ability to accurately interpret exchanged data and ensuring that the meaning of the data is preserved, regardless of the differences in data structure or format.
- **Pragmatic interoperability** gathers requirements ensuring that the data exchanged is interpreted and used in a manner consistent with the sender's intentions and the receiver's practical context.
- **Technical interoperability** gathers requirements referring to the ability of different systems to communicate and exchange data seamlessly by ensuring compatibility between software interfaces, and communication protocols, regardless of technological differences.
- **Data interoperability** focuses on requirements about the data exchanged. All transversal requirements related to data management must be met by all stakeholders: ensuring traceability, end-to-end security, and enabling configuration management across the entire data lifecycle.

5.1.1 Semantic interoperability requirements

5.1.1.1 Shared vocabulary for high-level platform concepts and technical terms

Rationale: A unified and shared vocabulary of high-level concepts across the SYNERGIES platform ensures consistent understanding of the purpose of the assets and their intended use. Furthermore, technical terms can be misinterpreted or lead to confusion, especially when dealing with complex, multidisciplinary systems. By establishing a common vocabulary, all stakeholders can interact with the platform efficiently, reducing ambiguities and improving collaboration.

State of the art: ISO 34501, GRVA 18-50, ISO 21448.

ID	Requirement
Req-Int-1-1-1	A glossary containing key concepts and technical terms should be implemented.
Req-Int-1-1-2	The SYNERGIES platform taxonomy should define terms related to high-level concepts such as "scenarios", "scenario catalogues", "scenario database", "federation layer" or "dataset" to have a common understanding of

	the elements being developed in the project and available on the platform at the end.
Req-Int-1-1-3	The glossary should define terms including scenario descriptor, data quality, accuracy, completeness.

Table 9: Shared vocabulary for high-level platform concepts and technical terms requirements.

5.1.1.2 Shared scenario descriptors

Rationale: Standardized scenario descriptors ensure a common understanding of the scenarios. These descriptors will serve as a shared language and ontology to specify scenarios, enabling consistent interpretation and use of scenarios across various stakeholders.

State of the art: ISO 34501, GRVA 18-50, ISO 21448, VVM methods: A.U.T.O. [7]

ID	Requirement
Req-Int-1-2-1	Standard scenarios descriptors should be specified.
Req-Int-1-2-2	The scenario descriptors should include essential elements such as actors, environment conditions, infrastructure, manoeuvres and actions, and share the semantic with descriptors used for Scenario Source Data.
Req-Int-1-2-3	Scenario descriptors should be designed for human-readable and machine-interpretable formats to enable their use across different use-cases.

Table 10: Shared scenario descriptors requirements.

5.1.2 Pragmatic interoperability requirements

5.1.2.1 Explicit assets scope

Rationale: It is critical to provide information that allow users to determine the intended use and limitations of assets (e.g., scenarios, datasets, models) to ensure that they are used efficiently and appropriately. Misuse or misinterpretation of assets could lead to inaccurate results or improper validation process. Clear asset's purpose, scope, and constraints will enable users to select and apply assets correctly within their respective processes.

State of the art: Metadata, Data coverage, Data traceability.

ID	Requirement
Req-Int-2-1-1	The description of assets should include a rationale for why it has been included.

Req-Int-2-1-2	The ontology or taxonomy should link particular traffic rules to relevant scenarios. This can be achieved through additions to the description and by the use of appropriate identifiers.
Req-Int-2-1-3	The user should be informed on the coverage of a set of scenarios for a scenario space defined in the project.

Table 11: Explicit assets scope requirements.

5.1.2.2 Availability of metadata

Rationale: Providing access to metadata describing the content of assets (such as raw data or scenarios) allows users to quickly and efficiently identify, categorize, and select relevant assets without needing to process or analyse the internal structure or data. This improves usability by saving time, reducing the computational overhead, and allowing for a more streamlined workflow in assets selection and management. It also supports better assets organization and retrieval, especially in environments where multiple sources contribute to a large scenario library.

State of the art: ASAM OpenLABEL, ASAM OpenODD, ASAM Open Simulation Interface, Ontology proposal from GaiaX project. [8]

ID	Requirement
Req-Int-2-2-1	Tags should be available to allow users to quickly filter and identify relevant scenarios based on specific attributes such as driving conditions, vehicle behaviours, environmental factors, or testing goals.
Req-Int-2-2-2	Tags should be standardized and shared tags should have a documented frame of reference.
Req-Int-2-2-3	Every scenario should have its abstraction level as a tag (functional, abstract, logical, concrete).

Table 12: Availability of metadata requirements.

5.1.3 Technical interoperability requirements

5.1.3.1 Resources unique identification

Rationale: Providing each asset with a unique identifier is essential for ensuring that users can reference them unambiguously. This is particularly important in a federated system where assets may come from various sources and need to be consistently referenced across platforms, tools, and workflows. A unique identifier allows for precise scenario retrieval, tracking, and management, minimizing the risk of confusion or errors when multiple similar scenarios exist.

State of the art: GUID.

ID	Requirement
Req-Int-3-1-1	This unique identifier must be immutable and persist across different versions or updates of the asset (e.g., scenarios).
Req-Int-3-1-2	Users must be able to reference and retrieve assets (e.g., scenarios) using its unique identifier via both the user interface and API.

Table 13: Resources unique identification requirements.

5.1.3.2 Availability of interfaces documentations within the platform

Rationale: Providing comprehensive and up-to-date documentation for all interfaces (APIs) ensures that developers and integrators can effectively utilize the platform's functionalities with third party's systems.

It also helps in onboarding new users, improving the overall user experience, and reducing the dependency on direct support.

State of the art: OpenAPI.

ID	Requirement
Req-Int-3-2-1	The documentation should be versioned.
Req-Int-3-2-2	The primary documentation language should be English.
Req-Int-3-2-3	The documentation should include working examples of the usage.

Table 14: Availability of interfaces documentations within the platform requirements.

5.1.3.3 Version management of assets

Rationale: Version management is a key to ensure traceability, consistency, and accountability of data. Without proper versioning, users may unknowingly work with outdated or incompatible assets, leading to discrepancies in results, inefficient workflows, and potential validation errors. By implementing a robust version management system, the platform can ensure that all users are accessing the correct versions of assets and can track changes over time.

State of the art: Semantic versioning, Traceability, Git.

ID	Requirement
Req-Int-3-3-1	Assets (including scenarios or tools from the marketplace) should be identified with a version number to allow identification of the scenario version that was used during the ADS assessment.
Req-Int-3-3-2	Older versions of assets (including scenarios or tools from the marketplace) should be available to allow them to be checked or used for testing.

Req-Int-3-3-3	A changelog should be maintained that provides a history of the assets and details of when and why changes were made.
----------------------	---

Table 15: Version management of assets requirements.

5.1.3.4 Availability of an Application Programming Interface (API)

Rationale: Providing API access ensures that users can seamlessly integrate the platform's resources with their own tools, enhancing flexibility, interoperability, and automation. This capability allows users to embed platform functionalities into their own workflows. By offering API access, the platform supports broader use cases and promotes collaboration between various tools and systems.

State of the art: OpenAPI.

ID	Requirement
Req-Int-3-4-1	API must comply with OpenAPI standards to ensure compatibility and automated documentation.
Req-Int-3-4-2	API must support integration with major third-party tools via RESTful and/or SOAP protocols.
Req-Int-3-4-3	API must support JSON or XML as data format.

Table 16: Availability of an Application Programming Interface (API) requirements.

5.1.3.5 Tooling governance

Rationale: New tools can become available for use in the SYNERGIES platform during platform operation. A module may substitute a part of the platform that would provide new or slightly altered technical capabilities.

State of the art: Backward compatibility.

ID	Requirement
Req-Int-3-5-1	Updated platform modules and tools shall maintain support for existing APIs, interfaces, and behaviours, unless explicitly deprecated.
Req-Int-3-5-2	Tools must be identified and properly versioned within the platform.

Table 17: Tooling governance requirements

5.1.4 Data interoperability requirements

5.1.4.1 Scenario Source Data Model(s)

Rationale: Scenarios will be extracted using WP5-developed tools / algorithms from WP3-provided "source data". Several "source data" categories exist: "fixed point reference" data (accident, drone & roadside observations) and "moving point reference" (moving vehicle data collection). Up to two different toolchains and different models and formats may therefore be developed.

However, data and algorithms must be compatible on a functional level, meaning that algorithms need to have all the information they need, in the suitable units, with suitable values mapping for enumerations, to detect scenarios from source data.

The Source Data Model(s) is (are) the description of the information that need to be contained in a "source data" file.

State of the art: SALSA Ontology, ADSCENE Data Model, Pre-Crash Matrix Model, Omega Format Model (incorporating ASAM Open Simulation Interface), Hi-Drive CDF.

ASAM Open Simulation Interface allows specification of object positions and trajectories and road information in fixed and moving coordinate frames.

Google Protocol Buffers are widely used in industry to (de)serialize such data.

ID	Requirement
Req-Int-4-1-1	Common file formats are required to store and exchange scenario source data.
Req-Int-4-1-2	Information necessary to extract scenarios from point reference" data (accident, drone & roadside observations) and "moving point reference" (moving vehicle data collection) need to be unambiguously specified.
Req-Int-4-1-3	Common scenario source data (de)serialization protocol is required to read, write and interpret scenario source data consistently.

Table 18: Scenario Source Data Model(s) requirements.

5.1.4.2 Scenario Source Data File Format(s)

Rationale: Scenarios will be extracted using WP5-developed tools / algorithms from WP3-provided "source data". Two "source data" categories exist: "fixed point reference" data (accident, drone & roadside observations) and "moving point reference" (moving vehicle data collection). Up to two different toolchains and different models and formats may therefore be developed.

However, data and tools running algorithms must be compatible on a technical level, meaning that the tools running the algorithms need to be able to read "source data" files to access the information they contain.

The Source Data Format(s) is (are) the description of the file format(s) used for a "source data" file.

State of the art: ADSCENE Data Format, Pre-Crash Matrix Format, Omega Format, Hi-Drive CDF, ASAM MDF (Measurement Data Format) and MCAP File Format are established file formats to store and exchange data in automotive/robotics applications.

ID	Requirement
Req-Int-4-2-1	A common scenario source data file format should be used throughout all data sources to enable development of common algorithms.
Req-Int-4-2-2	File format(s) used to provide data to scenario extraction tools need to be unambiguously specified.
Req-Int-4-2-3	File format(s) should rely on a standardised, free and open-source serialisation method (e.g. HDF5), with libraries freely available for every major language used in the industry for data processing (Python, MATLAB, C, C++, C#, Java)

Table 19: Scenario Source Data File requirements.

5.1.4.3 Scenario format

Rationale: Scenarios should have an open, established, standardized format.

State of the art: ASAM OpenSCENARIO XML is established in the automotive industry.

ID	Requirement
Req-Int-4-3-1	The platform shall have scenario representation features compatible with the different abstraction levels.

Table 20: Scenario format requirements.

5.1.4.4 Road network format

Rationale: The road network of the scenarios should be described in a standardized format.

State of the art: ASAM OpenDRIVE, Lanelet2, ASAM OSI.

ID	Requirement
Req-Int-4-4-1	The road network of the scenarios should be described in a standardized format
Req-Int-4-4-2	The road network logic should be explicitly contained in the format (following lanes).

Req-Int-4-4-3	The boundaries of a lane should be identifiable.
Req-Int-4-4-4	Real-world intersections should be representable in a detailed fashion (explanation: use LineStrings for geometry like Lanelet2 or OSI).

Table 21: Road network format requirements.

5.2 Data and Scenario Traceability

Data and scenario traceability refers to the ability to track the origin, changes, and flow of the data and the scenarios throughout their lifecycle. This ensures that they can all be traced back to their source and that the handling of data is transparent and auditable.

5.2.1 Lineage and Processing of Data and Scenarios

5.2.1.1 Scenario Methodology Documentation

Rationale: It ensures transparency, reproducibility, accuracy, consistency and compliance in the scenarios from various sources.

ID	Requirement
Req-Trac-1-1-1	Specific documentation of algorithms used for scenarios, the tools employed in the process, and the parameters applied must be implemented.

Table 22: Scenario Extraction Methodology Documentation requirements.

5.2.1.2 Data and Scenarios Lineage Recording

Rationale: It ensures transparency, traceability, and reliability in data-driven processes. It helps track the flow and transformation of data and scenarios from its origin to its final destination, providing insights into how data and scenarios are used, modified, and stored.

ID	Requirement
Req-Trac-1-2-1	The system must maintain complete records of scenarios and data lineage.
Req-Trac-1-2-2	The records should involve tracking the original source of the data and scenarios, documenting each transformation step applied to the data/scenario, and maintaining clear records of how different sources are combined or modified to create scenarios.
Req-Trac-1-2-3	The lineage documentation should enable any team member to understand how a scenario was created and trace it back to its original sources.

Table 23: Data and Scenario Lineage Recording requirements.

5.2.1.3 Data and Scenario Selection and Modification

Rationale: It refers to the process of choosing and adjusting specific data and scenarios for analysis or model training, based on the data and scenario origin, history, and traceability. It ensures relevance, representativeness, quality and regulatory compliance.

ID	Requirement
Req-Trac-1-3-1	The system must maintain documentation explaining why each data and scenario was selected for inclusion and justify any modifications made to the original scenario.
Req-Trac-1-3-2	It must include recording or justification that how modifications improve its effectiveness for these objectives.

Table 24: Scenario Selection and Modification requirements.

5.2.2 Data and Scenario Source Integration

5.2.2.1 Data and Scenario Source Credibility Assessment

Rationale: It ensures that the data and scenario used in decision-making processes or analysis is of high quality, dependable, and verifiable by evaluating the reliability, accuracy, and trustworthiness of data sources.

ID	Requirement
Req-Trac-2-1-1	The platform must implement a standardized process for assessing and documenting the credibility of data sources and scenarios.
Req-Trac-2-1-2	This process should evaluate the source's reputation, technical capabilities, data/scenario collection methodologies, and historical reliability.
Req-Trac-2-1-3	The assessment results must be documented and regularly reviewed.

Table 25: Source Credibility Assessment requirements.

5.2.2.2 Provider Credential Management

Rationale: It refers to the systematic handling, verification, and safeguarding of credentials associated with service providers, data sources, or any entities that interact within a system. It is vital because it ensures that the data used or exchanged in a system can be trusted, traced, and verified back to its source.

ID	Requirement
Req-Trac-2-2-1	The platform must maintain detailed records of data/scenario provider credentials and reliability metrics.

Req-Trac-2-2-2	The document must include the provider's expertise, certifications, quality assurance processes, and track record in providing similar data/scenario.
Req-Trac-2-2-3	The system should regularly update these records and assess provider performance.

Table 26: Provider Credential Management requirements.

5.2.3 Change Management System

5.2.3.1 Version Control and Modification Tracking

Rationale: It is a systematic approach to tracking and managing changes to data, scenarios or documents. It ensures traceability by maintaining a complete and accurate history of modifications—capturing the origin, purpose, and impact of each change. This supports integrity, accountability, and transparency by preventing unauthorized alterations, enabling error identification, and allowing changes to be monitored, reverted, or audited when necessary. As such, version control is essential for enhancing trustworthiness, improving auditability, and ensuring compliance with regulatory and governance requirements.

ID	Requirement
Req-Trac-3-1-1	The system must implement comprehensive tracking of all modifications made including creation and deletion to data/scenario and their related parameters.
Req-Trac-3-1-2	Each modification must be recorded with details about <i>what</i> was changed, <i>who</i> made the change, <i>when</i> it was made, and <i>why</i> it was necessary.

Table 27: Modification tracking requirements.

5.2.3.2 Change Rationale Documentation

Rationale: It provides a clear explanation of why a change was made, the decision-making process behind it, and its expected impact while ensuring that all modifications to data or scenarios are well-documented, transparent, and traceable. This helps maintain accountability, supports auditing processes, ensures compliance with regulations, and allows for informed decision-making in future changes.

ID	Requirement
Req-Trac-3-2-1	For each modification to data and scenario, the system must maintain clear documentation of the reasoning behind the change.
Req-Trac-3-2-2	The documentation must include the specific problems or requirements that prompted the change and how the modification addresses these issues.

Table 28: Change Rationale Documentation requirements.

5.2.3.3 Version Rollback Capabilities

Rationale: It allows the system to revert to a previous stable version in case of errors or undesired outcomes after updates or changes and ensures that any changes made to data, scenario or models can be tracked, and earlier versions can be restored if needed. This is crucial for maintaining data and scenario integrity, ensuring compliance, and supporting effective change management by providing a safety net against unintended consequences or failures in the system.

ID	Requirement
Req-Trac-3-3-1	The system must provide capabilities to roll back to previous state of scenarios or data.
Req-Trac-3-3-2	The system must include complete copies of previous versions and implementing procedures for safely reverting to earlier versions when necessary.

Table 29: Version Rollback Capabilities requirements.

5.2.3.4 Data and Scenario Modification Traceability

Rationale: It ensures that modifications are well-documented, enabling clear insights into how data and scenarios evolve over time. This traceability helps in identifying the impact of changes, ensuring compliance, maintaining data/scenario integrity, and facilitating troubleshooting or audits.

ID	Requirement
Req-Trac-3-4-1	The relationship between versions and maintaining metadata that explains how modifications affect the data and scenario's validation capabilities must be documented.

Table 30: Data and Scenario Modification Traceability requirements.

5.2.4 Data Traceability KPIs

Rationale: These KPIs help ensure that data is accurate, consistent, and accessible, allowing organizations to monitor data flow, identify any discrepancies, and ensure compliance with regulations. They provide transparency, facilitate auditing, and help maintain data integrity across processes, which is crucial for decision-making and accountability.

ID	Requirement
Req-Trac-4-1-1	Documentation Completeness: Extraction methodology documentation must be provided.
Req-Trac-4-1-2	Data Lineage Coverage: Scenarios must have complete lineage tracking. The measurement process should involve both automated lineages tracking tools and regular audits to verify tracking information completeness.

Req-Trac-4-1-3	Version Control Compliance : Version control compliance should ensure that all scenario modifications are properly tracked and reversible.
Req-Trac-4-1-4	External Source Credibility Score: <i>External source credibility score must be measured</i> considering factors of Documentation Completeness, Historical Data Lineage Coverage and Quality, Track Record of Timely Updates, Compliance with Data Format, and Provider Certification.

Table 31: Data Traceability KPIs requirements.

5.3 Data and Scenario Trustworthiness

Ensuring trustworthiness in raw data, scenario source data and scenarios is vital for the reliability of scenario generation, model development, and validation processes. Quality aspects such as completeness, accuracy, relevance, timeliness, and consistency form the foundation for building robust and credible AI-driven systems. The following framework outlines how each aspect should be addressed systematically.

5.3.1 Data and Scenario Completeness Assurance

5.3.1.1 Data and Scenario Coverage

Rationale: It ensures that data and scenarios are comprehensive, representative, and relevant. It helps validate data and scenario trustworthiness by minimizing biases, identifying gaps, and ensuring completeness across different conditions. This is crucial for making reliable decisions, improving model accuracy, and maintaining data integrity, ultimately fostering trust in data-driven insights.

ID	Requirement
Req-Trus-1-1-1	Data and defined set of scenarios should provide all necessary parameters required by the format selected by the SYNERGIES project.
Req-Trus-1-1-2	Automated completeness checks must be integrated into every step of data for defined sets of scenarios.
Req-Trus-1-1-3	Manual reviews should complement these checks during critical phases to validate data and scenario comprehensiveness.

Table 32: Data and Scenario Coverage requirements.

5.3.1.2 Comprehensive Metadata

Rationale: Metadata plays a crucial role in describing the context, sources, collection methods, and intended use of each data. It ensures traceability and facilitates the reproducibility of results.

ID	Requirement
----	-------------

Req-Trus-1-2-1	Metadata should include details on the type of sensors used for raw data generation, calibration information, data collection dates, and preprocessing steps applied.
-----------------------	---

Table 33: Comprehensive Metadata requirements.

5.3.1.3 Edge Case Inclusion

Rationale: Edge Case Inclusion ensures that rare, extreme, or unusual data points are considered in analysis, testing, and model training. This is crucial for data trustworthiness because it helps prevent biases, improves robustness, and ensures that systems perform reliably across diverse real-world scenarios. Ignoring edge cases can lead to inaccurate predictions, unfair outcomes, or security vulnerabilities, ultimately reducing confidence in the data and the models built upon it.

ID	Requirement
Req-Trus-1-3-1	Datasets should be regularly audited to ensure they include edge cases, such as low-visibility conditions, unexpected obstacles, or unusual traffic behaviors.
Req-Trus-1-3-2	Edge case scenarios should be reviewed periodically to confirm that they adequately reflect potential real-world challenges.

Table 34: Edge Case Inclusion requirements.

5.3.2 Data and Scenario Accuracy Assurance

5.3.2.1 Scenario Plausibility Checks

Rationale: It prevents errors, inconsistencies, or malicious alterations in the scenarios. It helps to detect and correct discrepancies early, reducing the risk of incorrect analyses, faulty decision-making, or security vulnerabilities.

ID	Requirement
Req-Trus-2-1-1	The system must implement verification of all scenario parameters to ensure they are relevant and fall within acceptable ranges. This includes checking physical constraints, behavioral parameters, temporal consistency, and logical relationships between different parameters.
Req-Trus-2-1-2	Any parameters outside expected ranges must be flagged for review and justified if accepted.

Table 35: Scenario Plausibility Checks.

5.3.2.2 Data and Scenario Source Verification

Rationale: Data and Scenario Source Verification ensures that the origins of data/scenario are credible, reliable, and valid. By confirming the authenticity and accuracy of the sources, organizations can improve accuracy assurance, leading to better decision-making, compliance with regulations, and enhanced operational efficiency.

ID	Requirement
Req-Trus-2-2-1	All data and scenario sources should follow the formats selected by the SYNERGIES project or accurate collection.
Req-Trus-2-2-2	All data and scenario sources and their data collection methodologies should be validated through published records or third-party assessments.

Table 36: Data and Scenario Source Verification requirements.

5.3.2.3 Error Detection and Correction

Rationale: It ensures data and scenario integrity by identifying and fixing errors that occur during transmission or storage. This is crucial for maintaining data and scenario trustworthiness, as errors can lead to misinformation, faulty decision-making, and security vulnerabilities.

ID	Requirement
Req-Trus-2-3-1	Automated test systems should be implemented to detect data and scenario inaccuracies, such as out-of-range values, conflicting data points, or sensor anomalies. For example, machine learning algorithms can be used to flag outliers.
Req-Trus-2-3-2	Detected errors should be logged, reviewed, and corrected on time to maintain data and scenario integrity.

Table 37: Error Detection and Correction requirements.

5.3.2.4 Validation by Experts

Rationale: Validation by experts ensures data and scenario trustworthiness and accuracy by leveraging domain-specific knowledge to assess the reliability, relevance, and correctness of data. Experts can identify inconsistencies, biases, or errors that automated processes may overlook, ensuring that the data and scenarios meet quality standards.

ID	Requirement
Req-Trus-2-4-1	For critical scenarios and data, expert review is necessary to confirm the plausibility of modeled behaviors, environmental conditions, and interactions.

Req-Trus-2-4-2	Subject matter experts should conduct periodic plausibility checks, especially for complex scenarios where automated checks may fall short.
Req-Trus-2-4-3	Expert reviews should be documented with detailed reports to support traceability.

Table 38: Validation by Experts requirements

5.3.3 Relevance Assurance

5.3.3.1 Scenario Relevance Filtering

Rationale: Data or scenario relevance refers to the degree to which a given dataset, individual data points, or task scenarios contribute to achieving the objective of SYNEGIES operational context. It ensures that only meaningful data and scenarios are stored. It helps to eliminate irrelevant, misleading, or outdated information that could compromise accuracy and reliability.

ID	Requirement
Req-Trus-3-1-1	Data and scenarios processes should include filters to ensure that only relevant data/scenario, including data from external sources, is retained.
Req-Trus-3-1-2	Data that no longer fits current project objectives should be archived or discarded to optimize storage and processing efficiency.

Table 39: Scenario Relevance Filtering requirements.

5.3.3.2 Dynamic Relevance Scoring

Rationale: It ensures that data/scenario is assessed in real-time based on evolving context, credibility, and accuracy, making it crucial for maintaining data and scenario trustworthiness and relevance assurance. This approach helps mitigate misinformation, enhances decision-making, and ensures that users receive the most relevant and trustworthy data at any given moment.

ID	Requirement
Req-Trus-3-2-1	Relevance scores to each scenario should be assigned. Scenarios that address key project functionalities or identified gaps should receive higher scores.
Req-Trus-3-2-2	Regularly update relevance scores based on evolving project goals to ensure alignment.

Table 40: Dynamic Relevance Scoring requirements.

5.3.3.3 Periodic Reassessment

Rationale: Periodic reassessment ensures that data and scenario remain accurate, relevant, and reliable over time. It ensures that the data continues to meet current analytical, regulatory, and operational needs. It prevents data degradation and enhances decision-making by keeping insights aligned with evolving requirements.

ID	Requirement
Req-Trus-3-3-1	Stored scenarios and data should be reassessed periodically to confirm their continued relevance, including those from external sources. For instance, if project goals shift due to changes in regulatory requirements or technological advancements, outdated scenarios should be identified and updated or removed as needed.

Table 41: Periodic Reassessment requirements.

5.3.3.4 Use Case Alignment

Rationale: Use Case Alignment ensures that data and scenarios are directly linked to objectives, making data/scenario more relevant and trustworthy. This alignment helps prevent data silos, reduces irrelevant or low-quality data, and enhances confidence in data-driven insights, ultimately supporting better governance and compliance.

ID	Requirement
Req-Trus-3-4-1	All data/scenario sources and parameters should be directly linked to predefined use cases.

Table 42: Use Case Alignment requirements.

5.3.4 Timeliness Assurance

5.3.4.1 Data Freshness Checks

Rationale: It ensures that data and scenarios are up-to-date as outdated data can lead to incorrect insights, poor decision-making, and reduced reliability.

ID	Requirement
Req-Trus-4-1-1	Automated checks should be implemented to periodically check the freshness of the data and scenarios. This is particularly important for scenarios that rely on time-sensitive data.
Req-Trus-4-1-2	If data has not been updated within a specified time window, it should be flagged for review.

Table 43: Data Freshness Checks requirements.

5.3.4.2 Update Automation

Rationale: Update automation ensures a proper job orchestration which maintains data integrity, consistency, and accuracy. It supports timeliness assurance by enabling dependent systems or records reflect the latest information without delays. This automation reduces human error and enhances operational efficiency, making data-driven decisions more reliable and up-to-date.

ID	Requirement
Req-Trus-4-2-1	Dependent assets such as data and scenarios must be refreshed automatically and in a timely manner if/when some of their ancestors' change.
Req-Trus-4-2-2	Update automation should be integrated into the project's data and scenario management pipeline.

Table 44: Update Automation requirements.

5.3.5 Consistency Assurance

5.3.5.1 Semantic Consistency

Rationale: Semantic consistency refers to ensuring that data is not only syntactically correct but also aligns with its intended meaning and context. It is crucial because it guarantees that the data accurately reflects real-world concepts and relationships, preventing misinterpretations and errors. This helps maintain the integrity and reliability of data, which is essential for informed decision-making and building trust in data-driven systems.

ID	Requirement
Req-Trus-5-1-1	Terminology, labels, and classifications should be uniform across data/scenarios to prevent confusion.
Req-Trus-5-1-2	Regular audits should be conducted to verify that naming conventions align with the project's data dictionary.

Table 45: Semantic Consistency requirements.

5.3.5.2 Temporal Alignment

Rationale: Temporal alignment refers to the synchronization of data across different systems or sources to ensure it reflects the same time frame or sequence of events. It ensures that the data is accurate, up-to-date, and aligned with real-world events or changes. Without temporal alignment, inconsistencies can arise, leading to misinterpretations and undermining confidence in data-driven decisions.

ID	Requirement
----	-------------

Req-Trus-5-2-1	For scenarios where time-sensitive data is collected from multiple sources, ensure that timestamps are synchronized to prevent temporal misalignment.
-----------------------	---

Table 46: Temporal Alignment requirements.

5.3.5.3 Data/Scenario Integration Protocols

Rationale: Data/scenario integration protocols refer to the structured set of rules, procedures, and technical standards that govern the collection, alignment, transformation, and unification of data from multiple heterogeneous sources into a coherent and interoperable dataset. They establish standardized methods for handling discrepancies, ensuring that integrated data/scenario is accurate, reliable, and consistent across systems. By addressing issues such as data duplication, format inconsistencies, and conflicting information, they help build trust in the data/scenario for decision-making and analysis.

ID	Requirement
Req-Trus-5-3-1	Structured set of rules or standards (e.g. data/scenario format consistency, temporal and spatial alignment) must be developed for integrating data/scenario from various sources, ensuring harmonization in terms of format, units, and structure.

Table 47: Data/Scenario Integration Protocols requirements.

5.3.6 Data trustworthiness KPIs

Rationale: They are essential metrics that help to evaluate the reliability, accuracy, and quality of the data used in decision-making. These metrics help to monitor data quality, mitigate risks, and maintain compliance with regulations.

ID	Requirement
Req-Trus-6-1-1	Scenario Parameter Completeness: This score should be measured to ensure data quality, mitigate risks, and maintain compliance. .
Req-Trus-6-1-2	Edge Case Coverage Rate: This rate should be calculated by dividing the number of documented edge cases by the total number of identified edge cases.
Req-Trus-6-1-3	Parameter Plausibility: This metric can be calculated by dividing the number of verified parameters by the total number of parameters. The verification process includes range validation for numerical parameters, format verification for structured data, logical consistency checks across related parameters, and temporal sequence validation.
Req-Trus-6-1-5	Scenario Relevance: This score can be calculated by averaging relevance scores across all scenarios, considering alignment with Use Case Requirements, Timeliness or Freshness of the Data, Consistency of the

	Data, Applicability to Current Testing Needs, and Coverage of Testing Scenarios.
Req-Trus-6-1-7	Semantic Consistency Score: This score checks all semantic elements (terms, classifications, labels) and their usage across different contexts and data/scenario. It should evaluate the uniform application of terminology and classifications across all data/scenarios. The score can be calculated by dividing the number of consistently used elements by the total number of semantic elements.
Req-Trus-6-1-8	Temporal Consistency Index: This score measures the alignment of time-sensitive data across multiple sources. The index can be calculated by dividing the number of temporally aligned data points by the total number of time-stamped data points.

Table 48: Data trustworthiness KPIs requirements.

5.4 Cybersecurity

To ensure the confidentiality, integrity, and availability of data, it is critical to implement strong cybersecurity measures. The following requirements and definitions shall be strictly enforced for all external users. Internal members, who are already authorised to exchange data under pre-established data sharing agreements, may be subject to adapted or less stringent requirements, in accordance with the existing policies.

5.4.1 Data Protection Framework

5.4.1.1 End-to-End Encryption Implementation

Rationale: End-to-End Encryption (E2EE) ensures that data is encrypted on the sender's device and only decrypted on the recipient's device, preventing third parties — including service providers, hackers, and government agencies — from accessing the communication. This implementation is crucial in cybersecurity as it safeguards sensitive information from interception, reduces the risk of data breaches, and enhances user privacy. By eliminating reliance on intermediaries for data security, E2EE strengthens confidentiality, integrity, and trust in digital communications.

ID	Requirement
Req-Cyb-1-1-1	The SYNERGIES platform must implement comprehensive encryption mechanisms that ensure complete protection of all sensitive data throughout its lifecycle.
Req-Cyb-1-1-2	This implementation shall utilize industry-standard encryption algorithms, with a minimum requirement of AES-256 for all data at rest. For data in transit, the system must enforce TLS 1.3 or higher protocols across all communication channels.

Req-Cyb-1-1-3	A robust encryption key management system must be established, incorporating regular key rotation policies that align with industry best practices.
Req-Cyb-1-1-4	It shall maintain distinct encryption mechanisms for different security levels, ensuring appropriate isolation and protection based on data sensitivity.
Req-Cyb-1-1-5	The SYNERGIES system must implement secure key storage and management systems that protect against unauthorized access while ensuring availability for authorized operations.

Table 49: End-to-End Encryption Implementation requirements

5.4.1.2 Secure Storage and Transmission

Rationale: Secure storage and transmission protect sensitive data from unauthorized access, theft, and tampering. In cybersecurity, this ensures confidentiality, integrity, and availability of data by using encryption, access controls, and secure communication protocols. It is crucial for preventing data breaches, complying with regulations, and maintaining user trust in digital systems.

ID	Requirement
Req-Cyb-1-2-1	Data storage and transmission systems must be architected with security layers that provide comprehensive protection against unauthorized access and data breaches.
Req-Cyb-1-2-2	The SYNERGIES platform shall implement secure storage zones that segregate data based on sensitivity levels, with appropriate controls and monitoring for each zone.
Req-Cyb-1-2-3	All data transfers must occur through established secure transmission channels that implement appropriate encryption and verification mechanisms.
Req-Cyb-1-2-4	It must maintain physically or logically separate storage areas for different security classifications, ensuring proper isolation of sensitive data.
Req-Cyb-1-2-5	SYNERGIES platform shall implement comprehensive backup encryption procedures that maintain security during disaster recovery scenarios, along with secure data disposal protocols that ensure complete and irrecoverable deletion of sensitive information when required.

Table 50: Secure Storage and Transmission requirements.

5.4.1.3 Security Audit Procedures

Rationale: It systematically evaluates an organization's security policies, controls, and practices to identify vulnerabilities, ensure compliance with regulations, and mitigate risks. These audits

help detect weaknesses before they can be exploited by cyber threats, ensuring data integrity, confidentiality, and availability. Security audits are crucial for preventing breaches, maintaining trust, and ensuring business continuity by proactively addressing security gaps.

ID	Requirement
Req-Cyb-1-3-1	A comprehensive security audit framework must be established to ensure continuous monitoring and improvement of security measures.
Req-Cyb-1-3-2	This audit framework shall mandate periodic internal security reviews that assess all aspects of the system's security implementation.
Req-Cyb-1-3-3	Annual external security audits must be conducted by qualified third-party organizations to provide independent verification of security controls.
Req-Cyb-1-3-4	The procedure shall implement continuous security monitoring mechanisms that provide real-time alerts for potential security incidents.
Req-Cyb-1-3-5	Detailed audit logs must be maintained for all security-relevant events, with appropriate retention periods and protection mechanisms.
Req-Cyb-1-3-6	The SYNERGIES platform shall require the creation and implementation of detailed action plans based on audit findings, with clear timelines and responsibility assignments.

Table 51: Security Audit Procedures requirements.

5.4.1.4 Incident Response Protocol (IRP)

Rationale: It is a structured approach to handling cybersecurity incidents, ensuring swift detection, containment, eradication, and recovery from threats. Its rationale lies in minimizing damage, reducing downtime, and preserving the integrity of systems and data. In cybersecurity, IRP is crucial for mitigating risks, ensuring compliance with regulations, and maintaining business continuity by preventing incidents from escalating into major breaches.

ID	Requirement
Req-Cyb-1-4-1	A comprehensive incident response system must be established to ensure rapid and effective response to security incidents. This system shall include detailed incident classification criteria that enable appropriate response prioritization and resource allocation.
Req-Cyb-1-4-2	The IRP must define specific response procedures for different incident types, including clear roles and responsibilities for response team members.
Req-Cyb-1-4-3	The IRP shall implement detailed recovery procedures that enable rapid service restoration while maintaining security controls.

Req-Cyb-1-4-4	The IRP must maintain comprehensive incident documentation and incorporate lessons learned into security improvement processes.
----------------------	---

Table 52: Incident Response Protocol requirements.

5.4.2 GDPR Compliance Framework

5.4.2.1 Personal Data Processing

Rationale: It ensures that data is collected, used, and stored in a lawful, fair, and transparent manner while protecting individuals' privacy rights. It is essential in cybersecurity because improper handling of personal data can lead to breaches, identity theft, and unauthorized access, compromising both individuals and organizations. Strong data processing practices, including encryption, access control, and compliance with regulations like GDPR, help mitigate risks and enhance overall security.

ID	Requirement
Req-Cyb-2-1-1	The SYNERGIES platform must implement strict controls for personal data handling that ensure full compliance with GDPR requirements. These controls shall include comprehensive mechanisms for identifying and classifying personal data across all system components.
Req-Cyb-2-1-2	Data minimization procedures must be implemented to ensure only necessary personal data is collected and processed.
Req-Cyb-2-1-3	The SYNERGIES platform shall establish clear processing purpose limitations that align with declared processing purposes and legal bases.
Req-Cyb-2-1-4	Detailed data retention schedules must be maintained, with automated enforcement of retention periods and secure deletion procedures.
Req-Cyb-2-1-5	It shall maintain comprehensive records of all processing activities, including detailed documentation of processing purposes, data categories, and security measures.

Table 53: Personal Data Processing requirements.

5.4.2.2 Data Subject Rights Management

Rationale: Data Subject Rights Management ensures individuals can exercise their rights over their personal data, such as access, rectification, erasure, and portability. It helps protect personal data from unauthorized access, misuse, or breaches, ensuring compliance with legal frameworks and fostering trust between organizations and users. Proper management mitigates legal risks, enhances transparency, and strengthens data governance.

ID	Requirement
Req-Cyb-2-2-1	Comprehensive procedures must be implemented for managing and fulfilling data subject rights under GDPR. These procedures shall include

	detailed mechanisms for handling data access requests, including verification of requestor identity and appropriate response timeframes.
Req-Cyb-2-2-2	The SYNERGIES platform must implement robust data portability mechanisms that enable the export of personal data in commonly used, machine-readable formats.
Req-Cyb-2-2-3	Clear protocols shall be established for executing data erasure requests, including verification procedures and documentation requirements.
Req-Cyb-2-2-4	It must include procedures for managing rectification requests, with appropriate validation and update mechanisms.
Req-Cyb-2-2-5	SYNERGIES platform shall maintain detailed documentation of all interactions related to data subject rights, including request handling timelines and outcomes.

Table 54: Data Subject Rights Management requirements.

5.4.3 Access Control and Authentication

5.4.3.1 Role-Based Access Control

Rationale: It restricts system access based on users' roles within an organization. It ensures that users only have the permissions necessary to perform their job functions, minimizing the risk of unauthorized access or data breaches. Role-Based Access Control enhances access control by enforcing the principle of least privilege, reducing the attack surface, and preventing privilege escalation. It also simplifies authentication management and compliance with regulatory requirements by standardizing access policies.

ID	Requirement
Req-Cyb-3-1-1	A role-based access control system must be implemented to ensure precise management of system access rights. This system shall define clear role hierarchies that reflect organizational structures and responsibilities.
Req-Cyb-3-1-2	Detailed permission matrices must be maintained, documenting the specific access rights associated with each role.
Req-Cyb-3-1-3	The SYNERGIES platform shall implement robust procedures for role assignment, including appropriate approval workflows and documentation requirements.
Req-Cyb-3-1-4	Regular role review processes must be conducted to ensure continued appropriateness of access rights.
Req-Cyb-3-1-5	It shall maintain comprehensive documentation of all role definitions and assignments, including historical records of changes and justifications.

Table 55: Role-Based Access Control requirements.

5.4.3.2 Multi-Factor Authentication

Rationale: Multi-Factor Authentication (MFA) enhances security by requiring users to verify their identity using multiple factors: something they know (password), something they have (security token or smartphone), or something they are (fingerprint or facial recognition). This reduces the risk of unauthorized access, even if one factor is compromised. In cybersecurity and access control, MFA is crucial for protecting sensitive systems and data from breaches, phishing attacks, and credential theft, ensuring stronger authentication beyond just passwords.

ID	Requirement
Req-Cyb-3-2-1	Strong multi-factor authentication mechanisms must be implemented for all sensitive operations within the system.
Req-Cyb-3-2-2	These mechanisms shall require at least two independent authentication factors, with additional factors required for highly sensitive operations.
Req-Cyb-3-2-3	SYNERGIES platform must support multiple authentication methods to ensure accessibility while maintaining security.
Req-Cyb-3-2-4	Robust backup authentication procedures shall be implemented to ensure system access in case of primary authentication method failure.
Req-Cyb-3-2-5	It must establish clear protocols for handling emergency authentication bypass situations, including appropriate logging and review requirements.
Req-Cyb-3-2-6	Comprehensive authentication audit logs shall be maintained, documenting all authentication attempts and outcomes.

Table 56: Multi-Factor Authentication requirements.

5.4.4 Cybersecurity KPIs

Rationale: They are essential metrics used to measure the effectiveness and efficiency of an organization's cybersecurity efforts. These KPIs help assess the organization's ability to protect its systems, data, and networks from cyber threats.

ID	Requirement
Req-Cyb-4-4-1	Encryption Coverage Ratio: This metric measures the proportion of sensitive data protected by encryption, both at rest and in transit, using AES-256 and TLS 1.3 standards. This can be calculated as the percentage of encrypted data relative to the total amount of sensitive data handled by the system. Regular monitoring is required to verify adherence to these standards.

Req-Cyb-4-4-2	Security Incident Response Time: This can be measured as the average time between incident detection and resolution, calculated by summing all response times and dividing by the number of incidents.
Req-Cyb-4-4-3	Secure Data Transfers Rate: This metric measures the proportion of data transfers that utilize secure transmission protocols, such as TLS 1.3. This can be calculated as the number of secure transfers over total transfers, and ensures that all communication channels are encrypted, mitigating the risk of data interception during transmission.
Req-Cyb-4-4-4	Personal Data Classified and Processed Correctly: This metric evaluates the proportion of personal data that is accurately classified and processed per GDPR guidelines. This KPI is calculated as correctly classified personal data over total personal data and proper classification is essential for ensuring compliance with GDPR and protecting user data.
Req-Cyb-4-4-5	Data Subject Rights Fulfillment Time: This metric tracks the average time taken to fulfill data subject rights requests, such as access, erasure, or rectification. This can be measured as the total response time in days divided by the number of requests.

Table 57: Cybersecurity KPIs requirements.

6 TECHNICAL REQUIREMENTS

6.1 Metadata format

Metadata properties have been categorized to distinguish those that are common to any objects on the SYNERGIES platform to those that requires specific properties according to the related object.

This typology is illustrated in the table Table 58: Metadata typology. It has been produced based on objects of the dataflow identified in section 3. Key concepts and terminology.

Metadata category	Raw data metadata	Scenario source data (SSD) metadata	Scenario metadata
Administrative <i>Ownership, licensing, IP, access policies, ...</i>			
Traceability <i>Author, timestamp, versioning, ...</i>			
Structural & referential <i>Format, schema, relationships between elements, linkages to other data, ...</i>	Common metadata properties		
Provenance / lineage <i>Transformations and processes applied, references to metadata of parent objects, ...</i>		Including reference to raw data metadata	Including reference to SSD metadata
Descriptive / content level <i>Keywords, thematic information, domain-specific descriptors, content-level details, ...</i>	Raw data metadata Including acquisition context descriptors and scenario domain specific descriptors	Scenario specific metadata Including scenario domain specific descriptors	
Quality metrics <i>Accuracy, completeness, known bias, integrity, ...</i>		Quality descriptors	

Table 58: Metadata typology.

The groups of metadata properties – common metadata properties, raw data metadata, scenario specific metadata and quality descriptors – are detailed in the sub-chapters below. For each group, expected information and interoperable formats are specified.

6.1.1 Common metadata properties

The table below (Table 59: Common metadata properties) lists the properties that the metadata format must be able to support. For each listed property, the requirement must be read as “*The metadata format should provide a placeholder for ...*”.

Req.	Metadata / Description / Interoperable format <i>The metadata should provide placeholder for ...</i>
Req-Tech-1-1-1	Traceability / ID Unique asset identifier. UUID (RFC 4122)
Req-Tech-1-1-2	Traceability / Title Name of the resource Plain text
Req-Tech-1-1-3	Traceability / Author & contributors Creators and contributors' identifiers. RFC 5322
Req-Tech-1-1-4	Traceability / Timestamps Creation and last modification time. ISO 8601
Req-Tech-1-1-5	Traceability / Version Version / revision of the asset. Semantic versioning v2.0.0
Req-Tech-1-1-6	Administrative / Ownership Identity of the owner of the authority over the data (legal and administrative responsibility). RFC 5322 Plain text
Req-Tech-1-1-7	Administrative / Licensing Licensing information specifying how the data can be used, distributed, or modified by others. SPDX identifier "Commercial" label
Req-Tech-1-1-8	Administrative / Intellectual property

	Any copyright, patents, or other IP protections associated with the data. Plain text
Req-Tech-1-1-9	Administrative / Classification of data Classification of data sensitivity/confidentiality. "Public" "Confidential" "Restricted"
Req-Tech-1-1-10	Structural / Format Name and version of the format used for the payload structure. Plain text
Req-Tech-1-1-11	Structural / Lineage Reference of parent object. UUID URI
Req-Tech-1-1-12	Structural / Frame of references Used frame of references (descriptors, ontologies). UUID URI
Req-Tech-1-1-13	Structural / Language Primary languages associated with the data. ISO 639-1 codes ; e.g. en, fr, de
Req-Tech-1-1-14	Provenance / Origin Description of the data provenance, including how, when, and by whom it was collected or generated. Plain text markdown URI
Req-Tech-1-1-15	Provenance / Input data Source material that has been used to produce the object's data. UUID URI
Req-Tech-1-1-16	Provenance / Transformation Documentation of the process used to generate the data from its source. Plain text markdown URI
Req-Tech-1-1-17	Descriptive / Abstract Description of the data's content, context and purpose. Plain text markdown

Req-Tech-1-1-18	Descriptive / Quality metrics Object dependent. <i>c.f. 6.1.4. Quality descriptors.</i>
Req-Tech-1-1-19	Descriptive / Domain specific descriptors Object dependent. <i>c.f. 6.1.2. Raw data metadata and 6.1.3. Scenario specific metadata.</i>

Table 59: Common metadata properties.

6.1.2 Raw data metadata

The table below (Table 60: Raw data metadata properties) lists all properties specific to raw data that must be provided in addition to the common metadata properties with raw data objects. For each listed property, the requirement must be read as "*The metadata format should provide a placeholder for ...*".

ID	Requirement
	<i>The metadata should provide placeholder for ...</i>
Req-Tech-1-2-1	Operation context <i>Environmental and operational context during data collection (e.g. weather conditions, location constraints, road type...).</i> Tags relying on scenario descriptors.
Req-Tech-1-2-2	Sensor setup <i>Types of sensors used for data recording, their technical parameters and mounting details (e.g. range, FOV, acquisition frequency, calibration, orientation, ...).</i> Plain text markdown URI
Req-Tech-1-2-3	Synchronization Details on how sensor data are timestamped and synchronized. Plain text markdown URI
Req-Tech-1-2-4	Known limitation Recognized technical or data gaps affecting usability. Plain text markdown URI
Req-Tech-1-2-5	Ground Truth Availability Description of whether and how ground truth data (e.g. annotations, labels, video or sensor overlays) is provided.

Plain text markdown URI	
Req-Tech-1-2-6	Assessment report Abstract of data audit process and outcomes. Plain text markdown URI

Table 60: Raw data metadata properties.

6.1.3 Scenario specific metadata

Scenario specific metadata is a crucial part of the datasets which are generated and used in the SYNERGIES project and finally published on the SYNERGIES platform. It is necessary to ensure uncomplicated work with the contained scenarios in terms of identifying those suitable for a specific use case, executing data analyses without analysing the contents of the scenario itself or simply structuring large datasets. Metadata does not contain new aspects of a scenario but is more a compact summary of it.

In section 5.1.2.2, requirements regarding the interoperability of scenario metadata were already elaborated. Going one step further for clear definition of scenario metadata also technical requirements must be discussed. This will be the objective of this section.

Table 61, shows the identified technical requirements for the scenario metadata.

ID	Requirement
Req-Tech-1-3-1	The scenario metadata format shall be documented.
Req-Tech-1-3-2	The scenario metadata format shall ensure traceability of format specification changes.
Req-Tech-1-3-3	The scenario metadata format shall define the set of tags that can be assigned.
Req-Tech-1-3-4	The tags shall be applied in accordance with the ISO 34504 standard.
Req-Tech-1-3-5	The scenario metadata format shall give the possibility to extend the set of specified tags as long as they are not conflicting with the specified tags from ISO 34504.
Req-Tech-1-3-6	If a scenario aspect is not covered by a tag, its applicability should be considered uncertain.
Req-Tech-1-3-7	It shall be possible to group tags. Example: tags related to a single entity, like "passenger car" and "decelerating".
Req-Tech-1-3-8	The scenario metadata format shall support descriptive information but also a further specification with numerical values (single value or ranges, specified units and data formats (time-format, GPS format, decimal separator)).

Req-Tech-1-3-9	The scenario metadata format shall define whether a tag can be extended by numerical values.
Req-Tech-1-3-10	The scenario metadata format shall support information on static elements, dynamic elements, environmental conditions, and additional information of a scenario (for example to determine representativeness of scenarios).
Req-Tech-1-3-11	The scenario metadata shall support administrative information (e.g., for the description of creator of scenario, date & time of creation, ...).
Req-Tech-1-3-12	Each scenario shall include metadata annotations identifying particular events, such as sudden pedestrian appearances or near-collisions.
Req-Tech-1-3-13	The platform shall facilitate scenario retrieval (relevance to SuT, ODDs, Actors traffic Behaviours) through metadata.

Table 61: Scenario metadata requirements.

The first two requirements (*Req-Tech-1-3-1* and *Req-Tech-1-3-2*) aim at a transparent dealing with how scenario metadata is defined in SYNERGIES. Firstly, complete documentation of the metadata format is essential. Nevertheless, in a multi-year project like SYNERGIES, there are changes expected with ongoing progress. To keep them traceable, the format must be versioned, and the current format version itself must be part of the set of tags.

Further, it is required that the space of possible tags is clearly defined in the scenario metadata format. Referring to this, the project partners came up with the decision to use a standardized framework for scenario tags. In SYNERGIES, the scenario metadata tags will be in conformance with ISO 35404 standard. Note that this means that it is possible to extend the space of tags, provided they do not conflict with the tags already provided in the ISO standard. (requirements *Req-Tech-1-3-3* to *Req-Tech-1-3-5*).

The tags assigned to a scenario should ease the analysis of them without the need to have a look at the data itself. Especially if there is information that is related to a scenario participant, it must be indicated to which the tagged information belongs to. To cover this requirement, it should be possible to group tags. This prevents information from being linked to the wrong entity when analysing tags. (requirement *Req-Tech-1-3-7*)

Requirements *Req-Tech-1-3-8* to *Req-Tech-1-3-11* are about the kind of information that can be addressed with scenario metadata. Besides descriptive information (e.g., decelerating, keeping speed, accelerating) the scenario metadata format should also support the assignment of numerical values. That means, for example, that the tag "accelerating" is extensible by a concrete value or a range of the acceleration. While such enrichment is not suitable for every tag, it should be defined whether it is possible or not. In a general overview, the scenario metadata format should support the tagging of information on static and dynamic elements but also on environmental conditions. Furthermore, administrative information and higher-level information should be a part of the scenario metadata format.

6.1.4 Quality descriptors

High-quality metadata is critical for ensuring datasets are properly identified, reused, and integrated with other datasets. Poor metadata quality, such as incomplete or inaccurate entries, can lead to misinterpretation, misclassification, and reduced usability, ultimately diminishing the impact and reliability of research outputs. Additionally, metadata plays a key role in ethical data handling by ensuring compliance with data protection regulations, safeguarding sensitive information, and aligning with EU-funded research mandates.

To achieve high-quality metadata, it is imperative to adhere to the requirements outlined in the Constraints (Section 5) of this project, particularly subsections 5.2, 5.3, and 5.4, and their associated Key Performance Indicators (KPIs). In this section, we elaborate on these metadata quality requirements, providing a detailed discussion of their importance and implementation.

ID	Requirement
Req-Tech-1-4-1	Completeness – Metadata records must include all required fields, ensuring that no essential information is missing. This includes attributes such as dataset title, creator(s), creation date, licensing terms, and applicable standards.
Req-Tech-1-4-2	Accuracy – All metadata entries must be precise and free from errors. This means that attributes such as geographic coordinates, timestamps, authorship details, and references must be correctly documented.
Req-Tech-1-4-3	Consistency – Metadata should follow a uniform structure and standardized vocabularies throughout the project. This includes adherence to predefined naming conventions, format standardization (e.g., ISO 8601 for dates), and compliance with established ontologies to ensure uniformity across datasets.
Req-Tech-1-4-4	Compliance with EU Regulations – Metadata must adhere to relevant EU policies, including GDPR, Open Science principles, and Horizon Europe data management guidelines.
Req-Tech-1-4-5	Machine-Readability – Metadata should be formatted in machine-readable standards like XML, JSON, or RDF to facilitate automated processing and integration into digital services. This enables advanced functionalities, such as automatic metadata harvesting, indexing, and semantic search.
Req-Tech-1-4-6	Versioning & Provenance – Metadata should document the history of dataset modifications, including version numbers, timestamps of updates, and contributors. Provenance information should detail the origin of the data, transformations applied, and any derivations to ensure reproducibility and transparency in research workflows.

Table 62: Quality descriptors requirements.

6.2 Scenario source data format

In the context of SYNERGIES, Scenario Source Data (SSD) corresponds to the data that is provided to WP5 for scenario identification and extraction. Such data can stem from multiple sources: in-traffic-vehicle data collection, fixed or mobile (e.g., drones) roadside observations, dynamic accident reconstruction, but also generative AI, traffic simulation.

The requirements for the SSD can be grouped into harmonization, tooling, traceability, content, quality exchange and metadata:

- **Harmonization** requirements will focus on a standardized framework that ensures consistency and compatibility across the various data sources.
- **Tooling** requirements will emphasize the development of software and applications that facilitate the collection, processing, and analysis of SSD efficiently.
- **Traceability** requirements will ensure that all data points can be tracked back to their original sources, maintaining transparency and accountability throughout the process.
- **Content** requirements will specify the types of information that must be included in the SSD to support effective scenario identification and extraction.
- **Quality** requirements will outline criteria that data must meet to be considered reliable and valid for use.
- **Exchange** requirements will focus on defining how SSD can be shared between different stakeholders while ensuring security and integrity of the data.
- **Metadata** requirements will dictate what descriptive information should accompany the SSD to provide context, origin, and relevance of the data collected.

ID	Requirement	Category
Req-Tech-2-1	SSD: Unified Data Format - The system shall define and use a standardized data format for scenario source data that supports all intended use cases within the SYNERGIES framework, eliminating the need for format-specific tooling.	Harmonization
Req-Tech-2-2	Tooling for reading and parsing data format - The system shall provide official libraries in a commonly used programming language (e.g., Python) to reliably read and parse the SSD without requiring custom implementations.	Tooling
Req-Tech-2-3	Tooling for writing data format - The system shall provide official libraries in commonly used programming languages (e.g., Python) to write SSD in the unified format, ensuring syntactic and semantic conformity.	Tooling
Req-Tech-2-4	Tooling for quality assurance of data format - The system shall include validation tools that automatically check for compliance with the SSD format and perform	Tooling

	basic plausibility checks on the content (e.g., physical feasibility, data consistency, data visualizations).	
Req-Tech-2-5	Traceability - Each SSD item shall include a globally unique identifier and version metadata to allow tracking of changes, identification of data sources, and avoidance of redundant processing.	Traceability
Req-Tech-2-6	SSD Capturing Condition - Each SSD dataset shall include structured metadata that describes the data acquisition context, including; <ul style="list-style-type: none"> - Raw data acquisition method (e.g., naturalistic driving, test track). - Data source (e.g., highD Dataset [9], inD Dataset [10]). - Legal and geographic context (e.g., country, traffic law). - Environment type (e.g., urban, rural, highway). 	Content
Req-Tech-2-7	SSD: Object List Data - Each scenario shall include an object list where each object contains: <ul style="list-style-type: none"> - A unique identifier - A time-series trajectory with position, velocity, and orientation - A geometric representation (e.g., bounding box) 	Content
Req-Tech-2-8	SSD: Map Information - Each scenario shall include absolute map data, using a standard such as Lanelet2 or OpenDRIVE, to enable detailed representation of complex road geometries.	Content
Req-Tech-2-9	SSD: Object and Map Link - The system shall ensure that object trajectories and map data are defined in the same coordinate system, enabling accurate spatial positioning and interaction.	Content
Req-Tech-2-10	SSD: Signal Frequency - All time-series data (e.g., trajectories, signals) shall be sampled at a minimum frequency of 10 Hz, and either synchronized or accompanied by a well-defined interpolation method.	Quality
Req-Tech-2-11	SSD: Compatibility - The system shall provide tools to convert existing datasets into the unified traffic data format to maximize reusability of prior data.	Harmonization
Req-Tech-2-12	SSD: Data of influencing factors - The system shall support inclusion of additional contextual data relevant	Content

	to driving behaviour, such as weather conditions, signage visibility, and digital messages, consistent with a multi-layer environment model.	
Req-Tech-2-13	Storage: Accessibility - Stored SSD shall be accessible via standardized APIs to support programmatic and automated data access for downstream applications.	Exchange
Req-Tech-2-15	Metadata: confidence metrics - Each SSD dataset shall associate confidence metrics with key parameters (e.g., object position, velocity, type), based on sensor accuracy and post-processing uncertainty.	Metadata
Req-Tech-2-16	Explicit assets scope - The description of assets shall include a rationale for why it has been included. Where scenarios are relevant to a particular national / provincial traffic rule, this should be included in the description and by the use of any appropriate identifiers in the ontology or taxonomy. The user shall have access to tags associated with test scenarios and test cases in order to understand the coverage of the asset.	Traceability
Req-Tech-2-17	Scenarios tagging - A frame of reference of shared tags shall be available within the platform. Tags may include the nature of the scenario: Nominal, critical, failure ² ; and comply with ISO 34504 [11].	Content
Req-Tech-2-18	Resources unique identification - Unique identifier shall be provided to be immutable and persist across different versions or updates of the SSD dataset.	Traceability
Req-Tech-2-19	Scenario Source Data Model(s) - The Source Data Model(s) is (are) the description of the information that need to be contained in a "source data" file.	Content
Req-Tech-2-20	Scenario Source Data File Format(s) - The Source Data Format(s) is (are) the description of the file format(s) used for a "source data" file.	Harmonization
Req-Tech-2-21	Road network format - The road network of the SSD should be described in a standardized format.	Content
Req-Tech-2-22	Data and format interoperability - The scenario source data generation system shall be able to handle data from different sources, formats, structure, frequency rates, resolution, accuracy levels and standards without loss of information or quality.	Harmonization

² Nominal, critical and failure scenarios are explained also in [14]

Req-Tech-2-23	Comprehensive Scenario Coverage - Comprehensive Scenario Coverage for Varying Road Environments, Edge Cases, and Complex Traffic Scenarios.	Content
Req-Tech-2-24	Standardized SSD Formats and Interfaces - The Data Formats and Interfaces shall be standardized.	Harmonization
Req-Tech-2-25	Consistent interfaces for object and trajectory - The interfaces for object and trajectory data recording shall be consistent.	Harmonization
Req-Tech-2-26	Input data quality (e.g., like Omega) - The platform shall provide guidelines about quality requirements for input data (content and accuracy).	Quality
Req-Tech-2-27	Open and detailed Guidelines for data tagging - The platform shall ensure minimum tagging set and a convenient guideline for tag generation.	Harmonization

Table 63: Scenarios source data formats requirements.

6.3 Scenario format

Scenarios form the content, communication medium and output of the SYNERGIES platform. The associated scenario format must enable communication between stakeholders, usability for the widest possible range of users and traceability and credibility for test and validation use-cases. Therefore, there are not only requirements concerning the scenario content, but also regarding scenario creation, documentation and integration:

ID	Requirement
Req-Tech-3-1	Interoperability - The scenario format shall be standardized and cover any required abstraction level (functional, abstract, logical and concrete).
Req-Tech-3-2	Harmonization - The scenario format shall support regional specifics and allow their abstraction.
Req-Tech-3-3	Coverage - The scenario format shall support coverage metrics defined in SYNERGIES and indicate a scenarios' contribution to coverage (e.g. by corresponding metadata).
Req-Tech-3-4	Vehicle-to-Everything (V2X) - The scenario format shall support V2X communication.
Req-Tech-3-5	Scenario-based Testing - The scenario format shall support scenario-based testing and corresponding validation. The testing will include concrete cases, mixed traffic and cooperative V2X scenarios.

Req-Tech-3-6	Evidence Dossiers – The scenario format (or corresponding metadata) shall support creation of evidence dossiers (e.g. by indication of a scenarios' validation state, credibility, versioning etc.).
Req-Tech-3-7	Export Formats – The scenario format shall be identical or convertible to standardized scenario formats (ASAM OpenSCENARIO XML and ASAM OpenSCENARIO DSL).
Req-Tech-3-8	Objects and Trajectories – The scenario format shall provide means to represent object and trajectory data.
Req-Tech-3-9	Technological Readiness Level – The scenario format shall be standardized, established and matured (e.g. ASAM OpenSCENARIO XML 1.3).
Req-Tech-3-10	Homologation – The scenario format shall support scenario spaces designed for homologation use-cases (e.g. by indicating its affiliation to the scenario space of a certain homologation use-case by corresponding metadata).
Req-Tech-3-11	Input Data Quality – The scenario format shall support providing information on input data quality (e.g. by corresponding metadata).
Req-Tech-3-12	Regulatory Institutions – The scenario format shall match the requirements of regulatory institutions (e.g. ALKS, Euro NCAP, NHTSA).
Req-Tech-3-13	Sources – The scenario format shall match the requirements of all relevant sources (e.g. nominal scenarios, scenarios resulting from accidents, scenarios resulting from risk analysis, scenarios resulting from driving, scenarios specific to priority vehicles, etc.).
Req-Tech-3-14	Traceability – The scenario format shall support traceability (e.g. by corresponding metadata) (see section 5.2).
Req-Tech-3-15	Versioning – The scenario format shall support traceable and incremental changes to scenarios.
Req-Tech-3-16	Dynamic Adjustment – The scenario format shall support dynamic scenario adjustment (e.g. by overloaded dynamic scenario parameters).
Req-Tech-3-17	History Logs – The scenario format shall support creation of history logs.
Req-Tech-3-18	Filtering – The scenario format shall support scenario filtering by regulation and system (e.g. by corresponding metadata).
Req-Tech-3-19	Test Environments – The scenario format shall support all relevant test environments (e.g. virtual test, track test and real-world testing).
Req-Tech-3-20	Consistency – The scenario format or corresponding metadata format shall provide metadata consistency with scenario source data.

Req-Tech-3-21	Sensor Technologies – The scenario format or corresponding metadata format should be independent of simulated sensors but should provide metadata on the sensor technologies used to record the scenario for traceability (e.g. test drive recording, drone recording, accident report).
Req-Tech-3-22	Regionality – The scenario format (or corresponding metadata) shall enable selection of scenarios based on a defined region.
Req-Tech-3-23	AI Act and GDPR Compliance – The scenario format (or corresponding metadata) shall provide means to indicate AI act and GDPR compliance (as defined by an overarching process) for traceability.
Req-Tech-3-24	AI-based Generation – The scenario format (or corresponding metadata) shall indicate AI-based generation of the scenario for traceability.
Req-Tech-3-25	GroundTruth Generation Guidelines – The scenario format (or corresponding metadata) shall indicate compliance with GroundTruth generation guidelines for traceability.
Req-Tech-3-26	Critical Zones – The scenario format (or corresponding metadata) shall support the description of "critical zones" for V2X communication.
Req-Tech-3-27	Open-Source and Open-Access – The scenario format shall be Open-Source and use Open-Access License and provide means to indicate that.
Req-Tech-3-28	Documentation – The scenario format shall be transparently documented.
Req-Tech-3-29	Uniqueness – The scenario format shall be unique - within the federated architecture of the SYNERGIES platform, only a single scenario format shall be used. Scenarios provided in other formats to the platform shall be converted to this format.
Req-Tech-3-30	Source and Generation Process – The scenario format or corresponding metadata shall provide information on the source data and the generation process.
Req-Tech-3-31	Privacy Laws – The scenario format (or corresponding metadata) shall provide means to indicate presence of personal data (according to privacy laws) in scenarios.
Req-Tech-3-32	Ontologies – The scenario format should allow the usage of standardized ontologies.

Table 64: Scenario format requirements.

6.4 Ontologies

This project requires a scalable, modular ontology for managing large number of testing scenarios. It should be in an open-source format like *.owl* or *.turtle* and unify scenario elements across different databases. The ontology must leverage existing standards, describe all tools and methods used, and support varying levels of detail. It should be capable of handling growing data volumes, allow for modular extensions by workgroups, and enable inferences between scenarios.

ID	Requirement
Req-Tech-4-1	Standardized format – Ontology should be available in one of standardized and open-source format (<i>.owl</i> or <i>.turtle</i>)
Req-Tech-4-2	Unified representation – Ontology should have a unified representation of scenario elements. Required to link content of different scenario databases to a single ontology.
Req-Tech-4-3	Abstraction level coverage – The ontology should allow varying levels of detail for different levels of abstraction in scenarios.
Req-Tech-4-4	Leveraging of existing standards – Ontology should rely on already established ontologies and related standards, when available. Including ISO34504 (as already referenced in previous requirements).
Req-Tech-4-5	Context relevance – Ontology should be able to describe all relevant elements in the context of tools, data and methods used in the project.
Req-Tech-4-6	Scalability – The ontology must be scalable to accommodate large volumes of data as the project grows, ensuring it can handle the increasing number of test scenarios and environmental factors.
Req-Tech-4-7	Modularity – The ontology should be modular to allow individual workgroups to manage and extend specific sections related to their domain.
Req-Tech-4-8	Inference of relations – The ontology must allow for inference of relations between two scenarios (of different abstraction levels).

Table 65: Ontologies requirements.

7 CONCLUSION

This deliverable represents a major milestone in the SYNERGIES project by laying the conceptual and technical foundations for the development and operation of the SYNERGIES Platform. Building upon Tasks T2.2, T2.3, and T2.4, it defines in detail the storylines that capture stakeholder interactions, the technical constraints guiding system development, and the specific interoperability, traceability, and cybersecurity requirements necessary for the platform to operate effectively across diverse user groups and data ecosystems.

A central achievement of this document is the formalization of storylines—structured narratives that map stakeholder needs into actionable system behaviours. These storylines articulate the various interactions between users (e.g., Data Providers, Tool Providers, Scenario Users) and platform components (e.g., Scenario Dataspace, Marketplace, Data Registry), and serve as the basis for functional requirements and high-level architecture. By grounding system specifications in actual user workflows and ecosystem demands, SYNERGIES ensures that platform development remains aligned with operational needs from the outset.

Furthermore, the deliverable introduces a comprehensive set of constraints grouped into four major dimensions: **interoperability**, **traceability**, **data and scenario trustworthiness**, and **cybersecurity**. These constraints are critical for ensuring that the platform not only performs its intended functions but also upholds regulatory and quality expectations.

- **Interoperability** is addressed at the semantic, pragmatic, technical, and data levels. The deliverable defines standardized terminology, metadata schemas, API requirements, and format specifications that enable seamless integration of heterogeneous data and tools within a federated ecosystem.
- **Traceability** requirements emphasize complete lineage tracking, transparent documentation of data transformations, and robust change management systems. This ensures that scenarios and datasets can be traced back to their sources, enhancing credibility, reproducibility, and regulatory compliance.
- **Trustworthiness** is assured through rigorous completeness checks, relevance scoring, expert validation, and accuracy verification. By codifying these measures as requirements, the deliverable enforces high-quality standards for both raw and derived data assets.
- **Cybersecurity** provisions—including access control, encryption, GDPR compliance, and audit mechanisms—ensure that the platform adheres to stringent data protection and privacy obligations, thereby safeguarding sensitive information and ensuring trust among contributors and users.

The technical requirements compiled in this deliverable are directly linked to the storylines and constraints and provide the blueprint for the SYNERGIES Platform's implementation. These include specifications for metadata structures, data and scenario formats, ontologies, interface documentation, and quality descriptors. They will inform the platform design and development activities in subsequent work packages, especially WP3 (Data Collection), WP4 (Data Processing and Tools), WP5 (Scenario Generation), and WP7 (Platform Development and Deployment).

8 REFERENCES

- [1] SYNERGIES, *Grant agreement - Project 101146542*, 2024.
- [2] ISO, *ISO 34501:2022, Road vehicles — Test scenarios for automated driving systems — Vocabulary*, 2022.
- [3] ISO, *ISO/IEC 25012:2008, Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model*, 2008.
- [4] SAE, *SAE J3016:2021 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*.
- [5] FAME, «FAME Glossary and Taxonomies,» [En ligne]. Available: <https://taxonomy.connectedautomateddriving.eu>.
- [6] SYNERGIES, «SYNERGIES Consortium Agreement, version 8, 2024/10/25».
- [7] VVM methods, «Automotive Urban Traffic Ontology (A.U.T.O.),» [En ligne]. Available: <https://github.com/lu-w/auto>.
- [8] GaiaX project, «GaiaX project,» GaiaX ontologies, [En ligne]. Available: <https://github.com/GAIA-X4PLC-AAD/ontology-management-base>.
- [9] levelXData, *The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems*, 2018.
- [10] levelXData, *The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections*, 2020.
- [11] ISO, *ISO 34504:2024, Road vehicles — Test scenarios for automated driving systems — Scenario categorization*, 2024.
- [12] UNECE, *GRVA 18-50e, Guidelines and recommendations for ADS safety requirements, assessments and test methods to inform regulatory development*, 2024.
- [13] ISO, *ISO/IEC 25022:2016, Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE) — Measurement of quality in use*, 2016.
- [14] CE-TRANS-WP.29-2022-1426, *Regulations..*

ABBREVIATIONS

Term	Definition
ADS	Autonomous Driving System
AI	Artificial Intelligence
API	Application Programming Interface
CA	Consortium Agreement
CCAM	Connected Cooperative & Automated Mobility
CDF	Compatible Data Format
E2EE	End-to-End Encryption
EU	European Union
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GRVA	Working Party on Automated/Autonomous and Connected Vehicles
GUID	Globally Unique Identifier
HMI	Human Machine Interaction
IP	Intellectual Property
KPI	Key Performance Indicator
MFA	Multi-Factor Authentication
ODD	Operational Design Domain
SCDB	Scenario Database
SSD	Scenario Source Data
SuT	System under Test
URI	Uniform Resource Identifier
UUID	Universally Unique Identifier
V2X	Vehicle-to-Everything