

# DIN SPEC 91491:2026-04 (E)

## AI-generated data schemas for the integration and consolidation of heterogeneous data sources; Text in English

---

### Contents

	Page
Foreword .....	5
Introduction.....	7
1 Scope.....	8
2 Normative references .....	8
3 Terms and definitions.....	8
4 Technological background and motivation .....	9
4.1 Current challenges in data integration .....	9
4.2 Limitations of existing standards (e.g., OPC-UA, EDI).....	9
4.3 Role of AI, LLMs, and prompting in schema generation.....	10
5 Core principles of this specification .....	10
5.1 General.....	10
5.2 AI-Assisted structuring and interpretation of input data .....	11
5.3 Prompt-enriched schema generation.....	11
5.4 Dynamic schema logic through natural language interfaces .....	11
5.5 Use of a universal intermediate schema.....	11
5.6 Role and relevance of synthetic data .....	12
6 System architecture.....	12
6.1 General.....	12
6.2 System overview.....	12
6.3 Components .....	13
6.3.1 Input parsers .....	13
6.3.2 Feature extractor.....	13
6.3.3 Prompt engine.....	14
6.3.4 Schema generator (LLM Execution Layer) .....	14
6.3.5 Schema evolution manager .....	14
6.3.6 Schema harmonizer .....	15
6.3.7 Schema class registry.....	15
6.3.8 Schema mapper (optional).....	16
6.3.9 Storage layer .....	16
6.4 Data flow model .....	16
7 Process description .....	17
7.1 General.....	17
7.2 Data ingestion.....	17
7.2.1 General.....	17
7.2.2 Implementation scope .....	17
7.2.3 Supported input formats.....	17
7.2.4 Step-by-step specification .....	18
7.2.5 Output of this stage .....	20
7.3 Structuring and feature extraction.....	20
7.3.1 General.....	20
7.3.2 Scope of application.....	20
7.3.3 Step-by-step specification .....	21
7.4 Prompt engineering and execution .....	24
7.4.1 General.....	24

7.4.2	Role in the system .....	24
7.4.3	Step-by-step specification .....	24
7.4.4	Output of this stage.....	28
7.5	Schema generation and mapping.....	28
7.5.1	General .....	28
7.5.2	Position in the pipeline .....	28
7.5.3	Step-by-step specification .....	28
7.5.4	Output of this stage.....	31
7.6	Schema harmonization and canonicalization .....	32
7.6.1	General .....	32
7.6.2	Position in the pipeline .....	32
7.6.3	Step-by-step specification .....	32
7.6.4	Output of this stage.....	35
7.6.5	System requirements and interfaces.....	35
7.7	Storage and usage of output data structures.....	35
7.7.1	General .....	35
7.7.2	Position in the pipeline .....	36
7.7.3	Step-by-step specification .....	36
7.7.4	Output of this stage.....	39
8	Quality criteria and requirements.....	39
8.1	General .....	39
8.2	Accuracy and completeness of generated schemas .....	39
8.3	Robustness and generalization through prompt variation.....	40
8.4	Traceability of schema generation via prompt definitions.....	41
8.5	Modularity and reusability of prompt templates .....	41
8.6	Data privacy, fairness, explainability, and security .....	42
9	Use cases and example applications .....	43
9.1	Financial data consolidation .....	43
9.1.1	General .....	43
9.1.2	Processing workflow and schema generation .....	44
9.1.3	Output and benefits.....	45
9.2	Logistics and vendor matching (e.g., invoice vs. order).....	46
9.3	Sensor data integration and machine telemetry structuring in manufacturing .....	47
9.4	Intelligent traffic management and integration of urban mobility data.....	48
9.5	Automated schema extraction for nested excel structures.....	50
9.6	Synthetic data-driven prototyping and testing .....	51
10	Validation and evaluation methods .....	52
10.1	General .....	52
10.2	Test strategy using synthetic and real-world data.....	52
10.3	Output consistency across prompt templates .....	53
10.4	Schema fidelity compared to human-defined gold standards.....	53
10.5	Robustness testing across data domains.....	53
11	Outlook and Future Development .....	54
11.1	Scalability of prompt-driven schema automation.....	54
11.2	Integration with domain-specific ontologies and taxonomies.....	54
11.3	Potential extensions into official industrial standards.....	54
11.4	Recommendations for further research and collaboration .....	55
Annex A (informative)	Schema change classification logic.....	56
A.1	General .....	56
A.2	Safe changes.....	56
A.3	Moderate changes .....	56
A.4	Critical changes.....	57
A.5	Threshold configuration .....	57
A.6	Example: Schema diff by category .....	57
Annex B (informative)	Reference prompt templates (per document type/use case).....	59

<b>B.1</b>	<b>Financial data consolidation</b> .....	<b>59</b>
<b>B.2</b>	<b>Logistics and vendor matching</b> .....	<b>59</b>
<b>B.3</b>	<b>Sensor data integration in manufacturing</b> .....	<b>59</b>
<b>B.4</b>	<b>Intelligent traffic management and urban mobility data</b> .....	<b>59</b>
<b>B.5</b>	<b>Excel-based schema extraction</b> .....	<b>59</b>
<b>B.6</b>	<b>Synthetic data-driven prototyping and testing</b> .....	<b>60</b>
	<b>Bibliography</b> .....	<b>61</b>

**Tables**

<b>Table 1</b>	<b>— List of supported input formats</b> .....	<b>18</b>
<b>Table 2</b>	<b>— List of parsing tasks</b> .....	<b>21</b>
<b>Table 3</b>	<b>— List of Example Error Handling and Edge Cases</b> .....	<b>24</b>
<b>Table 4</b>	<b>— List of Schema Change Classifications</b> .....	<b>30</b>
<b>Table 5</b>	<b>— Schema reclassification scopes</b> .....	<b>34</b>
<b>Table 6</b>	<b>— List of System Requirements and Interfaces</b> .....	<b>35</b>
<b>Table 7</b>	<b>— Taxonomy for storing schema artifact</b> .....	<b>36</b>
<b>Table 8</b>	<b>— Event Types for Audit Log</b> .....	<b>38</b>