TEMPLATE P16 · DATA AND MODEL LINEAGE

Data and model lineage

Records where the data come from, how they are transformed and which models and versions are involved, so that any output of the system can be explained, reproduced and rolled back.

Fernando García Varela Version 0.1 (working draft) 16-09-2026 Draft for review

IN THIS DOCUMENT

01 Instructions for use

02 Identification

03 Data sources

04 Datasets and transformations

05 Models and components

06 End-to-end journey

07 Reproducibility and traceability of outputs

08 Treatment in the event of rollback or retirement

09 Quality criteria

10 Approval and verification

11 Version control

Legal notice and disclaimer. SEVEN-G is a reference methodological framework provided "as is" and for information purposes only. It does not constitute legal, regulatory, financial or professional advice, nor does it guarantee compliance with any law or standard. References to general regulation (such as the EU AI Act, the GDPR, DORA or NIS2), to technical standards and to regulation specific to each sector or jurisdiction may be incomplete, may not apply to a particular case or may become out of date as a result of regulatory changes, interpretations or supervisory positions after their consultation date. **Each organisation that uses SEVEN-G is solely responsible for identifying the regulation that applies to it, verifying that it is current and certifying its own regulatory compliance**, with the appropriate qualified advice. This methodology is a generic, free aid shared with the community so that nobody has to start from scratch; each person or organisation can and should adapt it to its own use. It must not be inferred that its legally sensitive parts have been reviewed by legal counsel: those reviews, for each company or sector, are the ultimate responsibility of the company, consultant or organisation that uses it. Although every effort is made to keep it up to date, some regulation may have changed without being reflected here. To the fullest extent permitted by law, the author accepts no responsibility whatsoever for the effects of its application in any organisation or for its full applicability. The methodology does not grant certification of any kind. The author accepts no liability for any use made of this content or for any decisions taken on the basis of it. The data, figures, companies and cases in the examples are fictitious or illustrative.

01

Instructions for use

Aspect	Guidance
When it is used	Phase 4 (Solution design), before requesting G4. It is updated in phase 5 with the versions actually deployed and in phase 6 with every change to data, model, prompts or knowledge base.
Who completes it	AI Technical Owner, with the owners of the source data.
Who verifies it	AI Office under Lite; AI Auditor under Enterprise (01 §7.5). The AI Risk Owner reviews the legal basis and the restrictions on use.

Aspect	Guidance
At which *gate* it is reviewed	G4. It is checked again at G5 (deployed versions) and at every R6 continuity review.
Lite versus Enterprise	Lite may omit the fields marked **(Enterprise)**. If the system is classified as high-risk, all fields are completed regardless of intensity.
Relationship	P05 Inventory record · P11 Regulatory classification · P15 Architecture record · P19 Rollback plan · P27 Incident and change log · P30 Scale or retire decision · P64 Dataset record and knowledge source inventory · documents 51 and 53.

Completion rules:

- Each source, dataset, model and component has an **identifier, version and owner**. Generic descriptions such as "customer data" are not accepted: the table, file, API or service is identified.
- Third-party models, including general-purpose models consumed via API, are recorded with the supplier, commercial name and exact version.
- In generative AI and agents, **system prompts, knowledge bases and connected tools** are versioned components, just like a model.
- Every version change in production is recorded as a change in P27 and updates this template.
- "No data" is not zero: if a field is not known, *No data* and the reason are stated.

This template does not constitute legal advice. The legal basis for processing must be validated by data protection.

02

Identification

Field	Content	Guidance
Use case code		IA-AAAA-NNN from the use case register (T01).
Use case name		Short, understandable name.
AI system(s)		Identifier in the AI system inventory (T02, P05).
Intensity		Lite · Enterprise (P04).
Regulatory classification		Prohibited · High risk · Transparency obligations · Minimal risk · Out of scope · Pending classification (P11).

Field	Content	Guidance
Document version		0.1, 0.2…; 1.0 upon approval. Each change in production increments the version.
Author		Role and name.
Date		DD-MM-AAAA.
Document status		Draft · Submitted for verification · Verified · Approved · Superseded.

03

Data sources

One row per source. **Data type:** Personal · Special categories · Confidential · Internal · Public · Synthetic. **Legal basis (GDPR Art. 6):** Consent · Contract · Legal obligation · Vital interests · Public interest · Legitimate interest · Not applicable (no personal data).

ID	Source and source system	Data owner	Data type	Legal basis	Restrictions on use	Update frequency	Quality verified (Yes · No · Partial)
F-01 ***(illustrative example)***	Maintenance incident history, work order table of the asset management system	Head of maintenance	Internal	Not applicable	Internal use only; not shared with suppliers	Daily	Partial: 4 % of records without a fault code

Supplementary field	Content	Guidance
Third-party or acquired data		Supplier, licence and conditions of use for AI.
Transfers outside the European Economic Area (Enterprise)		Destination, safeguard applied and reference to the supplier assessment (P14).
Data protection impact assessment		Carried out · Not required (with justification) · Pending. Link to P11.

04

Datasets and transformations

Use: Training · Fine-tuning · Validation · Testing · Knowledge base (retrieval) · Inference · Monitoring.

Dataset ID	Use	Sources (ID)	Main transformations	Version or fingerprint	Period covered	Volume	Bias and representativeness controls (Enterprise)	Lo an ret
CD-01 ***(illustrative example)***	Training	F-01, F-02	Removal of duplicates; imputation of dates; pseudonymisation of the technician	v3 · fingerprint recorded in the data repository	01-2023 to 12-2025	180,000 records	Distribution by plant and asset type compared with the population	An rep 5 y

05

Models and components

Model type: In-house trained · In-house fine-tuned on a base model · Third-party via API · Third-party deployed on own infrastructure · General-purpose model · Rules (not AI). **Status:** In development · Validated · In production · Retired.

Model ID	Type	Supplier and exact version	Datasets used	Purpose	Reference metrics	Known limitations	Status
M-01 ***(illustrative example)***	In-house trained	Internal · v1.4	CD-01, CD-02	Estimate the probability of failure within 30 days	Discrimination and calibration metrics recorded in P22	Poor performance on assets with less than 6 months of history	Validated

Generative AI and agent components (if applicable):

Component	Identifier and version	Owner	Location	What a change requires
System prompts				Regression testing and recording in P27.
Knowledge base				Content review, cut-off date and access permissions.
Agent's connected tools				Permissions review in P18.
Generation parameters (Enterprise)				Regression testing.
Cost fallback model (document 52 §10.3), if planned				Provider and exact version; evaluated before G5 with the reference set and the bias tests of the primary model. Adding it or changing its conditions is a change (document 52 §6.4).

06

End-to-end journey

Describe, in order, how a source data item becomes an output of the system. One row per step; no text diagrams.

Step	Input	Process	Output	Record that evidences it
1 ***(illustrative example)***	F-01	Nightly extraction	Dated raw copy	Load process log
2 ***(illustrative example)***	Raw copy	CD-01 transformations	CD-01 v3	Dataset fingerprint
3 ***(illustrative example)***	Data for the day	Inference with M-01 v1.4	Prioritised list of assets	Inference log with model version

07

Reproducibility and traceability of outputs

Field	Content	Guidance
Can each output be linked to the version of the model, data and prompts?		Yes · Partially · No. If not, it is a finding for G4.
Automatic logs retained		What is logged (input, output, version, date, user). In high-risk systems, align with the record-keeping obligations of the EU AI Act (document 34).
Log retention period		Period set by the company, not shorter than that required by applicable regulation.
Ability to retrain or regenerate a previous version (Enterprise)		What is kept to reconstruct a version (code, data, parameters).
Reference for detecting drift		Dataset and period used as the reference for monitoring (P25). For generative AI and agents, usage reference: topics, intents and scope validated at G5 (document 52 §4.2.6).

08

Treatment in the event of rollback or retirement

Field	Content	Guidance
Previous stable version available		Identifier of the model and components to be reverted to (P19).
Data that must be retained		Due to legal obligation, audit or legal defence.
Data that must be deleted or anonymised		Includes copies held by suppliers.
Supplier obligations on termination (Enterprise)		Return, certified deletion, no reuse for training. Detailed in P30.

09

Quality criteria

The verifier checks, as a minimum, the following. The formal G4 criteria are in document 21 and checklist LV-G4 in document 22.

#	Check	Status (Met · Not met · Not applicable · Pending)
1	All sources have an owner, data type and legal basis where there are personal data.	
2	Each dataset has a version or fingerprint and a period covered.	
3	Each model and component has an exact version, including third-party ones.	
4	The end-to-end journey makes it possible to link an output to its versions.	
5	Known limitations are declared and carried over to P17 and P24.	
6	There is an identified previous version for rollback.	
7	The template existed and had been verified before G4 was requested.	

10

Approval and verification

Function	Role	Name	Date	Signature or reference
Author	AI Technical Owner			
Legal basis review	Data protection or AI Risk Owner			
Verifier	AI Office (Lite) · AI Auditor (Enterprise)			
***Gate* decision-maker**	According to 01 §7.5 (recorded in P29)			

Segregation of duties: the verifier is not part of the team that builds and is not the same person as the author (01 §7.4 and §8.2).

11

Version control

Version	Date	Changes
0.1	16-09-2026	First version of the template.
0.1	18-09-2026	Cost fallback model in the components and usage reference for drift in generative AI and agents (document 52 §4.2.6 and §10.3).

External sources cited

Links to the official or primary source, checked before publication. Always verify the version in force on the date of use.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) · https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng · Parlamento Europeo y Consejo de la UE · 27-04-2016

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) · https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng · Parlamento Europeo y Consejo de la UE · 13-06-2024