Data Contract Checklist: Your Blueprint for Trustworthy Data

The Data Product Owner (DPO) sits at the center of the Data Contract, translating business needs into reliable technical guarantees.

I. Defining the Product (For the DPO)

This section ensures the contract aligns with business value and audience needs.

- **Target Audience & Purpose:** Clearly state *who* the consumer is (analysts, AI agents, marketing team) and the **specific, limited business purpose** this data product is designed to support (e.g., *Personalized Offer Generation*).
- **Cost & Value Agreement:** Establish the cost or budget mechanism (chargeback) for consuming the data, formalizing the data product's value.
- **Communication & Deprecation Plan:** Define the official channel and policy for notifying consumers (and AI agents) about upcoming **major changes (breaking changes)**, ensuring a minimum grace period for migration.
- **Ownership & Accountability:** Clearly name the Data Product Owner and the Producer Team responsible for maintenance, support, and adherence to the contract terms.

II. Producer Guarantees (For the Engineers)

These are the technical warranties the Producer Team provides to ensure the data is reliable and ready.

- **Schema & Structure Guarantee:** Lock down the precise schema (field names, data types, and required formats). Use versioning (e.g., v1.2) to manage changes and ensure the pipeline breaks if the contract is violated.
- **Fitness for Use Metrics (Quality SLA):** Define measurable, non-negotiable standards for data quality. Specify thresholds for **Accuracy, Completeness, and Validity** (e.g., "Customer ID field must be 99.9% non-null").
- **Timeliness & Freshness SLA:** Define commitments for when and how often the data will be delivered. Specify the **Latency/Freshness** (e.g., "Data available within 5 minutes of transaction completion") and the scheduled **Update Frequency**.
- **Explicit Lineage & Provenance:** List the **exact upstream sources and versions** used to create this data product, documenting the transformation logic. This creates an auditable trail for troubleshooting and compliance.

III. Consumer Commitments (For the Business Users & Agents)

These are the guardrails and usage rules the Consumer must agree to for safe and compliant operation.

- **Purpose Limitation & Processing Permissions:** Consumers (including AI agents) must agree to use the data **only** for the explicitly defined purpose. Prohibit any use that violates the **Legal Basis** (GDPR) or **Service Provider** restrictions (CCPA).
- **Security & Scoped Access:** Define the security classification of the data (e.g., PII, Sensitive) and enforce strict access policies. **AI Agents** must be restricted to the **Principle of Least Privilege**, mapping their permissions directly to the human user's authorization.
- **Usage Monitoring & Auditability:** The consumer acknowledges and accepts that their access (and the access of any processes / agents via API / MCP) will be comprehensively logged for security and audit purposes, allowing for full traceability of all data actions.

Data Contract Example

Member 360 Health Engagement Data Product

Data Product Owner: Chief Medical Officer (CMO) / Member Engagement Team

Goal: To provide a single, trustworthy, and compliant view of member health status and contact data for targeted, high-impact engagement to improve health outcomes (e.g., HEDIS scores, care gap closures).

I. Producer Guarantees: Data Fitness & Timeliness (The Engineer's Commitment)

The Producer Team warrants that the data product meets the following automated, measurable specifications:

Component	Description & Requirement	Fitness/Timeliness Example
Schema & Structure	**Data is stable and machine-readable.** Guaranteed schema (JSON/Avro) with explicit, non-nullable fields for primary keys. All PII is tokenized or hashed.	**Member ID:** member_id_hashed (UUID, non-nullable). **PCP Status:** pcp_in_network_status (Boolean, non-nullable). **Contact:** member_email_hashed (masked PII).
Data Lineage & Provenance	**Transparency of sources.** The producer explicitly lists and commits to using only the following upstream data sources. Any change requires a Major Contract Version bump.	**Internal Sources:** Core Claims System (v2.1), Member Portal Activity Feed (v1.5), PCP Network Directory (v3.0). **External Sources:** NPI Registry (v4.2).
Accuracy & Validity (Fitness)	**Clinical and financial correctness.** Data must accurately reflect real-world status to prevent harm to the member or incorrect billing.	**Accuracy SLA:** Claims data must match source system totals with >99.99% fidelity. **Validity Check:** All CPT/ICD codes must conform to the current year's AMA standards. **In-Network:** The pcp_in_network_status field must match the source directory within 1 day.
Timeliness & Freshness SLA	**Data is current enough for clinical action.** Guaranteed data latency and update frequency to ensure clinical relevance.	**Critical Claims Data:** Available in the data product within **T+24 hours** of processing. **Member Portal Activity:** Available in the data product within **T+15 minutes** (for real-time engagement services).

II. Consumer Commitments: Proper Use & Compliance (The Consumers' Obligation)

The Consumer Team (including any downstream service or AI Agent) explicitly commits to the following restricted processing rules:

Area of Proper Use	Contract Component	Restriction / Enforcement
HIPAA/Privacy Rule	**Legal Basis for Processing**	The data (including ePHI) shall be used solely for **Treatment, Payment, and Healthcare Operations (TPO)** as defined under HIPAA. Any usage outside of TPO is a Major Contract Violation.
Purpose Limitation	**Usage Scope & Intent**	Usage is restricted to **Care Management, Quality Outreach, and Operational Efficiency.** Consumers are prohibited from using the data for non-TPO commercial purposes (e.g., selling member lists, unrelated cross-marketing).
AI/MCP Restrictions	**Data Agent Guardrails**	AI Agents (MCP Clients) must not attempt to **de-anonymize or link PII** beyond what is necessary for the defined purpose. The agent's session and all intermediate logs must be run within a **secure, auditable enclave** that adheres to the **Scoped Authorization** defined by the security policy.
Data Retention	**Lifecycle Policy**	Data may only be stored for **6 years** post-termination of the member relationship, or as otherwise required by state law. Consumers must ensure their copies and caches comply with this retention policy.
Audit & Accountability	**Traceability Mandate**	The Consumer must log every successful query and action taken by their application/agent, linking the action back to the **human user's identity** who authorized the underlying use case (e.g., linking the agent action to the responsible CMO analyst).