


# SOL ENGINEERING SERVICES

The Daniel Group LLC

---

## SUPPORTING DOCUMENTATION PACKAGE

### ALICE: Adaptive Labor Intelligence and Capacity Engine

SBIR Topic DLA26BZ03-NV011 | Defense Logistics Agency

## ARTIFACT 4 OF 11

# Data Catalog

Document Version 1.0 | July 2026

**Status: DRAFT — Pending Gate-4 Stakeholder Review**

**Classification: Unclassified**

## Document Control

| Field | Value |
|---|---|
| Artifact ID | SDP-ART-04 |
| Artifact Title | Data Catalog |
| Artifact Dependencies | SDP-ART-01 (Research Log) — Approved; SDP-ART-02 (ADR Register) — Approved; SDP-ART-03 (Component Catalog) — Approved Gate-3 |
| Package | ALICE Supporting Documentation Package (SDP) |
| Source Documents | ALICE TNA v1.0 Sections 12, 13, 16, 17; ALICE PRD v1.0 Section 17; SDP-ART-02; SDP-ART-03 |
| Prepared By | SOL Engineering Services \| The Daniel Group LLC |
| Prepared For | Defense Logistics Agency \| SBIR DLA26BZ03-NV011 |
| Version | 1.0 (Draft) |
| Date | July 2026 |
| Classification | Unclassified |
| Status | Draft — Pending Gate-4 stakeholder approval before Artifact 5 is produced |
| Entity Count | 12 data entities (DE-01 through DE-12) across four data zones |
| ID Convention | All IDs preserved from prior approved artifacts. Entity IDs (DE-XX) are introduced in this artifact and will be carried forward through all remaining artifacts. |

## 1. Purpose and Scope

This Data Catalog is Artifact 4 of the ALICE Supporting Documentation Package. It is the authoritative reference for all data entities in the ALICE system, providing the governance information required for data stewardship, privacy review, RMF authorization, CMMC assessment, and Phase I and Phase II implementation planning.

For each of the twelve data entities identified in TNA Section 12 and PRD Section 17, this catalog provides: a stable entity identifier (DE-01 through DE-12); the data zone in which the entity resides; its sensitivity classification; its authoritative source; its business and technical description; its complete key attribute set drawn from PRD Section 17.1; its data owner and data steward; its create, read, update, and delete access rights by component; its retention schedule with regulatory basis; its data quality validation rules; its data lineage chain; its disposal requirements; and its privacy and compliance obligations. The catalog concludes with the data zone summary, data governance rules, data lifecycle diagram, and data lineage register.

Items labeled [INFERRED] represent data architecture requirements reasoned from PRD and CONOPS content not explicitly stated but necessary for governance completeness. Items labeled [OPEN] identify unresolved questions requiring stakeholder resolution before Phase II design.

## 2. Data Entity Index

### 2.1 Entity Summary

| Entity ID | Entity Name | Zone | Sensitivity | Source System | Retention | Privacy Act Relevant | Phase I Required? |
|---|---|---|---|---|---|---|---|
| DE-01 | Position | Processed | CUI | DCPDS / DCHRMS | System operation + 3 years | No — organizational construct | Yes |
| DE-02 | Role | Processed / Analytical | CUI | ALICE-constructed (SME-validated) | System operation + 3 years | No — organizational construct | Yes |
| DE-03 | Person-Position Link (de-identified) | Processed | CUI — de-identified; no Privacy Act record | DCPDS / DCHRMS (de-identified at ingestion) | System operation + 3 years | Yes — derived from Privacy Act source; de-identified transformation is primary control | Yes |
| DE-04 | Certification | Processed | CUI | DLA Training Management System | System operation + 3 years | No | Yes |
| DE-05 | Certification Currency Record | Processed | CUI — de-identified | DLA TMS (de-identified at ingestion) | System operation + 3 years | Yes — derived from Privacy Act source; de-identified | Yes |
| DE-06 | Workload Metric | Processed | CUI | EBS (SAP DLA) | System operation + 3 years | No | Phase I if |

| Entity ID | Entity Name | Zone | Sensitivity | Source System | Retention | Privacy Act Relevant | Phase I Required? |
|---|---|---|---|---|---|---|---|
| | | | | | | | EBS access available; OPM FWD proxy other |

| Entity ID | Entity Name | Zone | Sensitivity | Source System | Retention | Privacy Act Relevant | Phase I Required? |
|---|---|---|---|---|---|---|---|
| | | | | | | | wise |
| DE-07 | Readiness Indicator | Analytical | CUI — internal analytical output | System-computed (ALICE) | System operation + 3 years | No | Yes |
| DE-08 | Synthetic Person | Synthetic | CUI by default pending PIA | ALICE Synthetic Engine (generated) | Default 12 months per generation cycle | Derived — not a Privacy Act record; CUI pending PIA | Yes |
| DE-09 | Scenario Run | Analytical | CUI | ALICE Simulation Engine (generated) | 3 years | No | Yes |
| DE-10 | Synthetic Data Export | Analytical / Audit | CUI — governance record | ALICE Export Workflow (generated) | 7 years | Yes — governance record for Privacy Act-derived synthetic data | Yes |
| DE-11 | Audit Log Entry | Audit | CUI — sensitive | All ALICE services (generated) | 7 years minimum | No — event records; no workforce data content | Yes |
| DE-12 | Generation Metadata | Synthetic / Audit | CUI | ALICE Synthetic Engine (generated) | Same as associated synthetic population + audit retention | Yes — compliance artifact for Privacy Act-derived data | Yes |

### 2.2 Data Zone Distribution

| Zone | Entities Residing in Zone | Persistence | Classification | Primary Access Controls |
|---|---|---|---|---|
| Raw Zone (Zone 1) | None (source records are transient in-flight; not assigned a DE-ID because they are never persisted in ALICE) | None — transient in-flight only | Privacy Act-protected (source records); in-flight CUI handling applies | LC-02 in-flight only; never written to storage; no zone credentials required |
| Processed Zone (Zone 2) | DE-01 (Position); DE-02 (Role, partial); DE-03 (Person-Position Link); DE-04 (Certification); DE-05 (Certification Currency Record); DE-06 (Workload Metric) | Persistent — retained for system operation + 3 years audit | CUI — de-identified; no Privacy Act records | Write: LC-02 only. Read: LC-04, LC-05. No access for LC-06 or LC-07. |
| Synthetic Zone (Zone 3) | DE-08 (Synthetic Person); DE-12 (Generation Metadata) | Persistent — 12-month default per generation; expired populations flagged | CUI by default pending formal PIA (ADR-007) | Write: LC-05 only (on generation). Read: LC-05 (export workflow), LC-06 (simulation). Data Steward: read for export authorization. |
| Analytical Zone (Zone 4) | DE-02 (Role, partial); DE-07 (Readiness Indicator); DE-09 (Scenario Run); DE-10 (Synthetic Data Export, partial) | Persistent — retention varies by entity (3 to 7 years) | CUI — no individual-identifiable data | Write: LC-04 (ontology/indicators), LC-06 (simulation results). Read: LC-07. No Raw or Processed Zone data. |
| Audit Log Store | DE-11 (Audit Log Entry); DE-10 (Synthetic Data Export, governance copy); DE-12 (Generation Metadata, compliance copy) | Persistent — 7 years minimum; non-deletable | CUI — sensitive; highest protection within ALICE | Write: LC-09 only (append-only). Read: System Administrator, designated Auditor. No delete capability for any role. |

## 3. Data Entity Detail Records

| DE-01: Position | | | |
|---|---|---|---|
| **Zone:** Processed | **Sensitivity:** CUI | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | An authorized billet within the DLA organizational structure. The fundamental node of the ALICE organizational ontology. Each Position record represents a single authorized duty position, independent of whether it is currently filled. Positions are organizational constructs, not records about individuals. |
| **Authoritative Source** | Defense Civilian Personnel Data System (DCPDS) in Phase I. Defense Civilian Human Resources Management System (DCHRMS) upon migration. Integration: INT-01. |
| **Key Attributes** | Position ID — unique identifier assigned by DCPDS/DCHRMS.<br>Occupational series — OPM occupational series code.<br>Grade — pay grade (GS/GG/WG or equivalent).<br>Duty station — geographic location code.<br>Supervisory indicator — whether the position is supervisory, managerial, or non-supervisory.<br>Mission function — ALICE-assigned mission function area classification.<br>Fill status — current fill state: filled, vacant, or partially filled.<br>Required certifications — list of Certification IDs (DE-04) required for this position.<br>[INFERRED from RL-06] Five leading degradation indicator attributes: vacancy duration (days); criticality rating; knowledge network density score; span-of-control metric; certification currency ratio. |
| **Data Owner** | DLA (position management authority). ALICE holds a read-only transformed copy. |
| **Data Steward** | ALICE Data Steward (transformed copy); DLA Human Resources (authoritative source). |
| **Producing Component** | LC-03 (schema translation from DCPDS/DCHRMS); LC-02 (de-identification and zone write). LC-04 enriches with ALICE-derived attributes (mission function, criticality rating). |
| **Consuming Components** | LC-04 (primary consumer — ontology construction); LC-05 (structural scope for generative model); LC-07 (position risk register display). |
| **CRUD Rights** | Create: LC-02 (on ingestion). LC-04 (ALICE-derived attributes).<br>Read: LC-04, LC-05.<br>Update: LC-02 (on delta ingestion). LC-04 (derived attribute refresh).<br>Delete: Automated retention enforcement only. No user-initiated deletion. |
| **Data Classification** | CUI. No individual-identifying data. Position is an organizational construct. |
| **Retention Schedule** | System operation plus 3 years post-decommission for audit purposes. Automated retention policy enforcement required (PRD Section 17.2). |
| **Data Quality Rules** | Position ID must be unique within the ALICE Processed Zone.<br>Occupational series must match OPM occupational series code list.<br>Grade must be within the valid DoD pay grade enumeration.<br>Fill status must be one of: filled, vacant, partially filled.<br>Mission function assignment must be validated against the configured mission function taxonomy.<br>Schema validation against the ALICE canonical data model must pass before write (enforced by LC-02). |
| **Privacy and Compliance** | Not a Privacy Act record. Position is an organizational construct. De-identification is not required for this entity, but the entity must never be |

| | |
|---|---|
| | linked to an identifiable individual within the ALICE environment. |
| **Data Lineage** | DCPDS/DCHRMS → LC-03 (schema translation) → LC-02 (canonical validation, zone write) → Processed Zone → LC-04 (ontology construction and attribute enrichment). |
| **Disposal** | Secure deletion per NIST SP 800-88 guidelines on system decommission. |
| **Related Information Flows** | IF-01 (source ingestion); IF-06 (de-identified write to Processed Zone); IF-07 (Processed Zone to Ontology Service). |
| **Related PRD Requirements** | REQ-F-001; REQ-F-007; REQ-F-008; PRD Section 17.1 (Position entity definition). |
| **Open Questions** | [OPEN] OQ-001: pilot sub-component selection determines the initial position count. ASM-003 requires minimum 300 positions for synthetic generation viability. |

## DE-02: Role

| | | | |
|---|---|---|---|
| **Zone:** Processed / Analytical | **Sensitivity:** CUI | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A functional capability definition associated with one or more Positions (DE-01). Roles are ALICE-constructed abstractions that model the functional dependencies between positions: which roles are prerequisites to which workflows, where single-point-of-failure conditions exist, and how workload is distributed across functional teams. Roles do not exist in source systems; they are defined by the ALICE ontology construction process with subject matter expert validation. |
| **Authoritative Source** | ALICE-constructed by the ontology team through SME interviews, organizational process documentation review, and workflow analysis. No direct source system equivalent. |
| **Key Attributes** | Role ID — ALICE-assigned unique identifier.<br>Function description — narrative description of the functional capability.<br>Workflow dependencies — list of prerequisite Role IDs and dependent Role IDs.<br>Criticality rating — SME-validated rating of the role's mission criticality (e.g., critical, high, moderate, standard).<br>Single-point-of-failure flag — boolean indicator whether this role represents a single point of failure in the workflow graph.<br>Associated position IDs — list of DE-01 Position IDs that carry this role. |
| **Data Owner** | ALICE Data Steward (ALICE-constructed entity with no external authoritative source). |
| **Data Steward** | ALICE Data Steward; validated by DLA subject matter experts. |
| **Producing Component** | LC-04 (Organizational Ontology Service — constructs and maintains the role graph). |
| **Consuming Components** | LC-04 (self-referential — reads role graph for readiness computation); LC-06 (agent behavioral rule basis); LC-07 (ontology visualization for scenario setup). |
| **CRUD Rights** | Create: LC-04 (during ontology construction, Phase I). |

| | |
|---|---|
| | Read: LC-04 (internal graph traversal); LC-06 (agent model basis); LC-07 (display).<br>Update: LC-04 (on SME validation revision or source data update).<br>Delete: Controlled process; requires SME confirmation and audit log entry. |
| **Data Classification** | CUI. Organizational construct; no Privacy Act implications. |
| **Retention Schedule** | System operation plus 3 years post-decommission. Historical versions archived for trend analysis. |
| **Data Quality Rules** | Role ID must be unique within the ALICE ontology.<br>Workflow dependency references must resolve to existing Role IDs (no dangling references).<br>Criticality rating must use the defined enumeration; free-text ratings are not permitted.<br>Single-point-of-failure flag derivation must be traceable to a defined graph analysis algorithm documented in the ontology specification.<br>All role definitions must be validated by a minimum of one DLA subject matter expert per pilot sub-component (PRD Section 12.2 MVP acceptance criteria). |
| **Privacy and Compliance** | Not a Privacy Act record. No individual data. SME validation documentation must be retained as a Phase I deliverable artifact. |
| **Data Lineage** | SME interviews + organizational process documentation → LC-04 ontology construction → Processed Zone (role graph) → LC-04 readiness computation → Analytical Zone (readiness indicators). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. |
| **Related Information Flows** | IF-07 (Processed Zone to Ontology Service — read); IF-12 (Ontology Service to Decision Support — role-scoped read). |
| **Related PRD Requirements** | REQ-F-007; REQ-F-008; PRD Section 17.1 (Role entity definition). |
| **Open Questions** | [OPEN] OQ-001: SME availability for role validation depends on pilot sub-component selection. [INFERRED] The role ontology specification — defining the algorithm for single-point-of-failure detection and criticality rating assignment — is a required Phase I design artifact not explicitly specified in the PRD. |

## DE-03: Person-Position Link (de-identified)

| | | | |
|---|---|---|---|
| **Zone**: Processed | **Sensitivity**: CUI — de-identified; no Privacy Act record | **Phase**: Phase I and II | **Type**: Source-Derived |

| | |
|---|---|
| **Description** | A tokenized association of a de-identified workforce member to an authorized position (DE-01). This entity is the central privacy-sensitive data element in the ALICE Processed Zone. It is derived from Privacy Act-protected DCPDS personnel records through a documented de-identification transformation at the ingestion boundary. It carries no direct identifiers (no name, Social Security Number, date of birth, or other field that would identify a specific individual). It enables readiness modeling and synthetic population generation without constituting a Privacy Act record. |

| | |
|---|---|
| **Authoritative Source** | DCPDS (Phase I) / DCHRMS (Phase II migration) and DPMAP (performance tier attribute). Both sources contribute to this entity. Integration: INT-01 (position data), INT-03 (aggregate performance tier only — no individual performance records). |
| **Key Attributes** | De-identified record ID — ALICE-assigned tokenized identifier replacing the DCPDS employee ID. The mapping from employee ID to token is maintained only by the de-identification service and is never stored in the ALICE analytical environment.<br>Position ID — foreign key to DE-01.<br>Grade — pay grade.<br>Series — occupational series code.<br>Tenure band — aggregated tenure category (e.g., 0-2 years, 3-5 years, 6-10 years, 11+ years). Not a specific hire date.<br>Certification currency flags — list of boolean currency indicators per required certification for the associated position.<br>Performance tier (aggregated) — aggregate performance tier category derived from DPMAP aggregate distribution data. Not an individual performance rating. |
| **Data Owner** | DLA (conceptual data owner of the underlying personnel records). ALICE holds a de-identified transformed copy. |
| **Data Steward** | ALICE Data Steward (transformed copy). DLA Human Resources (conceptual authority). |
| **Producing Component** | LC-03 (schema translation); LC-02 (de-identification transformation — this is the critical privacy control for this entity). |
| **Consuming Components** | LC-04 (readiness computation — reads de-identified person-position linkages); LC-05 (synthetic generation — training input). |
| **CRUD Rights** | Create: LC-02 only (on validated ingestion).<br>Read: LC-04 (readiness computation); LC-05 (generative model training).<br>Update: LC-02 (on delta ingestion — logical update of fill status and currency flags).<br>Delete: Automated retention policy only. No user-initiated deletion. |
| **Data Classification** | CUI — de-identified. This entity is not a Privacy Act record within ALICE because it carries no individual-identifying fields. However, it derives from Privacy Act-protected source records. The de-identification transformation is the primary privacy control. If the de-identification transformation is compromised, this entity may become a Privacy Act record. |
| **Retention Schedule** | System operation plus 3 years post-decommission for audit purposes. |
| **Data Quality Rules** | De-identified record ID must not correspond to any field in the source DCPDS/DCHRMS record that could identify an individual. Token generation must use a one-way or cryptographically sound method.<br>The token mapping table (employee ID to token) must never reside in the ALICE Processed, Synthetic, or Analytical Zones.<br>Tenure band must be a defined aggregate category, not a specific date or exact years-of-service value.<br>Performance tier must be an aggregate category derived from DPMAP aggregate distributions, not an individual rating. Individual DPMAP records must never be ingested (REQ-COMP-002; DoDI 1400.25 constraint).<br>Certification currency flags must be boolean (current/not current) derived from DE-05 Certification Currency Records — not raw completion dates. |

| | |
|---|---|
| | PII scan must confirm absence of: name fields; SSN or partial SSN; date of birth; home address; personal email; phone number; and any other direct identifier before zone write. |
| **Privacy and Compliance** | 5 U.S.C. § 552a (Privacy Act of 1974): The underlying source records are Privacy Act-protected. The de-identification transformation in LC-02 is the mechanism by which ALICE resolves the Privacy Act constraint. This transformation must be formally documented and validated.<br>5 C.F.R. Part 293: The de-identified extract must be produced or confirmed as de-identified by DLA before delivery to SOL's environment (RL-21).<br>DoDI 1400.25, Vol. 431: Performance tier data derived from DPMAP must be aggregate only (REQ-COMP-002; REQ-INT-003).<br>The de-identification specification (field removal list, tokenization method, aggregation approach) is a required Phase I design artifact and must be reviewed by the DLA Privacy Officer before Phase I ingestion begins. |
| **Data Lineage** | DCPDS/DCHRMS personnel record (Privacy Act-protected) → DLA de-identification (pre-delivery obligation per data use agreement) → LC-03 (schema translation to canonical model) → LC-02 (PII scan validation, zone write) → Processed Zone → LC-04 (readiness computation) / LC-05 (synthetic generation training input). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. Token mapping (if retained by DLA for re-identification capability) is subject to DLA's own data disposal obligations. |
| **Related Information Flows** | IF-01 (source ingestion); IF-06 (de-identified write to Processed Zone); IF-07 (Processed Zone to Ontology); IF-08 (Processed Zone to Synthetic Engine). |
| **Related PRD Requirements** | REQ-F-001; REQ-F-003; REQ-F-006; REQ-NF-008; REQ-COMP-001; REQ-COMP-002; PRD Section 17.1 (Person-Position Link entity definition). |
| **Open Questions** | [OPEN] OQ-002: Data use agreement must specify which party performs de-identification and how it is validated before delivery. [INFERRED] The de-identification specification document is a required pre-ingestion artifact and must be jointly approved by DLA Privacy Officer and SOL Privacy Officer before any DCPDS data flows to the ALICE environment. |

## DE-04: Certification

| | | | |
|---|---|---|---|
| **Zone:** Processed | **Sensitivity:** CUI | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A training or qualification credential definition required by one or more Positions (DE-01). Certifications are organizational reference data that describe what qualifications are required, who issues them, and when they expire. They are not records about individuals; they are definitions that are linked to individuals through DE-05 (Certification Currency Records). |
| **Authoritative Source** | DLA Training Management System. Integration: INT-02. |
| **Key Attributes** | Certification ID — unique identifier assigned by DLA TMS.<br>Certification name — human-readable name of the credential.<br>Issuing authority — the organization or body that issues the credential.<br>Expiration logic — whether the certification expires, and if so, the expiration period in months. |

| | |
|---|---|
| | Applicable positions — list of DE-01 Position IDs for which this certification is required. |
| **Data Owner** | DLA (DLA Training Management System authority). |
| **Data Steward** | ALICE Data Steward (transformed copy). |
| **Producing Component** | LC-03 (schema translation); LC-02 (validation and zone write). |
| **Consuming Components** | LC-04 (certification currency computation for readiness indicators); LC-07 (position risk register — certification currency display). |
| **CRUD Rights** | Create: LC-02 (on initial ingestion of DLA TMS data).<br>Read: LC-04; LC-07.<br>Update: LC-02 (on delta ingestion — when TMS updates certification definitions).<br>Delete: Automated retention policy only. |
| **Data Classification** | CUI. No individual data. Certification definitions are organizational reference data. |
| **Retention Schedule** | System operation plus 3 years post-decommission. |
| **Data Quality Rules** | Certification ID must be unique within the ALICE Processed Zone.<br>Expiration logic must specify a defined period in months or 'non-expiring'; ambiguous values are rejected.<br>Applicable positions list must reference valid DE-01 Position IDs.<br>Schema validation against the ALICE canonical data model must pass before zone write. |
| **Privacy and Compliance** | Not a Privacy Act record. No individual data. No special handling beyond CUI baseline. |
| **Data Lineage** | DLA TMS export → LC-03 (schema translation) → LC-02 (validation, zone write) → Processed Zone → LC-04 (certification currency computation). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. |
| **Related Information Flows** | IF-02 (DLA TMS ingestion); IF-06 (Processed Zone write); IF-07 (Processed Zone to Ontology). |
| **Related PRD Requirements** | REQ-F-002; REQ-F-007; REQ-F-009; PRD Section 17.1 (Certification entity definition). |
| **Open Questions** | [OPEN] OQ-008: DLA TMS extract format must be confirmed before LC-03 schema mapping for certifications can be finalized. |

## DE-05: Certification Currency Record

| | | | |
|---|---|---|---|
| **Zone**: Processed | **Sensitivity**: CUI — de-identified | **Phase**: Phase I and II | **Type**: Source-Derived |

| | |
|---|---|
| **Description** | A tokenized record linking a de-identified Person-Position Link (DE-03) to a specific Certification (DE-04) and that certification's current completion and currency status. This entity enables ALICE to model whether the workforce population currently meets certification requirements, supporting readiness indicator computation and synthetic population generation. It is derived from Privacy Act-protected DLA TMS training completion records through the same de-identification transformation |

| | |
|---|---|
| | applied to DE-03. |
| **Authoritative Source** | DLA Training Management System (certification completion records, de-identified at ingestion). Integration: INT-02. |
| **Key Attributes** | Record ID — ALICE-assigned tokenized identifier.<br>Person-Position Link ID — foreign key to DE-03 (tokenized, not the source employee ID).<br>Certification ID — foreign key to DE-04.<br>Completion date — [INFERRED: must be expressed as a relative period or date-only field, not a precise datetime that could be combined with other fields for re-identification. The de-identification specification must address this.]<br>Expiration date — derived from completion date and expiration logic in DE-04.<br>Currency status — computed boolean: current (within valid period) or not current. |
| **Data Owner** | DLA (DLA TMS authority for source records). ALICE holds de-identified transformed copy. |
| **Data Steward** | ALICE Data Steward. |
| **Producing Component** | LC-03 (schema translation); LC-02 (de-identification and zone write). |
| **Consuming Components** | LC-04 (certification currency computation for readiness indicators and position risk register); LC-05 (certification profile attribute in synthetic generation). |
| **CRUD Rights** | Create: LC-02 (on DLA TMS ingestion).<br>Read: LC-04; LC-05.<br>Update: LC-02 (on weekly delta — currency status changes as certifications expire or are renewed).<br>Delete: Automated retention policy only. |
| **Data Classification** | CUI — de-identified. Same privacy control logic as DE-03: derived from Privacy Act-protected training records; de-identification transformation is the primary control. |
| **Retention Schedule** | System operation plus 3 years post-decommission. |
| **Data Quality Rules** | Person-Position Link ID must reference a valid DE-03 record.<br>Certification ID must reference a valid DE-04 record.<br>Currency status must be a derived computed field, not a free-text entry.<br>PII scan must confirm absence of employee names, SSNs, or other direct identifiers before zone write.<br>Completion date handling must be defined in the de-identification specification to prevent quasi-identifier risk. |
| **Privacy and Compliance** | 5 U.S.C. § 552a applies to source records. De-identification transformation is the Privacy Act compliance mechanism. Data use documentation required per DoDI 1400.25, Vol. 431 for any DPMAP-derived attributes incorporated into this entity. |
| **Data Lineage** | DLA TMS completion records (Privacy Act-protected) → de-identification (pre-delivery or LC-02 in-flight) → LC-03 (schema translation) → LC-02 (PII scan, zone write) → Processed Zone → LC-04 (currency computation). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. |

| | |
|---|---|
| **Related Information Flows** | IF-02 (DLA TMS ingestion); IF-06 (Processed Zone write); IF-07 (Processed Zone to Ontology). |
| **Related PRD Requirements** | REQ-F-002; REQ-F-007; REQ-F-009; REQ-NF-013; PRD Section 17.1 (Certification Currency Record entity definition). |
| **Open Questions** | [INFERRED] The de-identification specification must define how completion dates are handled to prevent quasi-identifier risk. This is a required Phase I privacy design task. |

## DE-06: Workload Metric

| | | | |
|---|---|---|---|
| **Zone:** Processed | **Sensitivity:** CUI | **Phase:** Phase I and II (conditional) | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | Unit-level supply chain throughput data linked to organizational units within the DLA position structure. Workload Metrics enable ALICE to correlate workforce configuration (staffing levels, certification currency) with mission output (supply chain throughput), supporting the mission-output correlation capability described in the CONOPS. This entity contains no individual data; it is organizational and operational performance data at the unit level. |
| **Authoritative Source** | DLA Enterprise Business System (EBS, SAP-based). Integration: INT-04. |
| **Key Attributes** | Metric ID — ALICE-assigned unique identifier.<br>Organizational unit — DLA organizational unit code linked to the DE-01 position structure.<br>Metric type — type of throughput measure (e.g., order fill rate, processing volume, on-time delivery rate).<br>Value — numeric metric value for the reporting period.<br>Reporting period — the time period to which the metric applies. |
| **Data Owner** | DLA (EBS data authority). |
| **Data Steward** | ALICE Data Steward. |
| **Producing Component** | LC-03 (schema translation); LC-02 (validation and zone write). |
| **Consuming Components** | LC-04 (mission-output correlation in readiness computation). |
| **CRUD Rights** | Create: LC-02 (on weekly EBS extract ingestion).<br>Read: LC-04.<br>Update: LC-02 (on weekly delta).<br>Delete: Automated retention policy only. |
| **Data Classification** | CUI. No individual data. Operational performance data at organizational unit level. |
| **Retention Schedule** | System operation plus 3 years post-decommission. |
| **Data Quality Rules** | Organizational unit must reference a valid organizational unit in the DE-01 position structure.<br>Metric type must be from the defined metric type enumeration configured in LC-10.<br>Value must be numeric and within a defined plausible range (anomaly |

| | |
|---|---|
| | detection: values exceeding 3 standard deviations from historical mean are flagged for review before loading).<br>Reporting period must be a valid ISO date range. |
| **Privacy and Compliance** | No Privacy Act implications. CUI baseline controls apply. |
| **Data Lineage** | EBS extract → LC-03 (schema translation) → LC-02 (validation, zone write) → Processed Zone → LC-04 (mission-output correlation). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. |
| **Related Information Flows** | IF-04 (EBS workload ingestion); IF-06 (Processed Zone write); IF-07 (Processed Zone to Ontology). |
| **Related PRD Requirements** | REQ-F-004; PRD Section 17.1 (Workload Metric entity definition). |
| **Open Questions** | [OPEN] OQ-007: EBS access authorization may require a separate process. If EBS access is not available in Phase I, workload metrics are omitted from Phase I and the mission-output correlation capability is deferred. OPM FWD does not include an equivalent metric for proxy purposes. |

## DE-07: Readiness Indicator

| | | | |
|---|---|---|---|
| **Zone:** Analytical | **Sensitivity:** CUI — internal analytical output | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A computed readiness score for a specific mission function area or organizational unit within the DLA structure. Readiness Indicators are ALICE-generated outputs derived from the organizational ontology (position fill status, certification currency, workflow dependency health) and, where available, workload metrics (DE-06). They are the primary analytical product of the twin's continuous monitoring capability and the data underlying the enterprise readiness heat map and activity-level position risk register. |
| **Authoritative Source** | System-computed by LC-04 (Organizational Ontology and Twin State Service). Not sourced from any external system. |
| **Key Attributes** | Indicator ID — ALICE-assigned unique identifier.<br>Scope — the mission function area or organizational unit to which the indicator applies.<br>Computation timestamp — UTC datetime of most recent computation.<br>Score — numeric readiness score on a defined scale (e.g., 0 to 100). Scale definition is a configuration item in LC-10.<br>Component breakdown — decomposition of the overall score by contributing factor (fill rate, certification currency, dependency health, workload correlation).<br>Alert status — whether the indicator has crossed a configured threshold and triggered an alert event.<br>Data currency reference — timestamp of the source ingestion cycle that produced this indicator. |
| **Data Owner** | ALICE (system-generated analytical output). |

| | |
|---|---|
| **Data Steward** | ALICE Data Steward. |
| **Producing Component** | LC-04 (Organizational Ontology Service — computes and updates on each ingestion cycle). |
| **Consuming Components** | LC-07 (Decision Support Interface — enterprise heat map and risk register display); LC-04 (self-referential — historical snapshot archive for trend analysis and backtesting). |
| **CRUD Rights** | Create / Update: LC-04 only (on each successful ingestion cycle).<br>Read: LC-07; LC-04 (historical comparison).<br>Delete: Automated retention policy only. Historical snapshots retained for trend analysis. |
| **Data Classification** | CUI — internal analytical output. Contains no individual-identifiable data. Represents aggregate organizational readiness assessments. |
| **Retention Schedule** | System operation plus 3 years post-decommission. Historical snapshots retained for minimum 12 months for trend analysis (PRD REQ-F-021). |
| **Data Quality Rules** | Score must be computed from a formally documented algorithm maintained in the system design documentation.<br>Component breakdown must sum to the overall score within a defined tolerance.<br>Data currency reference must be populated on every record — readiness indicators without a currency timestamp must not be displayed (REQ-UX-008).<br>Indicators older than 24 hours from the last successful ingestion must be flagged as stale (REQ-NF-001).<br>Backtesting validation (Phase I deliverable): at least 80% of predicted degradation onset events must occur within 30 days of the predicted date when compared to historical records (PRD O-4). |
| **Privacy and Compliance** | No Privacy Act implications. Indicators are aggregate organizational metrics with no individual-identifiable data. CUI baseline controls apply. |
| **Data Lineage** | Processed Zone (DE-01, DE-02, DE-03, DE-04, DE-05, DE-06) → LC-04 (readiness computation algorithm) → Analytical Zone (DE-07). |
| **Disposal** | Secure deletion per NIST SP 800-88 on system decommission. Historical snapshots subject to same disposal requirements. |
| **Related Information Flows** | IF-12 (Readiness Indicators to Decision Support Interface). |
| **Related PRD Requirements** | REQ-F-009; REQ-F-019; REQ-F-020; REQ-F-021; REQ-F-022; REQ-NF-001; REQ-UX-008; PRD Section 17.1 (Readiness Indicator entity definition). |
| **Open Questions** | [INFERRED] The readiness score computation algorithm must be formally specified in Phase I design documentation. The algorithm definition is a required SME validation artifact and a Phase I deliverable prerequisite. |

## DE-08: Synthetic Person

| | | | |
|---|---|---|---|
| **Zone:** Synthetic | **Sensitivity:** CUI by default pending PIA (ADR-007) | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A statistically faithful, privacy-preserving synthetic representation of a |

| | |
|---|---|
| | workforce member, generated by the ALICE Synthetic Workforce Engine (LC-05) from de-identified Processed Zone records using a generative model with differential privacy noise injection. A Synthetic Person is not a record of any real individual. It is a statistically valid synthetic data point drawn from the learned joint distribution of the de-identified workforce population. The aggregate of all Synthetic Person records constitutes the synthetic workforce population used for simulation. |
| **Authoritative Source** | ALICE Synthetic Engine (LC-05). Derived from DE-03 (Person-Position Links) and DE-05 (Certification Currency Records) in the Processed Zone. |
| **Key Attributes** | Synthetic ID — ALICE-assigned unique identifier with no relationship to any real employee ID.<br>Grade band — aggregate pay grade category (not a specific grade).<br>Series family — occupational series family grouping.<br>Tenure band — aggregate tenure category.<br>Certification profile — set of boolean certification currency indicators per relevant certification type.<br>Performance tier — aggregate performance tier category.<br>Geographic region — regional grouping (not a specific duty station).<br>Generation version — reference to the DE-12 Generation Metadata record for this population version. |
| **Data Owner** | ALICE (generated by ALICE; no external authoritative source). |
| **Data Steward** | ALICE Data Steward (governs export authorization and population lifecycle). |
| **Producing Component** | LC-05 (Synthetic Workforce Engine). |
| **Consuming Components** | LC-06 (ABM Simulation Engine — reads population for agent instantiation). LC-05 (export workflow — reads population for authorized export generation). |
| **CRUD Rights** | Create: LC-05 only (during generation cycle; write to Synthetic Zone after validation passes).<br>Read: LC-06 (simulation); LC-05 (export workflow). Data Steward: read for export authorization review.<br>Update: Not applicable — synthetic populations are immutable once promoted. A new generation cycle creates a new population version, not an update to an existing one.<br>Delete: Automated expiration policy (default 12 months per generation). Expired populations are deleted after the expiration window per automated retention policy. |
| **Data Classification** | CUI by default pending formal Privacy Impact Assessment (ADR-007). The differential privacy guarantee (epsilon ≤ 1.0) provides strong mathematical privacy protection, but federal guidance does not currently authorize differential privacy as an automatic mechanism for reclassification below CUI (RL-23). |
| **Retention Schedule** | Configurable. Default: 12 months per generation cycle. Expired populations flagged; export blocked pending regeneration. Generation metadata (DE-12) retained per audit schedule regardless of population expiration. |
| **Data Quality Rules** | Fidelity validation: KS test for continuous variables and chi-squared test for categorical variables must confirm distributional deviation within 5% of the source population across all tracked attributes before promotion (REQ-F-016). |

| | |
|---|---|
| | Privacy validation: membership inference attack simulation must confirm attack success rate does not exceed 50% (random chance) before promotion (REQ-F-017).<br>Epsilon: differential privacy epsilon applied during training must not exceed 1.0 and must be documented in DE-12 (REQ-F-015).<br>Population size: must be sufficient for statistically meaningful simulation outputs. Minimum 300-position training corpus required per ASM-003.<br>No synthetic record may contain field values that could be linked to a specific real individual (re-identification risk validation required). |
| **Privacy and Compliance** | 5 U.S.C. § 552a (Privacy Act): Synthetic Person records are not Privacy Act records because they do not contain information about identifiable individuals. However, they are derived from Privacy Act-protected source records.<br>ADR-007: Treated as CUI pending formal PIA. PIA must be initiated in Phase II (TNA N-13) to determine whether the synthetic layer may be reclassified.<br>All exports require Data Steward authorization and carry CUI markings pending PIA determination. |
| **Data Lineage** | Processed Zone (DE-03, DE-05) → LC-05 (generative model training with DP noise injection, epsilon ≤ 1.0) → fidelity and privacy validation → Synthetic Zone (promoted population) + DE-12 (Generation Metadata). |
| **Disposal** | Secure deletion per NIST SP 800-88 on expiration and on system decommission. Generation Metadata (DE-12) is retained independently of population expiration. |
| **Related Information Flows** | IF-08 (Processed Zone to Synthetic Engine training); IF-09 (Synthetic Population to Synthetic Zone); IF-10 (Synthetic Zone to Simulation Engine); IF-15 (Synthetic Data Export Delivery). |
| **Related PRD Requirements** | REQ-F-015; REQ-F-016; REQ-F-017; REQ-F-018; REQ-NF-005; REQ-NF-008; REQ-COMP-004; PRD Section 17.1 (Synthetic Person entity definition). |
| **Open Questions** | [OPEN] OQ-006: NIST SP 800-188 finalization may affect PIA framework for the classification determination. [OPEN] OQ-001: Pilot sub-component position count must meet ASM-003 minimum (300 positions) for adequate training corpus. |

## DE-09: Scenario Run

| | | | |
|---|---|---|---|
| **Zone:** Analytical | **Sensitivity:** CUI | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A record of a single ABM simulation execution, including the parameters used, the synthetic population version applied, execution timing, and a reference to the simulation output files. Scenario Run records serve two functions: they are operational records enabling users to review and re-access prior simulation results, and they are audit artifacts for RMF and compliance purposes documenting when simulations occurred, who initiated them, and what parameters were applied. |
| **Authoritative Source** | ALICE Simulation Engine (LC-06) — generated on every scenario execution. |
| **Key Attributes** | Run ID — ALICE-assigned unique identifier. |

| | |
|---|---|
| | User ID — the authenticated user identity who initiated the scenario run.<br>Scenario type — from the defined scenario type enumeration (Key Departure, Workload Surge, Certification Shortfall, Geographic Reallocation, Pipeline Shortfall, or Custom).<br>Parameters — structured record of the scenario configuration: target position (for Key Departure), surge level (for Workload Surge), timeline, and response options evaluated.<br>Synthetic population version — reference to the DE-08 / DE-12 generation version used in this run.<br>Execution timestamp — UTC datetime of simulation initiation.<br>Duration — elapsed time in seconds from initiation to results delivery.<br>Output file reference — reference to the Analytical Zone storage location of simulation results. |
| **Data Owner** | ALICE. |
| **Data Steward** | ALICE Data Steward. |
| **Producing Component** | LC-06 (Simulation Engine). |
| **Consuming Components** | LC-07 (Decision Support Interface — historical scenario review); LC-09 (Audit Logging — scenario run event). |
| **CRUD Rights** | Create: LC-06 (on scenario execution).<br>Read: LC-07 (historical review); LC-09 (audit context).<br>Update: Not applicable — scenario runs are immutable records of past executions.<br>Delete: Automated retention policy only (3-year retention). |
| **Data Classification** | CUI. Parameters may contain activity-specific organizational data. Output references link to CUI simulation results. No individual-identifiable data in parameters or results. |
| **Retention Schedule** | 3 years (PRD Section 17.1). |
| **Data Quality Rules** | Run ID must be unique.<br>User ID must reference a valid authenticated user identity from LC-08.<br>Scenario type must be from the defined enumeration.<br>Synthetic population version must reference a valid, validated DE-12 generation record.<br>Duration must be a positive numeric value; scenario runs that did not complete must be logged with a failure status, not omitted. |
| **Privacy and Compliance** | No Privacy Act implications. Scenario parameters reference organizational constructs (position nodes), not individuals. CUI baseline controls apply. |
| **Data Lineage** | User scenario initiation (via LC-07) → LC-06 (simulation execution using DE-08 population) → Analytical Zone (DE-09 run record + simulation output files). |
| **Disposal** | Secure deletion per NIST SP 800-88 after 3-year retention period. |
| **Related Information Flows** | IF-10 (Synthetic Zone to Simulation Engine); IF-11 (Simulation Results to Analytical Zone). |
| **Related PRD Requirements** | REQ-F-012; REQ-F-013; REQ-F-014; REQ-F-023; PRD Section 17.1 (Scenario Run entity definition). |
| **Open Questions** | None specific to this entity. |

## DE-10: Synthetic Data Export

| Zone: Analytical / Audit | Sensitivity: CUI — governance record | Phase: Phase I and II | Type: Source-Derived |
|---|---|---|---|

| | |
|---|---|
| **Description** | An immutable governance record documenting every authorized export of a synthetic workforce population (DE-08) from the ALICE Synthetic Zone. The Synthetic Data Export record is the primary accountability artifact for the synthetic data export workflow. It documents who requested the export, the stated use case, the Data Steward's authorization decision, the timestamp of the export event, and the cryptographic hash of the delivered export file. It enables after-the-fact verification that every export was authorized, and that the delivered file has not been altered after delivery. |
| **Authoritative Source** | ALICE Export Workflow (LC-05) — generated on every export event regardless of approval or denial outcome. |
| **Key Attributes** | Export ID — ALICE-assigned unique identifier.<br>Requester ID — authenticated identity of the requestor.<br>Use case description — documented statement of the intended use of the exported data.<br>Data Steward ID — identity of the Data Steward who reviewed and decided the request.<br>Authorization decision — approved or denied.<br>Authorization timestamp — UTC datetime of the Steward's decision.<br>Export file hash — cryptographic hash of the delivered export file (for integrity verification). Populated only on approved exports.<br>Delivery method — the approved secure channel through which the file was delivered.<br>Synthetic population version — reference to the DE-12 generation version exported. |
| **Data Owner** | ALICE Data Steward. |
| **Data Steward** | ALICE Data Steward. |
| **Producing Component** | LC-05 (Export Workflow); LC-09 (Audit Logging — receives the export event record). |
| **Consuming Components** | LC-09 (audit log store — primary residence for this record); LC-07 (Data Steward interface — export request queue and history). |
| **CRUD Rights** | Create: LC-05 (on request submission) and LC-09 (on authorization decision and delivery).<br>Read: Data Steward; System Administrator (audit review); Authorized Auditor.<br>Update: Not applicable — Synthetic Data Export records are immutable.<br>Delete: Not permitted — retained for 7 years minimum (REQ-NF-012). Non-deletable per ADR-005. |
| **Data Classification** | CUI — governance record. The record itself does not contain workforce data; it documents the event of an export. |
| **Retention Schedule** | 7 years (PRD Section 17.1; REQ-SEC-010). |
| **Data Quality Rules** | Export ID must be unique. |

| | |
|---|---|
| | Export file hash must be a recognized cryptographic hash (e.g., SHA-256 or equivalent).<br>Denied export records must still be created and logged (denial is a governance event).<br>Authorization timestamp must precede the delivery timestamp for approved exports.<br>All fields except export file hash and delivery method must be populated on request submission; remaining fields populated on authorization and delivery. |
| **Privacy and Compliance** | Governance artifact for exports of Privacy Act-derived synthetic data (ADR-007). Required by REQ-F-018. The record enables after-the-fact verification of compliance with the synthetic data export governance framework. |
| **Data Lineage** | Export request (requester via LC-07) → LC-05 (request logging, Steward notification) → Data Steward decision → LC-05 (export generation if approved) → LC-09 (immutable export record with file hash). |
| **Disposal** | 7-year retention minimum. After retention period, secure deletion per NIST SP 800-88. |
| **Related Information Flows** | IF-15 (Synthetic Data Export Delivery); IF-14 (Audit Events to Audit Log). |
| **Related PRD Requirements** | REQ-F-018; REQ-NF-012; REQ-SEC-009; REQ-SEC-010; PRD Section 17.1 (Synthetic Data Export entity definition). |
| **Open Questions** | None specific to this entity. |

## DE-11: Audit Log Entry

| | | | |
|---|---|---|---|
| **Zone**: Audit | **Sensitivity**: CUI — sensitive | **Phase**: Phase I and II | **Type**: Source-Derived |

| | |
|---|---|
| **Description** | An immutable, append-only record of any user action or system event within the ALICE environment. Audit Log Entries are generated by every ALICE service component and stored exclusively by LC-09 (Immutable Audit Logging Service). They constitute the primary compliance evidence for RMF authorization, CMMC assessment, Privacy Act compliance demonstration, and operational security monitoring. No Audit Log Entry may be modified or deleted after creation. |
| **Authoritative Source** | All ALICE service components (LC-01 through LC-12) — each generates audit events per ADR-005 and the audit event catalog in SDP-ART-02 Section 3.5. |
| **Key Attributes** | Entry ID — LC-09-assigned unique identifier.<br>Timestamp — UTC datetime of the event (receipt timestamp assigned by LC-09; generating service also provides an origination timestamp).<br>Actor — the authenticated user identity or service identity responsible for the event.<br>Action type — the category of event (from the defined event type enumeration in the audit event catalog).<br>Affected entity — the ALICE entity or resource affected by the event (e.g., Position ID, Export ID, configuration parameter name).<br>Outcome — the result of the action (success, failure, denied, error). |

| | |
|---|---|
| | Source IP or system identifier — network or system identifier of the originating request.<br>Integrity hash — hash value enabling tamper detection of the log entry and its place in the hash chain. |
| **Data Owner** | ALICE (write authority); Authorized Auditors (read authority). |
| **Data Steward** | ALICE System Administrator (read access for operational review); DLA ISSO (Phase II — compliance review). |
| **Producing Component** | LC-09 (sole writer to the Audit Log Store; receives events from all other components). |
| **Consuming Components** | System Administrator (audit review); DLA ISSO (RMF continuous monitoring, Phase II); Authorized Auditor (compliance review). |
| **CRUD Rights** | Create: LC-09 only (append-only; each service delivers events to LC-09 via write-only interface).<br>Read: System Administrator (read-only); Authorized Auditor (read-only).<br>Update: Not permitted under any circumstances. Immutable by design (ADR-005).<br>Delete: Not permitted. Retained for 7 years minimum. Automated retention lifecycle managed by LC-09. |
| **Data Classification** | CUI — sensitive. Audit logs may reference user identities, system events, and data access patterns. Log entries must not contain raw workforce data content, but they may contain references to workforce data entities (e.g., Position IDs accessed). |
| **Retention Schedule** | 7 years minimum (REQ-NF-012; REQ-SEC-010). Non-deletable. |
| **Data Quality Rules** | Entry ID must be unique and assigned by LC-09, not by the generating service.<br>Timestamp must be UTC. Generating services must synchronize clocks with an authoritative time source.<br>Actor must reference a valid authenticated identity from LC-08 session records (for user events) or a valid service identity (for system events).<br>Action type must be from the defined event type enumeration. Free-text action types are not permitted.<br>Integrity hash must be computable for every entry, and the hash chain must be verifiable without access to any ALICE application service.<br>Events must be delivered to LC-09 within a defined delivery window (SLA to be defined in Phase I system design). Events that cannot be delivered must be buffered and delivered on LC-09 restoration. |
| **Privacy and Compliance** | Audit logs are the primary compliance evidence artifact for NIST SP 800-53 AU control family, CMMC Level 2 audit requirements, and DoDI 8510.01 RMF continuous monitoring. AU-9 requires that audit logs be protected against unauthorized modification — LC-09 architectural isolation is the enforcement mechanism (ADR-005). |
| **Data Lineage** | All ALICE services → write-only event delivery interface → LC-09 (entry assignment, hash-chain appending) → Audit Log Store. |
| **Disposal** | After the 7-year minimum retention period, secure deletion per NIST SP 800-88. Disposal must be documented as an administrative event itself, logged before the log segment is destroyed. |
| **Related Information Flows** | IF-14 (Audit Events to Audit Log — all services). |

| | |
|---|---|
| **Related PRD Requirements** | REQ-NF-012; REQ-SEC-009; REQ-SEC-010; REQ-F-018; REQ-ADM-006; PRD Section 17.1 (Audit Log Entry entity definition). |
| **Open Questions** | [INFERRED] Storage provisioning for 7-year retention at enterprise scale must be estimated in Phase I technical design. The event volume calculation depends on pilot sub-component scope. |

## DE-12: Generation Metadata

| | | | |
|---|---|---|---|
| **Zone:** Synthetic / Audit | **Sensitivity:** CUI | **Phase:** Phase I and II | **Type:** Source-Derived |

| | |
|---|---|
| **Description** | A record documenting every synthetic population generation event, including the differential privacy parameters applied, the fidelity and privacy validation results, the training dataset version used, and the generation timestamp. Generation Metadata records are compliance artifacts required for Privacy Impact Assessment support, RMF authorization evidence, CMMC assessment, and data lineage documentation. They are retained independently of the synthetic population they document: when a synthetic population expires or is superseded, the corresponding Generation Metadata record is retained for audit purposes. |
| **Authoritative Source** | ALICE Synthetic Engine (LC-05) — generated on every generation cycle. |
| **Key Attributes** | Generation version ID — ALICE-assigned unique identifier for this generation cycle.<br>Source dataset version — identifier of the Processed Zone dataset snapshot used as training input.<br>Generative model version — version identifier of the generative model architecture used.<br>Epsilon applied — the differential privacy epsilon value applied during training. Must be ≤ 1.0 per REQ-F-015.<br>Fidelity test results — KS test and chi-squared test results per attribute, with pass/fail determination.<br>Privacy test results — membership inference attack simulation results, with pass/fail determination (attack success rate ≤ 50%).<br>Validation timestamp — UTC datetime of validation completion.<br>Promotion status — whether the population was promoted to active status (pass) or quarantined (fail).<br>Population record count — number of Synthetic Person (DE-08) records in this generation. |
| **Data Owner** | ALICE Data Steward. |
| **Data Steward** | ALICE Data Steward; SOL Privacy Officer (compliance review). |
| **Producing Component** | LC-05 (Synthetic Workforce Engine). |
| **Consuming Components** | LC-05 (export workflow — references generation version for provenance metadata); LC-09 (audit log — generation event); DLA Privacy Officer and CDO (PIA evidence); DLA ISSO (RMF evidence). |
| **CRUD Rights** | Create: LC-05 (on each generation cycle, regardless of promotion outcome).<br>Read: LC-05 (export provenance); Data Steward; Auditor; DLA Privacy |

| | |
|---|---|
| | Officer (PIA review); DLA ISSO (RMF review).<br>Update: Not applicable — generation metadata records are immutable once created.<br>Delete: Not permitted during retention period. Retained per audit schedule (same as associated synthetic population plus audit retention). |
| **Data Classification** | CUI. Contains technical compliance information about the generation process. No workforce data content; records the parameters and results of the generation event. |
| **Retention Schedule** | Same as associated synthetic population plus applicable audit retention period (minimum 7 years from the later of population expiration or last export referencing this generation version). |
| **Data Quality Rules** | Epsilon must be a numeric value ≤ 1.0; values exceeding 1.0 must result in generation failure, not a logged exception.<br>Fidelity test results must include per-attribute test statistics, not summary indicators only.<br>Membership inference attack results must document the attack methodology used and the success rate achieved.<br>Promotion status must be binary (promoted or quarantined); no intermediate states.<br>Generation metadata must be created for every generation attempt, including failed and quarantined generations. |
| **Privacy and Compliance** | This record is the primary technical evidence artifact for the Privacy Act compliance claim that ALICE's synthetic population was generated with formal differential privacy guarantees.<br>REQ-COMP-004 requires that the synthetic data generation process document differential privacy parameters and maintain privacy audit records. DE-12 is the fulfillment of this requirement.<br>PIA evidence: DLA Privacy Officer and DLA CDO will review DE-12 records as part of the formal PIA process to determine whether the synthetic layer may be reclassified below CUI (ADR-007). |
| **Data Lineage** | LC-05 generation execution → DE-12 creation (concurrent with or immediately after DE-08 population generation) → Synthetic Zone (associated with DE-08 population) + Audit Log Store (compliance copy). |
| **Disposal** | Retained per audit schedule. Secure deletion per NIST SP 800-88 after retention period. |
| **Related Information Flows** | IF-09 (Synthetic Population to Synthetic Zone — DE-12 created concurrently); IF-14 (Audit Events — generation event logged to LC-09). |
| **Related PRD Requirements** | REQ-F-015; REQ-F-016; REQ-F-017; REQ-COMP-004; PRD Section 17.1 (Generation Metadata entity definition). |
| **Open Questions** | [OPEN] OQ-006: NIST SP 800-188 finalization may affect the content requirements for DE-12 if it prescribes specific privacy audit documentation standards. |

## 4. Data Governance Rules

The following data governance rules apply across all ALICE data entities and are binding on all ALICE components, operational users, and administrative personnel. These rules are Source-Derived from PRD Section 17.2 and the CONOPS privacy governance narrative.

| Rule ID | Rule Statement | Source | Affected Entities | Enforcement Mechanism | Validation Method |
|---|---|---|---|---|---|
| DGR-01 | No Privacy Act-protected individual record shall be stored in the simulation layer, decision support interface, or exported as any part of any ALICE output. | 5 U.S.C. § 552a; PRD Section 17.2; REQ-NF-008; REQ-COMP-001 | DE-03, DE-05 (source records); DE-08 (derived); all output entities | ADR-002 (four-zone data architecture) enforces this at the zone boundary. LC-02 PII detection enforces this at ingestion. LC-06 Synthetic Zone-only access (ADR-008) enforces this in simulation. | Automated PII scan of Processed Zone: zero PII records. Privacy review sign-off before Phase I demonstration. |
| DGR-02 | All de-identified extracts must have PII fields removed or tokenized before ingestion into the ALICE analytical environment. | 5 U.S.C. § 552a; PRD Section 17.2; REQ-F-006 | DE-03, DE-05 | LC-02 PII detection capability; de-identification specification document (required Phase I artifact). | Test: inject records with known PII fields; confirm rejection and logging; confirm zero PII records in Processed Zone. |
| DGR-03 | The synthetic population generation log (DE-12) must document the source dataset version, differential privacy epsilon applied, fidelity test results, and privacy audit results for every generation cycle. | PRD Section 17.2; REQ-COMP-004 | DE-12 | LC-05 generation workflow; required fields enforced at record creation. | Review DE-12 records: confirm all required fields populated for every generation cycle including quarantined generations. |
| DGR-04 | Data retention schedules must be enforced by automated policy, not manual deletion. | PRD Section 17.2 | All entities with defined retention schedules | Automated retention policy configured in LC-10; enforced by storage lifecycle management. | Audit review: confirm no manual deletion events in LC-09; confirm automated retention events are logged. |
| DGR-05 | All data lineage from source system to synthetic output must be documentable and | PRD Section 17.2; OMB M-19-15; TNA | All entities in the lineage chain: | Data lineage documentation maintained by ALICE Data Steward; | On-demand lineage demonstration: produce lineage chain for a specified synthetic generation |

| Rule ID | Rule Statement | Source | Affected Entities | Enforcement Mechanism | Validation Method |
|---|---|---|---|---|---|
| | auditable on demand. | Section 12.4 | DE-01 through DE-08 and DE-12 | lineage chain recorded in DE-12 and ingestion audit events in DE-11. | version from source extract to synthetic population. |
| DGR-06 | Synthetic data exports require Data Steward authorization, use case documentation, and generation of an immutable export record (DE-10) regardless of authorization outcome. | PRD Section 17.2; REQ-F-018 | DE-08 (exported); DE-10 (governance record) | LC-05 export workflow enforces Steward authorization gate. LC-09 records the outcome. | Audit review: confirm zero exports without a corresponding approved DE-10 record. |
| DGR-07 | DPMAP performance data ingested into ALICE must be aggregate only (performance tier distribution by occupational series and grade band). No individual performance rating records may be ingested. | DoDI 1400.25, Vol. 431; REQ-COMP-002; REQ-INT-003 | DE-03 (performance tier attribute) | LC-02 ingestion scope definition; data use agreement specification for INT-03. | Data use documentation review: confirm individual DPMAP records are excluded from data use agreement scope. Schema validation: confirm no individual-rating fields in schema. |
| DGR-08 | The synthetic population (DE-08) shall be treated as CUI by default until a formal Privacy Impact Assessment conducted by DLA and the DLA CDO determines otherwise. | ADR-007; IAPP November 2024 (RL-23) | DE-08; DE-10 (export governance) | CUI marking applied to all DE-08 records and all DE-10 export files at creation. PIA initiation required in Phase II. | PIA documentation review; CUI marking audit of exported files. |

## 5. Data Lineage Register

The following register documents the complete lineage chain for each data entity from its authoritative source through all ALICE transformations. This register fulfills the data lineage documentation requirements of TNA Section 12.4, PRD Section 17.2 (DGR-05), and OMB M-19-15.

| Entity ID | Lineage Stage 1: Source | Lineage Stage 2: Translation | Lineage Stage 3: Ingestion / De-identification | Lineage Stage 4: Zone Residence | Lineage Stage 5: Downstream Use | Key Transformation Controls |
|---|---|---|---|---|---|---|
| DE-01 | DCPDS / DCHRMS: authoritative position records. | LC-03: schema translation to ALICE canonical model. | LC-02: schema validation; PII scan (none expected for positions); zone write. | Processed Zone. | LC-04: ontology construction; LC-05: structural scope. | Schema validation; canonical model conformance. |
| DE-02 | Organizational process documentation; SME interviews. | No source system translation — ALICE-constructed. | LC-04: ontology construction with SME validation. | Processed Zone / Analytical Zone. | LC-04: readiness computation; LC-06: agent model. | SME validation (minimum one DLA SME per pilot sub-component). |
| DE-03 | DCPDS / DCHRMS: personnel records (Privacy Act-protected). | LC-03: schema translation to canonical model. | LC-02: PII scan (critical); de-identification (field removal, tokenization); zone write. | Processed Zone. | LC-04: readiness computation; LC-05: training input. | PII detection; de-identification specification; Privacy Act compliance review. |
| DE-04 | DLA TMS: certification definition records. | LC-03: schema translation. | LC-02: schema validation; zone write. | Processed Zone. | LC-04: currency computation; LC-07: risk register. | Schema validation; canonical model conformance. |
| DE-05 | DLA TMS: completion records (Privacy Act-adjacent). | LC-03: schema translation. | LC-02: PII scan; de-identification; zone write. | Processed Zone. | LC-04: currency computation; LC-05: certification profile in training. | De-identification; date field quasi-identifier review. |
| DE-06 | EBS: unit-level throughput metrics. | LC-03: schema translation. | LC-02: schema validation; zone write. | Processed Zone. | LC-04: mission-output correlation. | Schema validation; anomaly detection (value range check). |
| DE-07 | System-computed by LC-04 from Processed Zone entities. | No external source. | LC-04: readiness computation algorithm. | Analytical Zone. | LC-07: heat map and risk register display. | Algorithm documentation; backtesting validation (≥ 80% accuracy at Phase I); data currency timestamp. |
| DE-08 | Processed Zone: DE-03, | LC-05: generative | LC-05: fidelity validation (KS | Synthetic | LC-06: agent | Differential privacy |

| Entity ID | Lineage Stage 1: Source | Lineage Stage 2: Translation | Lineage Stage 3: Ingestion / De-identification | Lineage Stage 4: Zone Residence | Lineage Stage 5: Downstream Use | Key Transformation Controls |
|---|---|---|---|---|---|---|
| | DE-05 (training input). | model training with DP noise injection (epsilon ≤ 1.0). | test, chi-squared; ≤ 5% deviation); privacy validation (membership inference; attack success ≤ 50%); atomic promotion on validation pass. | Zone. | instantiation (simulation); LC-05: export workflow. | mechanism; fidelity and privacy validation documented in DE-12. |
| DE-09 | LC-06: generated on scenario execution. | No transformation. | LC-06: scenario run record creation. | Analytical Zone. | LC-07: historical scenario review. | Unique Run ID; valid user ID and population version references. |
| DE-10 | LC-05 export workflow: generated on export request and decision. | No transformation. | LC-05: export event record; LC-09: immutable log. | Analytical Zone / Audit Log Store. | Data Steward audit review. | Immutable record; file hash; non-deletable. |
| DE-11 | All ALICE services: event generation. | LC-09: entry ID assignment; hash-chain appending. | LC-09: append-only log store write. | Audit Log Store. | System Administrator and Auditor: compliance review. | Immutability; hash-chain integrity; 7-year retention. |
| DE-12 | LC-05: generation event documentation. | No transformation. | LC-05: metadata record creation concurrent with DE-08 generation. | Synthetic Zone / Audit Log Store. | PIA evidence; RMF evidence; export provenance. | All required fields populated; epsilon ≤ 1.0 enforced; quarantined generations documented. |

## 6. Data Lifecycle Summary

| Lifecycle Stage | Trigger | Entities Affected | Governing Component | Frequency | Audit Event Generated? |
|---|---|---|---|---|---|
| Source extraction | Scheduled ingestion | DE-01, DE-03, DE-04, | LC-02, LC-03 | Daily (position, | Yes — DE-11 ingestion event. |

| Lifecycle Stage | Trigger | Entities Affected | Governing Component | Frequency | Audit Event Generated? |
|---|---|---|---|---|---|
| | job; manual trigger by System Administrator. | DE-05, DE-06 | | certification); weekly (workload metrics); annually (performance calibration) | |
| De-identification transformation | Immediately upon receipt of source records from LC-03. | DE-03, DE-05 | LC-02 | Concurrent with each ingestion cycle. | Yes — DE-11 PII scan and de-identification event. |
| Zone write (Processed) | On successful de-identification and schema validation. | DE-01, DE-02 (partial), DE-03, DE-04, DE-05, DE-06 | LC-02 | Concurrent with each ingestion cycle. | Yes — DE-11 ingestion completion event with record counts. |
| Ontology refresh | On ingestion completion signal from LC-02. | DE-02 (full), DE-07 | LC-04 | Concurrent with each successful ingestion cycle. | Yes — DE-11 ontology update event. |
| Historical snapshot archive | On each ontology refresh. | DE-07 (historical copies) | LC-04 | Concurrent with each ingestion cycle. | Yes — DE-11 snapshot creation event. |
| Synthetic generation | Scheduled cycle (configurable; expected monthly or quarterly) or on-demand System Administrator trigger. | DE-08, DE-12 | LC-05 | Configurable; default monthly. | Yes — DE-11 generation event with DE-12 reference. |
| Synthetic validation | Immediately after generation. | DE-08, DE-12 | LC-05 | Concurrent with each generation cycle. | Yes — DE-11 validation pass/fail event with DE-12 reference. |
| Synthetic | After both | DE-08 | LC-05 | Concurre | Yes — DE-11 promotion |

| Lifecycle Stage | Trigger | Entities Affected | Governing Component | Frequency | Audit Event Generated? |
|---|---|---|---|---|---|
| population promotion | fidelity and privacy validation pass. | | | nt with each successful generation cycle. | event. |
| Synthetic population expiration | 12 months after promotion (configurable). | DE-08 | Automated retention policy (LC-10 configuration) | Automated. | Yes — DE-11 expiration event. DE-12 retained independently. |
| Synthetic data export | Data Steward authorization following requester submission. | DE-08 (read for export); DE-10 (created) | LC-05, LC-09 | On demand. | Yes — DE-11 export event and immutable DE-10 record. |
| Scenario execution | User initiation via LC-07. | DE-08 (read); DE-09 (created) | LC-06 | On demand. | Yes — DE-11 scenario run event. |
| Audit log retention enforcement | Automated after 7-year retention period. | DE-11, DE-10 | LC-09, LC-10 (retention policy) | Automated. | Yes — disposal event logged before deletion. |
| System decommission | Authorized decommission event. | All entities | All components | One-time. | Yes — secure deletion documented per NIST SP 800-88. |

## 7. Gate-4 Summary and Decision Required

### What Was Created

Artifact 4 (Data Catalog) provides the authoritative governance reference for all twelve ALICE data entities (DE-01 through DE-12). Each entity record includes its stable identifier, zone residence, sensitivity classification, source system, business and technical description, complete key attribute set drawn from PRD Section 17.1, data owner and steward, CRUD rights by component, retention schedule with regulatory basis, data quality validation rules, data lineage chain, disposal requirements, and privacy and compliance obligations. The catalog also provides the complete Data Governance Rule register (DGR-01 through DGR-08), the Data Lineage Register documenting the five-stage lineage chain for each entity, and the Data Lifecycle Summary table covering thirteen lifecycle events from source extraction through system decommission.

### Notable Decisions and Inferred Requirements Documented in This Artifact

- **[INFERRED — DE-02]** The role ontology specification — defining the single-point-of-failure detection algorithm and criticality rating assignment method — is a required Phase I design

artifact not explicitly specified in the PRD. It must be produced before ontology construction begins.

- **[INFERRED — DE-03]** The de-identification specification document — defining field removal list, tokenization method, aggregation approach, and date field quasi-identifier handling — is a required pre-ingestion artifact. It must be jointly approved by the DLA Privacy Officer and the SOL Privacy Officer before any DCPDS data flows into the ALICE environment.
- **[INFERRED — DE-05]** Completion date fields in DLA TMS data must be addressed in the de-identification specification to prevent quasi-identifier risk. This is not explicitly addressed in the PRD but is a necessary design task.
- **[INFERRED — DE-07]** The readiness score computation algorithm must be formally specified in Phase I design documentation. It is a required SME validation artifact and a Phase I deliverable prerequisite.
- **[INFERRED — DE-11]** Audit log storage provisioning for 7-year retention at enterprise scale must be estimated in Phase I technical design. The calculation depends on pilot sub-component scope and event volume.

### Open Issues

- **[Critical]** OQ-002: The data use agreement must specify which party performs de-identification of DCPDS records and how compliance is verified. This affects DE-03 and DE-05 ingestion design.
- **[High]** OQ-001: Pilot sub-component selection determines DE-01 scope and ASM-003 minimum position count for DE-08 training viability.
- **[High]** OQ-007: EBS access authorization affects DE-06 availability in Phase I. If EBS access is not available, mission-output correlation (LC-04 function) is deferred.
- **[Medium]** OQ-008: DLA TMS extract format must be confirmed before DE-04 and DE-05 schema mapping (LC-03) can be finalized.
- **[Medium]** OQ-006: NIST SP 800-188 finalization may affect DE-12 content requirements and DE-08 classification determination through the PIA process.

### Decision Required

**Gate-4 Decision:** Please approve Artifact 4 (Data Catalog) to proceed to Artifact 5 (Interface Catalog), or provide specific revision instructions. If any entity description, key attribute set, data quality rule, governance rule, or lineage chain requires modification, please identify the entity ID (DE-XX) or rule ID (DGR-XX) and the specific change required. All DE-XX IDs introduced in this artifact will be carried forward into all remaining SDP artifacts.