Extraction Report: 05_data_governance

Property	Value
Domain	Governance
Format	.xml
Text length	2,404 chars
Sentences	46

Entities

Spacy NER: 56 entities (788ms)

Entity	Label	Score
January 1, 2026		

| DATE | | | January 1, 2026 | DATE | | | 1 | CARDINAL | | | one | CARDINAL | | | four | CARDINAL | | | PII | ORG | | | PHI | ORG | | | the Data Privacy Officer | ORG | | | 2 | CARDINAL | | | 2.1 | CARDINAL | | | Confidential | ORG | | | Restricted | ORG | | | 2.2 | CARDINAL | | | quarterly | DATE | | | 90 days | DATE | | | 2.3 | CARDINAL | | | 1 hour | TIME | | | 2.4 | CARDINAL | | | Cross-department | ORG | | | 3 | CARDINAL | | | 3.1 | CARDINAL | | | 30 days | DATE | | | 3.2 | CARDINAL | | | 7 years | DATE | | | 3.3 | CARDINAL | | | 12 months | DATE | | | an additional 24 months | DATE | | | 3.4 | CARDINAL | |

Gliner NER: 70 entities (635ms)

Entity	Label	Score
data assets	data_structure	0.47
classified	constraint	0.46
sensitivity levels	constraint	0.59
access restrictions	constraint	0.96
Business data	database_table	0.81
employee authentication	requirement	0.94
Sensitive business data	constraint	0.80
role-based access	requirement	0.84
encryption	requirement	0.84
Restricted	constraint	0.33
Highly sensitive data	requirement	0.72
PII	requirement	0.88
PHI	requirement	0.90
financial records	requirement	0.95
encryption		
at rest and in transit	requirement	0.56
audit logging	requirement	0.90
approval	requirement	0.50
Data Privacy Officer	role	0.99
access	action	0.48
Confidential	constraint	0.85
approval	action	0.64
data owner	role	0.81
Access permissions	constraint	0.73
permissions	constraint	0.82
inactivity	action	0.43
Restricted data	constraint	0.68

Entity	Label	Score
short-lived tokens	constraint	0.79
maximum 1 hour	constraint	0.35

Stanza NER: 50 entities (1584ms)

Entity	Label	Score
January 1, 2026	DATE	
1	CARDINAL	
one	CARDINAL	
four	CARDINAL	
PHI	ORG	
the Data Privacy Officer	ORG	
2	CARDINAL	
2.1	CARDINAL	
2.2	CARDINAL	
quarterly	DATE	
90 days	DATE	
2.3	CARDINAL	
1 hour	TIME	
2.4	CARDINAL	
Data Sharing Agreement	LAW	
DSA	ORG	
3.	CARDINAL	
3.1	CARDINAL	
30 days	DATE	
3.2	CARDINAL	
7 years	DATE	
3.3	CARDINAL	
12 months	DATE	
an additional 24 months	DATE	
3.4	CARDINAL	
4	CARDINAL	
4.1	CARDINAL	
4.2	CARDINAL	

Flair NER: 5 entities (226ms)

Entity	Label	Score
Data Privacy Officer	ORG	0.85
Data Sharing Agreement	MISC	1.00
DSA	MISC	1.00
Governance Committee	ORG	1.00
GDPR	MISC	1.00

Relations (SRL)

27 SRL frames (100ms)

Verb	Agent	Patient	Other
classify	DATA CLASSIFICATION		

Verb	Agent	Patient	Other
All data assets		ARGM-MOD=must, ARG2=into one of four sen	
require		employee authentication	
require		role - based access and encryption	
require		2	ARGM-LOC=
log			ARGM-LOC=
require	All access to Confidential and Restricte	explicit approval from	
data owner and			
review	Access permissions		ARGM-MOD=
revoke	Unused permissions	2.3	ARGM-TMP=
account	Service	accessing	
use	Restricted data	short - lived tokens (maximum 1 hour)	ARGM-MOD=
require	Cross - department data sharing	Data Sharing Agreement (DSA) signed by	
delete	Customer PII		ARGM-MOD=
retain	Financial records		ARGM-MOD=
retain	System logs		ARGM-MOD=
archive			ARG2=for an
follow	Backup data	same retention rules	ARGM-MOD=
log	All access to Restricted data		ARGM-MOD=
be	Audit logs	immutable	
retain			ARGM-MOD=
trigger	Anomalous access patterns	4.4	ARGM-MOD=
complete	Quarterly access audits		ARGM-MOD=
report			ARG2=to the
document	All production datasets	data quality rules	ARGM-MOD=
require	Master data changes	approval from data steward and validatio	ARGM-TMP=
report	suspected data breach		ARGM-MOD=

Enriched Extraction (batch-enrich)

Single GPU call: 70 entities, 46 roles, 27 SRL frames (1299ms)

Latency Comparison

Method	Latency
Separate (GLiNER 635ms + GLiClass 585ms + SRL 100ms)	**1320ms**
batch-enrich (unified)	**1299ms**
Speedup	**1.0x**

Per-Sentence Enriched Results (sample)

Sentence	Entities	Role	SRL Frames
DATA GOVERNANCE POLICY — Effective January 1, 2026			

. | 0 | Condition (0.24) | 0 | | DATA GOVERNANCE POLICY — Effective January 1, 2026

1. | 0 | Condition (0.24) | 0 | | DATA CLASSIFICATION All data assets must be classified into | 3 | Type (0.71) | 1 | | Public: Information freely available. | 0 | Evidence (0.43) | 0 | | No access restrictions. | 1 | Condition (0.18) | 0 | | Internal: Business data for internal use. | 1 | Variable (0.50) | 0 | | Requires employee authentication. | 1 | Condition (0.22) | 1 | | Confidential: Sensitive business data. | 1 | Condition (0.29) | 0 | | Requires role-based access and encryption. | 2 | Permission (0.40) | 1 | | Restricted: Highly sensitive data (PII, PHI, financial recor | 5 | Condition (0.36) | 0 | | Requires encryption at rest and in transit, audit logging, a | 4 | Requirement (0.36) | 2 | | ACCESS CONTROL

2.1. | 0 | Permission (0.55) | 0 | | All access to Confidential and Restricted data requires expl | 6 | Permission (0.93) | 1 | | Access permissions must be reviewed quarterly. | 1 | Permission (0.98) | 1 | | Unused permissions are automatically revoked after 90 days o | 2 | Permission (0.98) | 1 |

Entity Type Distribution (enriched)

Label	Count
constraint	22
requirement	16
action	10
role	8
data_structure	4
database_column	2
temporal_concept	2
metric	2
database_table	1
system_component	1
process	1
event_type	1

Role Distribution (enriched)

Role	Count
Condition (State)	23
Permission (Normative)	8
Evidence (Discourse)	4
Requirement (Causal)	4
Variable (Scientific)	2
Trigger Rule (Other)	2
Type (Programming)	1
Action (Event)	1
State Change (Other)	1

QLang Sentences

Gliclass: 46 classifications (585ms)

Causal (4)

- [Requirement] (0.36) Requires encryption at rest and in transit, audit logging, and approval from the...
- [Requirement] (0.55) Customer PII must be deleted within 30 days of account closure unless legally re...
- [Requirement] (0.66) Financial records must be retained for a minimum of 7 years per regulatory requi...
- [Requirement] (0.51) Affected individuals must be notified within 72 hours per GDPR requirements....

Discourse (4)

- [Evidence] (0.43) Public: Information freely available....
- [Evidence] (0.65) System logs must be retained for 12 months and archived for an additional 24 mon...
- [Evidence] (0.58) Audit logs are immutable and must be retained for 5 years....
- [Evidence] (0.38) A breach investigation report must be completed within 14 days....

Event (1)

- [Action] (0.93) All access to Restricted data must be logged with user identity, timestamp, acti...

Normative (8)

- [Permission] (0.40) Requires role-based access and encryption....

- [Permission] (0.55) ACCESS CONTROL

2.1....

- [Permission] (0.93) All access to Confidential and Restricted data requires explicit approval from t...
- [Permission] (0.98) Access permissions must be reviewed quarterly....
- [Permission] (0.98) Unused permissions are automatically revoked after 90 days of inactivity.

2.3....

- [Permission] (0.51) Service accounts accessing Restricted data must use short-lived tokens (maximum ...
- [Permission] (0.45) Cross-department data sharing requires a Data Sharing Agreement (DSA) signed by ...
- [Permission] (0.37) Quarterly access audits must be completed and reported to the Governance Committ...

Other (3)

- [Trigger Rule] (0.86) Anomalous access patterns must trigger automated alerts within 5 minutes.

4.4....

- [Trigger Rule] (0.88) Data quality scores below 95% trigger a remediation plan within 10 business days...
- [State Change] (0.63) Master data changes require approval from the data steward and validation agains...

Programming (1)

- [Type] (0.71) DATA CLASSIFICATION All data assets must be classified into one of four sensitiv...

Scientific (2)

- [Variable] (0.50) Internal: Business data for internal use....
- [Variable] (0.22) 5. DATA QUALITY

5.1....

State (23)

- [Condition] (0.24) DATA GOVERNANCE POLICY — Effective January 1, 2026

....

- [Condition] (0.24) DATA GOVERNANCE POLICY — Effective January 1, 2026

1....

- [Condition] (0.18) No access restrictions....
- [Condition] (0.22) Requires employee authentication....
- [Condition] (0.29) Confidential: Sensitive business data....
- [Condition] (0.36) Restricted: Highly sensitive data (PII, PHI, financial records)....
- [Condition] (0.20) 2.4....
- [Condition] (0.26) 3. DATA RETENTION

3.1....

Qualtron: 8 classifications (12760ms)

Causal (4)

- [Requirement] (0.95) Requires encryption at rest and in transit, audit logging, and approval from the...
- [Requirement] (0.95) Service accounts accessing Restricted data must use short-lived tokens (maximum ...
- [Requirement] (0.95) Backup data must follow the same retention rules as the source data....
- [Requirement] (0.95) All production datasets must have documented data quality rules....

Other (2)

- [Document] (0.95) DATA GOVERNANCE POLICY — Effective January 1, 2026...
- [Section] (0.95) 3.2....

Programming (1)

- [Invariant] (0.95) Audit logs are immutable and must be retained for 5 years....

State (1)

- [Condition] (0.95) Internal: Business data for internal use....

QHG Process Models

FSM (2650ms)

Could not parse structured output

```
None
```

BPMN (3274ms)

Could not parse structured output

```
None
```

DFG (2438ms)

Could not parse structured output

```
None
```

KnowledgeState (8229ms)

Could not parse structured output

```
None
```

CNL / QNR2 Rules

18 rule patterns detected

Obligation (15)

- `must be classified into one of four sensitivity levels:`
- `must be reviewed quarterly.`
- `must use short-lived tokens (maximum 1 hour).`
- `must be deleted within 30 days of account closure unless legally required.`
- `must be retained for a minimum of 7 years per regulatory requirements.`

Requirement (3)

- `requires explicit approval from the`
- `requires a Data Sharing Agreement (DSA) signed by both`
- `require approval from the data steward and validation against`

Heuristic Facts & Rules

2 facts, 15 rules

Facts (sample)

- revoked after 90 days of inactivity....
- 5.2. Data quality scores below 95% trigger a remediation plan within 10 business days....

Rules (sample)

- [obligation] All data assets must be classified into one of four sensitivity levels:...
- [obligation] 2.2. Access permissions must be reviewed quarterly. Unused permissions are automatically...
- [obligation] 2.3. Service accounts accessing Restricted data must use short-lived tokens (maximum 1 hour)....
- [obligation, conditional, quantified] 3.1. Customer PII must be deleted within 30 days of account closure unless legally required....
- [obligation, quantified] 3.2. Financial records must be retained for a minimum of 7 years per regulatory requirements....
- [obligation, quantified] 3.3. System logs must be retained for 12 months and archived for an additional 24 months....
- [obligation] 3.4. Backup data must follow the same retention rules as the source data....
- [obligation] 4.1. All access to Restricted data must be logged with user identity, timestamp, action,...

Topics (Gensim LDA)

- **Topic 0**: {'word': 'data', 'weight': 0.0667}, {'word': 'restricted', 'weight': 0.0282}, {'word': 'within', 'weight': 0.0282}, {'word': 'requires', 'weight': 0.0154}, {'word': 'confidential', 'weight': 0.0154}, {'word': 'pii', 'weight': 0.0154}, {'word': 'use', 'weight': 0.0154}, {'word': 'privacy', 'weight': 0.0154}
- **Topic 1**: {'word': 'retained', 'weight': 0.0274}, {'word': 'quality', 'weight': 0.0273}, {'word': 'years', 'weight': 0.0273}, {'word': 'within', 'weight': 0.0273}, {'word': 'per', 'weight': 0.0273}, {'word': 'requirements', 'weight': 0.0273}, {'word': 'financial', 'weight': 0.0273}, {'word': 'records', 'weight': 0.0273}
- **Topic 2**: {'word': 'data', 'weight': 0.111}, {'word': 'rules', 'weight': 0.025}, {'word': 'access', 'weight': 0.025}, {'word': 'business', 'weight': 0.025}, {'word': 'requires', 'weight': 0.025}, {'word': 'months', 'weight': 0.0172}, {'word': 'approval', 'weight': 0.0172}, {'word': 'restricted', 'weight': 0.0172}
- **Topic 3**: {'word': 'data', 'weight': 0.0763}, {'word': 'department', 'weight': 0.0524}, {'word': 'sharing', 'weight': 0.0524}, {'word': 'officer', 'weight': 0.0286}, {'word': 'requires', 'weight': 0.0286}, {'word': 'signed', 'weight': 0.0286}, {'word': 'chief', 'weight': 0.0286}, {'word': 'cross', 'weight': 0.0286}
- **Topic 4**: {'word': 'access', 'weight': 0.0552}, {'word': 'within', 'weight': 0.038}, {'word': 'days', 'weight': 0.0379}, {'word': 'permissions', 'weight': 0.0379}, {'word': 'quarterly', 'weight': 0.0379}, {'word': 'completed', 'weight': 0.0379}, {'word': 'reported', 'weight': 0.0207}, {'word': 'breach', 'weight': 0.0207}

Summary (Sumy LexRank)

No access restrictions. Requires encryption at rest and in transit, audit logging, and approval from the Data Privacy Officer. All access to Confidential and Restricted data requires explicit approval from the data owner and the Data Privacy Officer. Quarterly access audits must be completed and reported to the Governance Committee. Master data changes require approval from the data steward and validation against business rules before propagation.