ComGen Annotation Manual for CBCMS+

Project Team: zhuangzhixian22s@ict.ac.cn

2025.3.16

Contents

1 Introduction

This annotation manual is designed to provide detailed guidance for annotating data in the context of the Compliance Policy Generator (ComGen). Unlike the Enhanced PDL Mapping Pipeline (EpMap), which maps unstructured legal text into structured Policy Definition Language (PDL) formats, ComGen focuses on generating compliance policies based on policy contexts such as data category, sensitivity, and jurisdictional scope.

The purpose of this manual is to standardize the annotation process, ensuring consistency and accuracy across all annotated datasets. Each PDL field is outlined with a clear definition, detailed annotation rules, illustrative examples, and common mistakes to avoid. Annotators are expected to use this manual to identify and annotate the appropriate PDL paths for a given set of input features, enabling ComGen to effectively generate compliance policies for diverse and dynamic data processing scenarios in global software development.

This structured approach not only ensures high-quality annotations but also facilitates the training and evaluation of ComGen, contributing to its high accuracy and efficiency in real-world compliance applications.

2 Annotation Guidelines

2.1 Action -> Security Measures -> Encryption

Definition: This field specifies the application of encryption techniques to secure data during processing or transfer.
Annotation Rules:

- Annotate this field when data sensitivity is high, such as financial data or health records.
- Encryption is typically required when the source or target domain has strict data security requirements.

Example:

- Input Metadata: Sensitivity: High; Data Category: Financial Data; Source: EU; Target: US.
- Annotation: `Action -> Security Measures -> Encryption`

Common Mistakes:

- Annotating low-sensitivity data (e.g., public data) unless explicitly required by jurisdictional rules.

2.2 Action -> Security Measures -> Pseudonymization

Definition: This field specifies the replacement of identifiable information with pseudonyms to protect personal data.
Annotation Rules:

- Annotate this field when processing sensitive data categories (e.g., health data) in jurisdictions requiring privacy-by-design.
- Ensure that pseudonymization is relevant to the data category and processing purpose.

Example:

- Input Metadata: Sensitivity: Medium; Data Category: Health Data; Source: EU; Target: US.
- Annotation: `Action -> Security Measures -> Pseudonymization`

Common Mistakes:

- Confusing pseudonymization with encryption; pseudonymization does not completely anonymize the data.

2.3 Action -> Security Measures -> Access Control

Definition: This field specifies measures to limit data access to authorized individuals or systems.
Annotation Rules:

- Annotate this field when the jurisdiction mandates strict role-based access to sensitive data.
- Ensure the metadata includes roles or access levels tied to the data.

Example:

- Input Metadata: Sensitivity: High; Data Category: Personal Data; Source: EU; Target: US.
- Annotation: `Action -> Security Measures -> Access Control`

Common Mistakes:

- Failing to annotate when jurisdictions explicitly require access control.

2.4 Action -> Security Measures -> Audit Logs

Definition: This field specifies the maintenance of audit logs to track data access and processing activities.
Annotation Rules:

- Annotate this field for high-sensitivity data or when processing is subject to regulatory audits.
- Audit logs are particularly relevant for compliance with cross-border regulations.

Example:

- Input Metadata: Sensitivity: High; Data Category: Financial Data; Source: US; Target: EU.
- Annotation: `Action -> Security Measures -> Audit Logs`

Common Mistakes:

- Omitting this field for jurisdictions that explicitly require audit trails.

2.5 Action -> Security Measures -> Data Masking

Definition: This field specifies the masking of sensitive data to limit its visibility during processing or sharing.
Annotation Rules:

- Annotate when data sharing involves third parties or non-secure environments.
- Focus on sensitive data categories, such as health or financial data.

Example:

- Input Metadata: Sensitivity: High; Data Category: Health Data; Source: EU; Target: Non-EU.
- Annotation: `Action -> Security Measures -> Data Masking`

Common Mistakes:

- Confusing masking with encryption; masking limits visibility, not securing data in transit.

2.6 Action -> Security Measures -> Testing

Definition: This field specifies the periodic testing of security measures to ensure their effectiveness.
Annotation Rules:

- Annotate when data sensitivity is high or when processing involves high-risk activities.
- Ensure testing is part of the jurisdiction's data protection requirements.

Example:

- Input Metadata: Sensitivity: High; Data Category: Financial Data; Source: EU; Target: US.
- Annotation: `Action -> Security Measures -> Testing`

Common Mistakes:

- Annotating without a clear requirement for security testing.

2.7 Action -> Security Measures -> Evaluation

Definition: This field specifies the regular evaluation of data security measures for effectiveness and compliance.
Annotation Rules:

- Annotate for jurisdictions that require periodic security evaluations.
- Ensure evaluation is linked to compliance requirements.

Example:

- Input Metadata: Sensitivity: Medium; Data Category: Personal Data; Source: US; Target: EU.
- Annotation: `Action -> Security Measures -> Evaluation`

Common Mistakes:

- Confusing evaluation with operational activities like access control or encryption.

2.8 Action -> Data Subject Rights -> Access

Definition: This field specifies the right of data subjects to access their personal data.
Annotation Rules:

- Annotate when data processing involves jurisdictions with strong user rights, such as the EU.
- Ensure the purpose explicitly states access rights.

Example:

- Input Metadata: Data Category: Personal Data; Source: EU; Target: US.
- Annotation: `Action -> Data Subject Rights -> Access`

Common Mistakes:

- Annotating for non-personal data.

2.9 Action -> Data Subject Rights -> Erasure

Definition: This field specifies the right of data subjects to request the erasure of their data.
Annotation Rules:

- Annotate when jurisdictions enforce data subject rights related to deletion.
- Ensure the metadata supports erasure requirements.

Example:

- Input Metadata: Data Category: Personal Data; Source: EU; Target: Non-EU.
- Annotation: `Action -> Data Subject Rights -> Erasure`

Common Mistakes:

- Confusing erasure with data masking.

2.10 Action -> Data Subject Rights -> Rectification

Definition: This field specifies the right of data subjects to correct inaccurate or incomplete data.
Annotation Rules:

- Annotate when jurisdictions prioritize data accuracy and allow corrections.
- Ensure rectification is applicable to the data category.

Example:

- Input Metadata: Data Category: Personal Data; Source: US; Target: EU.
- Annotation: `Action -> Data Subject Rights -> Rectification`

Common Mistakes:

- Failing to annotate when rectification is explicitly required.

2.11 Action -> Data Subject Rights -> Restriction

Definition: This field specifies the right of data subjects to restrict the processing of their personal data.
Annotation Rules:

- Annotate when the jurisdiction explicitly provides restriction rights (e.g., during disputes over data accuracy).
- Ensure restriction is relevant to the data category and jurisdiction.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: US.
- Annotation: `Action -> Data Subject Rights -> Restriction`

Common Mistakes:

- Confusing restriction with erasure or rectification.

2.12 Action -> Data Subject Rights -> Portability

Definition: This field specifies the right of data subjects to transfer their personal data to another controller.
Annotation Rules:

- Annotate when jurisdictions explicitly require data portability, often in consumer-facing scenarios.
- Ensure portability is relevant to the data type (e.g., financial or service data).

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: Non-EU.
- Annotation: `Action -> Data Subject Rights -> Portability`

Common Mistakes:

- Annotating data that is not suitable for transfer (e.g., aggregated data).

2.13 Action -> Data Subject Rights -> Object

Definition: This field specifies the right of data subjects to object to the processing of their personal data.
Annotation Rules:

- Annotate when jurisdictions enforce the right to object, especially for marketing or profiling activities.
- Ensure the purpose explicitly allows objections.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Low; Processing Purpose: Marketing; Source: EU; Target: US.
- Annotation: `Action -> Data Subject Rights -> Object`

Common Mistakes:

- Confusing objection with withdrawal of consent.

2.14 Action -> Data Subject Rights -> Withdraw

Definition: This field specifies the right of data subjects to withdraw previously given consent.
Annotation Rules:

- Annotate when data processing is based on consent, and the jurisdiction allows withdrawal.
- Ensure withdrawal applies specifically to consent-based activities.

Example:

- Input Metadata: Data Category: Personal Data; Processing Purpose: Service; Source: EU; Target: US.
- Annotation: `Action -> Data Subject Rights -> Withdraw`

Common Mistakes:

- Annotating non-consent-based processing activities.

2.15 Action -> Data Retention -> —

Definition: This field specifies the duration for which data is retained before being deleted.
Annotation Rules:

- Annotate when jurisdictions provide explicit data retention requirements.
- If no specific duration is provided, leave as —.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: US; Target: EU.
- Annotation: `Action -> Data Retention -> ---`

Common Mistakes:

- Assigning a retention period when none is specified in the metadata.

2.16 Action -> Compliance Requirements -> Minimization

Definition: This field specifies the principle of data minimization, ensuring only necessary data is collected and processed.
Annotation Rules:

- Annotate when jurisdictions explicitly require minimizing data collection and usage.
- Ensure the principle applies to high-sensitivity or unnecessary data.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Processing Purpose: Compliance; Source: EU; Target: Non-EU.
- Annotation: `Action -> Compliance Requirements -> Minimization`

Common Mistakes:

- Confusing minimization with purpose limitation or retention.

2.17 Action -> Compliance Requirements -> Purpose Limitation

Definition: This field specifies that data should only be used for the purposes explicitly stated at the time of collection.
Annotation Rules:

- Annotate when jurisdictions enforce clear restrictions on data usage purposes.
- Ensure the purpose aligns with metadata-defined processing activities.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Processing Purpose: Service; Source: EU; Target: US.
- Annotation: `Action -> Compliance Requirements -> Purpose Limitation`

Common Mistakes:

- Annotating activities unrelated to purpose limitation.

2.18 Action -> Compliance Requirements -> Transparency

Definition: This field specifies the requirement for transparency in data processing, ensuring data subjects are informed about how their data is used.
Annotation Rules:

- Annotate when jurisdictions mandate clear and accessible privacy notices.
- Ensure transparency aligns with user rights and data categories.

Example:

- Input Metadata: Data Category: Personal Data; Source: EU; Target: Non-EU.
- Annotation: `Action -> Compliance Requirements -> Transparency`

Common Mistakes:

- Annotating for metadata unrelated to user-facing communications.

2.19 Action -> Compliance Requirements -> Accuracy

Definition: This field specifies that data must be accurate and kept up-to-date.
Annotation Rules:

- Annotate when jurisdictions prioritize data accuracy for compliance.
- Ensure accuracy aligns with metadata indicating potential errors or updates.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: US; Target: EU.
- Annotation: `Action -> Compliance Requirements -> Accuracy`

Common Mistakes:

- Confusing accuracy with retention or minimization principles.

2.20 Action -> Third Party Sharing -> —

Definition: This field specifies data-sharing activities involving third parties.
Annotation Rules:

- Annotate when data sharing involves external entities, ensuring compliance with jurisdiction-specific requirements.
- If no specific sharing conditions are mentioned, leave as —.

Example:

- Input Metadata: Data Category: Health Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Action -> Third Party Sharing -> ---`

Common Mistakes:

- Annotating when no sharing is indicated in the metadata.

2.21 Action -> Breach Notifications -> Immediate

Definition: This field specifies the requirement for immediate notification in the event of a data breach.
Annotation Rules:

- Annotate when jurisdictions mandate immediate notification to data subjects or authorities.
- Ensure the breach notification requirement explicitly mentions urgency.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Action -> Breach Notifications -> Immediate`

Common Mistakes:

- Confusing "immediate" with other timelines, such as "within 72 hrs."

2.22 Action -> Breach Notifications -> Within 72 hrs

Definition: This field specifies the requirement for breach notifications to be sent within 72 hours.
Annotation Rules:

- Annotate when legal requirements mention a 72-hour notification window.
- Ensure the timeline is explicitly stated in the jurisdiction.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: EU; Target: Non-EU.
- Annotation: `Action -> Breach Notifications -> Within 72 hrs`

Common Mistakes:

- Annotating when no specific timeline is provided.

2.23 Action -> Breach Notifications -> Periodic Reporting

Definition: This field specifies the requirement for periodic reporting of breach incidents.
Annotation Rules:

- Annotate when jurisdictions require periodic breach updates or summaries.
- Ensure periodic reporting is explicitly mandated by the jurisdiction.

Example:

- Input Metadata: Data Category: Health Data; Sensitivity: Medium; Source: US; Target: EU.
- Annotation: `Action -> Breach Notifications -> Periodic Reporting`

Common Mistakes:

- Confusing periodic reporting with immediate or one-time notifications.

2.24 Action -> Data Localization -> —

Definition: This field specifies the requirement for data to be stored within specific jurisdictions.
Annotation Rules:

- Annotate when legal frameworks mandate data storage within certain regions.
- If no localization requirement is mentioned, leave as —.

Example:

- Input Metadata: Data Category: Sensitive Data; Sensitivity: High; Source: China; Target: US.
- Annotation: `Action -> Data Localization -> ---`

Common Mistakes:

- Annotating when no explicit localization rules are applicable.

2.25 Liability -> Accountability -> Controller

Definition: This field specifies the accountability obligations of the data controller.
Annotation Rules:

- Annotate when jurisdictions assign accountability explicitly to controllers.
- Ensure accountability involves decision-making responsibilities.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: Non-EU.
- Annotation: `Liability -> Accountability -> Controller`

Common Mistakes:

- Annotating processors or joint controllers as "Controller."

2.26 Liability -> Accountability -> Processor

Definition: This field specifies the accountability obligations of the data processor.
Annotation Rules:

- Annotate when jurisdictions assign accountability explicitly to processors.
- Ensure the processor acts on behalf of a controller.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Low; Source: US; Target: EU.
- Annotation: `Liability -> Accountability -> Processor`

Common Mistakes:

- Confusing processors with independent controllers.

2.27 Liability -> Accountability -> Joint Controller

Definition: This field specifies the accountability obligations of joint controllers.
Annotation Rules:

- Annotate when jurisdictions assign shared accountability to multiple controllers.
- Ensure joint decision-making responsibilities are explicitly stated.

Example:

- Input Metadata: Data Category: Health Data; Sensitivity: High; Source: EU; Target: Non-EU.
- Annotation: `Liability -> Accountability -> Joint Controller`

Common Mistakes:

- Annotating individual controllers as "Joint Controller."

2.28 Liability -> Accountability -> Supervisory Authority

Definition: This field specifies the accountability obligations of supervisory authorities.
Annotation Rules:

- Annotate when jurisdictions assign specific roles to regulatory authorities.
- Ensure supervisory authorities are clearly mentioned in the metadata.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: Non-EU.
- Annotation: `Liability -> Accountability -> Supervisory Authority`

Common Mistakes:

- Confusing supervisory authorities with controllers or processors.

2.29 Liability -> Supervision Requirements -> Mandatory Reporting

Definition: This field specifies the requirement for mandatory reporting to authorities.
Annotation Rules:

- Annotate when jurisdictions mandate regular or ad-hoc reporting to authorities.
- Ensure the reporting requirement explicitly aligns with legal obligations.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: EU; Target: Non-EU.
- Annotation: `Liability -> Supervision Requirements -> Mandatory Reporting`

Common Mistakes:

- Annotating when reporting is optional or periodic.

2.30 Liability -> Supervision Requirements -> Periodic Audits

Definition: This field specifies the requirement for periodic audits to ensure compliance.
Annotation Rules:

- Annotate when legal frameworks require periodic auditing of data practices.
- Ensure the frequency of audits is aligned with jurisdictional requirements.

Example:

- Input Metadata: Data Category: Health Data; Sensitivity: Medium; Source: US; Target: EU.
- Annotation: `Liability -> Supervision Requirements -> Periodic Audits`

Common Mistakes:

- Annotating ad-hoc reporting requirements as "Periodic Audits."

2.31 Liability -> Supervision Requirements -> Incident Reporting

Definition: This field specifies the requirement for reporting data-related incidents to relevant authorities.
Annotation Rules:

- Annotate when jurisdictions mandate incident reporting in the event of a data breach or non-compliance.
- Ensure the incident reporting requirement explicitly aligns with jurisdictional obligations.

Example:

- Input Metadata: Data Category: Sensitive Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Liability -> Supervision Requirements -> Incident Reporting`

Common Mistakes:

- Confusing periodic audits with incident-specific reporting.

2.32 Liability -> Data Protection Impact Assessment -> —

Definition: This field captures the requirement for conducting data protection impact assessments (DPIAs).
Annotation Rules:

- Annotate when legal frameworks mandate DPIAs for high-risk processing activities.
- If no specific assessment is required, leave as —.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Liability -> Data Protection Impact Assessment -> ---`

Common Mistakes:

- Annotating assessments unrelated to data protection.

2.33 Extension -> Custom Clauses -> Targeted Advertising Restrictions

Definition: This field specifies restrictions on using data for targeted advertising purposes.
Annotation Rules:

- Annotate when jurisdictions or user preferences explicitly restrict targeted advertising.

- Ensure the restriction specifically mentions advertising activities.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: US.
- Annotation: `Extension -> Custom Clauses -> Targeted Advertising Restrictions`

Common Mistakes:

- Confusing advertising restrictions with general marketing limitations.

2.34 Extension -> Custom Clauses -> Financial Risk Assessment

Definition: This field specifies clauses related to assessing financial risks associated with data processing activities.
Annotation Rules:

- Annotate when legal or organizational requirements specify financial risk assessments.
- Ensure the clause explicitly mentions financial implications or risk evaluations.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Extension -> Custom Clauses -> Financial Risk Assessment`

Common Mistakes:

- Annotating general risk assessments unrelated to financial factors.

2.35 Extension -> Custom Clauses -> Data Protection Officer

Definition: This field specifies the requirement to appoint a Data Protection Officer (DPO).
Annotation Rules:

- Annotate when legal frameworks mandate the appointment of a DPO.
- Ensure the requirement is explicitly mentioned in the jurisdictional guidelines.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Extension -> Custom Clauses -> Data Protection Officer`

Common Mistakes:

- Annotating optional roles as mandatory DPO requirements.

2.36 Extension -> Custom Clauses -> Cookie Consent

Definition: This field specifies the requirement to obtain explicit user consent for using cookies.
Annotation Rules:

- Annotate when jurisdictions or user preferences explicitly require cookie consent.
- Ensure the consent requirement specifically mentions cookies.

Example:

- Input Metadata: Data Category: Web Data; Sensitivity: Medium; Source: EU; Target: US.
- Annotation: `Extension -> Custom Clauses -> Cookie Consent`

Common Mistakes:

- Confusing cookie consent with broader data consent requirements.

2.37 Extension -> Cross Border Compliance -> —

Definition: This field specifies compliance requirements for cross-border data transfers.
Annotation Rules:

- Annotate when legal frameworks explicitly address cross-border compliance.
- If no specific compliance requirement is mentioned, leave as —.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Extension -> Cross Border Compliance -> ---`

Common Mistakes:

- Annotating generic compliance requirements unrelated to cross-border transfers.

2.38 Extension -> Sanctions -> Fines

Definition: This field specifies sanctions in the form of fines for non-compliance.
Annotation Rules:

- Annotate when legal frameworks impose monetary penalties for violations.
- Ensure the fine-related sanctions are explicitly stated.

Example:

- Input Metadata: Data Category: Sensitive Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Extension -> Sanctions -> Fines`

Common Mistakes:

- Confusing fines with other forms of sanctions, such as legal proceedings.

2.39 Extension -> Sanctions -> Legal Proceedings

Definition: This field specifies sanctions involving legal proceedings for non-compliance.
Annotation Rules:

- Annotate when legal frameworks allow initiating court actions for violations.
- Ensure the proceedings are explicitly mentioned as part of the compliance requirements.

Example:

- Input Metadata: Data Category: Personal Data; Sensitivity: Medium; Source: EU; Target: US.
- Annotation: `Extension -> Sanctions -> Legal Proceedings`

Common Mistakes:

- Annotating monetary fines as "Legal Proceedings."

2.40 Extension -> Sanctions -> Operational Suspension

Definition: This field specifies sanctions involving operational suspension for non-compliance.
Annotation Rules:

- Annotate when legal frameworks allow suspension of business operations for violations.
- Ensure the suspension is explicitly mentioned as a possible consequence.

Example:

- Input Metadata: Data Category: Financial Data; Sensitivity: High; Source: EU; Target: US.
- Annotation: `Extension -> Sanctions -> Operational Suspension`

Common Mistakes:

- Confusing operational suspension with monetary penalties or legal actions.

3 Acknowledgments

This manual has been designed to ensure consistent and high-quality annotations for ComGen. For questions or clarifications, contact the project team: zhuangzhixian22s@ict.ac.cn.