CDSCO Health Innovation Acceleration Hackathon: FAQs

1. **Is it permissible to develop the solution from a consumer-centric perspective?**
   Solutions must be designed from a regulatory perspective; the primary objective is to enhance regulator efficiency, which subsequently benefits the patient.

2. **Are applicants permitted to focus exclusively on specific use cases?**
   While addressing all six use cases as a comprehensive module is the preferred approach, focused submissions targeting individual problem statements are acceptable during this initial phase.

3. **Is a separate report required for each use case during Stage 1?**
   Yes. Each use case has a specific, predefined output format to which all solutions must strictly adhere.

4. **Which infrastructure will the solution be hosted on?**
   Solutions must be hosted on CDAC servers, specifically the CDAC Data Center Airawat located at CDAC Pune.

5. **Are the templates required for verification available to the public?**
   Yes. All necessary checklist items, forms, and documentation requirements are accessible via the Sugam portal.

6. **How does the submission deadline relate to the 45-day timeline?**
   The final date for submission is May 10, 2026. The 5-day on-premises hackathon (Stage 2) is a separate phase that follows the conclusion of the 45-day build period.

7. **What specific elements should be included in the demonstration video?**
   The demonstration must showcase the outputs for all six use cases in their respective predefined formats.

8. **Is it permissible to use synthetic training data derived from research-exclusive datasets?**
   Participants are required to utilize only publicly available regulatory domain datasets

9. **Does the solution need to encompass drugs, cosmetics, and medical devices?**
   The solution is required to cover all three verticals as managed within the Sugam and Medical Device portals.

10. **Is the execution of a Non-Disclosure Agreement (NDA) mandatory?**
    A signed NDA is required prior to being granted access to restricted regulatory and applicant data.

11. **Can model access be restricted to a specific ID rather than being hosted on a public GitHub repository?**
    Participants may add ‘indiaaihackathon25' as a collaborator on a private GitHub repository. Alternatively, source code may be submitted via email to fellow3.gpai-india@meity.gov.in using the subject line: "CDSCO-IndiaAI Health Innovation Acceleration Hackathon – Applicant Name - Source code."

12. **Will GPU server resources be provided for the duration of the hackathon?**
    Access to CDAC Param Rudra at CDAC Pune will be provided during the Stage 2 on-premises round.

13. **What are the guidelines regarding the use of Indian versus foreign open-source AI models?**
    Participants are encouraged to use Indian open source models. If foreign models are utilized, they must be hosted locally to ensure the security and confidentiality of data.

14. **Question: Where can one find the templates for the expected output formats?**
    Yes. All necessary checklist items, forms, and documentation requirements are accessible via the Sugam portal. Output templates are also attached in Annexure – 1 to this document.

Annexure I: Output Formats

Use Case 1: AI-powered data anonymization tool (web or API-based)

Participants are required to build a web-based or API-based tool that detect and classify Personally Identifiable Information (PII) or Protected Health Information (PHI) in both structured and unstructured data using a hybrid rule-based and NLP approach for contextualised anonymization.

Implement a two-step process: de-identification (Pseudonymization), replace identifiers with secure tokens and irreversible anonymization, generalize and normalize sensitive information, in compliance with Indian and International data protection laws and guidelines (DPDP Act 2023, NDHM, ICMR, and CDSCO guidelines) with the ability to streamline CDSCO's regulatory evaluation using AI to achieve faster, more consistent approvals and enhance public health access.

Illustration:

The example below illustrates how the tool should process a typical SAE narrative:

"Patient Sunita Devi, Aadhar 9876-5432-1012, 38F, diagnosed with Type 2 Diabetes, admitted to Max Hospital, New Delhi on 02-Jan-2026 following severe allergic reaction to trial drug. Treating physician Dr. Rakesh Gupta reported the event. Sponsor: ABC Pharma Pvt Ltd."

2-step process:

Output after Step 1 (De-identification): *"SUBJ-1103, AADHAR-TOKEN-7741, AGE-SEX-TOKEN-082, diagnosed with COND-TOKEN-019, admitted to SITE-TOKEN-022 on DATE-TOKEN-A1 following severe allergic reaction to trial drug. INV-TOKEN-014 reported the event. SPONSOR-TOKEN-07."*

Output after Step 2 (Irreversible Anonymization)**:** *"Patient, 35–40F, diagnosed with a metabolic condition, admitted to a hospital in North India in early 2026 following severe allergic reaction to trial drug. Treating physician reported the event. Sponsor details suppressed."*

Use Case 2: Document Summarization

Participants are required to build a system that extracts and synthesizes critical regulatory information from diverse, high-volume source documents and converts it into a precise, standardized text format for reviewer consumption. The system must handle three distinct types of source material:

I. **Application data corresponding to each item across multiple checklists on the SUGAM portal**

II. **SAE case narration to facilitate efficient case-resolution workflows**

III. **Meeting summaries to synthesize key decisions, actionable items and next steps from lengthy meeting transcripts or audio files**

The final output across all document types must be a concise, standardized text format that enables due diligence and decision-making by Reviewing Officers.

Illustrative Summary Format 1: Application Data - Module 3 (Quality, 3.2.P.8 - Stability)

For each module-level checklist item on SUGAM, the system should produce a structured summary in the following format:

Summary Structure:

1. Study Design / Conditions

- Long-term, intermediate, and accelerated conditions (e.g., 25°C/60% RH, 30°C/65% RH, 40°C/75% RH)
- Photostability / stress studies (if done)
- Packaging system tested (e.g., PVdC-Alu blister, HDPE bottle)

2. Duration & Sampling Points

- Time-points (0, 3, 6, 9, 12, 18, 24, 36 months)
- State whether studies are ongoing or completed

3. Batch Information

- Number of batches on stability (pilot/commercial)
- Batch size, manufacturing scale, site
- Strengths/formulations tested (5, 10, 20 mg)

4. Parameters Tested

- Physical appearance (colour, integrity, coating)
- Assay / potency

- Degradation products / impurities
- Dissolution
- Moisture content (if relevant)
- Microbiological attributes (if relevant)

5. Results & Interpretation

- Data trends across time and conditions
- Whether OOS/atypical results observed
- Shelf-life justified (e.g., 36 months at 25°C/60% RH)
- Recommended storage statement (e.g., "Store below 30°C, protect from light")

6. Supporting / Special Studies (if available)

- Photostability (ICH Q1B)
- In-use stability (multi-dose bottles)
- Stress studies to establish degradation pathways

Critical Requirements:

- Preserve all critical numerical data (storage conditions, time points, assay values, impurity levels, dissolution %, batch numbers, shelf-life duration)
- Retain all statistical data (regression analysis, confidence intervals, degradation rates)
- Include all regulatory compliance indicators (ICH Q1A(R2), Q1B compliance status)
- Maintain technical precision in terminology (use exact analytical method names, specification limits, acceptance criteria)
- Preserve risk assessments and justifications for storage conditions and shelf-life

Regulatory Compliance Focus:

- Ensure alignment with ICH Q1A(R2) - Stability Testing, Q1B - Photostability, and Q1E - Evaluation of Stability Data
- Verify compliance with NDCT Rules for stability data
- Confirm adherence to WHO stability testing guidelines for tropical zones (Zone IV conditions)
- Include statistical evaluation methods as per regulatory expectations

Language and Style Requirements:

- Use clear, factual, regulator-ready language (avoid speculative or promotional terms)
- Employ precise pharmaceutical terminology (ICH-compliant nomenclature)
- Present definitive conclusions with supporting data
- Emphasize scientific rationale for decisions and recommendations

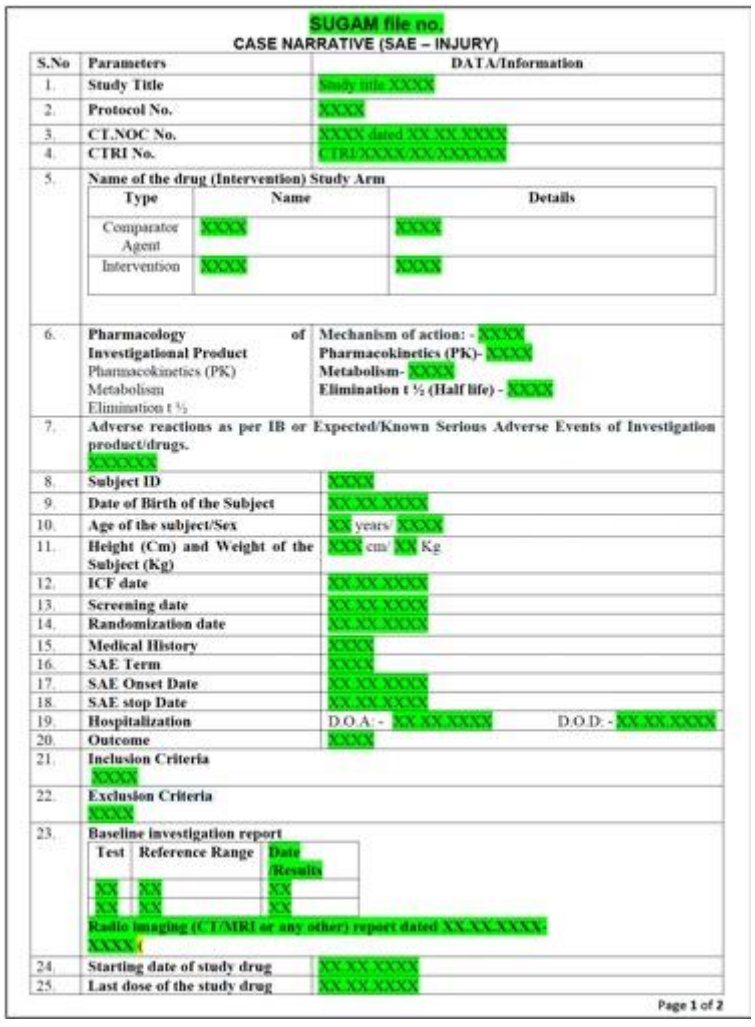
- Exclude procedural descriptions unless they impact product quality or regulatory compliance

Quality Standards:

- Technical Accuracy: All numerical values, units, and specifications must be exact
- Completeness: Each of the 6 sections must be populated with relevant data
- Regulatory Readiness: Summary must be suitable for direct inclusion in CDSCO dossier
- Scientific Integrity: Conclusions must be supported by presented data
- Risk Communication: Clearly identify any stability risks or limitations

Illustrative Summary Format 2: SAE Case Narration

The SAE case narration summary format is as given in the CDSCO SAE Case Narrative template:



SUGAM file no.

CASE NARRATIVE (SAE – INJURY)

S.No	Parameters	DATA/Information
1.	Study Title	Study title XXXX
2.	Protocol No.	XXXX
3.	CT.NOC No.	XXXX dated XX.XX.XXXX
4.	CTRI No.	CTRI/XXXX/XX/XXXXXX
5.	Name of the drug (Intervention) Study Arm	
6.	Pharmacology of Investigational Product Pharmacokinetics (PK) Metabolism Elimination t ½	Mechanism of action: - XXXX Pharmacokinetics (PK)- XXXX Metabolism- XXXX Elimination t ½ (Half life) - XXXX
7.	Adverse reactions as per IB or Expected/Known Serious Adverse Events of Investigation product/drugs. XXXXXX	
8.	Subject ID	XXXX
9.	Date of Birth of the Subject	XX.XX.XXXX
10.	Age of the subject/Sex	XX years/ XXXX
11.	Height (Cm) and Weight of the Subject (Kg)	XXX cm/ XX Kg
12.	ICF date	XX.XX.XXXX
13.	Screening date	XX.XX.XXXX
14.	Randomization date	XX.XX.XXXX
15.	Medical History	XXXX
16.	SAE Term	XXXX
17.	SAE Onset Date	XX.XX.XXXX
18.	SAE stop Date	XX.XX.XXXX
19.	Hospitalization	D.O.A: - XX.XX.XXXX D.O.D: - XX.XX.XXXX
20.	Outcome	XXXX
21.	Inclusion Criteria XXXX	
22.	Exclusion Criteria XXXX	
23.	Baseline investigation report	
24.	Starting date of study drug	XX.XX.XXXX
25.	Last dose of the study drug	XX.XX.XXXX

5. Name of the drug (Intervention) Study Arm

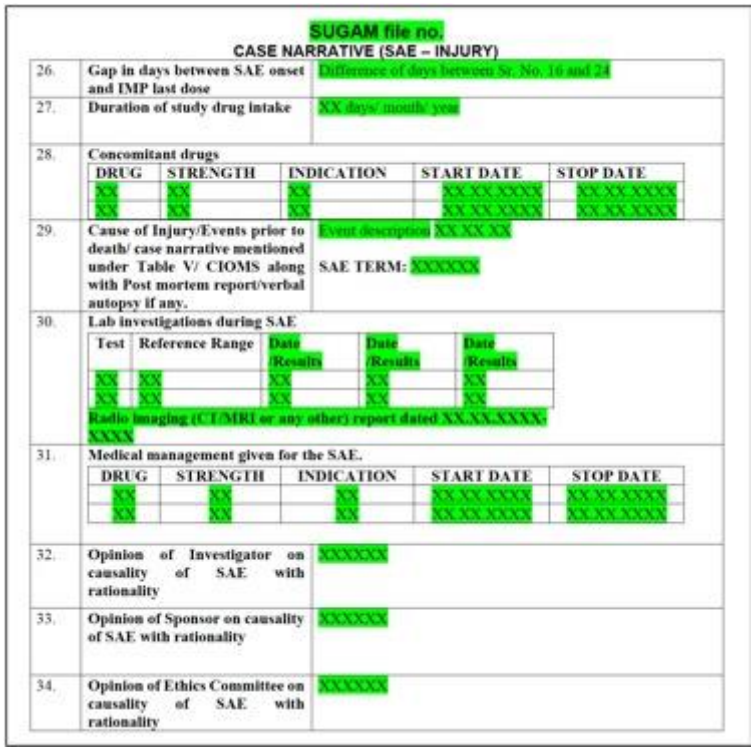
Type	Name	Details
Comparator Agent	XXXX	XXXX
Intervention	XXXX	XXXX

23. Baseline investigation report

Test	Reference Range	Date /Results
XX	XX	XX
XX	XX	XX

Radio imaging (CT/MRI or any other) report dated XX.XX.XXXX: XXXX.(

Page 1 of 2



SUGAM file no.

CASE NARRATIVE (SAE – INJURY)

26.	Gap in days between SAE onset and IMP last dose	Difference of days between Sr. No. 16 and 24
27.	Duration of study drug intake	XX days/ month/ year
28.	Concomitant drugs	

DRUG	STRENGTH	INDICATION	START DATE	STOP DATE
XX	XX	XX	XX.XX.XXXX	XX.XX.XXXX
XX	XX	XX	XX.XX.XXXX	XX.XX.XXXX

29.	Cause of Injury/Events prior to death/ case narrative mentioned under Table V/ CIOMS along with Post mortem report/verbal autopsy if any.	Event description XX.XX.XX SAE TERM: XXXXXX
30.	Lab investigations during SAE	

Test	Reference Range	Date /Results	Date /Results	Date /Results
XX	XX	XX	XX	XX
XX	XX	XX	XX	XX

Radio Imaging (CT/MRI or any other) report dated XX.XX.XXXX-XXXX

31.	Medical management given for the SAE.	

DRUG	STRENGTH	INDICATION	START DATE	STOP DATE
XX	XX	XX	XX.XX.XXXX	XX.XX.XXXX
XX	XX	XX	XX.XX.XXXX	XX.XX.XXXX

32.	Opinion of Investigator on causality of SAE with rationality	XXXXXX
33.	Opinion of Sponsor on causality of SAE with rationality	XXXXXX
34.	Opinion of Ethics Committee on causality of SAE with rationality	XXXXXX

Illustrative Summary Format 3: SEC Meeting Summary

For meeting transcripts or audio files, the system should produce a structured summary in the following format. The format differs depending on whether the meeting concerns a Clinical Trial (CT) or Marketing Authorisation (MA) application.

For CT:

1. Details about proposal: Brief description about the proposal to be discussed
    a) Application number
    b) Name of the applicant
    c) Name of the product/ device/ vaccine
    d) Therapeutic division
    e) Division handling (e.g., CGT/ IND etc.)
    f) Product details (route of administration, strength, dosage etc.)

2. Details about the deliberation: Upto max 200 words description on what is to be deliberated
    a) Objective of the study
    b) Details about the study protocol
        i) Number of participants
        ii) Inclusion and exclusion criteria
        iii) Number of sites
        iv) Any regulatory precedents cited by applicant or division
    c) Safety/ efficacy concerns raised, if any
    d) Risk based assessment

3. Reasons for CT approval / CT waiver/ CT protocol amendment - Justification for safety and efficacy

4. Recommendation (with conditions, if any)
    a) Approval or Rejection with reason

For MA:

1. Details about proposal: Brief description about the proposal to be discussed
    a) Application number
    b) Name of the applicant
    c) Name of the product/ device/ vaccine
    d) Product details (route of administration, strength, dosage etc.)

2. Details about the deliberation: Upto max 200 words description on what is to be deliberated
    a) Objective of the study conducted for MA
    b) Global approval status, if applicable
    c) Clinical literature/ study data
    d) Safety/efficacy concerns raised
    e) If any comparator products were referenced or meets analysis presented

3. Reasons for MA: Justification for safety and efficacy

4. Recommendation (with conditions, if any)
    a) Approval or Rejection with reasons and conditions
    b) Conditions (e.g., post-marketing surveillance, Phase IV, limited shelf-life, label change required etc.)

Use Case 3: Completeness Assessment

Participants are required to build a tool that verifies the completeness, consistency and accuracy of administrative sections, mandatory forms and regulatory checklists for each clinical approval application, as well as for SAE reports, with detailed flagging of missing or inconsistent fields.

Illustration:

The checklist for Form CT-04 (Application for Grant of Permission to Conduct Clinical Trial in Form CT-06) is shown below:

Checklist for Application made in Form CT-04 for grant of permission to conduct Clinical trial in Form CT-06

Checklist Point Number	Particular/ Heading
1.	**Module - Administrative/Legal Information**
1.1	Covering letter
1.2	Comprehensive table of contents (Modules 1 to 5)
1.3	Administrative information
1.3.1	**Legal and statutory documents (1.2.1, 1.2.2, 1.2.3, 1.2.4 for licence and approvals and 1.2.5, 1.2.6, 1.2.7, 1.2.8, 1.2.9, 1.2.10, 1.2.11 for legal documents pertaining to application)**
1.3.1.1	Copy of Form 11 for imported drug product
1.3.1.2	Form-29 for indigenous drug
1.3.1.3	Application in Form 44
1.3.1.4	Treasury Challan (fee)
1.3.1.5	Form12 & Challan Details
1.3.2.1	A copy of Certificate of Pharmaceutical Product (COPP) as per WHO GMP certification scheme for imported drug
1.3.2.2	A copy of Free Sale Certificate (FSC) from the country of origin for imported drug products.
1.3.2.3	A copy of Site Master File
1.3.2.4	Certificate of Analysis for three consecutive batches
1.3.2.5	Product Permission Document (PPD) as per Annex B
1.3.2.6	Copy of Form 11 for imported drug product
1.3.2.7	Clinical Trial no objection letters / approval
1.3.2.8	A copy of plant registration / approval certificate issued by the Ministry of Health / National Regulatory Authority of the country of origin.
1.3.2.9	A copy of approval, if any, showing the drug is permitted for manufacturing and/or marketing in the country of origin.
1.3.2	**Legal documents pertaining to application (to be notarized):**
1.3.3	Name, address, telephone, fax, e-mail of the authorized agent in India: (for imported drug products)
1.3.3.1	Name, address, telephone, fax, e-mail of manufacturer of drug product
1.3.3.2	Authorized Coordinates related to the application
1.3.3.3	Name, address, telephone, fax, e-mail of the manufacturing premises holding Market Authorization of the drug product (for imported drug products)
1.3.3.4	Name, address, telephone, fax, e-mail of manufacturer of drug substance
1.3.3.5	Name, address, telephone, fax, e-mail of other manufacturer(s) involved in the production process
1.3.3.6	Name, address, telephone, fax, e-mail of the responsible official
1.3.3.7	Non-proprietary name or common name of drug product
1.4.	Product Labeling (should conform to the specifications under the Drugs and Cosmetics Rules 1945)
1.4.1	Package insert (in English) Monograph for health professionals or information for prescription.
1.4.2	Secondary package label
1.4.3	Primary package label
1.4.4	Summary of product characteristics As per Annex C

Example Checks:

For a CT-04 application, the tool should perform checks such as the following:

- Covering Letter (Item 1.1): Verifies that a covering letter is present
- Comprehensive Table of Contents (Item 1.2): Verifies that the table of contents covers all 5 modules and that every document listed is actually present in the submission
- Treasury Challan (Item 1.3.1.4): Verifies that the fee payment proof is attached via Form 12 and Challan details are filled
- Clinical Trial No Objection Letter (Item 1.3.2.7): Verifies that the clinical trial no objection letter or approval is present and valid
- Authorized Agent Details (Item 1.3.3): Checks that name, address, telephone, fax and email of the authorized agent in India are fully populated (mandatory for imported drug products)

Any missing or inconsistent field is flagged with the specific item number, the nature of the deficiency and the action required, enabling the Reviewing Officer to send a targeted query to the applicant rather than reviewing the entire dossier manually.

Use Case 4: Classification Tool

Participants are required to build a tool that uses AI to classify SAE cases by severity, detect duplicates and generate recommendations for officer review. The objective is to enable data-driven workload allocation and faster resolution of critical cases.

Classification: Incoming SAE reports should be automatically classified into the following categories, as defined under NDCTR 2019:

- Death: Patient died during or following the trial
- Life-threatening: Adverse event that places the patient at immediate risk of death
- Inpatient hospitalisation: Admission to hospital as a result of the adverse event
- Prolonged hospitalisation: Extension of an existing hospital stay due to the adverse event
- Persistent or significant disability or incapacity: Substantial disruption to the patient's ability to carry out normal life functions
- Congenital anomaly or birth defect: Adverse event resulting in a structural or functional abnormality present at birth

Cases should then be ranked in a priority queue, with the most critical cases surfacing automatically for immediate review.

Auto Flag: The tool should automatically flags:

- Cases requiring immediate attention based on severity
- Duplicate submissions: same event submitted multiple times across sponsors or sites
- SAE reports with missing mandatory fields

Recommendations: For each classified case the tool suggests a recommended action based on severity category and case details.

Case ID	Severity	Priority	Duplicate	Recommended Action
SAE-2024-001	Death	Critical	No	Immediate review - request cause of death report from sponsor
SAE-2024-002	Hospitalisation	High	Yes	Merge with SAE-2024-198; seek follow-up on recovery status
SAE-2024-003	Minor	Low	No	Standard review queue

Use Case 5: Document Comparison

Participants are required to build a tool that identifies and highlights substantive and specific changes between different versions of an applicant's filing, producing a visualization or report detailing significant changes in text, data or tables across document versions.

It distinguishes between:

- **Minor changes** like formatting, typos, rephrasing with no regulatory significance
- **Substantive changes** like changes that affect safety, efficacy or quality claims

Illustration:

Version 1: *"The recommended dose of Compound-X is 10mg once daily for adult patients above 18 years."*
Version 2: *"The recommended dose of Compound-X is 20mg once daily for patients above 12 years."*
Changes detected:

- Dose: 10mg to 20mg *(substantive)*
- Age group: above 18 to above 12 *(substantive)*

Output Format (Illustrative)

Section	Change Type	Version 1	Version 2
Dosage	Substantive	10mg once daily	20mg once daily
Age group	Substantive	Above 18 years	Above 12 years

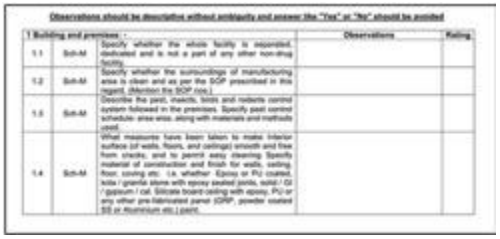
Use Case 6: Inspection Report Generation

Participants are required to build a tool that converts unstructured and handwritten observations gathered during site inspections into standardized, formal reports conforming to CDSCO inspection templates.

Illustration:

The CDSCO guidance document for risk-based inspections requires inspectors to record their remarks in multiple formats:

- Observations table: Checklist-based with remarks and ratings per item



- List of deficiencies: categorized as Critical, Major and Others

- Quality rating table: comparative ratings across drug categories

These are currently filled manually by DIs during and after site visits in handwritten notes, loose observations and informal formats, which then need to be converted into the structured templates.

Solution:

An AI-powered tool processes raw field observations and generates formal reports through the following steps:

Step 1: Digitize: OCR converts handwritten notes into machine-readable text

Step 2: Extract & classify: AI reads the digitized text and maps each observation to the relevant checklist item - categorizing as Critical, Major or Other

Step 3: Populate template: Extracted observations are automatically populated into the standard CDSCO inspection report template fields