

Standard Operating Procedure

on

Data-Driven Compliance Audit

Contents

1. **Objective**
2. **Scope**
3. **Core Directives for All Compliance Audits**
4. **Operational Structure**
5. **Conduct of Data-Driven Compliance Audits**
6. **Verticals of Data-Driven Compliance Audits**
7. **Implementation Plan**
8. **Annexure I: Data Ecosystem**
9. **Annexure II: Salary, Pension, Establishment Audit Case Study**
10. **Annexure III: Social Sector Audit Case Study**
11. **Annexure IV: Works Audit Case Study**
12. **Annexure V: Problem-Statement Based Audit**

Standard Operating Procedure on Compliance Audit

Applicability: All CAG State Audit and A&E offices
Version: SOP/CA/01 – Version 1.0
Effective Date / Review Cycle: Effective from 01.04.2065 – Review Annually

As per the Constitution of India, it is the duty of the Comptroller and Auditor General to provide **assurance** to the Legislature that **public funds have been spent for the purpose for which they were appropriated.** This assurance forms the **core objective of compliance auditing** and guides all audit planning and risk assessment activities.

1. Objective

This SOP seeks to standardize the methodology for **data-driven problem-statement-based compliance audits**. All compliance audits undertaken after this document comes into effect shall be planned and executed through such defined analytical and problem-centric approaches. It also aims to **move away from the conventional unit/DDO-based, generalized transaction audit method**, establishing data analysis and focused inquiry as the foundation of every compliance audit.

2. Scope

This SOP shall apply to **all compliance audits conducted by all state offices of the Comptroller and Auditor General of India (CAG)**. It is intended to be uniformly followed across sectors and departments, ensuring that every compliance audit - irrespective of subject or scale - is designed and executed through a **data-driven and problem-statement-based framework** as prescribed herein.

Background for the SOP: The Rich Data Ecosystem

At the outset, it is important to recognise that the government today has **heavily digitised the majority of its financial and administrative processes**, creating an extensive and credible data ecosystem. Platforms such as **Direct Benefit Transfer (DBT)**, **Public Financial Management System (PFMS)**, **Works and Asset Management Information Systems (WAMIS)**, and **Goods and Services Tax Network (GSTN)** have transformed the way funds, beneficiaries, and projects are tracked.

- DBT alone has processed over 18,600 crore transactions amounting to **₹44 lakh crore** for 176 crore beneficiaries, while PFMS has enabled real-time fund monitoring across thousands of agencies, disbursing over ₹20 lakh crore.

- State-level systems like WAMIS in Bihar and Maharashtra have digitised project approvals, e-measurements, and billing, improving accountability in works and infrastructure.

- Likewise, GSTN provides a continuous digital audit trail for more than **1.4 crore** taxpayers, with annual collections exceeding ₹22 lakh crore.

- Another major development has been the rollout of the **SNA–SPARSH framework**, which has brought complete transparency and granularity to fund flows under Centrally Sponsored Schemes (CSS). For the first time, transactional-level data of over **₹5.4 lakh crore** under CSS is available to Audit and A&E offices through the SNA–SPARSH dashboard. Earlier invisible to AG accounts, this data now provides a full financial footprint - linking every release to implementing agencies, down to the local body and payment account integrated with **PFMS**. CSS has thus become a co-driver of state expenditure analytics, enabling real-time tracking of central assistance across programmes and districts.

- A parallel transformation has occurred through the **Finance Commission fund tracking architecture**, where local bodies - Panchayati Raj Institutions and Urban Local Bodies - are now fully mapped to dedicated **Finance Commission-linked accounts** on PFMS. Every payment made from these accounts is digitally traceable to the project ID created at the PRI level, ensuring audit visibility for over **₹1.5 lakh crore** of annual grants and direct linkage between local governance and financial accountability.

- Additionally, the **MCA-21 database** under the Ministry of Corporate Affairs provides granular, entity-level information for more than **20 lakh registered companies**, covering filings, directors, ownership, and compliance histories. This dataset serves as a critical external reference point for verifying contractor authenticity, identifying shell entities, and cross-validating vendors engaged in public works and procurement audits.

With the wide-scale digitization of government functions, a vast and reliable data ecosystem has taken shape across departments and institutions. The need now is to make effective use of this data to strengthen the foundation of every compliance audit. This ecosystem spans three levels (detailed in Annexure-I):

- **Data within the IA&AD**, such as accounting compilations, VLC data, and inspection reports available with A&E and Audit offices;
- D**ata with government departments**, including beneficiary and payment databases, MIS/WAMIS records, and portal-based activity data;
- **Open-source data**, such as NITI Aayog indicators, MOSPI, Census, RBI, and SECC datasets that offer valuable context and validation.

At the state level, several audit offices have already begun leveraging departmental and financial datasets, demonstrating the potential of data analytics and cross-data intersection to strengthen assurance outcomes:

- **Uttar Pradesh:** Data analysis of a major social-sector scheme of about ₹600 crore flagged nearly **25 percent of transactions** as high-risk; field verification confirmed that around **85 percent** of these were **fraudulent, bogus, or ineligible**, establishing strong evidence for data-based selection.

- **Odisha, Jaipur, and Chitrakoot: Accounts and Entitlement (A&E) offices** analysed **salary and pension datasets** and uncovered **widespread irregularities** - including duplicate records, double draws, and irregular arrear payments - long-standing issues now traceable through structured data review.

- **Kolkata:** Analytical audit of two **women-oriented schemes** flagged **over 1,700 suspect beneficiaries**, revealing **suspected irregularities exceeding ₹5 crore**, through extensive dataset examination and cross-verification

These examples reflect the growing capacity of audit offices to operate within a digitised governance environment. With most government processes now digitised, field offices have begun adapting audit methods to match this transformation. The need now is to **scale and institutionalise these practices** across the country - ensuring that **data-driven and problem-anchored audits** become the standard approach. The present guidelines have been framed to enable this transition and embed analytical audit practices across all IA&AD offices

3. Core Directives for All Compliance Audits

- **Compliance Audits will henceforth be Data-driven and Problem-Statement Based:** Every compliance audit shall be compulsory anchored in 2 things – data analysis and problem statement. At the headquarters stage itself, selection of departments, schemes, activities, and audit units must be driven by structured analysis of data. Data analysis should guide the formulation of audit objectives, risk parameters, and test checks, ensuring that field verification merely validates data-derived insights. This shift embeds objectivity, consistency, and scalability in audit planning, making data the central axis of all compliance audit work.

- **Purpose-Linked Field Audit:** Field audits shall be undertaken only when supported by a defined analytical output and a clearly established problem statement specifying the exact purpose, scope, and issues to be verified. No audit party shall be deputed to the field without an approved statement of verification points and objectives.

- **Group Officer Accountability:** The Group Officer shall be fully accountable for the audit design and field execution under his or her charge. The officer shall remain answerable to the Head of Department and senior management for where, why, and what each field party is verifying, and shall ensure that all verification work remains strictly within the approved audit plan.

- **Time-Bound Completion:** Every compliance audit shall be completed within a defined time frame, ordinarily within **two to three months** from the date of commencement. Any extension beyond this period shall require prior written approval of the competent authority, with reasons duly recorded.

- **Justification and Intimation to Auditee:** Each field deployment shall be supported by a written justification clearly linking it to the analytical results and the approved problem statement. The concerned department or auditee unit shall be formally intimated that the audit will examine only those specific records and processes that fall within the defined audit scope.

- **Applicability of MSO (Audit):** The provisions of the **Manual of Standing Orders (Audit)** shall apply to all compliance audits **insofar as they are not in conflict with the directives contained in this Standard Operating Procedure**. In the event of any inconsistency, the provisions of this SOP shall prevail.

- **A&E-Audit Joint Efforts:** To advance the paradigm of data-driven compliance audits, it is essential that the analytical strength of Audit and the data depth of Accounts function in complete alignment. The vast, structured datasets maintained in A&E offices - across VLC, IFMS, and related systems - offer a foundation for this transformation. Heads of Departments of Accounts and Audit in each State shall ensure the creation of **joint audit teams** that work in continuous coordination to extract, curate, and utilize these datasets for audit purposes.

4. Operational Structure

In light of the evolving governance landscape and the rapid infusion of technology across government systems, audit institutions must evolve in parallel. The government today operates through integrated digital ecosystems - real-time fund flows, automated approvals, and data-driven decision-making supported by robust MIS platforms. To remain an effective instrument of assurance and accountability, the compliance-audit process must operate with the same digital readiness and analytical depth.

Every department functions around a **core mandate** - such as implementing welfare schemes, constructing infrastructure, or delivering essential public services. Most of these functions have undergone significant **digitization**, generating structured datasets that capture approvals, fund flows, and outcomes. Compliance audit must therefore be designed at the **intersection of this core mandate and its digital systems**, assessing whether the digital evidence aligns with the department's stated objectives and performance.

Within each department, a set of **clear problem statements** should be framed around this core mandate - identifying specific risks, bottlenecks, or outcomes that require assurance. These problem statements, drawn from data analysis, prior audit findings, or stakeholder inputs, will guide what aspects are tested and how field verification is directed.
In essence, data provides the analytical foundation, the departmental mandate defines what must be assured, and the problem statement establishes **why** it needs to be examined. Together, these elements ensure that compliance audits are purpose-driven, evidence-based, and contextually relevant to the functioning of a digitized government.

Thus, every compliance audit shall now be firmly anchored in two compulsory elements that together ensure analytical depth, contextual understanding, and relevance to each department's core mandate.

1. **Data-Driven Foundation:** Every audit must begin with structured analysis of available datasets - accounting, beneficiary, expenditure, or MIS records. Data shall guide the identification of audit units, selection of focus areas, and definition of risk parameters, ensuring objectivity and coverage across the department's key functions.

2. **Defined Problem Statement:** Each audit shall be centred around a clear problem statement linked to the department's core responsibilities - such as delays in service delivery, poor maintenance of assets, or shortfalls in scheme outcomes. These statements shall emerge from data analysis, prior audit insights, or program-specific risks.

No generalized or routine transaction checks shall be undertaken*. The data foundation shall determine* ***what*** *is to be examined, and the problem statement shall define* ***why*** *it must be examined - together forming the twin anchors of every compliance audit.*

5. Conduct of Data-Driven Problem-Statement Based Compliance Audits

Data-driven compliance audits shall be undertaken through systematic analysis of departmental and financial datasets to assess the robustness of internal controls and identify problem areas in service delivery and project management.

A. Data Analysis

- Comprehensive analysis of all relevant datasets shall be conducted to assess internal controls in the financial management system.
- **Rule-based checks** shall be applied to 100% of transactions (entire population) to test compliance with applicable financial and procedural rules.
- **Outlier detection** shall be carried out using 10–15 defined features (risk parameters) at the transaction level, covering areas such as sanction, approval, payment timing, beneficiary duplication, and fund utilization.
- **Intensive examination** of identified outlier transactions shall be conducted with reference to supporting records and, where necessary, through **joint physical verification** to confirm deviations.
- **Root-cause analysis** shall be undertaken to identify systemic gaps or control weaknesses.
- **Recommendations** shall be framed based on analytical findings to strengthen internal controls and improve compliance.

B. Identification of Problem Areas

- Problem areas shall be identified through cross-analysis of multiple datasets and sources.
- The convergence of insights from these sources shall guide the formulation of **problem statements** for each department, focusing on issues of efficiency, equity, and effectiveness in service delivery.
- These identified problem statements will form the foundation for audit design, field verification, and recommendation framing.

6. Verticals of Data-Driven Compliance Audits

Background: Financial Profile of States

The financial profile of the States, as detailed in the Comptroller and Auditor General's **"State Finances 2022-23"** publication, shows that total expenditure of all States for FY 2022-23 was **₹42,43,920 crore**.

We can simplify this complex profile by breaking down the expenditure into three key, high-volume verticals:

1. Social Sector
This sector represents the government's direct investment in human capital and welfare, which is the largest functional component of States' expenditure.

- **Coverage:** These areas account for **38.24%** of the Total Expenditure.
- **Core Departments/Functions:** This expenditure is driven primarily by departments managing Education, Health & Family Welfare, Social Welfare & Nutrition, Welfare of vulnerable groups (SCs/STs/OBCs/Minorities)
- **Expenditure Type:** This includes the massive volume of **Revenue Expenditure** related to scheme operations, subsidies, and grants-in-aid for welfare programs.

2. Works Sector
This sector represents the government's investment in **physical and non-financial assets**, which is the primary driver of capital outlay and is directly linked to infrastructure development.

- **Coverage:** These areas account for **28.90%** of the Total Expenditure.
- **Core Departments/Functions:** This expenditure is concentrated in areas like: Transport (Roads, Bridges), Rural Development (Infrastructure, Assets), Housing and Urban Development, Irrigation and Flood Control, Public Works
- **Expenditure Type:** This is primarily driven by **Capital Expenditure** and associated grants given for asset creation (Grants-in-aid for Capital Assets).

3. Salary, Pension, and Establishment Sector
This vertical represents the government's contractual and statutory **committed expenditure** that forms the fixed, recurring cost of running the administration across all departments.

- **Coverage:** These expenditures account for approximately **37% to 43%** of the total Revenue Expenditure (which is the major part of total expenditure).
    - When calculated as a direct percentage of the total spending, these object heads (Salaries, Wages, Pensionary Benefits, Office Expenses, Travel Expenses) alone total **16,09,159 crore**, representing **37.92%** of the Total Expenditure.
- **Applicability:** Unlike the other two verticals, which are concentrated in specific functional departments, the expenditure on **Salary, Pension, and Establishment is applicable to a**ll **government departments and entities** as it constitutes the fundamental cost of personnel.

Basis the financial profile of states, to operationalise data-driven problem statement based compliance audits, they shall be organised into **four verticals**, each reflecting the nature of expenditure, data environment, and core service objectives.

Every audit under these verticals shall be designed with a **data-driven foundation** and a **problem-based focus**, ensuring analytical depth and contextual relevance even in data-deficient environments.

Vertical 1 – Salary, Pension and Establishment Audits
(Refer Annexure II)

This vertical covers payroll, pension, and establishment-related expenditure that together form the financial backbone of government operations.

- **Scope:** Salary disbursement, arrear payments, pension sanction, leave encashment, and other establishment-related transactions.
- **Joint A&E–Audit Approach:** A&E offices hold the source accounting data, while Audit analyses these records to identify irregularities. The process shall be continuous and collaborative, creating a feedback loop between data compilation and data use.
- **Data Sources:** Treasury and VLC data, GPO and PPO records, arrear files, sanction registers, and related expenditure statements.
- **Analytical Method:**
    - Convert legacy PDF/Word records through OCR for use in analysis tools.
    - Apply rule-based checks on 100 % of transactions to detect breaches of sanction and payment rules.
    - Use 10–15 risk parameters to flag outliers - duplicate or ghost employees, double salary draws, irregular arrears, or misclassifications.
    - Examine high-risk cases through supporting records and field verification.
    - Conduct root-cause analysis and frame recommendations to strengthen internal controls.
- **Audit Focus:**
    - Accuracy and validity of payroll and pension payments.
    - Timeliness of arrear clearances.
    - Effectiveness of internal checks within establishment accounting systems.
- **Illustrative Case:** Pilots in Jaipur (Rajasthan), Odisha, and Chitrakoot (UP) are demonstrated in Annexure-II. They have shown that OCR-enabled payroll analysis can uncover irregularities invisible in manual scrutiny.

Vertical 2 –Audit of Beneficiary Oriented Departments
(Refer Annexure III)

This vertical covers departments with a **beneficiary-oriented mandate** - such as Health, Education, Women & Child Development, Social Welfare, and Labour - where funds are released directly to citizens through schemes and service programmes. These departments collectively handle large volumes of transactions and generate structured, high-value datasets suitable for analytical audit.

- **Scope:** Direct Benefit Transfer (DBT) schemes, scholarships, pensions, subsidies, maternity and welfare payments, and other citizen-facing programmes.

- **Data Sources:** Beneficiary registers, Aadhaar-linked payment files, PFMS/treasury data, sanction orders, and scheme-specific MIS dashboards.
- **Analytical Method:**
    - Integrate financial and beneficiary databases to ensure one-to-one matching between sanctioned and disbursed amounts.
    - Use rule-based and outlier analysis to detect duplicate identities, recycled documents, payments to ineligible beneficiaries, and stage-gap violations.
    - Employ data visualisation tools to map disbursement trends, coverage, and leakages across districts and schemes.
    - Validate analytical findings through field verification, sample beneficiary checks, and reconciliation with departmental records.
- **Problem-Statement Focus:**
    - Each audit shall define clear problem statements around service delivery, inclusion, timeliness, or administrative efficiency - e.g., "Are eligible beneficiaries excluded?", "Are benefits reaching the intended stage within prescribed timelines?", "Are scheme databases cleaned and updated regularly?"
    - These problem statements shall emerge from data trends, public grievances, or previous audit and PAC observations.
- **Audit Focus:**
    - Accuracy and integrity of beneficiary databases.
    - Efficiency and timeliness of benefit delivery.
    - Alignment between financial releases (A&E data) and scheme-level disbursement.
- **Illustrative Example:** The audit of *Mukhyamantri Kanya Sumangala Yojana* by the Office of the Principal AG (Audit-I), Uttar Pradesh, analysed over 30 lakh records and revealed duplicate beneficiaries, recycled documents, and stage mismatches - showcasing the power of a data-driven audit anchored in a clear problem statement.

Vertical 3 – Works, Procurement, and Infrastructure Audits
(Refer Annexure IV)

This vertical covers departments engaged in **construction, procurement, and infrastructure creation** - such as **Public Works, Rural Development, Irrigation, Housing, and Transport**— where large public funds are deployed through contracts and physical assets. These sectors collectively accounted for **₹2.88 lakh crore in capital outlay (≈45 % of total capital expenditure)** and **₹1.20 lakh crore in revenue expenditure** in FY 2022-23, underscoring their fiscal significance and the high level of assurance required.

- **Scope**: Civil & mechanical works, e-procurement, contract management, project execution, maintenance, and related expenditure.
- **Data Sources**: VLC / IFMS data, GST TDS & Form-7 returns, CPWA forms (61, 64, 65, 79), e-Procurement and WAMIS portals, contractor registration databases, BOCW records, and GSTN / Income-Tax cross-links.
- **Analytical Methodology**
    - **Approach 1 – VLC-Based Analysis (States without Works Accounting System):** Utilise **transaction-level VLC data** (treasury, DDO, voucher, contractor fields) to

conduct **time-series anomaly detection** over three years using mean, standard deviation, and z-score metrics. Identify expenditure spikes, irregular payment patterns, or outlier divisions/DDOs.

- **Approach 2 – CPWA + VLC Combined Analysis (States with Works Accounting System):** Integrate **CPWA forms (61, 64, 65, 79)** with **VLC datasets** to achieve full coverage of budgeted and deposit works. Apply statistical trend analysis and **DDO profiling** (remoteness, staff dependency, e-MB use) to detect abnormal spending, transfer-entry anomalies, and concentration of expenditure.
- **Approach 3 – GST-TDS / GSTR-7 Based Indirect Verification:** Use **contractor-wise GSTR-7 returns and GST-TDS data** as proxy indicators of works payments. Analyse invoice-level payment patterns, contractor exposure, and TDS deductions to infer irregularities. Combine with **temporal and DDO risk features** to detect collusive or shell entities and validate statewide payment flows.

Vertical 4 – Data-Deficit Departments (Problem-Statement-Centric Audits)
(Refer Annexure V)

This vertical covers departments and functions where structured datasets are limited, fragmented, or unavailable at the **centralized level**. In such cases, data may not serve as the primary deciding factor for audit selection, yet the audit approach must remain analytical, evidence-based, and guided by clearly defined **problem statements** rooted in the department's mandate and operational realities.

- **Analytical Orientation:** Even in data-deficit environments, audits shall be designed through reasoned analysis - reviewing available financial, administrative, and narrative records to identify control risks and operational gaps. The objective is not random verification, but targeted examination of well-understood problem areas.
- **Problem-Statement Definition:** Each audit shall revolve around a well-defined, contextually grounded problem statement. Such problem statements must arise from informed understanding rather than assumption and shall reflect issues already visible in governance discourse - e.g., staff shortages, non-functional assets, weak supervision, or delayed service delivery.
- **Sources for Identifying Problem Statements:**
  - **Departmental Audit Trail:** Recurring observations from *Inspection Reports (IRs)* can point to persistent control weaknesses (e.g., repeated observations on dysfunctional health infrastructure like PHC, CHC).
  - **Field Knowledge:** Insights gathered from *audit parties and local officers* often reveal implementation bottlenecks (e.g., non-functional equipment despite recorded expenditure).
  - **Legislative and PAC Inputs:** Discussions in *PAC proceedings or Assembly Committees* highlight chronic policy or implementation gaps (e.g., absence of teachers, doctors in schools and hospitals are issues often raised by PAC members)
  - **Press and Media Reports:** Investigative or thematic reports in credible outlets can flag emerging risks (e.g., irregularities in procurement or public delivery systems).
  - **Collaboration with Civil Society:** Interactions with *reputed NGOs and research institutions* can help identify systemic issues not yet reflected in departmental records (e.g., exclusion of vulnerable groups from welfare schemes).

- **Data at Decentralised Levels:** These departments may not have data consolidated at the state or headquarters level, yet **unit-level or problem-specific datasets** - such as local registers, spreadsheets, or portal data maintained at district or field offices - may exist. Such data shall be used to conduct **data-driven analysis at the unit level**, ensuring that even where central systems are weak, audits remain analytical and evidence-backed.

- **Audit Focus:** The audit design shall integrate these insights with whatever data exists - centralized or local - such as partial MIS records, treasury extracts, or performance indicators, to select specific units, schemes, or districts for in-depth review. Every test should link back to the identified problem statement and conclude with actionable recommendations.

Departments with some of their known and recurring problem areas are listed in **Annexure V**, along with indicative problem statements for reference and future audit planning.

7. Implementation Plan

I. Annual Audit Plan and Allocation of Resources

The Annual Audit Plan shall ensure a balanced distribution of audit effort across all three categories of audit - Financial, Performance/Horizontal, and Compliance - while progressively enhancing the analytical depth of each stream.

Table 1: Distribution of Total Audit Effort

Category of Audit	Indicative Share of Total Mandays	Key Focus / Coverage
Financial Audit	15% – 20%	Statutory and certification audits, annual accounts, and finance-based assurance functions.
Performance / Horizontal Audits	20% – 25%	Thematic or cross-cutting studies across departments such as health, education, and infrastructure.
Compliance Audit (Total)	~50%	Detailed examination of compliance with rules, procedures, and programme implementation.

The first two categories represent **committed statutory and thematic responsibilities**. All **remaining audit resources (approximately 50%)** shall be utilised for **Compliance Audits**.

Table 2: Allocation within Compliance Audits

Type of Compliance Audit	Indicative Share	Approach
Data-Driven Problem-Statement-Based Audits	**~90%**	To be undertaken using structured datasets and defined problem statements, ensuring analytical and contextual depth.
Transaction-Based Audits	~10%	To be undertaken only in exceptional cases, with a reasoned and documented basis for selection.

This approach ensures that audit resources are aligned with the strategic shift towards evidence-based, high-impact compliance audits.

II. Joint A&E–Audit Teams

To make the data-centric audit model sustainable, **Joint A&E–Audit Teams** shall be constituted in every state. These teams will operate under the **co-management of both Heads of Departments (Audit and A&E)** and will form the operational backbone for all data-driven audits.

Their key responsibilities will include:

- Coordinated access to accounting, VLC, treasury, and scheme-level data.
- Joint creation and maintenance of structured audit datasets.
- Validation of analytical results before field verification.
- Development of shared dashboards and analytical templates for use across offices.

Their functions will include:

- A&E offices shall leverage **OCR, DataSnipper, and other modern data-processing tools** to convert unstructured datasets - such as GSTN records, voucher data, and IFMS extracts - into analyzable formats.
- In this process, they shall be **supported by Audit teams**, ensuring that data transformation aligns with analytical requirements and audit objectives.
- Audit offices, in coordination with A&E offices, shall progressively develop **analytical modules and learning-based systems**, incorporating techniques such as **outlier detection, AI-assisted pattern recognition, and automation** to enhance audit efficiency and risk identification.

This joint framework ensures that data creation, validation, and analytical use remain part of a single, continuous oversight process rather than isolated exercises, strengthening coordination between accounting and audit functions at all levels.

III. Structured Audit Planning Framework

Every compliance audit - whether data-driven or problem-statement-based - shall be anchored in a **Structured Audit Planning Framework (SAPF)**.The SAPF shall serve as the **core planning and accountability document** for each compliance audit.

Each SAPF shall clearly specify:

- **Audit Objectives:** What is being examined and why.
- **Audit Criteria:** Rules, procedures, or performance standards against which compliance is tested.
- **Sources of Evidence:** Datasets, documents, field verification inputs, or stakeholder consultations.
- **Analytical Methods:** Nature of data analysis, AI/ML models, or risk parameters to be applied.
- **Validation Chain:** Steps for exception validation, including sampling, reconciliation, or physical verification.
- **Findings and Recommendations Linkage:** Ensuring every audit observation is traceable to its analytical origin.

ANNEXURE – 1
Comprehensive List of Datasets

Data Source / System	Dataset / Module	Dataset Description	Custodian / Access Point	Remarks / Audit Use
VLC (Voucher Level Compilation)	Head of Account–wise Receipt & Payment	Contains month-wise, head-wise receipts and payments for all divisions.	AG (A&E)	May require OCR/DataSnipper conversion where records are not digitised. Enables financial reconciliation and expenditure analysis.
	Income-Tax TDS / GST TDS	Division-wise, transaction-wise deductions enabling assessment of actual payments by divisions.	AG (A&E)	Schedule of GST TDS includes contractor names—usable for vendor analysis and reconciliation.
	GST Challan Data	Depositor-wise, date-wise, transaction-wise data with GST No. and TAN.	AG (A&E)	Captures both Treasury and non-Treasury transactions; directly accessible via GSTN or Cyber Treasury.
	GST TDS DDO Return (Form 7)	DDO-wise, contractor-wise, invoice-wise TDS deduction return.	AG (A&E)	Useful for contractor-wise tax reconciliation and payment validation.
Monthly Works Accounts (Compiled)	CPWA Forms 64, 65, 79 etc.	Form 64: Work-wise, month-wise expenditure; Form 65: Deposit works; Form 79: Consolidated summary.	AG (A&E)	Basis for analysing cumulative work expenditure and deposit accounting.
	Form 61 with Vouchers and Bills	Includes (i) Payments to contractor (ii) Stock charged (iii) Expenditure through transfer entry.	AG (A&E)	Expenditure via transfer entry is high-risk; must be analysed for possible misclassification.

IFMS / Treasury Systems	Voucher Details	Treasury-integrated works contain 15-digit Head of Account, division, contractor name, deductions, etc.	AG (A&E) / Director (Treasury)	Allows project-wise analysis and detection of split transactions. For non-integrated states, eKuber or consolidated treasury data may be obtained.
WAMIS (Works & Accounts Management Info System)	Project / Contract Modules	Digitised system covering approvals, estimates, measurements, financial & physical progress, and extensions of time.	Works Dept / State Admin	Enables tracking of progress and cost/time overruns; integrates monthly accounts to AG (A&E).
e-Procurement Portals	Tender, Bid & Award Data	Contains tender conditions, qualification criteria, disqualification list, bid values, Letter of Acceptance, EMD details.	State Works / IT Dept	Used for vendor validation, collusion analysis, and compliance of bid procedures.
BOCW (Building & Other Construction Workers)	Labour Cess Deduction and Transfer Data	NEFT/Bank drafts transferred by divisions to welfare boards with contractor/work details.	BOCW Board	Links labour cess with work execution; helps trace off-treasury payments.
Contractor Registration Database	Registration & Qualification Records	Contains PAN, experience, solvency, machinery, and validity details.	Concerned Works / PWD Dept	Used for verifying eligibility and detecting repeated or blacklisted contractors.
Mining Data (e-Rawanna / e-Transit Pass)	Material Transit Data	Captures transport permits for material (e.g., bitumen, aggregates) used in works.	State Mining Dept	Enables matching of material movement with sanctioned works to detect diversion or misreporting.
CRC (Consignee	Receipt of Material	Details of contractor, work,	Division / Stores Dept	Verifies receipt of materials against

Receipt Certificate)		destination, and material quantity (e.g., bitumen).		issue orders and payments.
PFMS (Public Financial Management System)	Fund Flow and Payment Data	Tracks fund flow from sanction to final payment across agencies and schemes.	Controller General of Accounts / State Finance Dept	Enables real-time reconciliation with AG accounts and identification of delays or diversions.
DBT (Direct Benefit Transfer) Portal / State Dashboards	Beneficiary & Payment Data	Aadhaar-linked records for scheme disbursements.	Relevant departments	Used for duplication, eligibility, and stage-gap analysis.
SNA–SPARSH Portal	CSS Transactional Data	Provides full footprint of ₹5.4 lakh crore Centrally Sponsored Scheme funds.	A&E / State Finance Dept	Transaction-level data accessible to A&E/Audit; allows cross-matching with scheme outcomes.
GSTN (Goods & Services Tax Network)	Invoice & Return Data	Taxpayer-wise indirect-tax trail; billions of invoices annually.	A&E / State Tax Dept	Used to verify vendor authenticity and detect shell entities in procurement audits.
MCA-21 Corporate Database	Company Filings & Ownership Records	Statutory filings for 20 lakh companies, including directors and financials.	Ministry of Corporate Affairs	Enables contractor background checks and related-party analysis.
NITI Aayog / Aspirational District Portal	Socio-Economic Indicators	District-level data on health, education, infrastructure, etc.	NITI Aayog	Provides contextual benchmarks for performance audits.
MOSPI / RBI / Census / SECC / Open Data Portal	National Statistics & Surveys	Macro- and micro-level indicators.	Govt of India	For contextual comparison, sampling design, and validation.
Field Level Registers / Unit-Level Data	Scheme / Division Records	Local records such as beneficiary lists, work registers, stock ledgers.	District / Implementing Units	Used in data-deficit contexts for problem-specific analysis.

ANNEXURE II
Salary, Pension and Establishment Audits

Odisha Hospital Salary Fraud – Case Study Establishment Audit

Subject: Fraudulent drawl of Salary and Contingent Expenditure - ₹ 2.79 crore
Unit: District Headquarters Hospital (DHH), Angul
Key Vouchers on Record: TV No. **54** (June 2025); TV No. **410** (31.07.2025)

How the fraud unfolded?

While examining vouchers of DHH Anugul, we observed that office contingent expenditure was being paid in the account of the Junior Assistant of DHH Anugul for ₹1,34,092 which is a deviation from the financial rules as no government is to be paid in the personal account of any official other his salary and related expenditures. Subsequently, we examined the payments received by the Junior assistant from 2017-25 in his account. We noticed that though he was a Level 4 officer, he was drawing the salary of Level 16. We noticed that sometimes the DA drawn by him went up to ₹14 lakhs per month and sometimes his basic pay went up to ₹10 lakh. Once these details emerged from the IFMS data and e-HRMS data, we sent a team to examine the documents in the Anugul district treasury which confirmed the fraudulent drawl of ₹2.45 crore (which included continent expenditure drawl of ₹22 lakh.

(1) DDO level – CDMO Anugul failed to see that his Junior Assistant, who was preparing salary bills was drawing more than him (DDO) and as per the spot assessment of the AG team, DDO had shared his username and password to operate IFMS on his behalf to the Junior Assistant. Thus, a key internal control was by passed. The officer who committed the fraud was posted in the same section since 2017 and started over drawls in few thousand, which went on up to ₹14 lakh. He also alternately changed the basic pay and DA far excess than his actual entitlement and several DDOs who were posted in the DHH Anugul during 2017-25 could detect his fraud. He also authorized more salary to other officials who were working in DHH Anugul, which came to light when all records of Anugul DHH were examined.

(2) Treasury Level- The Treasury Officers since 2017-25 failed to detect the abnormal salary drawl and the total salary in month went beyond ₹14 lakh which should have raised their eyebrows as the highest salary received by the Chief Secretary also cannot go beyond ₹3.5 lakh per month.

(3) Weakens in e-HRMS and IFMS- The basic pay field and DA field in e-HRMS appears to be editable because of which the Junior Assistant was able to change his basic pay and dearness allowance at will. In IFMS due to introduction of global budgeting, the unit level limitation of budget which existed earlier, appears to be non-existent. Therefore, the Treasury Officers could not detect the abnormal drawl.

Method of detection

We examined the documents especially the Bank Account No of the Junior Assistant and ran a query on the IFMS treasury data of Anugul Treasury initially for one year and subsequently all the years from 2017 to 2025 as the officer was posted in the DHH Anugul since 2017. Then we got the service book to verify his pay level, designation and date of appointment in DHH Anugul and we examined the Acquittance roll with reference to the Service book data. Based on the above data we prepared due and drawn salary statement for the officer. The same was verified by the AG team in DHH Anugul and District Treasury Anugul which confirmed the fraud beyond doubt.

Action taken by PAG office

Once the AG team completed its examination, we reported the matter to the Chief Secretary and Principal Secretary Finance along with all documents that we had examined for unearthing the fraud.

Action taken by the State Government

1. CDMO has placed the Junior Assistant under suspension on 14/10/2025 and FIR has also been lodged against the errant official.
2. The Principal Secretary – Finance Department has issued instruction on 24/10/2025 that all salary bills shall be processed parallelly in the HRMS/IFMS and also in physical mode and submit to the Treasury.
3. The Finance Department on 23/10/2025 directed the Director of Treasury and Inspection on the following.

i. To conduct an internal inquiry by a team of officers headed by an officer not below the rank of Additional Director and identify the person(s) responsible for the fraudulent drawal.

ii. To identify loopholes both in the IFMS and HRMS system which fails to detect the fraud.

iii. Role of global allotment system in enabling these kinds of manipulations at the DDO level.

iv. Identify the reason of failure of the Internal Audit to detect the fraud and measures to strengthen it.

We have mentioned process flow chart below to examine whether similar instances are there in other departments of the State Government.

WAY FORWARD

FLOW Of SOP: OC–Payroll Cross-Link & Pay-Level Mismatch

Start

Begin the process with initial preparations and gather necessary information.

Pick Office Contingency Voucher

Open voucher and sanction view using VLC/IFMS for further review.

Beneficiary Check

Compare sanction order vs beneficiary list; check bank A/c if employee is the beneficiary.

Scrutiny of Sub-vouchers

Open scanned invoices; check if marked as “Paid by me” noted by employee.

Bank A/c Cross-link

Retrieve e-Disbursement details from IFMS, confirm salary + Office Contingency credits to same A/c.

Verification of Pay-level

Verify HRMS acquittance; compare with cadre pay matrix values.

More than Five-Year Pattern

Examine records for >5 years for detection of abnormal cumulative trends.

Confirmation

Write to DDO/TO regarding excess drawal of employee

Frame Objections

Tie each objection to evidence supporting discrepancies noted in previous steps.

Conclude the process with archive of KD pack and issue audit memo/para.

Flow-SOP: OC–Payroll Cross-Link & Pay-Level Mismatch (Repeatable Template)

0) Start
Action: Pick an OC voucher for scrutiny (month/district wise).
Tool/Data: VLC/IFMS – voucher & sanction view.
We saw: Voucher breakup + sanction orders.
So we: Noted beneficiaries, designations shown

1) Beneficiary check (sanction vs voucher)
Action: Compare sanction order with Beneficiary List.
Tool/Data: VLC/IFMS sanction PDF + voucher.
We saw: One sub-component sanctioned to a Employee (designation "Accountant").
So we: Flagged "Employee-as-beneficiary" and captured the **bank A/c number** from the Beneficiary List.
Record: KD-1 (sanction & voucher excerpts), KD-2 (noting on staff beneficiary).

2) scrutiny of sub-vouchers
Action: Open invoices backing each Office Contingency.
Tool/Data: Scanned invoices (IFMS/ Voucher along with Sub voucher).
We saw: Vendor invoices endorsed "Paid by me" by the same Employee.
So we: Flagged breach of role segregation (Employee endorsing vendor bills).
Record: KD-3 (invoice images with highlights).

3) Bank account cross-link
Action: Retrieve all transactions for the **bank A/c captured in Step 1**.
Tool/Data: IFMS E Disbursement Report.
We saw: The **same A/c** received **salary credits** and **OC credits**.
So we: Established common beneficiary flow and elevated the risk.
Record: KD-4 (account mapping extract).

4) verification of pay-level
Action: Verify the salary bill for that name/EmpID and read Basic/DA/HRA/Gross.
Tool/Data: HRMS/ Acquittance Roll.
We saw: Figures aligning to a **much higher pay level** than the cadre on Service book (Level-16-paid to a junior cadre).
So we: Flagged designation/pay-level mismatch.
Record: KD-5 (salary bill), KD-6 (Odisha 7thpay matrix).

5) More than Five-year drawal pattern
Action: Examine due drawn statement for more than 5-year for the said employee (name/EmpID).
Tool/Data: e-HRMS/Acquittance roll (FY-wise/month-wise).
We saw: Large cumulative drawal under the same name.
So we: Treated it as a sustained-pattern risk.
Record: KD- 9 (more than 5-year sheet).

6) Confirmation
Action: Seek clarifications on (a) employee-as-beneficiary, (b) account ownership, (c) pay-level basis.
Tool/Data: Formal letter to DDO and TO.
We saw: Reply citing "typographical error" with partial recovery.
So we: Reconciled reply vs evidence (Steps 1–6) and found gaps (routing, endorsements, pay-level, multi-year pattern).
Record: KD-7 (letter), KD-8 (reply)

7) Frame objections (tie each to hard evidence)

-
- **O1—OC routed to staff A/c without proper GST:** KD-1, KD-4.
- **O2—Same A/c for salary and OC:** KD-4.
- **O3—Pay-level far above cadre:** KD-5, KD-6.
- **O4—Vendor invoices endorsed by staff:** KD-3.
- **O5—DDO reply inadequate vs 5-year pattern:** KD-5 (history), KD-8.

Chitrakoot Treasury Fraud - Case Study Establishment Audit

1. Background of Special Inspection of Treasury

During the financial year 2024–25, Inspection Team No. 2 conducted an inspection of the Chitrakoot Treasury. During the inspection, while reviewing the 'List of Deleted Pensioners' issued at the treasury level, it was found that the list included the name of a pensioner whose Life Certificate was still valid, yet he had been marked as 'Deleted'.

This raised suspicion among the inspection team, and the pension file of the concerned pensioner was requested from the treasury. Even after two written requests, the file was not provided. Consequently, the team downloaded the 'Cumulative Pension Payment Statement (CPPS)' of the said pensioner from the CTS Pension Module.

On analysis, it was observed that the concerned pensioner had been paid arrear pensions repeatedly and in unusually large amounts. Considering the possibility of financial irregularities, the matter was reported to Headquarters. In view of the likelihood of extensive financial irregularities, the competent authority constituted a Special Inspection Team and directed it to conduct a detailed examination of pension arrear payments.

2. First Stage – Data Download and Conversion (Data Extraction & Conversion)

i. The 'Cumulative Pension Payment Statement (CPPS)' report of all category-wise pension payments from the CTS Pension Module of the Chitrakoot Treasury was downloaded in text format.
ii. In this report, each pensioner is uniquely identified by an INDEX NUMBER, and the month-wise chronological payment details are recorded accordingly.
iii. The PAY MODE column of the report indicates the type of payment as follows:
    - ‘A’ = Arrear Pension, ‘R’ = Regular Pension, ‘F’ = First Payment, ‘G’ = Gratuity Payment. (etc.)
iv. The downloaded CPPS data was then converted via Excel.

3. Second Stage – Data Analysis:

General Note:

i. Since the total number of pensioners in Chitrakoot Treasury was relatively small, most of the data analysis work was effectively performed in Excel (filter, pivot table, lookup etc) , and the report was prepared on that basis.
ii. Wherever it was necessary to join datasets, extract specific data, or identify cases of gratuity payments exceeding the prescribed ceiling, the IDEA software was utilized. Additionally, wherever required, Summarization, Filter Criteria, and other analytical procedures were also performed through IDEA to ensure the accuracy and reliability of the data analysis.

3.1 Analysis of Pension Arrear Payments:

i. Initially, out of a total of 5,008 pensioners in Chitrakoot Treasury, data related to the three major categories — E-01, E-11, and C-01 (comprising a total of 4,589 pensioners) — was segregated for focused analysis.
ii. Within these three categories, all payment data related to Arrear Payments (A’) was filtered out.

iii. For each pensioner (Index-wise), data on the total amount paid and the frequency (count) of payments was summarized using Excel Pivot Tables and IDEA, primarily for frequency analysis.

iv. The analysis revealed the following:
- Certain pensioners had received arrear payments an abnormally high number of times - **one pensioner received such payments up to 104 times.**
- Some pensioners had received excessively large sums, **with one pensioner receiving about ₹3.58 crore as arrear payments.**

v. Based on this:
- A list of pensioners who had received arrear payments of ₹10 lakh or more was prepared (96 cases in total).
- Subsequently, based on the summarized data, all pension arrear payment cases below ₹10 lakh, in which any arrear payment was made during FY 2024–25 and 2025–26, were also sampled and listed (21 cases in total).

3.2 Analysis of Gratuity Payments:

i. After the detection of irregularities in the arrear pension analysis, the Gratuity Payments (G') were also examined.

ii. As per current rules, the maximum permissible limit for gratuity payment is ₹25 lakh. During IDEA/Excel analysis, **three cases** were identified where this limit had been exceeded.

iii. Further, to verify the overlap between arrear and gratuity payments, data relating to gratuity payments of those pensioners who had received arrear pensions was matched using the IDEA Join Function. Cross-verification was also conducted using Excel Filters, and a consolidated list was prepared.

iv. From the data analysis, 27 such cases were initially identified, and upon scrutiny of their physical files, 18 cases of unauthorized gratuity payments were confirmed.

4. Third Stage – File Scrutiny and Physical Verification:

1. A total of 117 pensioner files, selected through the sampling process described above, were obtained for examination.
   - Out of these, unauthorized payments were confirmed in 93 cases.

2. The physical scrutiny of pension files revealed the following key irregularities:
   i. Payment of arrear pension was made without valid documents, authorization, or approval orders.
   ii. Pension payments continued even after the death of the pensioners, amounting to several crores of rupees.
   iii. Negligence was observed in the process of pension discontinuation or revision, and serious lapses were found in record management.
   iv. Gratuity payments were made without proper authorization, sanction orders, or payment approvals.
3. In cases where pensioners had died, posthumous payments were identified using Excel filters based on the date of death (also cross-verified using criteria in IDEA).

5. Preliminary Findings:

Based on data analysis and subsequent scrutiny of pension files, the **Special Inspection Team** categorized the **unauthorized pension payments** into four distinct groups as shown in the table below:

Sl. No	Category	Unauthorized Payment (Rs.)	No. of Cases
i.	Unauthorized pension payments made to deceased pensioners[1]	₹ 15,30,63,408	07 cases
i.	Unauthorized pension arrear payments made to pensioners[2]	₹ 25,80,69,699	79 cases
i.	Unauthorized gratuity pension payments made to pensioners[3]	₹ 98,42,919	
v.	Unauthorized pension payments made to pensioners without updated life certificates[4]	₹ 2,29,67,730	07 cases
	Total estimated loss to the Government exchequer	**₹ 44,39,43,756**	**93 cases**

[1] This category includes unauthorized arrear pension payments made before and after the death of the pensioner, as well as any other types of pension payments made after death, such as regular pension and gratuity, etc.

[2] This category includes unauthorized arrear pension payments made to the pensioner.

[3] This category includes unauthorized gratuity payments made to the pensioners covered under category (ii).

[4] In this category, all unauthorized pension arrear payments made to the pensioner in the absence of an updated life certificate attached to the pension file, as well as any other types of pension payments (such as regular pension and gratuity, etc.) made after the month corresponding to the signature date on the last life certificate attached to the file, have been treated as unauthorized and included accordingly.

ANNEXURE III:
Beneficiary Scheme Audit

Office of the Principal Accountant General (Audit-I) , Uttar Pradesh, Prayagraj

Part 1.

Introduction

Social-sector departments in the state perform a wide range of welfare functions aimed at supporting vulnerable sections of society.

Broadly, their activities can be grouped into **three components**.

1. **Direct Benefit Transfer (DBT)-based Schemes:** This forms the largest and most critical part of the social sector. These schemes provide direct financial assistance to citizens - particularly to the poor, disadvantaged, and marginalised groups. They include *scholarships for SC/ST/minority students*, *pensions for old age, widows, and persons with disabilities*, and *targeted schemes for the girl child and economically weaker students*. These are high-volume, high-value programmes with direct cash flow to beneficiaries, forming the core of what is referred to as *social-sector transfers*.

2. **Service-based Interventions:** Apart from direct transfers, social-sector departments also operate physical or institutional services. These include *One-Stop Centres for women in distress*, *shelter homes for the elderly and differently abled*, *residential hostels for financially weak students across communities*, and *rehabilitation or welfare centres*. Here, expenditure is not paid directly to individuals but incurred on facilities and service delivery infrastructure.

3. **Other Community or Welfare Schemes:** This covers miscellaneous welfare initiatives that do not fall neatly under DBT or institutional categories - for example, livelihood or rehabilitation grants, skill programmes, or state-funded welfare trusts.

The guiding philosophy of compliance audit in social-sector departments is to *assess whether welfare schemes are truly delivering the intended benefits to the people they are meant for and operating in line with prescribed rules and objectives.* These audits must recognise that government programmes here deal with the most marginalised and scattered sections of society, where identities can be easily misused and leakages can occur at scale. The central purpose, therefore, is not only to check financial compliance but to strengthen systems, ensure genuine reach of benefits, and improve the overall quality and integrity of public service delivery.

This document outlines the **strategy for conducting Focus-Based Compliance Audits** of such schemes. It explains how audits can use data analytics to identify high-risk transactions and geographies, design Audit Design Matrices (ADMs) for both standard and scheme-specific checks, plan targeted field verification, and ensure that reporting and recommendations are uniform and evidence-driven.
While the document uses examples from two major social-sector programmes in Uttar Pradesh - the *Mukhyamantri Kanya Sumangala Yojana* and the *Scholarship Schemes* - the principles and formats described here are designed to be applicable to any beneficiary-based scheme implemented through Direct Benefit Transfer systems.

Notes before Reading the Document

1. **On the Use of Analytical Tools and Models** The document refers extensively to analytical tools and models that use platforms such as Jupyter, Python, SQL, and allied systems. These references are not meant to suggest a high technical barrier. In practice, field experience has shown that these tools are now easily usable by audit officers without formal backgrounds in data science. With the growing integration of **AI assistants and general-purpose coding models**, even complex analyses - such as anomaly detection, clustering, or pattern recognition - can be performed through guided automation. The emphasis, therefore, is not on mastering programming syntax, but on understanding **what these tools can achieve** for audit: how they help detect duplication, analyse large beneficiary datasets, and bring precision to audit sampling and risk profiling. In short, the tools should be seen as **enablers of sharper audit judgement**, not as barriers to it.

2. **On the Importance of Source-Level and Cross-Cutting Data** A consistent finding across field audits has been that the real strength of data-driven auditing lies in obtaining datasets **directly from the source systems**, rather than relying only on departmental portals or limited audit requests. Audit offices are encouraged to proactively collect such base datasets - even when a department is not under immediate audit - because they often have recurring and cross-cutting relevance. For instance, **income and caste certificate databases, educational board records, and SECC datasets** are central to verifying eligibility in multiple social-sector schemes such as pensions, scholarships, and girl-child incentives. Likewise, some of the most critical datasets for auditing a department may not originate within that department at all; they may lie with Revenue, Education, or IT directorates. The guiding principle is to **build and maintain a shared pool of foundational data** that enriches every future audit exercise.

3. **On Coverage of the Full Administrative Chain** Social-sector schemes operate through a layered chain of verification, sanction, and disbursement. While Drawing and Disbursing Officers (DDOs) are important points of transaction, the real locus of accountability often lies higher up the functional hierarchy. In departments such as Social Welfare, Backward Classes, Minorities, and Women & Child Development, the DDOs - typically Program Officers or District Probation Officers - act primarily as implementing agents under directions from **Block Development Officers, Sub-Divisional Magistrates, and District Magistrates**, who perform verification and approval functions. Audit, therefore, needs to look beyond the financial node and examine this entire control structure to understand where compliance lapses originate and where corrective action is most feasible. The goal is not to overburden the implementing officers but to ensure that audit findings connect meaningfully with **decision-making and accountability levels** within the system.

4. **Note on Confidentiality and Responsible Use of GPTs:** The use of GPT-based assistants in audit work may significantly ease the process of writing and executing analytical code and could serve as a practical aid for officers working with tools such as SQL, Python, and Jupyter. It may be noted, however, that **no government or audit data should be uploaded to external AI platforms**. Confidential data must always remain within secure departmental systems. At the same time, the use of GPTs for **non-data-dependent assistance** - for example, obtaining installation steps for analytical tools, learning syntax, generating sample code structures, or seeking help in debugging generic scripts - may be considered a legitimate and useful form of support. Such use does not involve sharing any sensitive information and could in fact **alleviate concerns of confidentiality breaches** by allowing officers to learn and operate analytical tools confidently without relying on external contractors or unsafe workarounds.

Part 2.

Master Framework for Use of Data in Analytics

2.1 Understanding the Three-Layer Data Architecture for Social-Sector Audits

Effective focus-based compliance audits depend on a layered data ecosystem. For any social-sector scheme, three core categories of data are required for meaningful analytics:

Layer	Type of Data	Primary Custodian	Purpose / Use in Audit
Layer 1 – Internal Financial Data	*VLC data, Integrated Financial Management System (IFMS)* data, Treasury data, budget allocations, DDO-wise expenditure details.	Accountant General (A&E), State Finance Dept.	To verify actual release and expenditure against the scheme, trace fund flow, reconcile payments, and flag unusual patterns of withdrawals or booking.
Layer 2 – Scheme-Specific Operational Data	Scheme portal data: beneficiary lists, payment approvals, sanction orders, uploaded documents, eligibility status.	Implementing Department (e.g., Social Welfare, Minority Welfare, Women & Child).	To test compliance with scheme guidelines, identify duplication, ineligible or fake entries, and confirm progression of beneficiaries through various stages.
Layer 3 – Cross-Verification / Supporting Data	Caste & income certificates, school/board records, GST, ration cards, social census data, disability certificates, bank & IFSC master, etc.	Other departments (Revenue, Education, NIC).	To validate authenticity of information used in scheme databases and ensure inter-departmental consistency.

Example: To audit the *Scholarship* schemes, the audit could combine:

- **Scholarship Dump Data** → shows beneficiary-level sanction, total applications.
- **Income & Caste Certificate Data (Revenue Dept)** + **Board Roll Data (Education Dept)** → verifies eligibility.

2.2 Tools and Technology Stack

The tools required could vary by complexity and data volume. A standard three-tier workflow is proposed:

Stage	Tool / Platform	Purpose / Function
Data Extraction & Loading	SQL Server / PGAdmin (PostgreSQL) / SSMS / DBeaver	Import and store large departmental dumps (CSV, DBF, Excel, API extracts). Maintain schema documentation and indexing.
Cleaning & Pre-Processing	IDEA / Jupyter (Python pandas)	Identify duplicates, nulls, invalid dates, inconsistent codes; format transformations; creation of unique beneficiary keys.
Analytical Processing	SQL (for structured supervised queries) + Python (for unsupervised analysis).	- **Supervised analysis (SQL):** joins, conditional filters, aggregation, and pattern-based checks. - **Unsupervised anomaly detection (Python):** clustering, outlier detection (Isolation Forest, DBSCAN, etc).

Illustration: Using GPT to Run Large-Volume Data Checks in Jupyter

An Audit Office has received a **CSV file for the Scholarship Scheme** containing over **25 lakh rows**. Performing even basic checks - like identifying duplicates in *bank account numbers*, *usernames*, or *student names* - is impossible in Excel or other spreadsheet tools due to file size limits. However, the same checks **may easily be done** in **Jupyter Notebook** by taking the assistance of **GPT-based tools**, without needing deep technical knowledge.

Step 1 – Identify Columns in the Dataset

The auditor may first tell GPT: "The file path is '/Users /Desktop/ Scholarship Audit/Scholarship Analysis/Raw Data/scholarship_rawdata_22_Shahjahanpur.xlsx'. Please give me Python code to read this file and list all the column names." GPT will instantly produce executable code such as:

import pandas as pd

 # Read the file and display column names (Python Code given by GPT)
df = pd.read_excel(file_path)
print(df.columns.tolist())
Running this in Jupyter will list every column in the dataset, which forms the basis for the next step.

Step 2 – Check for Duplicates

Once the auditor knows which columns to test, they may again ask GPT:
"Please give me code to check for duplicate records based on *student_name*, *father_name*, *dob*, and *bank_account_no*."
GPT may respond with:

 # Check for duplicates on selected columns
duplicate_records = df[df.duplicated(subset=['student_name', 'father_name', 'dob', 'bank_account_no'], keep=False)]

print("Number of duplicate records found:", len(duplicate_records))
duplicate_records.to_excel('duplicates_found.xlsx', index=False)
The result file duplicates_found.xlsx may be reviewed for field verification or sampling.

[4]:
```
import pandas as pd

file_path = '/Users/amritesh/Desktop/Scholarship_Audit_Districtwise_1110/Scholarship Audit/Scholarship Analysis/Raw Data/scholarship_rawdata_2

# Load the dataset
df = pd.read_excel(file_path)

# Count how many times each name appears
name_counts = df['name'].value_counts()

# Filter only names that appear more than once
repeated_names = name_counts[name_counts > 1]

print("Total unique names repeated:", len(repeated_names))
print("Sample repeated names and their counts:")
print(repeated_names.head(10))
```

```
Total unique names repeated: 8449
Sample repeated names and their counts:
name
Muskan            121
Shivani           113
Abhishek Kumar     96
SHIVANI            88
AMIT KUMAR         86
Anuj Kumar         84
Sumit Kumar        84
Pooja Devi         83
Kajal              79
Anjali             78
Name: count, dtype: int64
```

2.3 Audit Queries and Objectives

There are two broad types of analytical queries in any focus-based audit: **standardized queries** (common across schemes) and **scheme-specific rule-based queries**.

(A) Standardized Data Checks – Applicable to All Schemes

Objective	Typical Query Example	Insight / Purpose
Duplicate Beneficiaries	Identify same Aadhaar / bank account / mobile appearing under multiple beneficiaries or years.	Detect recycled or fake entries.
Duplicate Payments	Match transaction ID + Aadhaar + scheme year + amount for repetition.	Capture duplicate disbursals.
Missing Supporting Records	Left-join with income/caste data; filter nulls.	Identify beneficiaries without required certificates.
Age / Eligibility Mismatch	Compute derived age from DoB and compare with scheme limits.	Flag underage/overage claimants.
Temporal Gaps	Compare payment dates and stage progression intervals.	Verify required gap between scheme stages.
Geographical Anomalies	Group by district / block / institute; flag extreme deviation from mean.	Identify localized concentration of irregularities.
Bank Account Patterning	Detect multiple beneficiaries linked to same account or IFSC out of expected geography.	Identify collusive networks.
Document Integrity	Check file name hashes, timestamps, and size uniformity.	Indicate potential manipulation or overwriting.

(B) Scheme-Specific Rule-Based Queries

Scheme	Example Queries	Audit Objective
Mukhyamantri Kanya Sumangala Yojana (MKSY) **Scheme gives financial assistance to girl child at birth, class I, class VI, class IX and graduation**	1. Verify chronological gap of ≥ 5 years between Stages 1 → 3 and 3 → 5. 2. Confirm consistent DoB, Aadhaar, and guardian name across all six stages.	Detect false progression claims and fake child profiles.
Scholarship Schemes	1. Check if a student is enrolled in multiple institutions in same year. 2. Verify attendance ≥ 75 % and marks ≥ 50 % before sanction (mandatory criteria for scholarships). 3. Detect same Aadhaar used for different caste categories.	Ensure eligibility compliance, prevent multi-claim fraud.
Old Age Pension / Disability Pension	1. Cross-match beneficiaries with deceased persons (civil registration data). 2. Check duplication across pension types.	Identify payments to deceased or dual beneficiaries.

Note:
Scheme-specific rules can now be semi-automatically extracted from government orders and notifications using **AI summarization models**, allowing faster ADM creation for future audits.

2.4 Key Data Fields for Integration and Cross-Verification

To ensure interoperability across datasets and meaningful joins, the following fields must be mandatorily captured and validated:

Category	Critical Fields	Remarks / Use in Audit
Beneficiary Identity	Aadhaar / Aadhaar-tokenized ID, Name, Gender, DOB, Mobile No.	Uniqueness & linkage across schemes.
Scheme Data	Scheme Code, Stage, Application No., Sanction Order ID, Sanction Date, Status.	Core keys for cross-year tracking.
Financial Data (IFMS/VLC)	DDO Code, Voucher No., Head of Account, Date of Drawal, Amount, Transaction ID.	Reconciliation of sanction vs drawal.
Geographic Data	District / Block / Institution Code, Institute Name, IFSC of Bank.	For spatial analysis and red-flag heat maps.
Validation Data	Income Cert No., Caste Cert No., Education Board Roll No., Disability Cert No.	Cross-verification with external databases.

2.5 Application of IFMS and VLC Data for Remote Risk Flagging

The **VLC** and **IFMS** databases form the backbone of financial analytics for remote risk assessment. By linking DDO-wise expenditure to scheme datasets, audit can:

1. Identify **DDOs with unusually high average payment size** per beneficiary.
2. Detect **non-existent beneficiaries** where expenditure exists but no corresponding record in the scheme portal.
3. Match **fund releases and drawals** with approved sanction IDs to ensure authenticity.
4. Use **temporal analysis** (month-wise drawals) to spot year-end expenditure bunching, indicative of mechanical sanctioning.

These red flags directly feed into the **selection of DDOs for field verification**.

2.6 Implementation Flow (Summary)

Step	Activity	Responsible Unit
1	Identify available datasets (Layer 1–3).	DAC + Group Officer
2	Extract & store in PostgreSQL environment.	Data Steward
3	Conduct QC & indexing; design standard + scheme-specific queries.	DAC
4	Run analytical scripts; flag anomalies.	DAC / Data Scientist
5	Validate anomalies with DDO / scheme records.	Field Audit Team
6	Document deviations and monetary impact in reporting matrix.	Audit Officer

2.7 Outcome

A uniform master analytics layer ensures:

- Comparable and repeatable audit results across schemes.
- Quantifiable basis for DDO and district selection.
- Integration of data science with field audit to produce verifiable, evidence-based audit findings.

Part 3.

Identification of DDOs and Units for Field Audit

3.1 Why Field Audit Remains Central

Even in a data-driven audit environment, **field audit is the decisive layer** of assurance. Data can highlight where patterns appear abnormal or where probabilities of misuse seem high, but it cannot by itself confirm fraud, fabrication, or collusion. A beneficiary record may be complete in every column yet false on the ground; conversely, incomplete data may still represent genuine cases.

Thus, data analytics must be seen as a **directional tool**, not a replacement for the auditor's judgment.

The field is where:

- Authenticity of uploaded documents is verified through physical inspection.
- Beneficiaries, institutions, and DDOs can be questioned directly.
- Evidence that lies **beyond digital footprints** - such as signatures, manual registers, or social validation - is collected.

At the same time, data allows audit to enter the field **with focus** rather than randomness. Machine learning, statistical scoring, and rule-based checks can narrow millions of records into a few hundred *high-probability targets*. This synthesis - **AI for discovery, human audit for confirmation** - is at the heart of focus-based compliance audits.

3.2 Two Components of Field Audit

Field audit in social-sector schemes has **two clear parts**. Data analytics helps in identifying both, but their purpose in the field is different.

Component A – Data-Flagged Verification (Direct Transactions): These are **beneficiaries or transactions directly flagged by data** as suspicious. Examples include duplicate payments, same bank account used for many beneficiaries, age or income mismatch, or voucher drawn without a sanction. These cases already have a high level of confidence from data. The work of the field team is to **verify documents and records**, confirm that the irregularity exists, and record the money value involved.

Component B – Risk-Area Audit (High-Risk Geographies): Here, **individual transactions are not directly flagged**, but **some areas or DDOs show abnormal patterns**. This could be a district or block with unusually high approvals, sharp spikes in drawals, or unusual age or gender mix of beneficiaries. Data only shows that these areas behave differently; it does not confirm any wrong-doing. The field team must **audit** such areas in detail - study the process, test a sample of cases, and see whether the abnormal pattern has any systemic reason or misuse behind it.

Component	What It Covers	Confidence from Data	Main Field Task	Example
A – Data-Flagged Verification	Direct transactions or beneficiaries found irregular in data.	High	Verify supporting documents, and beneficiary presence.	Same bank account used twice for scholarship payment; verified in field.
B – Risk-Area Audit	DDOs, districts, or institutions showing unusual data patterns.	Medium	Audit processes, sample records, and control systems.	One block showing 25 % graduation-stage claims vs 5 % elsewhere; audit finds fake certificates.

3.3 Analytical Basis for Selecting DDOs, Districts, or Geographies

Risk ranking for field audit should be flexible in **granularity** - it can operate at DDO level, district level, block level, or even institutional cluster level, depending on data density and scheme design. Analytics should therefore output not just "risky DDOs" but also "risky geographies."

3.3.1 Quantifiable Risk Indicators

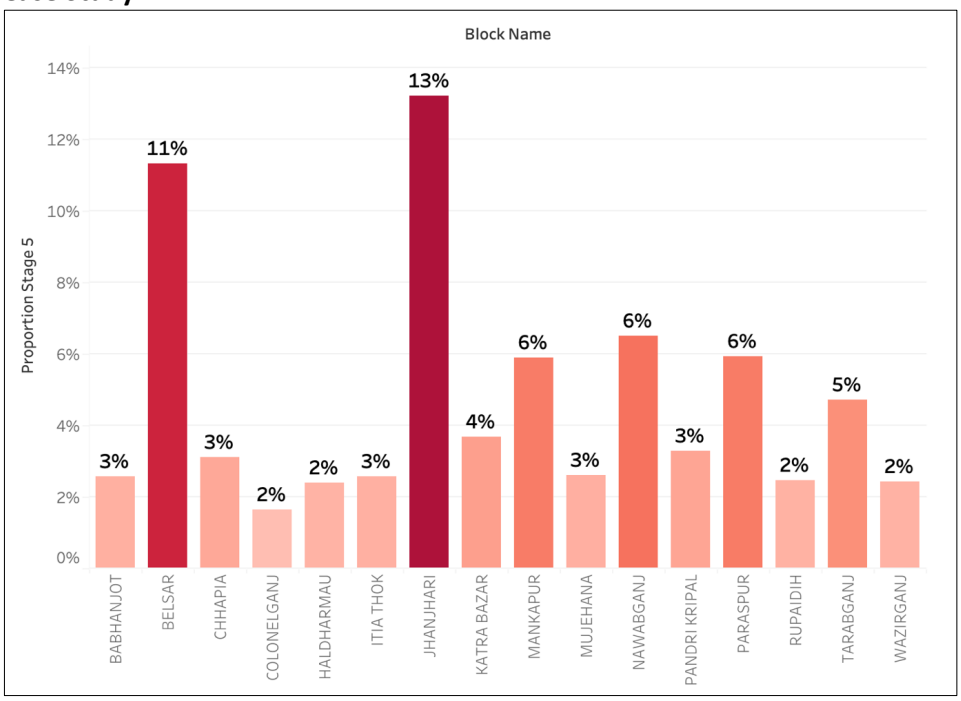
The following indicators - computable from VLC, IFMS, and scheme databases - should form the minimum analytic base for selecting audit units.

Each indicator can be expressed as a deviation from the state mean (z-score, percentile rank) or as a categorical red-flag (High/Medium/Low).

Risk Indicator	Data Source	Analytical Method	Audit Interpretation / Use
Beneficiary–Drawal Ratio Outlier	Scheme + IFMS	Compare average drawal per beneficiary (₹) across DDOs/districts.	High values suggest inflated beneficiary numbers or fund diversion.
Temporal Drawal Bunching	IFMS	Month-wise trend analysis; flag if >35–40 % of payments cluster in last quarter.	Indicates mechanical sanctioning or forced utilisation.
Head-of-Account Deviation	VLC	Cross-tabulate heads used by DDO vs scheme-notified heads.	Mis-classification or mis-booking of expenditure.
Repeated Sanction IDs / Voucher Overlaps	IFMS + Scheme	Detect reuse of same sanction order or voucher number.	Duplicate payments or data tampering.
Rejection / Approval Ratio Anomaly	Scheme portal	Compute approval : rejection ratio per DDO vs state mean.	Extremely high or low ratios may signal manipulation.
Demographic Skew	Scheme + Census	Compare age/gender composition of beneficiaries with district demographic profile.	Disproportionate representation suggests selective targeting.
Bank / IFSC Concentration	Scheme + Bank master	Identify IFSC codes receiving abnormally high share of payments.	Proxy or ghost accounts concentrated under same branch.
Dormant or Zero-Transaction Beneficiaries	Scheme + Bank reconciliation	Beneficiaries sanctioned but showing no withdrawal activity.	Indicative of non-existent or inactive beneficiaries.
Document Pattern Anomaly	Scheme + Cross-Verification Layer	File size, timestamp or hash uniformity analysis.	Mass upload from single source, possible fabrication.
Inter-Scheme Duplication	Multiple schemes	Cross-match Aadhaar or bank accounts across schemes.	Same person drawing benefits from multiple overlapping programmes.

These indicators may be weighted (for example, 0.2 × financial anomalies + 0.3 × beneficiary anomalies + 0.5 × document integrity anomalies) to generate a **Composite Risk Index (CRI)** at any chosen level - DDO, district, or cluster.

Case Study:



How risky areas were identified for field audit in a social sector scheme in Gonda district UP: Under the data-driven audit approach, block-wise claimant patterns were analysed to detect outliers. In Gonda district, for example, the share of Stage-5 (Class IX) claimants was found abnormally high in two blocks compared to the district average. Such concentration anomalies were used to select priority blocks for targeted field verification.

3.3.2 Technical Implementation

1. **Data Aggregation Layer:** Merge VLC/IFMS with scheme-portal data at DDO + month + beneficiary granularity.
2. **Normalization:** Calculate metrics such as drawal per beneficiary, sanction-to-drawal lag, and approval ratios.
3. **Statistical Scoring:** Apply z-scores or percentile ranking for each metric; label top 5 % as "high-risk."
4. **Cluster Detection:** Use unsupervised models (DBSCAN, K-Means) on indicators to detect natural clusters of anomalous DDOs.
5. **Composite Scoring:** Aggregate indicators into a single CRI (0–100).
6. **Visualization:** Produce heat-maps showing high-risk zones - districts, DDO clusters, or blocks.

3.3.3 Example of Output Dashboard

Unit Type	Identifier	Composite Risk Score (0–100)	Dominant Anomaly	Recommended Action
DDO	DDO-001 (Minority Welfare, Lucknow)	89	Duplicate sanctions + year-end bunching	Include in primary field sample
District	Gonda	84	High graduation-stage ratio in MKSY	Target 3 blocks for anomaly verification
Block	Tarabganj	82	Excessive IFSC concentration	Verify college-linked payments
Institute	Govt. Inter College No. 5	76	Duplicate bank accounts	Physical verification of student list

Interpretation:

- Units with **CRI > 80** = *mandatory field verification.*
- Units between **60–80** = *include sample for comparative study.*
- Units below **60** = *retain as control set.*

This scoring system allows planning at any operational level - from individual DDOs to whole districts - ensuring that both **financial anomalies (transaction side)** and **behavioural anomalies (geographic or demographic side)** feed directly into field-audit planning.

3.4 Field Audit Toolkits

To operationalise this model, two **standardised toolkits** can be issued to every field party before departure. These are auto-generated from the analytics dashboard.

Toolkit Type	Purpose	Contents
Toolkit 1 – Data-Flagged Beneficiary Pack	To verify specific beneficiaries or transactions identified by analytics with high probability of irregularity.	– Beneficiary ID, scheme, anomaly description (e.g., duplicate bank, invalid age). – Expected documents to inspect: sanction order, income/caste cert., school record, passbook. – Field checklist: confirm existence, physical address, document authenticity.
Toolkit 2 – Outlier-Area Verification Pack	To investigate blocks/DDOs flagged as statistical anomalies by ML or outlier analysis.	– Summary table of anomalies (approval ratio, sanction density, etc.). – Random beneficiary subset for spot verification. – Interview guide: DDO process flow, verification bottlenecks, certificate-checking mechanism.

Example of Toolkit 1 Entry Scheme: Post-Matric Scholarship

	A	B	C	D	E	F	G	H	I	J
1	inst_code	inst_name	app_id	session_year	name	fname	birthdt	overall_risk_type	flag_sources	Field Audit Observation?
2	08102	BABU JAIPAL SINGH PRIVATE ITI, HASAN PUR JIWANI	081020102100010	2022-23	ARUN KUMAR	RAMKUMAR	1999-01-07 00:00:00.000	HIGH RISK	1G	
3	08102	BABU JAIPAL SINGH PRIVATE ITI, HASAN PUR JIWANI	081020102200009	2022-23	SANJEET JANGRA	VINOD JANGRA	2000-02-02 00:00:00.000	HIGH RISK	1G	
4	08102	BABU JAIPAL SINGH PRIVATE ITI, HASAN PUR JIWANI	081020102200023	2022-23	KAPIL KUMAR	RAJ KUMAR	1999-04-01 00:00:00.000	HIGH RISK	1G	
5	08102	BABU JAIPAL SINGH PRIVATE ITI, HASAN PUR JIWANI	081020502200019	2022-23	NARENDRA KUMAR	SARDAR SINGH	1986-09-17 00:00:00.000	HIGH RISK	1G	
6	08089	BAGHPAT I T I BAMNAULI	080890102100010	2022-23	ASHISH KUMAR	ARUN KUMAR	1998-12-01 00:00:00.000	HIGH RISK	1G	
7	08089	BAGHPAT I T I BAMNAULI	080890102100101	2022-23	SHUBHANSHU MALIK	MANOJ MALIK	1998-11-25 00:00:00.000	HIGH RISK	1G	
8	08089	BAGHPAT I T I BAMNAULI	080890102200022	2022-23	SUNNY KUMAR	SURESH PAL	1999-05-24 00:00:00.000	HIGH RISK	1G	
9	08089	BAGHPAT I T I BAMNAULI	080890102200024	2023-24	ANKUR RATHI	AMARPAL SIGH	1997-05-05 00:00:00.000	HIGH RISK	1G	
10	08089	BAGHPAT I T I BAMNAULI	080890102200049	2023-24	SAGAR	RAMPAL	2000-06-20 00:00:00.000	HIGH RISK	1G	
11	08089	BAGHPAT I T I BAMNAULI	080890102200053	2022-23	VINIT KUMAR	NARENDRA KUMAR	2000-08-07 00:00:00.000	HIGH RISK	1G	
12	08089	BAGHPAT I T I BAMNAULI	080890102200053	2023-24	VINIT KUMAR	NARENDRA KUMAR	2000-08-07 00:00:00.000	HIGH RISK	1G	
13	08089	BAGHPAT I T I BAMNAULI	080890102200058	2022-23	PRINCE TOMAR	ANIL KUMAR	2001-07-02 00:00:00.000	HIGH RISK	1G	
14	08089	BAGHPAT I T I BAMNAULI	080890102200058	2023-24	PRINCE TOMAR	ANIL KUMAR	2001-07-02 00:00:00.000	HIGH RISK	1G	
15	08089	BAGHPAT I T I BAMNAULI	080890102300134	2023-24	Anuj Kumar	C/O: Sukhpal Singh	2000-05-12 00:00:00.000	HIGH RISK	1G	
16	08089	BAGHPAT I T I BAMNAULI	080890102300134	2024-25	Anuj Kumar	C/O: Sukhpal Singh	2000-05-12 00:00:00.000	HIGH RISK	1G	
17	08089	BAGHPAT I T I BAMNAULI	080890102400021	2024-25	Goutam Kumar	S/O: Rantej Singh	2000-02-10 00:00:00.000	HIGH RISK	1G	
18	08089	BAGHPAT I T I BAMNAULI	080890102400022	2024-25	Pratap Kumar	S/O: Arun Kumar	2000-06-16 00:00:00.000	HIGH RISK	1G	
19	08089	BAGHPAT I T I BAMNAULI	080890102400025	2024-25	Gaurav Panwar	S/O Naresh Pal	1996-10-15 00:00:00.000	HIGH RISK	1G	
20	08089	BAGHPAT I T I BAMNAULI	080890102400026	2024-25	Arpit Baliyan	S/O: Pratap Singh	2002-05-05 00:00:00.000	HIGH RISK	1G	
21	08089	BAGHPAT I T I BAMNAULI	080890102400037	2024-25	Sagar	Devendra Singh	2000-01-11 00:00:00.000	HIGH RISK	1G	
22	08089	BAGHPAT I T I BAMNAULI	080890102400050	2024-25	Anuj Baliyan	Satish Baliyan	2000-01-05 00:00:00.000	HIGH RISK	1G	
23	08089	BAGHPAT I T I BAMNAULI	080890102400071	2024-25	Antriksh Chaudhary	S/O: Anil Kumar	2000-01-05 00:00:00.000	HIGH RISK	1G	
24	08020	BARAUT COLLEGE OF EDUCATION GURANA ROAD BARAUT	080200102400025	2024-25	Nikunj Chauhan	S/O: Rakesh Kumar	2007-02-12 00:00:00.000	HIGH RISK	1C	
25	08020	BARAUT COLLEGE OF EDUCATION GURANA ROAD BARAUT	080200202400051	2024-25	Jatin	S/O: Ajab Singh	2005-09-10 00:00:00.000	HIGH RISK	1C	
26	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102200055	2023-24	AKASH NAIN	KALURAM NAIN	1996-10-13 00:00:00.000	HIGH RISK	1G	
27	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102300126	2023-24	Brijpal Kumar	S/O: Roshan Lal	1992-03-08 00:00:00.000	HIGH RISK	1G	
28	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102300126	2024-25	Brijpal Kumar	S/O: Roshan Lal	1992-03-08 00:00:00.000	HIGH RISK	1G	
29	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102300127	2023-24	Prashant Panchal	C/O Rammeher Panchal	1999-12-04 00:00:00.000	HIGH RISK	1G	
30	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102300127	2024-25	Prashant Panchal	C/O Rammeher Panchal	1999-12-04 00:00:00.000	HIGH RISK	1G	
31	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490102400041	2024-25	Ashish Kumar	RAMESH CHAND	2003-03-07 00:00:00.000	HIGH RISK	1G	
32	08049	BHARTIYA RISHIKUL ITI CHHAPROLI ROAD BARAUT	080490502200015	2022-23	KULDEEP SHARMA	OMPAL SHARMA	2000-05-23 00:00:00.000	HIGH RISK	1G	

Figure: Toolkit for Field Verification under Scholarship Audit. This toolkit lists individual scholarship applications flagged through data analytics for specific risk indicators. Each row represents one unique application. The column titled *Flag Code* records the reason for flagging - for instance, **"1G"** denotes cases where the same applicant appeared to be linked with multiple bank account numbers. The toolkit was provided to field audit parties, who were required to verify each flagged case on-site and record whether the anomaly was actually observed or not.

3.5 Role of Machine Learning and Human Judgment

Machine learning models - such as clustering, isolation forests, or time-series deviation scores - can **highlight "super-wild" patterns** that conventional analytics might miss: sudden bursts of activity, micro-collusion among DDOs, or improbable progression rates. However, these are **probabilistic signals**. They cannot substitute audit reasoning. For instance, a block with unusually high rejections could reflect stricter local scrutiny rather than fraud. Therefore, **ML-flagged anomalies must always feed into a human-verified field audit cycle.**

In practice:

1. ML algorithms identify statistical outliers.
2. DAC reviews and prioritises plausible ones.
3. Audit field parties validate on-ground authenticity and document evidence.
4. Verified cases feed back into model training for improved precision.

This iterative cycle ensures that **data analytics stays grounded in field reality** and field audit benefits from continuously refined risk intelligence.

3.6 Integrating All Layers for DDO Prioritisation

A combined dashboard integrating **Scheme MIS + IFMS + VLC + Cross-Verification Layer** should produce:

- A **DDO Risk Heatmap** (colour-coded by composite anomaly index).
- A **Scheme-wise Deviations Chart** (count of red flags by DDO).
- A **Temporal Pattern Graph** (month-wise drawal concentration).

Such visual summaries allow senior officers to **plan audit tours strategically**, ensuring maximum coverage of high-risk zones within available time and manpower.

3.7 Expected Outcomes

- Field teams work with **clarity, not randomness** - knowing which transactions to check and why.
- Data-confirmed cases yield **quantifiable money-value audit observations.**
- Outlier-based checks uncover **systemic or procedural weaknesses** invisible to data alone.
- Machine learning enhances **risk detection**, while auditors retain control over **verification and judgment.**

Ultimately, this two-track, data-enabled field audit model ensures that the CAG's compliance audit of social-sector schemes remains **evidence-based, risk-focused, and human-centred** - combining the precision of analytics with the credibility of physical verification.

Part 4.

Audit Design Matrix (ADM)

4.1 Purpose of ADM

The Audit Design Matrix translates audit objectives into **specific, testable questions** that can be answered directly through data and verified in the field. In focus-based audits of social-sector schemes, two categories of ADMs are needed:

1. **Standard ADM** – questions and checks applicable to *all* beneficiary-based schemes.
2. **Scheme-Specific ADM** – checks basis actual **rules and eligibility conditions** of scheme.

4.2 Part A - Standard ADM for Beneficiary-Based Schemes

Every beneficiary-linked dataset - scholarship, pension, girl-child incentive, housing - contains one record per payment or sanction. In theory, each row is unique, but in practice the same individual may appear multiple times under slightly different identifiers. Hence, the first step of analytics is to **create a reliable "identity key"** for each person and test for duplication or mismatch.

Best Practice: Creating Composite Identity Keys for Duplicate Detection

Although most Direct Benefit Transfer (DBT) schemes are Aadhaar-linked, experience from field audits shows that claimants have often managed to bypass or distort Aadhaar tagging. In several datasets, Aadhaar details may also be absent or masked. To ensure meaningful duplicate testing, audit teams may therefore create **composite identity keys** by combining multiple personal and contextual fields.

Recommended approach:

- Combine fields such as *Name + Father's Name + Gender + Date of Birth + District/Block + Bank Account Number*.
- Where available, include *IP address* or *mobile number* as secondary markers to strengthen matching.
- For education-linked schemes, use *Board Name + Year + Roll Number* as an additional cross-identifier.
- Relying solely on the system-generated beneficiary or application ID is insufficient, as each new portal entry is treated as unique even when the same person has reapplied.

Using such multi-field composite identifiers may significantly improve the accuracy of detecting duplicate or recycled claims across DBT schemes.

Common Data-Level Questions for All Schemes

Area	Audit Question / Check	Data Fields Required	Risk Indicated
Duplication of Beneficiaries	Are there multiple payments to the same person (based on identity key)?	Name, DOB, Father's Name, Bank A/c	Duplicate or fake claim.
Bank Account Mismatch	Are payments going to accounts mapped to multiple beneficiaries?	Bank A/c No., IFSC, Beneficiary ID	Ghost or proxy accounts.
Income / Caste Certificate Mismatch	Are income or caste certificates invalid, expired, or repeated across beneficiaries?	Certificate No., Issue Date, Department Code	Certificate reuse or forgery.

Document Integrity	Are uploaded files identical in size/time/hash across many records?	File metadata	Mass upload of forged docs.
Geographical Outlier	Are certain blocks showing abnormal concentration of beneficiaries or payment value?	District/Block Code, Payment Amount	Collusion or weak scrutiny.
Temporal Outlier	Are most sanctions clustered at year-end months?	Sanction Date, Payment Date	Mechanical or back-dated approvals.
Cross-Scheme Overlap	Are same identity keys found in multiple welfare schemes?	Aadhaar/Name/Bank A/c across datasets	Double-benefitting.
Beneficiary Dropouts	Are beneficiaries missing in continuation stages of multi-year schemes?	Year, Stage Code	Poor tracking or fake renewals.
IP / Device Pattern	Are many applications uploaded from same IP or device ID?	IP Address, Timestamp	Bulk upload by intermediaries.

These checks form the **baseline ADM** to be applied to any scheme before designing scheme-specific queries.

4.3 Part B - Scheme-Specific ADM

Each scheme has its own business rules - eligibility criteria, stage gaps, income limits, academic conditions, etc. - that must be converted into testable data questions.
Below are examples from two major schemes in Uttar Pradesh.

(i) Mukhyamantri Kanya Sumangala Yojana (MKSY)

The scheme provides conditional transfers to girl beneficiaries at **six life stages**: birth, immunisation, school admission (Class 1, 6, 9), and graduation.

Typical eligibility rules:

Rule / Condition	Focused ADM Question	Required Fields
Minimum age gap between stages (e.g., Birth → Class 1 = 5 years)	For each beneficiary, check if time gap between Stage 1 and Stage 3 ≥ 5 years.	DOB, Stage Dates
Only two girl children per family eligible	Identify families with > 2 beneficiaries sharing same parent name/address.	Parent Name, Address
Same child cannot claim higher-education stage without earlier stages	Check if any Stage 6 payment exists without matching Stage 3–5 record.	Beneficiary ID, Stage Code
Payment only for valid institutes	Verify if institute codes in data match state-approved list.	Institute Code, Institute Master

(ii) Scholarship Scheme (Pre-Matric / Post-Matric)

Scholarship data are annual and depend on enrolment and academic performance. Common rules and how they become ADM questions:

Rule / Condition	Focused ADM Question	Required Fields
Student must be enrolled in a recognised institution	Cross-check college/school code against education department master.	Institute Code

Student cannot receive two scholarships in same year	Check same student (Name + DOB + Father's Name) across scheme and year.	Name, DOB, Father's Name, Year
Attendance ≥ 75 % required	Compare attendance field with threshold; flag below-limit cases.	Attendance %
Scholarship discontinued on dropping out	Detect beneficiaries appearing in previous year but missing in current year.	Year, Beneficiary ID
Income below notified ceiling	Verify income certificate value ≤ threshold.	Income Amount

4.4 Using the ADM in Practice

1. **Run Standard ADM first** on the entire dataset to remove clear duplications and errors.
2. **Apply Scheme-Specific ADM** next to identify violations of the scheme's own rules.
3. **Summarise Results:** count number of irregular cases and total payment value per ADM question.
4. **Send High-Risk Records to Field Teams** as part of *Toolkit 1* (data-flagged verification).
5. **Update the ADM Template** after field feedback so that new risk patterns can be added for the next cycle.

Part 5.

Reporting Framework

5.1 The Need for Standardisation Before Reporting

Before reporting can even begin, it is necessary to ensure that the information coming from multiple field audit parties is **collected in a standard, comparable form**. In conventional audits, much time is lost because different parties record findings in different formats, and aggregation becomes manual.

To avoid this, **pre-formatted audit toolkits** must be used by all teams.

These toolkits, generated from the analytical system, serve two purposes:

1. They provide each team with **a uniform set of cases or areas to examine** (as discussed under Components A and B).
2. They allow every observation to be recorded in the same columns and codes, so that results from all districts or DDOs can later be compiled automatically.

When every audit party works on the same structure - question, finding, supporting document, money value, and status - the process of data consolidation and report generation becomes mechanical rather than manual. This step is essential because the purpose of focus-based audits is not only to find individual irregularities but also to **generate a state-level analytical picture** from field verifications.

5.2 Structure of the Reporting Framework

Reporting should have **two distinct components**, each reflecting a different lens of analysis. Together they help audit portray both the *macro picture* (what problems recur everywhere) and the *micro picture* (what problems are unique to a region or implementation pattern).

5.2.1 Component 1 – Issue-Based Consolidated Reporting

This part of reporting captures **common problem areas** found across districts and schemes. The idea is to identify what types of deviations repeatedly occur and to quantify them.

A consolidated table should be maintained for roughly **ten major issue-heads**, such as:

- Duplicate or multiple payments to same beneficiary.
- Payments made to invalid or proxy bank accounts.
- Invalid or reused income/caste certificates.
- Beneficiaries drawing from multiple schemes simultaneously.
- Missing or tampered documents.
- Payments sanctioned without proper verification.
- Payments made to ineligible age or class groups.
- Concentration of payments in a single IFSC or institution.
- Absence of field verification at district level.

Each audit party will simply record **Yes/No** against each case, with supporting evidence and amount involved. The reporting tool can then auto-aggregate these entries to a **state-wide summary**, like:

Issue Head	No. of Cases Confirmed	Money Value (₹ lakh)	Schemes Involved
Duplicate Payments	152	54.7	Scholarship, MKSY
Invalid Income Certificates	93	27.1	Scholarship
Ghost Beneficiaries	61	19.3	Pension
Ineligible Age Group	47	15.8	MKSY

This uniform format ensures that results from every district feed into one consolidated analytical dashboard without manual re-tabulation.

5.2.2 Component 2 – Region-Specific / Area-Specific Reporting

Alongside the consolidated issue tables, field parties must also prepare **area-specific summaries** highlighting patterns that are peculiar to a region.

These may include:

- Clusters of similar irregularities in adjoining districts.
- Local administrative practices leading to systematic deviations.
- Influence of intermediaries, NGOs, or cyber-cafés assisting in fraudulent applications.

For instance, in both the *Mukhyamantri Kanya Sumangala Yojana* and *scholarship schemes*, similar misuse patterns - reused school certificates and proxy bank accounts - were found across several north-eastern districts of Uttar Pradesh. This shows that a single fraudulent technique can diffuse regionally through informal channels. Capturing such regional variations helps the audit build a **spatial risk map**, enabling targeted administrative action.

Each district's reporting sheet should therefore have two short sub-tables:

1. *Common deviations confirmed* (linked to the ten issue heads above).
2. *Region-specific findings* - new patterns, local issues, or procedural weaknesses unique to that geography.

This dual approach ensures that the final CAG report can present both **thematic insights** and **regional differentiation**, rather than scattered lists of local irregularities.

Part 6.

Recommendations Framework

6.1 Purpose

The recommendations section is where a compliance audit becomes meaningful for governance. An audit that merely lists irregularities does not add value unless it also identifies **why** those irregularities occurred and **how** they can be prevented. The goal of this framework is to ensure that every audit report ends with clear, actionable recommendations that connect field findings to systemic improvements.

6.2 The Need for Root-Cause Analysis

Before any recommendation is written, there must be an explicit **root-cause analysis**. This means understanding *why* a pattern of irregularities occurred - whether it arose from a design flaw in the scheme, a procedural gap, weak supervision, or misuse by intermediaries.
Root-cause analysis must avoid two extremes:

- Merely restating the irregularity ("documents were fake") - which adds no new insight, and
- Making broad administrative judgments ("officials failed to check properly") - which offer no direction for reform.

Instead, it should identify specific, traceable weaknesses that can be corrected. Typical root causes observed in social-sector audits include:

Type of Weakness	**Illustration**
Design and IT Gaps	Portals not linked to validation datasets like income, caste, or education board records; lack of automated checks for duplicate certificates.
Procedural Weakness	Sanctions issued without cross-verification, mechanical approvals during year-end rush, lack of second-level review before disbursal.
Field-Level Constraints	Verification workload exceeding capacity; one officer verifying thousands of applications within short timelines.
External Exploitation	Cyber-cafés or intermediaries uploading multiple forged applications using common IPs or devices.

Identifying such reasons turns an audit paragraph from a statement of loss into a **learning tool** for the department.

6.3 Structure of Recommendations

Recommendations should always be written at **two levels** - **Policy / System Level** and **Field / Implementation Level** - to distinguish where corrective action lies.

6.3.1 (A) Secretariat / Directorate Level Recommendations

These are **policy and system improvement measures** to address weaknesses found across the state or across multiple schemes. They should focus on prevention through design improvement rather than retrospective fault-finding.

Key elements may include:

- **Data Linkages:** Departments should integrate their beneficiary portals with external databases such as education board rolls, income and caste certificate repositories, and population records. This will ensure that eligibility is verified automatically rather than manually.

- **Built-In Risk Analytics:** Departmental MIS systems should embed simple data-integrity checks - duplicate bank accounts, repeated certificate numbers, same IFSC concentration - so that red flags appear *before* payment, not after audit.

- **Workflow Reinforcement:** Introduce mandatory dual verification for each stage: one officer to verify documents and another to approve payment. This distributes accountability and reduces mechanical sanctioning.

- **Periodic Data Review:** Every six months, conduct internal analytics using the same ADM questions used in audit to self-test data quality. Such pre-audit testing improves departmental readiness and reduces recurring audit objections.

These recommendations are not about creating new administrative layers, but about embedding **basic validation logic** into the scheme's own systems.

6.3.2 (B) Field / District Level Recommendations
These address **implementation and supervision gaps** visible on the ground. The intent is to make local monitoring more data-aware and time-bound without adding excessive burden on staff.

- **Use of Verification Dashboards:** Each District Magistrate or departmental head should have a simple dashboard showing pending verifications, rejected applications, and high-risk flags. This allows quick reviews rather than paperwork-based inspections.

- **Focused Verification Drives:** Where audit finds repetitive anomalies (for example, same IFSC or repeated certificates), targeted verification drives should be launched in those blocks rather than state-wide random checks.

- **Improved Record Authenticity:** Each DDO should maintain a digital "beneficiary verification register" linking field visit reports, photo evidence, and supporting document numbers. This creates traceability for future audits.

- **Capacity and Awareness:** Short training modules for data entry operators and field verifiers should be conducted to explain how misuse occurs - e.g., how duplicate certificates look or how IP patterns indicate proxy uploads. Awareness itself becomes a deterrent.

6.4 Implementation and Feedback Loop
Recommendations should not end with the report. After acceptance by the department, each key recommendation must be **mapped to a measurable follow-up indicator** (for example, "linkage with board data completed," "dual verification module activated"). These indicators should be reviewed in the next audit planning cycle to measure improvement. This feedback loop ensures that the compliance audit evolves continuously, learning from past findings and adapting its data analytics and field focus areas accordingly.

ANNEXURE IV

Data Driven Approach for Works Audit

As per the *CAG's State Finance Report for FY 2022-23*, the **Works sector** - comprising *Public Works, Transport, and Irrigation & Flood Control* - emerges as one of the most capital-intensive and fiscally significant segments of government expenditure. Collectively, these three sectors accounted for an extraordinary **₹2.88 lakh crore in capital outlay**, representing nearly **45 percent of the State's total capital expenditure**, and an additional **₹1.20 lakh crore in revenue expenditure**. Within the capital segment, **Transport alone absorbed ₹1.65 lakh crore (25.4%)**, followed by **Irrigation & Flood Control ₹1.06 lakh crore (16.4%)** and **Public Works ₹16,785 crore (2.6%)**.

These massive allocations reflect the sheer fiscal weight of physical infrastructure creation - roads, bridges, canals, flood-control works, and related assets - executed across thousands of divisions and contracts. Given such financial magnitude, complexity, and the volume of transactions, the Works domain represents a high-value and high-risk area demanding **focused, data-driven audits**.

1. Objectives of Data-Driven Works Audit

1. **To identify divisional formations and responsible officers** associated with each work and payment, ensuring clear accountability across the execution chain.
2. **To establish a complete audit trail** covering all stages of a work - from tender initiation, bid evaluation, and award to final payment and closure.
3. **To enable risk profiling of contractors** based on factors such as work concentration, bid participation, financial exposure, and payment behaviour.
4. **To detect patterns and anomalies** in fund release, deduction practices (GST, IT, labour cess), and repeated award of works to select contractors.
5. **To link financial data with administrative and physical progress**, ensuring that expenditure corresponds with actual on-ground execution.
6. **To reduce manual dependence** by digitising legacy records through OCR and converting them into structured, analyzable datasets.
7. **To strengthen audit selection and field verification** by prioritising high-risk works, divisions, or DDOs identified through integrated data analysis.

2. Data Sources for Data-Driven Works Audit

For a **data-driven works audit approach**, a rich volume of data is already available across multiple platforms maintained by both the **Accountant General (A&E)** and various **executing departments**. These datasets - routinely generated as part of accounting, treasury, procurement, and tax processes - offer extensive coverage of financial, contractual, and physical aspects of works. Since much of this data is already being produced and maintained in digital form, gaining access and integration is both practical and feasible. Thus a **high-quality, data-driven audit** can certainly be conducted - one that meaningfully enhances assurance, transparency, and accountability in the management of public works.

Dataset / System	Description	Custodian / Source	Remarks / Audit Use
VLC (Voucher Level Computerisation)	Contains Head-of-Account-wise, month-wise receipts and payments.	AG (A&E)	In cases where these forms are not digitised, OCR / DataSnipper tools may be used for conversion.
Income Tax TDS / GST TDS	Division-wise, transaction-wise deductions enabling assessment of actual payment by each division.	AG (A&E)	GST TDS schedules also contain contractor names - useful for linkage and validation of payments.
GST Challan Data	Depositor-wise, date-wise, transaction-wise details of each division. Key fields include GST Number and TAN. Captures both Treasury and Non-Treasury transactions.	AG (A&E)	Directly available through GSTN portal or e-Treasury / Cyber Treasury interfaces.
GST TDS DDO Return (Form 7)	DDO-wise, contractor-wise, invoice-wise record of TDS deducted.	AG (A&E)	Enables reconciliation of contractor-wise payments and deductions.
Monthly Works Accounts (Compiled)	CPWA Forms 64, 65, 79 etc.: • Form 64 – work-wise, month-wise expenditure for budgeted works and cumulative totals. • Form 65 – details of deposit works. • Form 79 – summary of both budgeted and deposit works.	AG (A&E)	Foundational dataset for analysing expenditure at the work level.
Form 61 (Schedule Docket) with Vouchers and Bills	Contains three expenditure components – (1) payments to contractors, (2) stock charged to the work, and (3) charges through transfer entries.	AG (A&E)	*Transfer-entry expenditure* is a high-risk area and must be mandatorily analysed.
IFMS (Integrated Financial Management System)	Voucher-level details: 15-digit Head of Account, Treasury Code, Division, Deductions, Contractor Name (if available). In some states (e.g. Uttarakhand), project details with unique IDs enable detection of splitting of works.	AG (A&E) / State Treasury Directorate	Where Works is not integrated with Treasury, compiled records from all treasuries or e-Kuber data may be used to capture payment, payee, and bank details.

WAMIS (Works and Accounts Management Information System)	Automated Works/Contract Management System in some states; includes preparation of monthly accounts. Contains measurement books, work status, financial & physical progress, extension of time records.	State Works Department Administrator	Critical for linking financial data with physical progress and measurement books.
e-Procurement / Tendering System	Digital repository of tender documents, bid submissions, EMD details, evaluation, and award (Letter of Acceptance).	State e-Procurement Department Administrator	Enables analysis of competition, bidder qualification, tender timelines, and EMD refunds.
BOCW (Building and Other Construction Workers) Data	Records of labour-cess deductions transferred by divisions to the BOCW Board; includes contractor name, work name, and amount deducted.	BOCW Welfare Board	Helps identify off-treasury or parallel payments related to works.
Contractor Registration Database	Contains contractor details - PAN, financial solvency, machinery availability, registration date, and past experience.	Concerned Department	Facilitates risk profiling and detection of contractor concentration across divisions.
Mining Data (e-Rawanna / e-Transit Pass)	Tracks movement of construction material for specific works via contractor.	Department of Mines & Geology	Cross-check with divisional records to verify authenticity of material usage certificates.
CRC (Consignee Receipt Certificate)	Lists contractor name, work name, destination, and quantity of bitumen received.	Division	Useful for verifying delivery of bituminous materials in road works.
e-HRMS (Human Resource Management System)	Captures officer identity, tenure, and posting details.	State Finance / Personnel Department	Enables mapping of DDOs and Executive Engineers to periods of expenditure or anomalies.

3. Approaches for Conducting Works Audit through a Data-Driven Methodology

This paper proposes **three analytical approaches** for undertaking *Works audit* through a data-driven framework. These approaches are based on the nature and maturity of information systems operating across States and the type of datasets available to the Accountant General (A&E). While the three methods presented below are indicative, **field offices are encouraged to innovate and develop additional methodologies** suited to their local data ecosystems and audit priorities.

The **core principle** across all approaches is the application of **anomaly detection (outliers) in time-series and transactional data** to identify *outlier transactions* or *divisions/DDOs* exhibiting unusual expenditure patterns. Such analytical detection would form the basis for selecting units and transactions for detailed field audit. Given the variation in IT system maturity and data completeness across States, the concept of outlier analysis would be applied flexibly to whichever datasets are available.

Approach 1: States where Works Accounting System has been Dispensed with (VLC-Based Audit)

In States where the Works accounting system has been discontinued and data is fully captured in the **Voucher Level Computerisation (VLC)** system, audit analysis can be performed using transaction-level details available therein.

- **Data fields** include Treasury, DDO, Transaction ID, Voucher Number, Voucher Date, Head of Account, Gross Amount, Net Amount, Deductions, Contractor Name, and GSTN.
- The dataset should be **supplemented with a DDO profile**, covering parameters such as whether the DDO is attached to a sub-treasury, its remoteness of location, dependency on outsourced staff, and implementation status of e-Measurement Book (e-MB).
- **Analytical features** will be computed using data for at least the past three financial years - including *mean, median, standard deviation, z-score, and rolling mean* - to detect expenditure spikes and outlier behaviour over time.

Approach 2: States where Works Accounting System is still in Existence (CPWA + VLC-Based Audit)

In States where the Works Accounting System (based on CPWA Forms like form 61 which shows voucher wise/workwise transactions) continues to operate alongside Treasury-linked datasets:

- Analysis can be conducted on **VLC data fields** such as Treasury, DDO, Transaction ID, Voucher Number, Voucher Date, Head of Account, and Gross Amount.
- Similar to Approach 1, the analysis will be **supplemented with DDO profile attributes** - linking sub-treasury attachment, staff dependence, remoteness, and e-MB adoption.
- Time-series analytics will again utilise *three-year expenditure profiles* with statistical features (mean, median, standard deviation, z-score, rolling mean) to highlight abnormal transaction behaviour or expenditure surges indicative of risk.

Approach 3: States with Access to GST-TDS and GSTR-7 Data (Indirect Verification Approach)

In States where **GST-TDS data or GSTR-7 returns** are available, these can serve as an effective proxy dataset for assessing works-related payments and contractor activity.

- State AG offices already have access to GSTR-7 returns filed by Tax Deductors (Divisions/DDOs), containing **contractor-wise payment details and TDS deductions**.
- Since April 2025, GSTR-7 includes **Division-wise contractor details, invoice-wise gross amounts, and TDS values**, allowing direct analytical verification of payment flows.
- This data can be **augmented with temporal and DDO profile features** (as described in Approaches 1 and 2) for enhanced anomaly detection.
- A compilation of GSTR-7 returns across **100 – 800 Divisions** can be achieved within a period of roughly three months, and in some cases, data may also be obtained **centrally through the GST Wing or DGACR** for statewide analysis.

4. Implementation Methodology and Technological Support

For operationalising the data-driven audit of Works, it is essential to translate the conceptual framework into an implementable, technology-enabled workflow. The following methodology outlines the key stages and systems support required to execute the proposed approaches effectively.

1. **Selection of Divisions / DDOs:** Using the risk parameters and profiling techniques described in para 3 above, audit teams will identify Divisions and DDOs where data-driven review and OCR-based digitisation will yield maximum audit value. Priority will be given to divisions with high expenditure, recurring contractors, or weak data controls.

2. **Data Access and Compilation:** Multiple datasets - such as VLC, IFMS, GST-TDS, e-Procurement, WAMIS, and BOCW - will be obtained from the Accountant General (A&E) and concerned State departments. These datasets, though diverse in origin, are largely **readily available in digital form** and can be consolidated through structured extraction and mapping using primary keys like *Head of Account, DDO Code, GSTIN, and Contractor PAN.*

3. **Digitisation and OCR Conversion:** In cases where critical records such as **Form 61 dockets, vouchers, measurement books, or CPWA forms** exist only in paper or scanned-PDF format, Optical Character Recognition (OCR) or data-snipping tools will be deployed to convert them into analyzable datasets.
    - OCR software shall extract key fields - *Work ID, Contractor Name, PAN/GST, Bill Amount, Deductions, and Dates* - and link them with IFMS and GST data.
    - GPU or high-performance processing infrastructure may be used by the HoD within delegated powers to ensure timely and accurate extraction at scale.

4. **Data Integration and Analytical Review:** Once datasets are structured, they will be merged into a **composite analytical model**, enabling outlier detection, temporal trend analysis, and contractor risk clustering. Analytical tools such as Python-based models or visual analytics platforms may be used to detect anomalies (e.g., March-end expenditure spikes, irregular deduction patterns, or concentration of works).

5. **Validation and Reporting:** Audit observations emerging from analytics will be validated through divisional accounts, tender records, and field verification. This ensures that data-based inferences are corroborated by on-ground audit evidence before reporting.

6. **Technological and Institutional Support:**
    - **OCR / Data Extraction Tools:** For digitisation of physical and scanned documents.
    - **Analytical Platforms:** Python, Orange, or equivalent open-source tools for statistical anomaly detection.
    - **Hardware / Server Infrastructure:** Local GPU servers or cloud-based solutions for bulk OCR and data processing.
    - **Capacity Building:** Field offices may require short training modules on dataset handling, risk-based audit planning, and interpretation of analytical outputs.

Together, these measures establish a practical pathway for implementing large-scale, data-driven audits of Works, ensuring both analytical depth and operational feasibility.

Conclusion

The proposed framework represents a shift from **document-centric auditing** to **data-centric assurance**, integrating financial, contractual, and physical dimensions of Works expenditure. By combining datasets across VLC, IFMS, GST-TDS, e-Procurement, and OCR-converted documents, audit teams can generate a near-complete digital audit trail for every work - linking payments, officers, contractors, and physical progress.

This transition will not replace traditional verification but will **amplify audit effectiveness** - allowing limited field resources to focus on high-risk divisions and transactions identified through analytics. As more datasets become available and tools mature, the model can evolve into a **scalable, replicable, and technology-driven audit framework** for Works, enabling superior accountability and transparency in public infrastructure governance.

ANNEXURE V

Illustrations of Identified Problems in Data-Deficient Departments

In the **Forest Department,** which has lot of operations in data-deficient setups, following problem statements may be looked for compliance audit:

1. **Rising Human - Wildlife Conflict:** Escalating incidents due to fragmented habitats, encroachments, and weak restoration planning.

2. **Mismanagement of CAMPA Funds:** Delays, inadmissible expenditures, and poor monitoring of compensatory afforestation projects.

3. **Encroachment & Outdated Land Records:** Non-reconciled forest land records leading to diversion and loss of forest area.

4. **Cost Overruns in Conservation Projects:** Weak project appraisal and fragmented governance causing time and cost escalations.

5. **Low Plantation Survival Rates:** Target-driven plantations with poor ecological suitability and lack of post-plantation maintenance.

6. **Illegal Mining & Non-Recovery of Dues:** Unregulated operations and weak coordination with mining and revenue authorities.

7. **Manpower Shortage & Skill Gaps:** Chronic vacancies and inadequate training in GIS, enforcement, and ecological management.

8. **Weak Monitoring & Technology Adoption:** Absence of real-time MIS or GIS systems for plantation, finance, and enforcement tracking.

9. **Poor Legal Enforcement:** Low conviction rates in forest offences due to poor documentation and lack of case tracking.

10. **Financial Irregularities & Weak Internal Audit:** Unapproved expenditures and poor procurement oversight from limited internal controls.

Same way, in the **Police Department**, where current compliance audits have largely focused on **transactional or financial irregularities** rather than core policing outcomes, can be examined through a **problem-statement-based approach**. The following problem areas may be explored to assess operational efficiency, service delivery, and systemic accountability.

1. **Delayed or incomplete FIR registration**, especially in gender, SC/ST and cyber-crime cases, indicating non-adherence to procedural timelines (e.g., under Lalita Kumari vs State of Uttar Pradesh guidelines).

2. **Non-compliance with investigation protocols:** inadequate evidence collection, weak forensic linkages, and inconsistent documentation of investigations.

3. **Inadequate surveillance infrastructure**: non-functional CCTVs, absent integration with control centres, and gaps in monitoring public spaces.

4. **Obsolete weaponry and ammunition shortages**: poor maintenance, outdated firearms, delayed allocation of new arms, and weak inspection protocols.

5. **Shortage or non-availability of operational vehicles** and delayed repairs, undermining patrolling efficiency and crime-response capability.

6. **Infrastructure deficiencies**: substantial delays in construction, renovation and maintenance of police stations, training institutes, and residential quarters.

7. **High vacancies and skill-gaps** in key areas such as cyber crime, forensic investigation and specialised units, affecting investigative capacity and responsiveness.

8. **Poor digitisation and record-management**: incomplete adoption of systems such as e-FIR, CCTNS, online case tracking and integrated data platforms.

9. **Weak procurement, inventory control and distribution systems**: overstocking of uniforms, arms and equipment at central depots with ineffective field distribution and utilisation.

10. **Weak internal controls and governance structures:** inadequate segregation of duties, lack of internal audit oversight, weak tendering controls and incomplete documentation of works and maintenance activities.