Office of Data Governance and Analytics

Data Ethics Review Checklist

Purpose

This checklist is a pragmatic tool for project leads, data analysts, and IT staff to implement the [Agency] Data Ethics Policy. It must be completed for any new project, system, or software procurement that involves the collection, processing, or dissemination of personal information.

Project Information

Project Name:

Project Lead:

Date of Review:

Phase 0: Initial Risk Triage

Instructions: Complete this phase first to determine the required level of review.

1. Risk Level Triage

Does this project involve any of the following, check Y/N:

□ An automated system to make a high-stakes decision about an individual (e.g., benefits, fines, risk scores)? ***Example:*** *"A system that flags a citizen for a fraud investigation."*

Please Explain:

□ Collection of new sensitive data types (e.g., biometrics, genetic data, health data not already managed by a covered entity)? ***Example***: *"Using facial recognition on security footage."*

Please Explain:

□ Novel matching or combination of large datasets that was not previously possible? ***Example***: *"Combining VDOT toll records with tax records to verify residency."*
Please Explain:

□ Processing data from minors or other vulnerable populations? ***Example***: *"An app for students in the K-12 system."*
Please Explain:

□ A third-party vendor (SaaS) processing or storing Commonwealth PII? ***Example:*** *"Hiring a vendor to run our new case management system in the cloud."*
Please Explain:

2. Triage Outcome

If "No" to all of the above: The project is **Low-Risk**. The project lead may proceed after completing the rest of this checklist.

If "Yes" to any of the above: The project is **High-Risk**. A full Data Protection Impact Assessment (DPIA) must be scheduled with the Data Governance/Privacy office before proceeding.

Phase 1: Collection and Purpose (The “Why” and the “What”

1. Lawlessness and Public Benefit

□ Has the specific Code of Virginia § authorizing this data collection been identified? ***Example***: *"Yes, Code of Virginia § 32.1-39 (VDH) authorizes collection of disease data to protect public health."*

Please Explain:

□ Is the public benefit clearly defined and documented? ***Example***: *"The benefit is to reduce grant fraud by 15%, ensuring funds are distributed to eligible citizens."*

Please Explain:

2. Purpose Limitation

□ Is the purpose for collection specific and limited? ***Example:*** *"Data will only be used for verifying eligibility for the XYZ program, not for any other marketing or outreach."*

Please Explain:

□ Are there controls to prevent the data from being used for other “incompatible" purposes? ***Example***: *"The system is firewalled. Any new use would require a formal governance review and a new data sharing agreement."*

Please Explain:

3. Data Minimization

□ Is every single data field being collected absolutely necessary for the stated purpose? ***Example***: *"We are collecting date of birth for age verification, but not social security number, as a driver's license number is sufficient."*

Please Explain:

□ Have we considered if aggregated or de-identified data could achieve the same goal? ***Example***: *"For our dashboard, we only need a count of applicants per zip code, so we will not collect or store individual addresses for this purpose."*

Please Explain:

4. Transparency

□ Will the individual be informed at or before the time of collection? ***Example***: *"Yes, a clear 'Privacy Notice' is displayed on the webform before the user enters any data."*

Please Explain:

□ Does the notice clearly state the purpose, the legal authority, and whether disclosure is mandatory or voluntary? ***Example:*** *"The notice states, 'Providing your SSN is voluntary, but failure to do so may prevent us from processing your application.'"*

Please Explain:

Phase 2: Use, Processing, and Storage (The “How)

1. Fairness and Non-Discrimination

□ If this system uses an algorithm or AI to make decisions (e.g., scoring, risk-assessment), has it been tested for unfair bias? ***Example**: "The fraud-detection model was tested for disparate impact on applicants from rural vs. urban zip codes and showed no statistical bias."*

Please Explain:

□ Is there a plan for *ongoing* monitoring of the system for bias that may emerge over time? ***Example**: "A report on model-decision demographics will be sent to the Data Governance office quarterly for review."*

Please Explain:

2. Human Agency and Oversight

□ Is there a "human-in-the-loop" for critical or high-stakes decisions? ***Example**: "The system can flag an application as 'high-risk,' but only a human case manager can issue a final denial of benefits."*

Please Explain:

□ Is there a clear process for an individual to appeal an automated decision? ***Example**: "The denial letter includes a phone number and web link for a 'Request for Reconsideration' by a human supervisor."*

Please Explain:

3. Data Quality and Integrity

□ Are there validation rules at the point of entry to ensure data is accurate? ***Example**: "The system checks against the official USPS database to validate all addresses entered."*

Please Explain:

□ Is there a "clearly prescribed and uncomplicated procedure" (per GDCDPA § 2.2-3800) for individuals to access and correct their information? ***Example**: "Users have a 'Profile' page where they can see and edit their own contact information at any time."*

Please Explain:

4. Security and Confidentiality

□ Is the data encrypted at-rest and in-transit per SEC 530 standards? ***Example**: "Yes, using AES-256 (at-rest) and TLS 1.3 (in-transit)."*

Please Explain:

□ Are technical safeguards in place to prevent misuse or unauthorized access? ***Example**: "The data is stored in a secured Commonwealth data center, and access requires multi-factor authentication (MFA)."*

Please Explain:

5. Vendor and “Black Box” Risk

□ Is this system or algorithm proprietary to a third-party vendor? ***Example***: *"Yes, we are procuring a COTS 'student success' model from Vendor X."*

Please Explain:

□ Has the vendor provided documentation on how their system was tested for fairness and bias? ***Example***: *"No. This is a red flag. We must demand this in the RFP or contract."*

Please Explain:

□ Can the vendor's decisions be audited, appealed, or explained in plain English? ***Example***: *"Vendor says the AI model is proprietary. This is a 'No.' The system is an unexplainable 'black box' and presents significant risk."*

Please Explain:

□ Does the contract/RFP include our data ethics/security/GDCDPA requirements as binding clauses? ***Example***: *"Yes, our standard data protection addendum is attached to the contract."*

Please Explain:

□ Does the contract/RFP include our data ethics/security/GDCDPA requirements as binding clauses? ***Example***: *"Yes, our standard data protection addendum is attached to the contract."*

Please Explain:

Phase 3: Access, Sharing, & Lifecycle (The “Who” and “When”)

1. Access Control

□ Is data access limited by "role-based access control" (RBAC)? ***Example***: *"Only 'Caseworkers' can see PII. 'Analysts' can only see de-identified data. 'IT Admins' can manage the system but not view data."*

Please Explain:

□ Are access logs maintained and periodically audited? ***Example***: *"All access to sensitive records is logged. Logs are reviewed weekly by the security team for anomalous activity."*

Please Explain:

2. Data Sharing

□ If sharing data with other agencies or partners, is there a formal Data Sharing Agreement (DSA) in place? ***Example***: *"Yes, both agencies participate in the Commonwealth Data Trust (CDT) and the data to be shared is included in the exhibits. VDH explicitly limits their use of the data to the COVID-19 public health study only."*

Please Explain:

□ Does the DSA legally bind the other party to the same (or stricter) data protection standards? ***Example**: "The CDT includes a clause requiring the partner to adhere to all GDCDPA provisions and report any breaches within 24 hours."*

Please Explain:

3. Retention and Deletion

□ Has the official Library of Virginia (LVA) retention schedule for this data been identified? ***Example**: "Per LVA General Schedule 101, these 'Application Records' will be retained for 5 years after the case is closed."*

Please Explain:

□ Is there an automated or documented process to securely destroy the data once the retention period ends? ***Example**: "A quarterly script automatically identifies and securely purges records that have passed their 5-year retention date."*

Please Explain:

Phase 4: Review and Sign-Off

□ **Project Lead:** I affirm that this review has been completed thoroughly and that all answers are accurate to the best of my knowledge.

Signature: Date:

□ **Data Governance / Ethics Lead:** I have reviewed this checklist and confirm the project meets the ethical standards of the agency.

Signature: Date: