DATA GOVERNANCE MANUAL

Harris, Ben
The City of Tulsa 175 East 2nd Street, Suite 1405 Tulsa, OK 74103

TABLE OF CONTENTS

Data Analytics Manager Ethical Guidelines

The Data Analytics Manager has access to all the data within the City of Tulsa. This level of access should be held to the highest standard. The Data Analytics manager shall guard the privacy and security of the data with utmost diligence. This role will protect how decisions are made using the data. Facts instead of exaggeration will be presented to media and key decision makers. Failure to follow the guidelines below will result in their immediate removal from the position.
The Data Manager

- will **not** intentionally share data - with a public or private party - that has been classified as protected or sensitive.
- will **not** intentionally share sensitive data - that hasn't completed the data governance process -without an NDA agreement or the explicit approval of the Major.
- will **not** intentionally exaggerate outcomes of analytical projects.
- will always provide documentation to explain their analytical process.

In the case of data being stolen from The City of Tulsa, there will be no repercussions for the Data Analytics Manager.

Artificial Intelligence Policy

.1 The Artificial Intelligence (AI) Policy defines how advanced automation techniques are used within the City of Tulsa. It ensures AI is not used nefariously.

.1 Artificial intelligence is a computer science model or algorithm that fits within the following fields of study.

.1 Machine Learning
.2 Natural Language Processing
.3 Computer Vision
.4 Speech
.5 Planning
.6 Robotics

.2 The goal of AI is to automate processes or repetitive tasks, ultimately leading the organization to become more efficient.

.3 AI’s purpose is not to replace jobs, but to use human capital most effectively. For example, if an employee retires, their position may be reimagined.

.4 Each AI model developed should strive to remove bias. No model will go into production until equality and equity outcomes have been considered.

.5 A production AI project must be approved by the Mayor, Deputy Mayor, Chief of Staff or Deputy Chief of Staff.

Data Analytics Self-Service Policy

.1 The Data Analytics Self-Service Policy defines the steps necessary to guarantee analytics can be trusted. It covers the review of analytical methods, data sources, and the publishing and organization of dashboards. The Data Steward will work with his/her department to facilitate data use.

- .1 Data must be sourced from the Central Data Repository.
- .2 Traditional testing methods
  - .1 When building an automated report, it is essential to compare results with a known good report.
  - .2 If a data issue is found, it will be submitted to the Data Governance Committee.
- .3 Dashboards
  - .1 Dashboards will be built using data from the Central Data Repository.
  - .2 Once a dashboard has been built and tested, an extract should be created and added to the Tableau server. The extract should be scheduled for refresh.
  - .3 The dashboard's data source should be updated to the Tableau servers extract. This will allow the dashboard to reflect changes in the data over time.
  - .4 The dashboard will then be published to Tableau server in the creator's department site.
- .4 Advanced analytical testing methods
  - .1 For regression analysis, R2, MAE, and RMSE are valid measures of success.
    They are often used together. One is typically not sufficient by itself.
  - .2 For classification analysis, Precision, Recall, F1-score, Support, and Accuracy are valid measures of success. They are often used together. One is typically not sufficient by itself.
  - .3 Train, Test and Validation model testing ensure the results are not biased and extrapolate outside the dataset.
- .1 Often data is split into 70% train, 20% test, 10% validation data sets.
  - .4 An alternative to Train, Test and Validation sets is cross-validation.
    - .1 Instead of splitting the data, a k-folds method is used to train and test the data. It runs the training k times over k different data subsets. The advantage is the ability to use the whole dataset. The disadvantage is training the model k times.
  - .5 Testing methods must be documented and then reviewed by another person.

Data Classification Policy

.1 The purpose of this policy is to establish a framework for classifying City of Tulsa ("City") Data based on its Disclosure Risk and Impact Risk. Data Classification facilitates the disclosure of City Data internally and to the public when the risk of disclosure is outweighed by the City's commitment to transparency. This policy is intended to provide guidance to City personnel when deciding how to generate, collect, process, disseminate, or destroy City Data.

.1 1 The Data Classification Policy applies to all City of Tulsa Data or Data Sets as defined in this policy. The Data Classification Policy governs all permanent and temporary City of Tulsa employees, contractors, subcontractors, consultants, and vendors who are permitted to use or access City Data for any reason.

.12 Data Stewardship is the careful and responsible management of City Data belonging to the City as a whole, regardless of the entity or source that may have originated, created, or compiled the Data. Data Stewards provide maximum access to City Data internally and to the public, balanced by the obligation to protect the information in accordance with the policies established by the City of Tulsa and any other law or regulation. Any Data generated, collected, processed, disseminated, or disposed of by the City of Tulsa is an asset of the City, not of the particular department or subordinate organization which acts on the City's behalf. Departments developing policies, procedures, practices, and training should avoid the mindset of Data ownership and implement the practice of Data Stewardship.

.2 Definitions

.21 Availability: the characteristic of Data that enables users' access to that Data in a useable format without interference or obstruction.

.22 Classification: the act or process by which Data is determined to be of a described disclosure or Impact Risk, or criticality and value.

.23 Classifying Authority: The right or ability to establish a Disclosure Risk and Impact Risk for any Data generated within the scope of that Authority
as defined in this policy.

.24 Confidentiality: the characteristic of Data whereby only those with sufficient privileges and a demonstrated need may access that Data.

.25 Correlation Risk: the disclosure or Impact Risk of Data inferred from aggregated individual Data or Data Sets which would not reasonably exist in the individual Data or Data Sets when viewed separately.

.26 Correlative Classification: the classification of aggregated individual Data or Data Sets which reflects their Correlation Risk.

.27 Data: final versions of statistical or factual, quantitative, or qualitative information that: (1) is in alphanumeric form reflected in a list, table, graph, chart, image, or other non-narrative form, that can be digitally transmitted or processed; (2) is regularly received, created or maintained by or for a City department, office, administrative unit, commission, board, advisory committee or other subdivision of City government; (3) records a measurement, transaction or determination related to the business of the City and mission of such City subdivision; and (4) is inclusive of software source code developed or maintained by or on behalf of the City of Tulsa.

Data shall not include information provided by other governmental entities or image files, such as designs, drawings, photos or scanned copies of original documents; provided, however, that Data does include statistical or factual information about image files and geographic information system (GIS) Data.

.28 Data Set: any Data which is created in an iterative manner which retains nearly identical characteristics to the previously generated Data, relates to a particular subject or function, results from the same activity, or has some other relationship arising out of their creation, receipt, or use such that it could reasonably be considered the same Data.

.29 Derivative Classification: the classification that results from incorporating, paraphrasing, restating, or generating in new fom Data that is already classified, consistent with the classification markings that apply to the Source Information. Derivative Classification is intended to ensure field level Data Classifications are maintained through Data aggregation, disaggregation, and Data Set creation including the specified Data field.

.30 Disclosure Risk: the magnitude of harm that can be expected to result from the consequences of unauthorized disclosure of Data. Disclosure Risk dictates the methods of disclosure and the requirements for authorization to view Data or Data Sets and the degree of safeguarding therefore required.

.31 Impact Risk: the magnitude of hatm that can be expected to result from the consequences of unauthorized use, modification, or loss of Data. Impact Risk dictates the priority of the Data or Data Set to City operations and the degree of safeguarding therefore required.

.32 Implementation Guidance: any policy, procedure, practice, instruction, or training that prescribes how any portion of the policy is executed.

.33 Integrity: the quality or state of being whole, complete, and uncorrupted; exists when Data is unchanged from its source and has not been accidentally or intentionally modified, altered, or destroyed without authorization.

.34 Media: the physical method by which digital Data is stored or transmitted, regardless of particular form.

.35 Originating Authority: any individual authorized in writing to classify Data in the first instance by virtue of it having been created within their Classifying Authority.

.36 Record: all documents, including, but not limited to, any book, paper, photograph, microfilm, Data files created by or used with computer software, computer tape, disk, record, sound recording, film recording, video record or other material regardless of physical form or characteristic, created by, received by, under the authority of, or coming into the custody, control or possession of public officials, public bodies, or their representatives in connection with the transaction of public business, the expenditure of public funds or the administering of public property.

.37 Source Information: any existing Data from which Data or Data Sets may originate.

.3 Classification Standards

.31 New data collection programs. At the beginning of a new Data collection effort, Data shall be originally classified under this policy if any of the following conditions are met:

.311 A Record is maintained containing Data,

.312 A Record may be created from Data, or

.313 An individual with the authority to classify Data determines that unauthorized access, disclosure, or alteration of Data could reasonably result in damage to the City and the individual is able to identify or describe the damage.

.32 Existing data collection programs. For Data which the City has collected prior to implementation of this Policy, Data may be classified under this policy if any of the following conditions are met:

.321 A request for disclosure of Data is received

.322 An individual with the authority to classify Data determines that unauthorized access, disclosure, or alteration of Data could reasonably result in damage to the City and the individual is able to identify or describe the damage

.33 Implementation guidance for classification standards. Individuals with Originating Authority to classify Data may create Implementation Guidance defining Data within their Classification Authority which is presumed to satisfy any of the requirements of Section .31.

.4 Classification Categories

.41 For the purposes of this policy, all Data shall be classified according to: a) Disclosure Risk and b) Impact Risk.

.42 Data may be classified at one of the following categories:

DISCLOSURE CLASSIFICATIONS

PUBLIC	DATA WHICH DOES NOT CONTAIN PROTECTED INFORMATION OR SENSITIVE INFORMATION.
SENSITIVE	DATA, WHICH, IF MADE PUBLIC, COULD RAISE PRIVACY, CONFIDENTIALITY OR SECURITY CONCERNS OR HAVE THE POTENTIAL TO JEOPARDIZE PUBLIC HEALTH, SAFETY OR WELFARE TO AN EXTENT THAT IS GREATER THAN THE POTENTIAL PUBLIC BENEFIT OF MAKING THE INFORMATION PUBLIC.
PROTECTED	DATA WHICH IS SPECIFICALLY REQUIRED BY LAW TO BE KEPT CONFIDENTIAL, INCLUDING, BUT NOT LIMITED TO, DATA PROTECTED BY A STATE EVIDENTIARY PRIVILEGE, SOCIAL SECURITY NUMBERS, AND PERSONAL FINANCIAL INFORMATION.

IMPACT CLASSIFICATIONS

LOW	OPERATIONAL DEGRADATION TO SECONDARY FUNCTIONS OF THE DEPARTMENT, NO IMPACT TO PRIMARY FUNCTIONS NO DAMAGE TO ORGANIZATIONAL ASSETS NO FINANCIAL LOSS NO HARM TO INDIVIDUALS
(2)	OPERATIONAL DEGRADATION TO THE PRIMARY FUNCTIONS OF THE DEPARTMENT WITHIN ACCEPTABLE PARAMETERS
	MINOR DAMAGE TO DEPARTMENTAL ASSETS MINOR FINANCIAL LOSS (EXPENDITURES WITHOUT NEED FOR ADDITIONAL AUTHORIZATION) HARM TO INDIVIDUALS WHICH WOULD NOT RESULT IN FINANCIAL LOSS (PRIVACY, EMBARASSMENT)

MODERATE (3)	SIGNIFICANT OPERATIONAL DEGRADATION TO PRIMARY FUNCTIONS OF THE DEPARTMENT BEYOND ACCEPTABLE PARAMETERS BUT WITHIN INTERNAL CAPACITY TO REMEDY MODERATE DAMAGE TO DEPARTMENTAL ASSETS OR MINOR DAMAGE TO ORGANIZATIONAL ASSETS FINANCIAL LOSSES WITHIN DEPARTMENTAL AUTHORIZATION HARM TO AN INDIVIDUAL WHICH WOULD REASONABLY RESULT IN FINANCIAL LOSS TO THE ORGANIZATION
(4)	OPERATIONAL DEGRADATION TO PRIMARY FUNCTIONS OF THE DEPARTMENT BEYOND INTERNAL ABILITY TO REMEDY SIGNIFICANT DAMAGE TO DEPARTMENTAL ASSETS OR MODERATE DAMAGE TO ORGANIZATIONAL ASSETS FINANCIAL LOSSES WITHIN
	ORGANIZATIONAL AUTHORIZATION EXPOSURE COULD REASONABLY RESULT IN IDENTITY THEFT, FRAUD, OR OTHER FINANCIAL LOSS, OR ANY PHYSICAL HARM TO AN INDIVIDUAL

HIGH (5)	CATASTROPHIC OPERATIONAL DEGRADATION TO PRIMARY FUNCTIONS OF THE ORGANIZATION BEYOND INTERNAL CAPACITY TO REMEDY SIGNIFICANT DAMAGE TO ORGANIZATIONAL ASSETS CATESTROPHIC FINANCIAL LOSSES BEYOND ORGANIZATIONAL AUTHORIZATION SEVERE OR CATASTROPHIC FINANCIAL HARM TO INDIVIDUALS OR PHYSICAL HARM INVOLVING SERIOUS INJURY OR DEATH

.42 Implementation Guidance for classification levels. Individuals with Originating Authority to Data may create Implementation Guidance defining City Data or Data Sets within their classification authority which is presumed to fall within any of the categories of Section .31.

.43 Prohibitions and Limitations. In no case shall Data be classified in order to:

.431 Conceal violations of law, inefficiency, or administrative error

.432 Prevent embarrassment to a person, organization, or Department beyond the purpose of this policy

.433 Prevent or delay the release of Data beyond the purpose of this policy

.44 Identification and markings. The Information Technology Governance Board Security Sub-Committee, as provided for by the Information Technology Governance Charter, shall issue Implementation Guidance establishing required Classification markings of Media.

.5 Classifying Authority

.51 The authority to classify Data may be exercised only by:

.511 The Mayor, by executive authority, or an employee to whom the Mayor has delegated authority

.512 Directors of Departments, concerning all Data generated within their department, or an employee to whom the Director has delegated authority

.513 Individuals with authority as delegated pursuant to Section .52

.52 Delegated Authority

.521 Delegations of classification authority shall be limited to the minimum required to administer this policy. No delegated authority shall exist without training in classification as required by Implementation Guidance.

.522 Delegation of the executive authority. The Mayor may delegate his or her Classifying Authority to a subordinate individual within the Office of the Mayor, to act as the Mayor's representative, if that individual meets the requirements of Section .521 of this policy.

.523 Delegation of total departmental level authority to subordinate officials within the department. Directors of departments may delegate their Classifying Authority to a subordinate individual within their department, to act as their representative, if that individual meets the requirements of Section .521 of this policy.

.524 Delegation of total departmental level authority to individuals outside the department. If the individual with Classifying Authority per Section .51 of this policy seeks to delegate that authority to an individual outside their respective department, that individual shall be required to show cause before the Information Technology Govemance Board Security Sub-Committee, which may authorize the delegation by an affirmative vote by a majority of the voting members of the Board. The authority to delegate shall remain with the individual with Classifying Authority per Section .51 of this policy.

.525 Record of delegation. Each delegation of Classifying Authority shall be in writing.

.526 Exceptions. Reserved.

.6 Derivative Classification, Correlative Classification, and Redaction

.61 Derivative Classification. Persons who only reproduce, extract, or summarize Data which has been classified do not need to possess original classification authority to apply a Derivative Classification in the performance of such actions.

.62 Persons who apply Derivative Classifications shall:

.621 Observe and respect original classification decisions, and

.622 Carry forward to any newly created Data, Data Set, Record, or Media the pertinent classification and markings

.63 Correlative Classification

.631 Where Data or a Data Set is generated from multiple sources, or, is presented or produced with other Data or Data Sets in a manner which could reasonably allow for the individual presentations or productions to be aggregated into a single form which could reasonably allow the inference of Data not intended to be classified at the level of the individual Data or Data Sets or disclosed in the manner that the classification, presentation, or production of the individual or would allow, the classification and presentation or production of the individual Data or Data Set shall be reviewed as part of the aggregated Data or Data Set and shall be given the classification and disclosure treatment required for the whole.

.632 Individuals with authority to classify Data may create Implementation Guidance defining Correlative Classifications for Data or Data Sets within their classification authority.

.64 Redaction

.641 Authority to determine which portions of Data or a Data Set, as presented, establish that Data or Data Set's classification shall remain with the Classifying Authority. The Classifying Authority shall review redacted Data or Data Sets for residual disclosure and Impact Risk, including derivative or correlative risk, upon request by individuals authorized to make such redactions.

.642 Individuals with authority to classify Data may create Implementation Guidance for the Redaction of Data or Data Sets within their classification authority.

.7 Hierarchy of Classification Authority

.71 Derivative and Correlative Classifications. For Data or Data Sets generated from Derivative Sources or Data which is considered for Correlative risk, the disclosure classification and impact classification of any part bearing the worst-case impact shall establish the lowest classification of the Data or Data Sets as a whole regardless of whether derivative or correlated Data or Data Sets are generated by different Classifying Authorities.

.72 Transfer of Data or Data Sets

.721 Transfer for Storage. In the case of Data or Data Sets transferred from one Classifying Authority to another Classifying Authority for the purposes of storage only, the Classifying Authority in which the Data or Data Set originated shall remain the Originating Authority to classify the and and shall retain responsibility to review and maintain that classification.

.722 Transfer of function. In the case of Data or Data Sets transferred to another Classifying Authority in conjunction with the transfer of the functions for which or in which the Data or Data Set originated, the receiving Classifying Authority shall be deemed to be the Originating Authority for the purposes of this policy.

.723 Transfer in cases of abolishment or cessation. In the case of Data or Data Sets that are not officially transferred by methods described in Section .721 or .722, but that originated in a Classifying Authority that has ceased to exist and for which there is no successor Classifying Authority, each Classifying Authority in possession of that Data or Data Sets shall then be deemed the Originating Authority for the purposes of this policy. Review of such Data or Data Sets shall be done in consultation with any other Classifying Authority with an interest in the subject matter contained or expressed in the Data or Data Sets.

.8 Classification Review, Oversight, and Dispute Resolution

.81 Reclassification review. Classifying Authorities shall review all Data Classifications made within their authority from time to time.

.82 Uniform guidance. The Information Technology Security Sub-Committee shall from time to time as established in Policy review all Implementing Guidance, policies, procedures, practices, instructions, or training created by any Classifying Authority under the provisions of this policy for the purpose of recommending Uniform Guidance for all Classifying Authorities to the Mayor, for the promulgation of standardized policy, procedure, practice, instruction, or training across the organization with regard to Data Classification.

.83 Oversight. The Information Technology Security Sub-Committee shall have oversight responsibility for the implementation of this Policy.

.84 Challenges and Dispute Resolution

.841 Classifying Authorities who, in good faith, believe that the classification of or whether original, derivative, or correlative, or Originating Authority to make such Classification, is improper, may challenge the Classification to the Information Technology Security Sub-Committee.

.842 Classifying Authorities shall establish Implementation Guidance under which authorized holders or users of Data may challenge the classification of Data they believe is improperly classified before the Information Technology Security Sub-Committee.

.843 The Information Technology Security Sub-Committee shall ensure by policy and procedure that challengers or disputants are not subject to retribution for bringing such actions and challenges and disputes are given an opportunity for impartial review.

.9 Safeguarding Data

.91 Reserved.

Data Integration Policy

.1 The purpose of this policy is to establish a framework for how the City of Tulsa - with the assistance of IT - plans to integrate data from new software applications. Its purpose is not to control the purchase of new software, but to review a series of questions that will help plan and mitigate cost of integration.

.2 Questions to be discussed

- .1 Will the data created by the software be needed for analytics or dashboarding now or in the future?
  - .1 If so, will the database technology integrate with our system?
  - .2 Will the City of Tulsa own the data?
  - .3 Will the City of Tulsa be able to download all the raw data?
    - .1 How frequently?

Data Provenance Policy

.1 The purpose of the Data Provenance Policy is to set the framework for how changes in data are tracked. If the City of Tulsa knows how data is moved and modified, then it will have a higher degree of confidence in the use of the data.

  .1 Data used to make decisions will be accessed through a Central Data Repository.

  .2 Data existing in the Central Data Repository will have been submitted to the Data Governance Committee for review.

  .3 Per the Data Governance Review Process, it will have been classified, organized, and the metadata documented before being added to the Central Data Repository.

  .4 The automated ETL process –created by IT– must save changes made between the source data and the Central Data Repository.

<u>Data Quality Policy</u>

.1 The Data Quality Policy defines how data inconsistencies are handled. It ensures data quality will continuously improve the categories below.

.1 Through the Data Governance process the following sub categories of Data Quality are considered.

.1 Timeliness – The data is updated enough for the required uses.

.2 Completeness – Percentage of needed data available.

.3 Uniqueness – The data can be distinguished from other data sets.

.4 Consistency – The data is equal across different data sets.

.5 Validity – The data meets certain defined requirements (date formats, type and range)

.6 Accuracy – The data accurately reflects reality.

.2 If a data quality issue is discovered, then a form is completed and submitted to the Data Governance Committee.

.3 The issue will be forwarded to the department’s Data Steward for review.

.4 If a problem is found, the Data Steward will work with a data developer to fix the issue.

Data Security Policy

.1 The Data Security Policy governs the layers of protection applied to the storage and retrieval of data. It also covers security auditing.

- .1 Data security will be designed to mimic the data classification folders structure. It will use Active Directory (AD) groups to secure data.
- .2 The first level of the data permission hierarchy is the Data Classification Folder. It exists of three top level folders: Public, Sensitive, and Protected. Each Folder will have an independent AD group.
- .3 The second level of the data permission hierarchy is the Department Folder. It breaks the datasets by which department owns the data.
- .4 The third level of the data permission hierarchy is the Application Folder. Each application will have a folder and a security group. Each dataset owned by an application will be labeled by its identifier and stored in the Application Folder.
- .5 Each Data Security Request will need to be submitted to the user's supervisor and department's Data Steward for review and approval.
- .6 Once the security request has been approved, IT will add the user to the correct AD group.
- .7 The requests will be periodically reviewed and audited by IT security.