datafirst

The City’s Data Challenge
DATA FIRST CHALLENGE
CRITERIA & APPLICATION PROCESS
DATA FIRST

MAIN PURPOSE OF THE CHALLENGE

Increase city data availability in the city

Enhance data governance & compliance

Nurture a strong data community
DATA FIRST

CRITERIA MEASUREMENT

1. Comply with Dubai Data Law - 30%
2. Participate in Smart Dubai Data Events - 15%
3. Collaborate on data requests - 25%
4. Participate in each data ingestion cycle - 25%
5. Support a data driven culture - 5%
WHY

OBJECTIVE

- To improve data compliance with Dubai Data Law
- To open motivational channel between entities, create the spirit of health competition

WHAT

“SUCCESS MEASUREMENT AGAINST CRITERIA”

- Entities who are compliant by 70% and above
- 50% improvements in overall compliance

HOW

“MEASUREMENT METHOD”

- Measure the average of 2019 data compliance reports
15% PARTICIPATE IN SMART DUBAI DATA EVENTS

WHY
OBJECTIVE

▪ To enhance smart data ecosystem engagement
▪ To create a channel to showcase best practices

WHAT
“SUCCESS MEASUREMENT AGAINST CRITERIA”

▪ Entities who participated in more than 50% of Smart Dubai Data events in 2019

HOW
“MEASUREMENT METHOD”

▪ Track attendance matrix of 2019 events managed by Smart Dubai Data
25% COLLABORATE ON DATA REQUESTS

WHY
OBJECTIVE

- To improve the collaboration between the government entities to meet all data requests
- To encourage cross-functional projects between government entities

WHAT
“SUCCESS MEASUREMENT AGAINST CRITERIA”

- Entities with 100% of closed and completed data requests

HOW
“MEASUREMENT METHOD”

- Track all data requests status in 2019 for each entity
25% PARTICIPATE IN EACH DATA INGESTION CYCLES

WHY OBJECTIVE

- To increase city “open data” availability for the public on DubaiPulse
- To increase city “shared data” accessibility for the government entities on DubaiPulse

WHAT “SUCCESS MEASUREMENT AGAINST CRITERIA”

- Entities who participated in minimum 5 ingestion cycles with at least 5 datasets each cycle in 2019

HOW “MEASUREMENT METHOD”

- Track all ingestion cycles report to verify number of datasets per entity
5% SUPPORT A DATA DRIVEN CULTURE

WHY
OBJECTIVE

- To act as a change agent in raising the awareness of the importance of exchange open / shared data
- To drive own entity toward becoming a data driven organization

“SUCCESS MEASUREMENT AGAINST CRITERIA”

Entities who support Dubai Data initiative by:

- Organizing internal workshops
- Encouraging / Sponsoring employees to attend RIT Data Analytics workshops and Master of Data Analytics
- Organize awareness programs within own entity

“MEASUREMENT METHOD”

- Submit evidence of conducting data awareness sessions and workshops
DATA FIRST

SELECTION PROCESS

THE ANNOUNCEMENT

July 14  Announce “Data First” the city data challenge and apply for the competition

KEEPING TRACK

July - December  Smart Dubai will monitor all entities progress

PANEL REVIEW

December  Smart Dubai will conduct a panel review to select winners based on scores against criteria

CELEBRATION

January  Announce the winners
To measure the percentage of compliance to Dubai Data Law, two factors will be measured for each element of compliance.

**Reporting Frequency**

**Factors Measured**

Each element of compliance will be scored from 0 to 2 on two factors:

1. **Timeliness**
   - Score: 0-2
2. **Completeness**
   - Score: 0-2
% OF COMPLIANCE TO DUBAI DATA LAW

To measure the percentage of compliance to Dubai Data Law, two factors will be measured for each element of compliance.

<table>
<thead>
<tr>
<th>ELEMENT OF COMPLIANCE</th>
<th>1 - Designate Team</th>
<th>2 - Determine Governance</th>
<th>3 - Inventory Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>3%</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELEMENT OF COMPLIANCE</th>
<th>4 - Prioritize Data</th>
<th>5 - Classify Data</th>
<th>6 - Prepare Data Release Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>4%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELEMENT OF COMPLIANCE</th>
<th>7 - Complete Data Cataloguing</th>
<th>8 - Ingest Data on Platform</th>
<th>9 - Maintain data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>25%</td>
<td>33%</td>
<td>12%</td>
</tr>
</tbody>
</table>
DATA FIRST
% OF COMPLIANCE TO DUBAI DATA LAW

1 DESIGNATE TEAMS

Effective/empowered Data teams formed through internal decree and with clearly communicated roles and responsibilities.
TEAM STRUCTURE

ORGANIZATION STRUCTURE:
4 ROLES

- **DUBAI DATA LEADER**
  - Executive sponsor
  - Single point of contact for entity’s Dubai Data effort
  - Decision making responsibility

- **DUBAI DATA ADMINISTRATOR**
  - Working level, day-to-day point of contact

**DATA STEWARD**
- SPECIFIC COMMITTEES
- BUSINESS UNITS

**DATA SPECIALIST**
- IT
- DBA

INVENTORY
CLASSIFICATION

ACCESS
INGESTION
DATA MANAGEMENT
DETERMINE GOVERNANCE
% OF COMPLIANCE TO DUBAI DATA LAW

2 DETERMINE GOVERNANCE

Formally documented, approved and implemented RACI model defining approval process
## TEAM STRUCTURE
### DUBAI DATA ESTABLISHMENT SUGGESTED RACI MATRIX

<table>
<thead>
<tr>
<th></th>
<th>DIRECTOR GENERAL</th>
<th>DUBAI DATA LEADER</th>
<th>DATA ADMINISTRATOR</th>
<th>DATA STEWARD/SPECIALIST</th>
<th>DUBAI DATA ESTABLISHMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>START DATA INVENTORY PROCESS</td>
<td>I</td>
<td>I</td>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>REQUEST ENTITY DATA INVENTORY</td>
<td>I</td>
<td>I</td>
<td>A</td>
<td>R</td>
<td>C</td>
</tr>
<tr>
<td>PREPARE DATA INVENTORY</td>
<td>I</td>
<td>I</td>
<td>A</td>
<td>R</td>
<td>I</td>
</tr>
<tr>
<td>REVIEW DATA INVENTORY</td>
<td>I</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>APPROVE DATA INVENTORY</td>
<td>R</td>
<td>A</td>
<td>C</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>FINALISE DATA INVENTORY</td>
<td>I</td>
<td>C</td>
<td>R</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>PREPARE DATA CLASSIFICATION</td>
<td>I</td>
<td>A</td>
<td>R</td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>REVIEW DATA CLASSIFICATION</td>
<td>I</td>
<td>C</td>
<td>R</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>APPROVE DATA CLASSIFICATION</td>
<td>R</td>
<td>A</td>
<td>C</td>
<td>I</td>
<td>R</td>
</tr>
<tr>
<td>DETERMINE DATA INGESTION METHOD</td>
<td>I</td>
<td>C</td>
<td>A</td>
<td>R</td>
<td>I</td>
</tr>
<tr>
<td>REVIEW DATA INGESTION METHOD</td>
<td>I</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>APPROVE DATA INGESTION METHOD</td>
<td>I</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>R</td>
</tr>
</tbody>
</table>
INVENTORY DATA
3 INVENTORY DATA

• Data inventory sheet is submitted
• Written statement is sent by Data Leader confirming that:

  • List has been reviewed/approved by Data Leader
  • The list includes all obvious and high value datasets, already published/shared data and all existing identified lists of this entity
  • The process has been followed and checked by Data Administrator
  • List of all departments involved in creating the draft lists and name/role of data lead responsible
Data Inventory is an inventory of the Dubai Data that a Dubai Government Entity has in its possession, including any Dubai Data which the Dubai Government Entity produces or controls.

<table>
<thead>
<tr>
<th>DATASET NAME</th>
<th>DATASET DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal name of the dataset that will be used to reference the dataset between SDO &amp; Government Entity.</td>
<td>Statement about the dataset describes the contents and the fields included</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATA ATTRIBUTES</th>
<th>ATTRIBUTE DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes are the columns or fields that make up the data set</td>
<td>Statement describing the contents and the nature of data in the attribute</td>
</tr>
</tbody>
</table>
PRIORITIZE AND CLASSIFY DATA
% OF COMPLIANCE TO DUBAI DATA LAW

4 PRIORITIES

ALL DATASETS:

• Have a prioritization score
• Overall prioritization is reviewed and signed off by Data Leader
• DDE comments are reflected in final prioritization (if any are received)
Prioritizing Data

Data Prioritization Criteria

There are two broad sets of criteria:

1. Benefit Criteria
   For evaluating the potential value of opening a particular dataset to citizens and companies, or sharing it with other Dubai Government Entities.

2. Readiness Criteria
   For evaluating the effort involved in getting the dataset ready for publication or sharing (through the Smart Dubai Platform).
### DATA PRIORITIZATION RESULTS (SAMPLE)

<table>
<thead>
<tr>
<th>DATASET NAME</th>
<th>DATASET DESCRIPTION</th>
<th>BENEFITS</th>
<th>QUALITY AND READINESS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>User demand</td>
<td>Economic Impact</td>
</tr>
<tr>
<td>Formal name of the dataset that will be used to reference the dataset between SDO &amp; Government Entity.</td>
<td>Statement about the dataset describes the contents and the fields included</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
% OF COMPLIANCE TO DUBAI DATA LAW

5 CLASSIFY

All prioritized datasets:
- Are classified, and
- Reasons for classification are documented
- DDE feedback on prioritization reflected
PRACTICAL CLASSIFICATION

CLASSIFY DATA SETS

DATASET IDENTIFICATION

• Identify priority datasets for classification

DATASET CLASSIFICATION

• Classify the attributes
• Split dataset based on attribute
• Identify attributes causing a secret, sensitive, or confidential scoring in order to arrive at open classification

REVIEW & APPROVAL

• Review classification by smart Dubai office
• Review and agree on final classification by smart Dubai and department
DATA SETS IDENTIFICATION
IDENTIFY PRIORITY DATASETS FOR CLASSIFICATION

DO ANY OF THE FOLLOWING APPLY TO DATASETS?

1. Part of Primary Registry
2. Strategic alignment with Smart Dubai priorities (in particular Smart Economy and Smart Living)
3. User demand for the data
4. Government efficiency benefits
5. Service improvement benefits
6. Economic growth benefits
7. Transparency/accountability benefits
The relationships between the different categories of Dubai Data are illustrated in this figure:

**DUBAI DATA**
Owned by entities is made available for sharing and re-use by other entities

**OPEN DATA**
Openly disclosed to everyone

**CONFIDENTIAL**
Shareable across entities according to professional responsibilities

**SENSITIVE**
Shareable within certain groups subject to strict controls

**SECRET**
Shareable in a limited way between certain individuals under strict controls
DATA FIRST

% OF COMPLIANCE TO DUBAI DATA LAW

6 DATA RELEASE PLAN

Data Release Plan completed as per the standard and approved by management of entity
PRACTICAL CLASSIFICATION

INITIATION AND DATA RELEASE PLAN

CREATE AN INITIAL ROADMAP FOR YOUR ENTITY’S DATA WORK THAT MEETS WITH DDE REQUIREMENTS

Q1 | Q2 | Q3 | Q4 | Q1 | Q2
---|----|----|----|----|----
|    |    |    |    |    |    
|    |    |    |    |    |    
|    |    |    |    |    |    
|    |    |    |    |    |    
|    |    |    |    |    |    

[Diagram showing Q1, Q2, Q3, and Q4]
7 CATALOGUING
(INCLUDES METADATA, QUALITY, FORMAT STANDARD REQUIREMENTS)

EACH DATASET:
• has been classified,
• has data quality assessment report,
• has sample dataset,
• has core/optional metadata provided,
• compliance statement has been signed off by Data Leader
# META DATA ITEMS

<table>
<thead>
<tr>
<th>BUSINESS</th>
<th>TECHNICAL</th>
<th>DATA QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dataset Name</td>
<td>1. Data Provenance</td>
<td>1. Data Quality – Completeness</td>
</tr>
<tr>
<td>2. Contributor/Custodian</td>
<td>2. Format (MIME)</td>
<td>2. Data Quality – Uniqueness</td>
</tr>
<tr>
<td>5. Coverage (Geographic area)</td>
<td>5. Attribute Delimiter</td>
<td>5. Data Quality – Consistency</td>
</tr>
<tr>
<td>8. Attribute Description</td>
<td>8. Attribute Join Rules to Related Datasets</td>
<td></td>
</tr>
<tr>
<td>10. Attribute Classification</td>
<td>10. Source Platform Type</td>
<td></td>
</tr>
<tr>
<td>11. Attribute Range of Values</td>
<td>11. Preferred Ingestion Method</td>
<td></td>
</tr>
<tr>
<td>12. Primary Identifier Attribute for the dataset</td>
<td>12. Description for Method</td>
<td></td>
</tr>
<tr>
<td>13. Dataset Supplementary Reference Data</td>
<td>13. Connectivity Option</td>
<td></td>
</tr>
<tr>
<td>14. Related datasets</td>
<td>14. Description for Connectivity</td>
<td></td>
</tr>
<tr>
<td>15. Related Identifier</td>
<td>15. Frequency of Update on Source</td>
<td></td>
</tr>
<tr>
<td>16. Dataset Point of Contact</td>
<td>16. Frequency of Update to SDP</td>
<td></td>
</tr>
<tr>
<td>17. Dataset POC Email</td>
<td>17. Scheduling of SDP Update</td>
<td></td>
</tr>
<tr>
<td>18. Dataset POC Phone</td>
<td>18. Expected Volume/Ingestion Cycle (Daily if Real-time)</td>
<td></td>
</tr>
<tr>
<td>19. Rights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Subject Level 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Subject Level 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Subject Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Tags / Keywords</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Already Published?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. If Published, where?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Already Shared?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. If Shared, with which Entities?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. Expected Data Set Utilization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Remarks/Special Business Rules</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8 INGEST DATA
Data ingested on the platform with no technical issues from entity as per agreed ingestion method

9 DATA MAINTENANCE
Data quality maintained as per agreed data contract (i.e. frequency of publication, attributes etc) for all datasets