DUBAI DATA COMPLIANCE FRAMEWORK AND KPIS
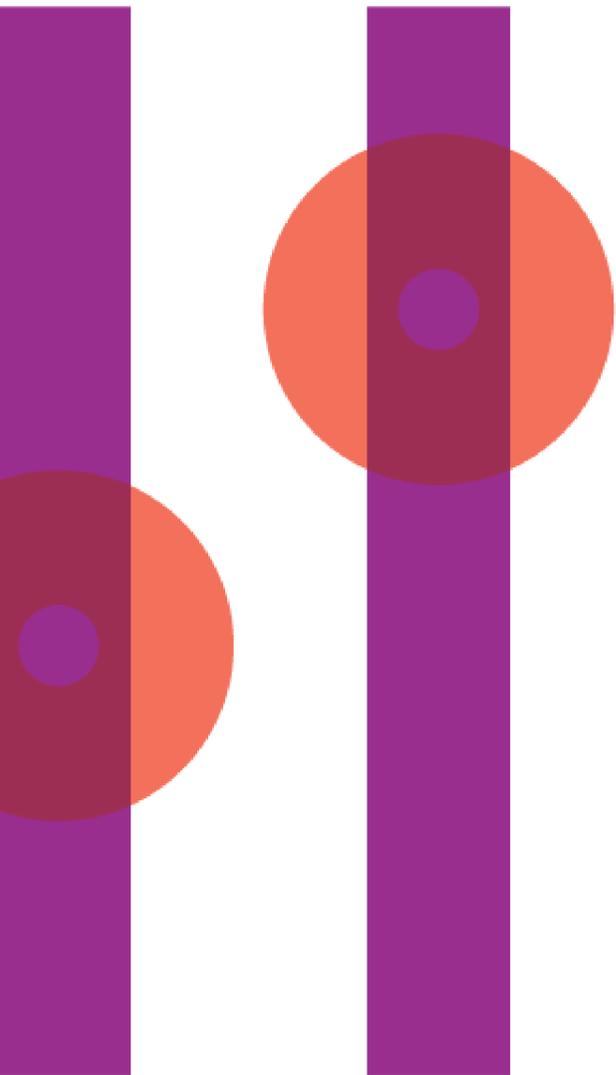
SMART DUBAI

دبـــي الذكية

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

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

Elements of Compliance
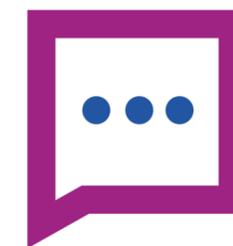
Each element of compliance is assigned a different weight depending on importance:

1 Designate Team	**2** Determine Governance	**3** Inventory Data
Weight: 3%	Weight: 3%	Weight: 10%
4 Prioritize Data	**5** Classify Data	**6** Prepare Data Release Plan
Weight:4%	Weight: 5%	Weight: 5%
7 Complete Data Cataloguing	**8** Ingest Data on Platform	**9** Maintain data
Weight:25%	Weight:33%	Weight:12%

SDF TIMELINESS WILL BE SCORED FROM 0 TO 2

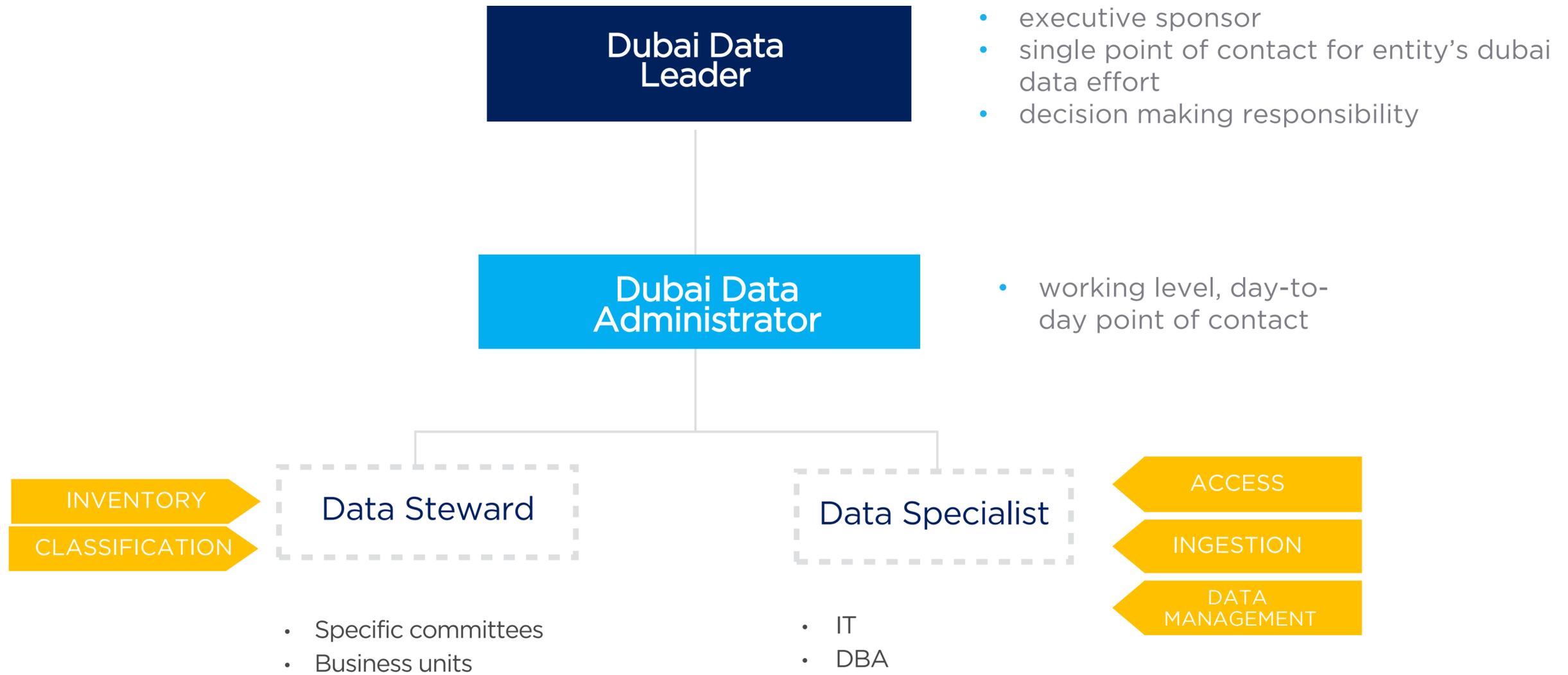| SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
TIMELINESS	Not submitted to Dubai Data Establishment	Submitted after set deadline	Submitted on time

DESIGNATE DUBAI DATA TEAM

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED ON REQUIREMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
1 DESIGNATE TEAMS	Not completed yet	Team members nominated but not formally formed within entities and/or team list is not updated	Effective/ empowered Data teams formed through internal decree and with clearly communicated roles and responsibilities

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

INVENTORY

CLASSIFICATION

Data Steward

Data Specialist

ACCESS

INGESTION

DATA MANAGEMENT

- Specific committees
- Business units

- IT
- DBA

COMPETENCY PROFILES: ENTITY-LEVEL ROLES

Competency Cluster	Competency	Dubai Data Leader	Dubai Data Advisor	Dubai Data Administrator	Data Steward	Data Specialist
1. Data Stewardship	1.1 Data leadership	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	1.2 Managing the business context for data	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	1.3 Managing data as a collective asset	3-Entity leader			2-Practitioner	
2. Applying Open Standards	2.1 Managing data inventories	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	2.2 Standards-based data preparation	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	2.3 Standards-based procurement	3-Entity leader		3-Entity leader	2-Practitioner	
3. Publishing Open Data	3.1 Data classification	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	3.2 Data publishing		3-Entity leader	3-Entity leader	2-Practioner	2-Practitioner
	3.3 Dataset compliance and quality assurance	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	
4. Protecting Privacy	4.1 Privacy by design	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
	4.2 Operationalising privacy	3-Entity leader	3-Entity leader	3-Entity leader	2-Practioner	2-Practitioner
	4.3 Managing privacy compliance	3-Entity leader	3-Entity leader	3-Entity leader		
5. Acting as a Smart User of Data	5.1 Data availability	3-Entity leader		2-Practitioner	2-Practitioner	1-Basic
	5.2 Data analytics	3-Entity leader		2-Practitioner	2-Practitioner	2-Practitioner
	5.3 Data insights	3-Entity leader		2-Practitioner	2-Practitioner	
6. Promoting Re-Use of our Data	6.1 User analysis	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner	1-Basic
	6.2 Market engagement	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner	
	6.3 User-centric development of data-driven product	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner	
7. Helping Deliver Collaborative Governance	7.1 Continuous improvement of the Dubai Data Manual	3-Entity leader		3-Entity leader	2-Practitioner	2-Practitioner
	7.2 Collaborative development of shared services	3-Entity leader		2-Practitioner	2-Practitioner	2-Practitioner
	7.3 Leadership of the Dubai Data Standards Advisory Board	3-Entity leader				
8. Managing Change	8.1 Roadmap development and implementation	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	1-Basic
	8.2 Managing skills for Dubai Data	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner	1-Basic
	8.3 Championing change and demonstrating impact	3-Entity leader	3-Entity leader	3-Entity leader	2-Practitioner	2-Practitioner
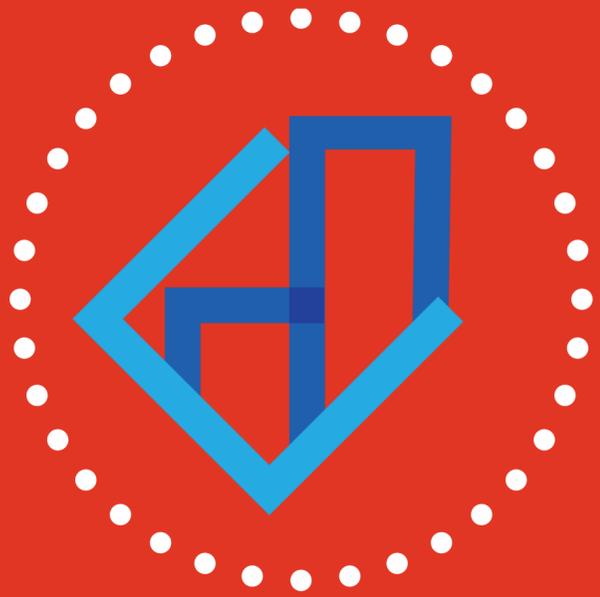
DETERMINE GOVERMANCE

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED ON REQUIREMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
2 DETERMINE GOVERNANCE	Not completed yet	Some practices exists but are not formally documented	Formally documented, approved and implemented RACI model defining approval process

RESPONSIBILITY ASSIGNMENT SYSTEM

The Dubai Data Law does not impose internal approval processes of each entity.

It was written to respect the existing process of each entity. A clear internal approval process will help build trust and confidence in the data published by the entity among data users, other data providers, and team members within the entity.

Our goal is to publish trustworthy and secure data that adds value for the city.

RACI at a glance

Responsible	Does the work to achieve the task
Approver	Ultimately answerable for the correct and thorough completion of deliverable; delegates the work to those responsible
Consulted	In the loop, typically subject matter experts
Informed	Kept up-to-date on progress

DUBAI DATA ESTABLISHMENT SUGGESTED RACI MATRIX

	Director General	Dubai Data Leader	Data Administrator	Data Steward/Specialist	Dubai Data Establishment
Start data inventory process	I	I	R	I	I
Request entity data inventory	I	A	R	C	I
Prepare data inventory	I	I	A	R	I
Review data inventory	I	R	R	A	C
Approve data inventory	R	A	C	I	I
Finalise data inventory	I	C	R	A	R
Prepare data classification	I	A	R	C	I
Review data classification	I	R	A	I	C
Approve data classification	R	A	C	I	R
Determine data ingestion method	I	C	A	R	I
Review data ingestion method	I	C	R	C	I
Approve data ingestion method	I	C	A	C	R

INVENTORY DATA

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED IN REQUIRMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2

3 INVENTORY DATA	Not completed yet	Inventory prepared but not as per Data Inventories Module and/or or does not include all obvious high value datasets	• 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

DUBAI DATA INVENTORY PROCESS

1. Identity a data representative per department within the Entity

2. Ensure each department produces a draft list of datasets

   - Look at existing data lists
   - Brainstorm other datasets
   - Create a table of datasets

3. Integrate and validate datasets from departments within the entity

4. Prioritize integrated Data Inventory based on Primary Registry and Dubai Dashboard Priorities

5. Assessment and approval

6. Dubai Data Establishment sign off

Dataset Name	Contributor/Custodian	Creator Business Unit within Entity	Dataset Description	Attribute	Attribute Description	Attribute Classification
Traffic_Accidents	Dubai Police	Traffic Dept.	the accual accidents happen between viechle, after investigation	ID	System Generated ID for the Accident	Open
				ACD_DATE	Accident Date	Open
				ACD_TIME	Accident time	Open
				Acc Location	statement describes the location of the accident	Open
				ACC Type	Accident category (i.e. minor, major, etc.)	Open
				ACC Cause	The cause of the accident	Open
				weather	Description of the weather	Open
				road status	Description of the road status	Open
				AGE	Main actor age	Open
				GENDER	Main actor gender	Open
				INJURY_SEVERITIES	Driver injury level	Open
				DRIVING_LISCENCE_ISSUE_DATE	DRIVING_LISCENCE_ISSUE_	Open
				OCCUPATION	Driver occupation	Open
				INTOXICATION	INTOXICATION	Open
				SEAT_BELT_STATUS	Flag	Open
				YEAR_MANUFACTURED	YEAR_MANUFACTURED	Open
				INSURANCE_COMPANY_NAME	INSURANCE_COMPANY_NAM	Open
Traffic_incidents	Dubai Police	Operation Dept.	collected from operation contact center this is captured in real-time with coordinates.	ACCI_ID	GENERATED BY OPERATION SYSTEM	Open
				ACCI_TIME	Accident time	Open
				ACCI_NAME	categorization of the accident	Open
				ACCI_X	Latitude	Open
				ACCI_Y	Longitude	Open
				ACCI_NUM	SEQUENCE NUMBER	Open

IDENTIFYING DATA SETS

| Type of Dataset	Definition	Example	Metro Example
Real Time Data	Constantly updating data — often high volume and high velocity	• Weather • Airport footfall • Toll volume • Pollution • Location-based data • Electricity usage	• Current location of metros and trams • current passenger numbers
Operational Data	These are the records that are made as part of an organisation carrying out its day-to-day business. Often high volume and high velocity data	• Entity organization chart • Forecast or modelling data • Buildings owned/maintained • Budget • Spending • Staff levels • Performance against metrics	• Staff numbers • Duty rota for staff at metro stations
Reference Data	Authoritative or definitive data that rarely changes about things, and that is necessary to help understand other data – often because it includes identifiers for those other things. Often produced by the public sector as a service in itself due to its high importance and value	• Timetables • Names and locations of schools, hospitals, bus stops, metro stations • Tax codes • Land holdings • Mapping data • Indicators • Address data • Citizen id	• Metro timetables • Capacity per metro line • Metro station locations
Aggregated Data	Analyzed and summarized data, which provides overview information of reference or administrative data	• Hospital operation success rates • School exam pass rates • Population statistics • Housing • Tourist numbers by month, year • Nationalities of visitors	• Metro usage per month • Metro trends over time

DATA INVENTORY EXAMPLE

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.

Dataset name	Dataset Description	Data Attributes	Attribute Description
Formal name of the dataset that will be used to reference the dataset between SDO & Government Entity.	Statement about the dataset describes the contents and the fields included	Attributes are the columns or fields that make up the data set	Statement describing the contents and the nature of data in the attribute
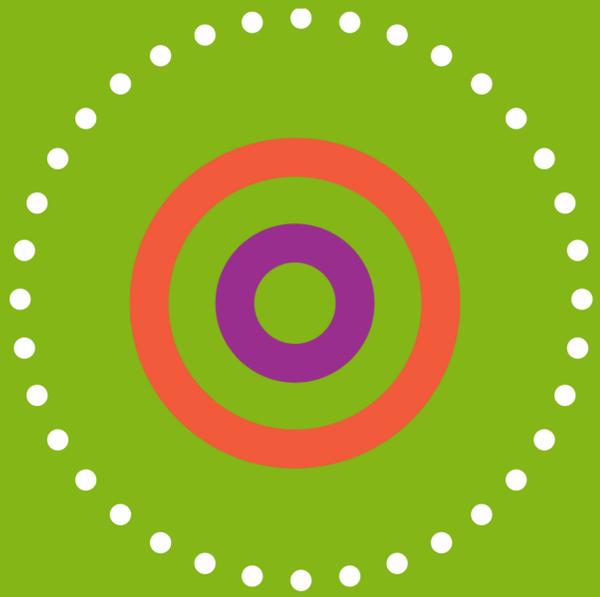
PRIORITIZE AND CLASSIFY DATA

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED IN REQUIRMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
4 Prioritize	Not completed yet	• Partial completion of prioritization of datasets; and/or • prioritization not signed off by Data Leader; and/or • DDE coordinator comments were not reflected in prioritization results	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)

DATA PRIORITISATION CRITERIA

There are two broad sets of criteria:

1

2

Benefit Criteria

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

Readiness Criteria

For evaluating the effort involved in getting the dataset ready for publication or sharing (through the Smart Dubai Platform).

DATA PRIORITISATION CRITERIA - BENEFIT CRITERIA

This provides a simple way to evaluate the comparative impact that publishing different data would have on the strategic goals of Smart Dubai. Each dataset should be given a score from 1 to 5 for each of the following four questions:

1. User demand for data: How likely is it that citizens, Government Entities, external companies or organisations would want to use or have access to this data?

2. Economic impact: If we open up this data, how likely is it Private-Sector companies could use it - perhaps "mashed up" with other data - to create commercially valuable products and services? (Smart Economy)

3. Better services: How likely is it that publishing this data will lead to innovations and services that improve the quality of life for people in Dubai? (Smart Living)

4. Better governance: How likely is it that will publishing this data will improve the efficiency, transparency and accountability of Government Entities? (Smart Governance)

Benefit Scoring Scale

5. Definite: We have clear evidence of efficiency; transparency or accountability gains that could be made.

4. Highly likely

3. Possible

2. Unlikely

1.Highly unlikely

Most benefit

Least benefit

DATA PRIORITISATION - READINESS CRITERIA

This is to establish how much work is needed to prepare it for publication (ensuring it meets minimal requirements). Data which is of high quality, already documented, up do data and with a clear owner can be easily published or shared. These criteria help to identify 'quick wins' for the Entity. Please assess each dataset for the following:

1. **Data quality: please choose <u>high,</u> <u>medium</u> or <u>low</u> for the following:**

   1. How accurate is the data? (high/medium/low)
   2. How complete is the data? (high/medium/low)
   3. How up to date is the data? (high/medium/low)
   4. Does the data use a schema or is standardized? (choose one)

      - **Yes, data is published with same headings / fields (schema) each time (high)**
      - **The data does not use a schema AND is not published regularly (i.e. it is one-off data) (medium)**
      - **Data is regularly updated, but does not use a set schema (low)**

Readiness Scoring Scale

1. **Data Quality - Data Accuracy**

   a. High Accuracy (we review and check accuracy) = 2 pts

   b. Medium Accuracy (not sure) = 1 pt

   c. Low Accuracy (there are known errors in the data) = 0 pts

Each (a) is awarded 2 points, (b)'s 1 point and (c)'s 0 points, giving a maximum total score out of 8 for data quality. (This will be automatically calculated by the tool.)

DATA PRIORITISATION - READINESS CRITERIA

This is to establish how much work is needed to prepare it for publication (ensuring it meets minimal requirements). Data which is of high quality, already documented, up do data and with a clear owner can be easily published or shared. These criteria help to identify 'quick wins' for the Entity. Please assess each dataset for the following:

1. **Ease of preparation: please answer <u>yes</u> or <u>no</u> to the following:**

   1. Is there a clear specific data owner? (yes/no)

   2. Does the data have existing metadata - that is, information on what the data is about, how it was generated etc.? (yes/no)

   3. Is the data already published somewhere or available on the web / through an API? (yes/no)

   4. Is the data in an open machine-readable format? (yes/no)

Readiness Scoring Scale

2. **Ease of Preparation**
   a. Yes = 1 pt
   b. No = 0 pts

Each 'Yes' is awarded (1 pt) and each 'No' is awarded (0 pts), giving a maximum total score of 4 for ease of preparation.

DATA PRIORITISATION – RESULTS (SAMPLE)

Dataset name	Dataset Description	Benefits				Quality and Readiness			
		User demand	Economic Impact	Better services	Better Governance	Defined Owner	Existing Metadata	Already Published	Open Format
Formal name of the dataset that will be used to reference the dataset between SDO & Government Entity.	Statement about the dataset describes the contents and the fields included	3	2	4	3	YES	YES	NO	YES

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED IN REQUIRMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
5 Classify	Not completed yet	• Partial classification of prioritized datasets; and/or • reasons for classification are not documented; and/or DDE feedback not reflected	All prioritized datasets: ✓ are classified, and ✓ reasons for classification are documented ✓ DDE feedback on prioritization reflected

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

IDENTIFIED DATASET

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

YES

PRIORITY DATASET

NO

REGULAR DATASET

DATA SETS CLASSIFICATION
TYPES OF CLASSIFICATION

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
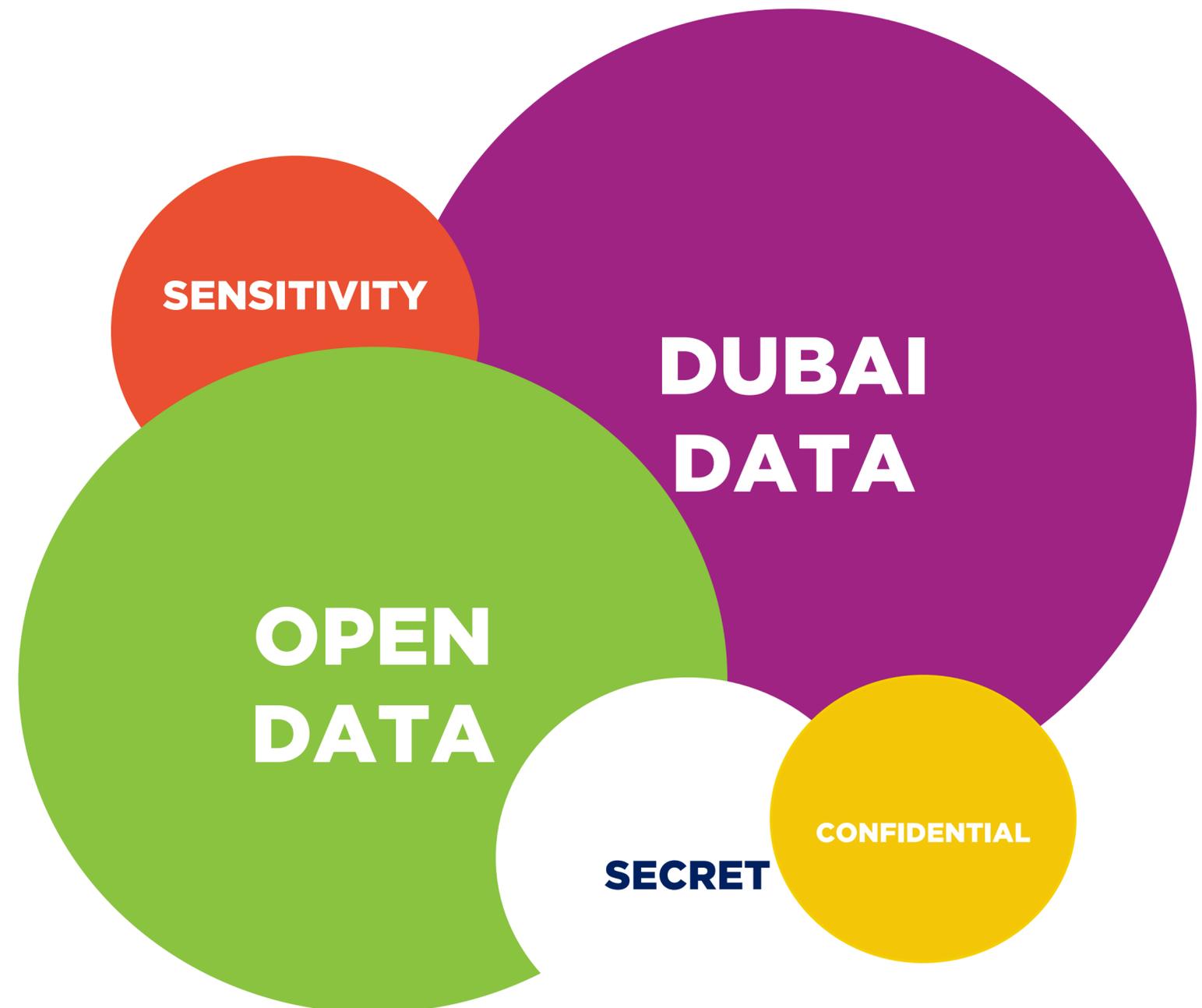Openly disclosed to everyone

Confidential
Shareable across entities according to professional responsibilities

Sensitive
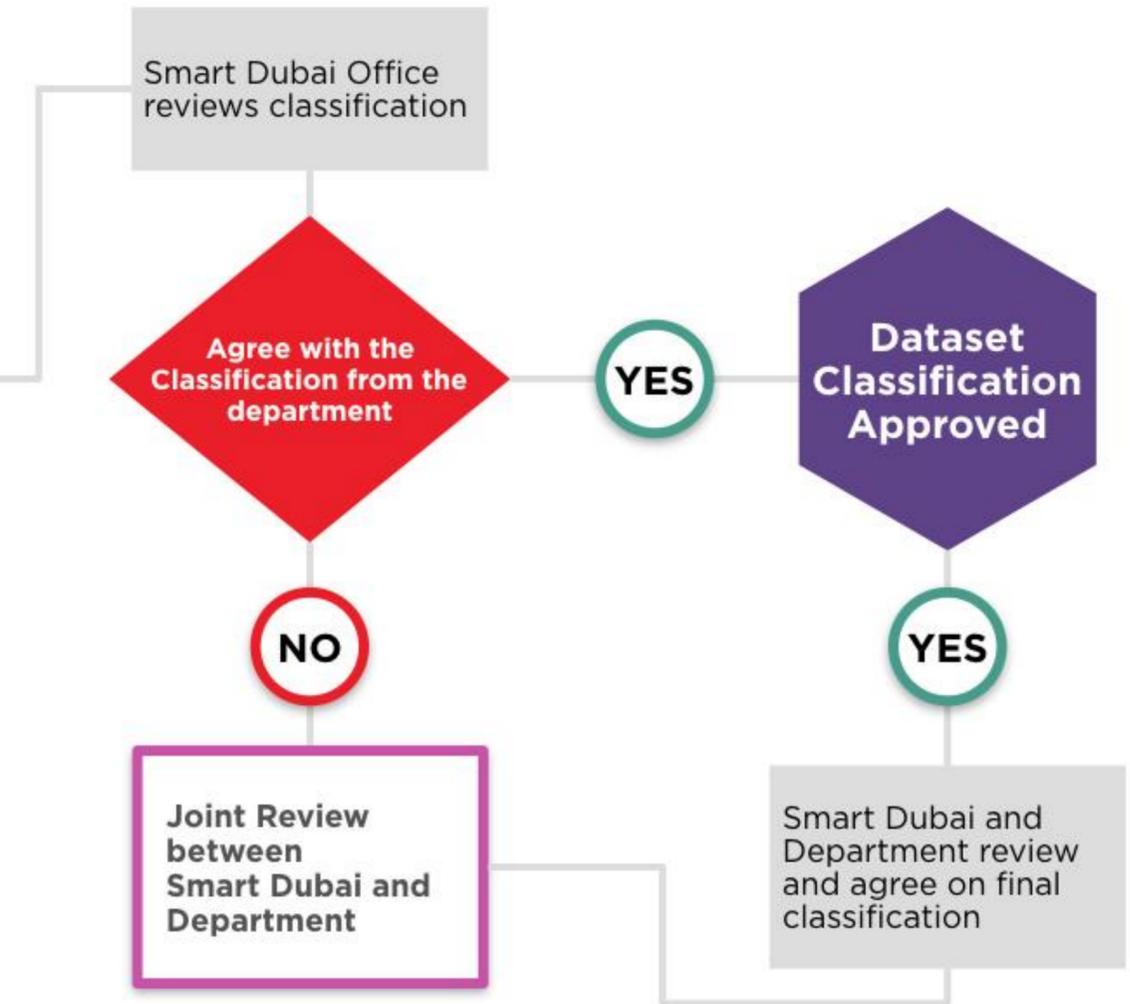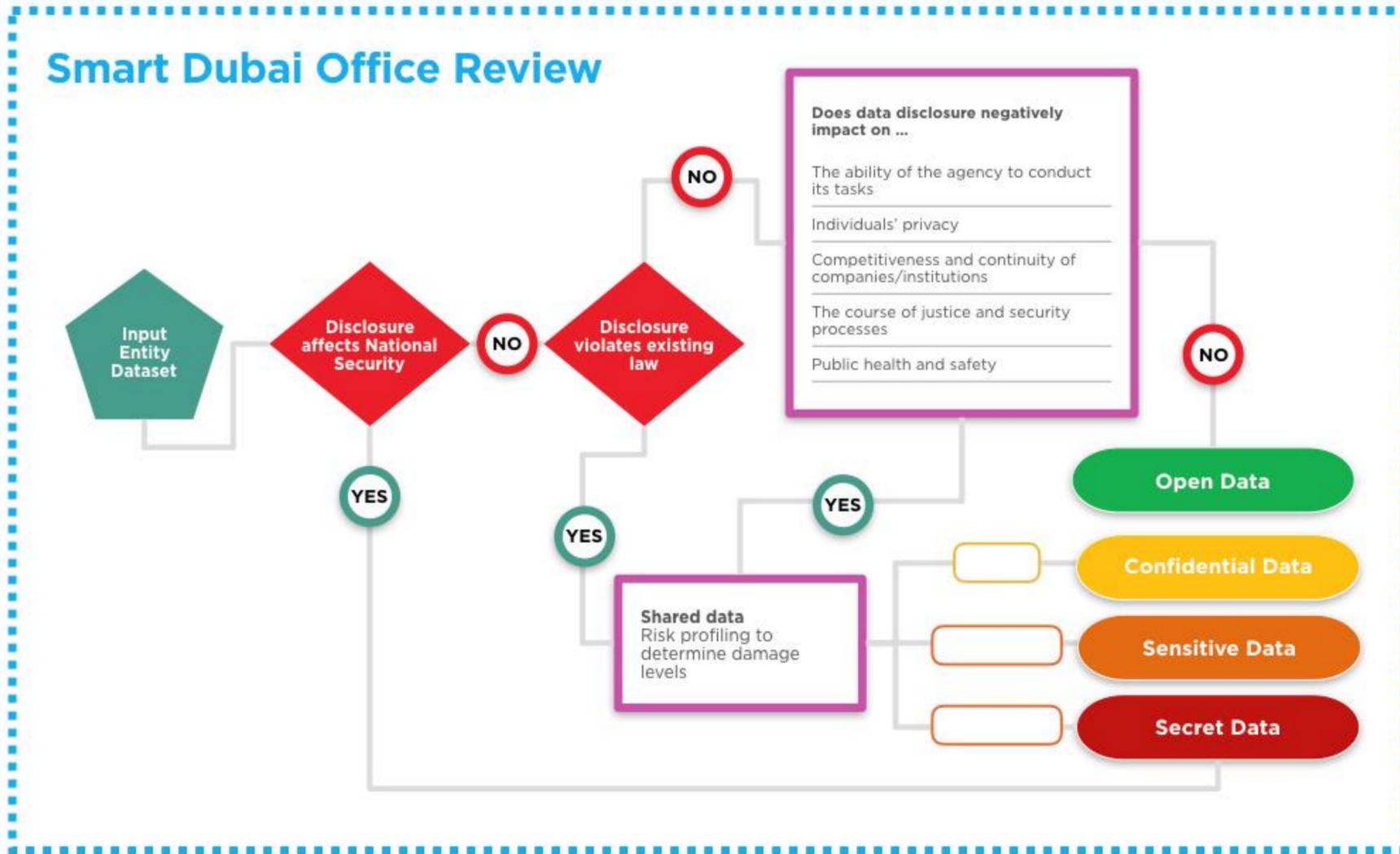Shareable within certain groups subject to strict controls

Secret
Shareable in a limited way between certain individuals under strict controls

CLASSIFICATION REVIEW & APPROVALS

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED ON REQUIREMENTS SET IN THE DUBAI DATA MANUAL
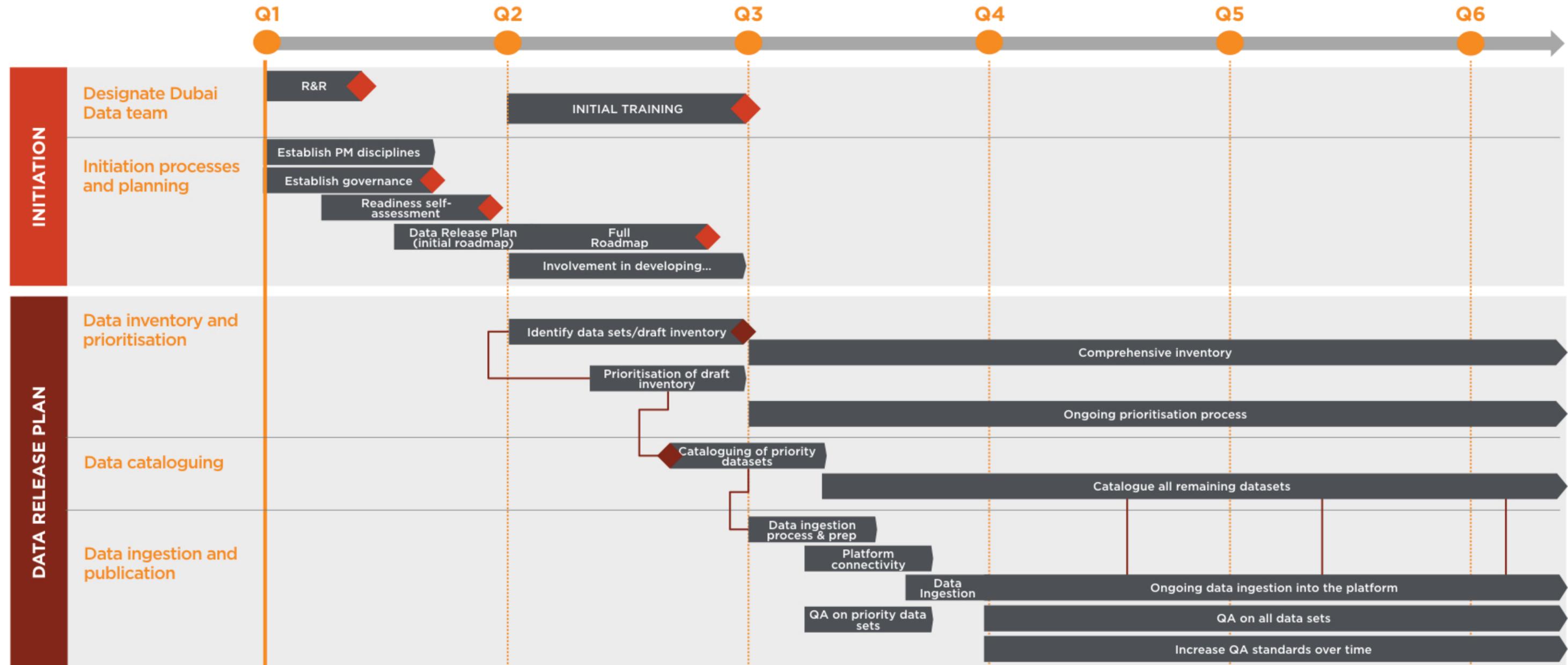
SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
6 Data release plan	Not completed yet	Partial completion of the data release plan and/or not approved by management of entity	Data Release Plan completed as per the standard and approved by management of entity

INITIATION AND DATA RELEASE PLAN

CREATE AN INITIAL ROADMAP FOR YOUR ENTITY'S DATA WORK THAT MEETS WITH DDE REQUIRMENTS

	Q1	Q2	Q3	Q4	Q5	Q6

INITIATION

Designate Dubai Data team
- R&R
- INITIAL TRAINING

Initiation processes and planning
- Establish PM disciplines
- Establish governance
- Readiness self-assessment
- Data Release Plan (initial roadmap)
- Full Roadmap
- Involvement in developing...

DATA RELEASE PLAN

Data inventory and prioritisation
- Identify data sets/draft inventory
- Comprehensive inventory
- Prioritisation of draft inventory
- Ongoing prioritisation process

Data cataloguing
- Cataloguing of priority datasets
- Catalogue all remaining datasets

Data ingestion and publication
- Data ingestion process & prep
- Platform connectivity
- Data Ingestion
- Ongoing data ingestion into the platform
- QA on priority data sets
- QA on all data sets
- Increase QA standards over time

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED ON REQUIREMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
7 Cataloguing (includes metadata, quality, format standard requirements)	Not completed yet	• Partial completion of cataloguing requirements for datasets; • and/or data contract not signed off by Data Leader	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

Business	Technical	Data Quality

Business

1. Dataset Name
2. Contributor/Custodian
3. Creator Business Unit within Entity
4. Dataset Description
5. Coverage (Geographic area)
6. Dataset Temporal Windows
7. Attributes
8. Attribute Description
9. Language
10. Attribute Classification
11. Attribute Range of Values
12. Primary Identifier Attribute for the dataset
13. Dataset Supplementary Reference Data
14. Related datasets
15. Related Identifier
16. Dataset Point of Contact
17. Dataset POC Email
18. Dataset POC Phone
19. Rights
20. Subject Level 1
21. Subject Level 2
22. Subject Level 3
23. Tags / Keywords
24. Already Published?
25. If Published, where?
26. Already Shared?
27. If Shared, with which Entities?
28. Expected Data Set Utilization
29. Remarks/Special Business Rules

Technical

1. Data Provenance
2. Format (MIME)
3. Attribute Datatype
4. Attribute Size
5. Attribute Delimiter
6. Attribute Delimiter- Other
7. Attribute Foreign Key
8. Attribute Join Rules to Related Datasets
9. Dataset Source
10. Source Platform Type
11. Preferred Ingestion Method
12. Description for Method
13. Connectivity Option
14. Description for Connectivity
15. Frequency of Update on Source
16. Frequency of Update to SDP
17. Scheduling of SDP Update
18. Expected Volume/Ingestion Cycle (Daily if Real-time)

Data Quality

1. Data Quality – Completeness
2. Data Quality – Uniqueness
3. Data Quality – Timeliness
4. Data Quality – Accuracy
5. Data Quality – Consistency
6. Data Quality – Reconciliation
7. Data Quality Issue Resolution SLA

COMPLETENESS OF EACH ELEMENT OF COMPLIANCE IS BASED ON REQUIREMENTS SET IN THE DUBAI DATA MANUAL

SCORE SCALE	SCORE=0	SCORE=1	SCORE=2
8 Ingest Data	Not completed yet	Data ingestion attempted but not successful	Data ingested on the platform with no technical issues from entity as per agreed ingestion method
9 Data Maintenance	Not completed yet	Partial maintenance of data quality	Data quality maintained as per agreed data contract (i.e. frequency of publication, attributes etc) for all datasets