



AGID

Agenzia per l'Italia Digitale

FormezPA

GESTIONE E VALIDAZIONE DELLA QUALITÀ DEI DATI

Franco Francia – Vicepresidente DAMA Italy

Data Management Association – Chi siamo



Il **Data Management Association International (DAMA)** è un'associazione internazionale con più di 40 anni di storia: il primo chapter (capitolo nazionale) è nato nel 1980 a Los Angeles (CA).

L'associazione ad oggi conta oltre **71 capitoli nazionali**, per un totale di oltre 100.000 membri.

È un'associazione **no-profit, no-vendor related**, composta da volontari (professionisti, esperti, accademici, studenti) interessati ad approfondire le tematiche di Data Management e a diffondere le sue best practices (<https://dama.org>).



Il **Data Management Association Italy Chapter (DAMA Italy)** è il capitolo italiano affiliato ufficialmente, attivo e riconosciuto dell'International DAMA dal Novembre 2018 (<https://dama-italy.org/>)

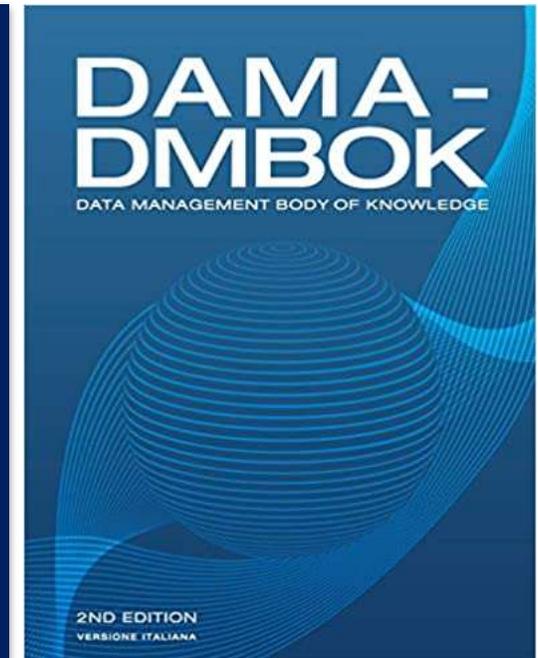


Il **Data Management Body of Knowledge (DMBoK®2)** è la seconda edizione della pubblicazione ufficiale del DAMA International, in cui viene definito e descritto approfonditamente il *framework* di riferimento e le *guidelines* chiave di Data Management, fondamentali per affrontare le sfide del presente per le organizzazioni in ambito di gestione e governance dei dati.

Contenuti del DMBok®2

La pubblicazione contiene molteplici materiali ed esemplificazioni che descrivono in modo esauriente:

- ✓ gli **standard** di Data Management e le modalità di applicazione di questi principi all'interno delle organizzazioni, nelle diverse aree funzionali
- ✓ un quadro funzionale per l'**implementazione** del Data Management, incluse pratiche, metodi e tecniche ampiamente utilizzate, funzioni, ruoli, risultati e metriche
- ✓ un **vocabolario comune** per i concetti di Data Management, funzionale come base per le migliori pratiche per i professionisti della gestione dei dati

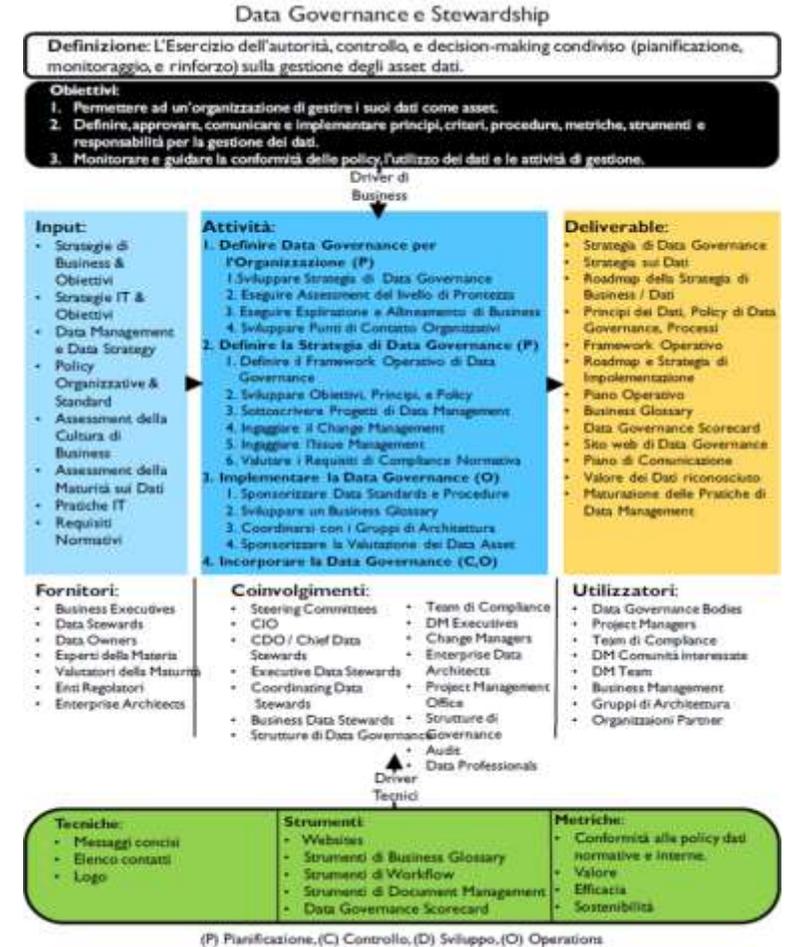




Il framework di DAMA-DMBOK2 (DAMA Wheel)



Environmental Factors Hexagon



Knowledge Area Context Diagram (SIPOC)

Data Quality Management secondo il DMBOK2

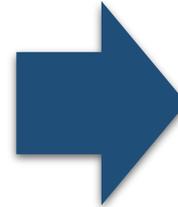
Data Quality Management

Definizione: La pianificazione, l'implementazione e il controllo delle attività che applicano tecniche di gestione della qualità dei dati, al fine di garantire che siano adatti allo scopo e soddisfino le esigenze degli utilizzatori.

Obiettivi:

1. Sviluppare un approccio regolato per rendere i dati adatti allo scopo in base alle esigenze degli utilizzatori di dati.
2. Definire standard, requisiti e specifiche per i controlli della qualità dei dati come parte del ciclo di vita dei dati.
3. Definire e implementare processi per misurare, monitorare e creare report sui livelli di qualità dei dati.
4. Identificare e promuovere opportunità per migliorare la qualità dei dati, attraverso miglioramenti dei processi e dei sistemi.

Drivers di Business



Drivers Tecnici



(P) Pianificazione, (C) Controllo, (D) Sviluppo, (O) Operations

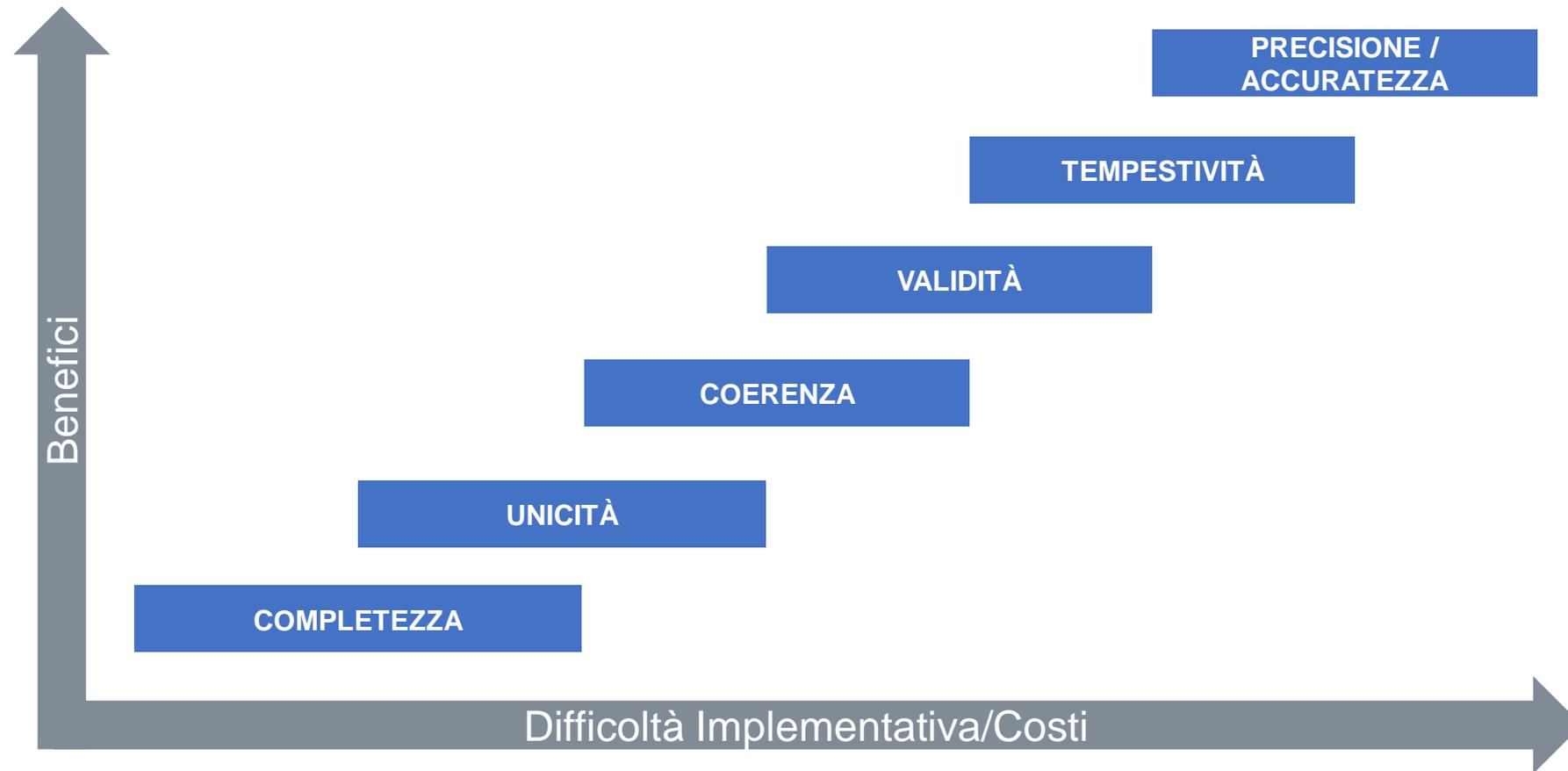
Obiettivi di un programma di Data Quality Management

- Generare e promuovere una «Data Quality Awareness» e Cultura
- Definire il Valore, Costi e Impatti della Data Quality
- Definire i requisiti di Data Quality
- Verificare e Validare il Modello/i dei Dati
- Fare un Assessment del livello di Maturità e Qualità dei Dati (incluso un Capabilities Assessment per correttamente definire ruoli e responsabilità)
- Definire Dimensioni, Misure ed Indicatori di Data Quality
- Profilare, Analizzare e Valutare la Data Quality
- Definire le «Regole» di Business per la Data Quality
- Test e Validazione dei requisiti di Data Quality
- Definire e Valutare i Data Quality Service Levels Agreements
- Misurazione e Monitoraggio continuativo
- Gestione delle problematiche relative alla DQ

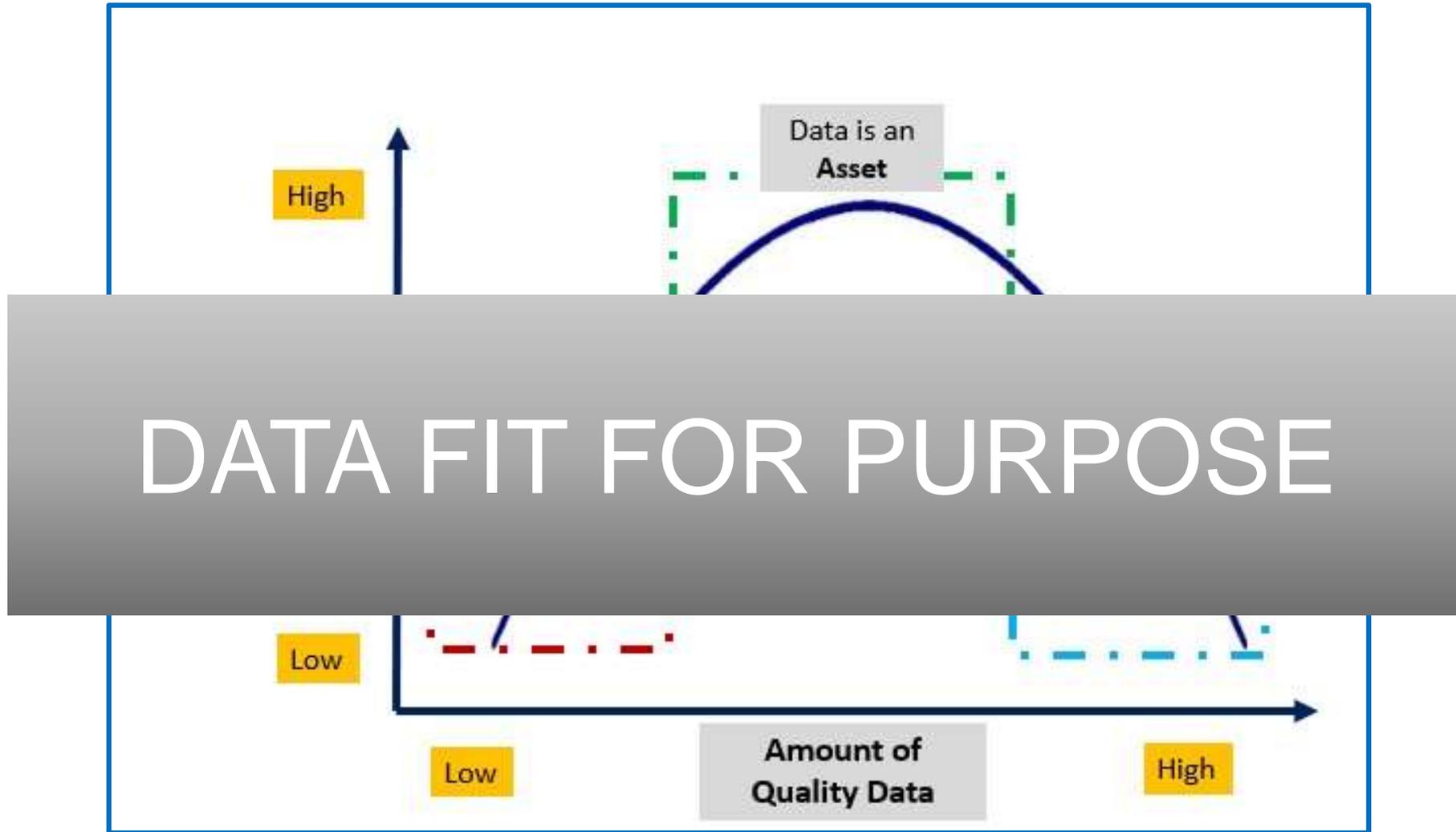
Dimensioni della Data Quality (DMBOK2)

Dimensione di Qualità	Descrizione
Precisione / Accuratezza	La precisione si riferisce al grado in cui i dati rappresentano correttamente le entità della "vita reale".
Completezza	La completezza si riferisce alla presenza di tutti i dati richiesti.
Tempestività	Il concetto di tempestività dei dati si riferisce a diverse caratteristiche dei dati. Le misure di tempestività devono essere comprese in termini di volatilità attesa.
Coerenza	La coerenza può fare riferimento a garantire che i valori dei dati siano rappresentati in modo coerente all'interno di un set di dati e tra set di dati, e associati in modo coerente tra i set.
Integrità	L'integrità dei dati include idee associate a completezza, accuratezza e coerenza. Nei dati, l'integrità di solito si riferisce all'integrità referenziale (coerenza tra oggetti dati tramite una chiave di riferimento contenuta in entrambi gli oggetti) o coerenza interna all'interno di un set di dati in modo tale che non vi siano buchi o parti mancanti.
Ragionevolezza	La ragionevolezza chiede se un modello di dati soddisfa le aspettative.
Unicità / De duplicazione	L'unicità afferma che nessuna entità esiste più di una volta all'interno del set di dati.
Validità	La validità indica se i valori dei dati sono coerenti con un dominio definito di valori.

Dimensioni: difficoltà implementative vs Benefici

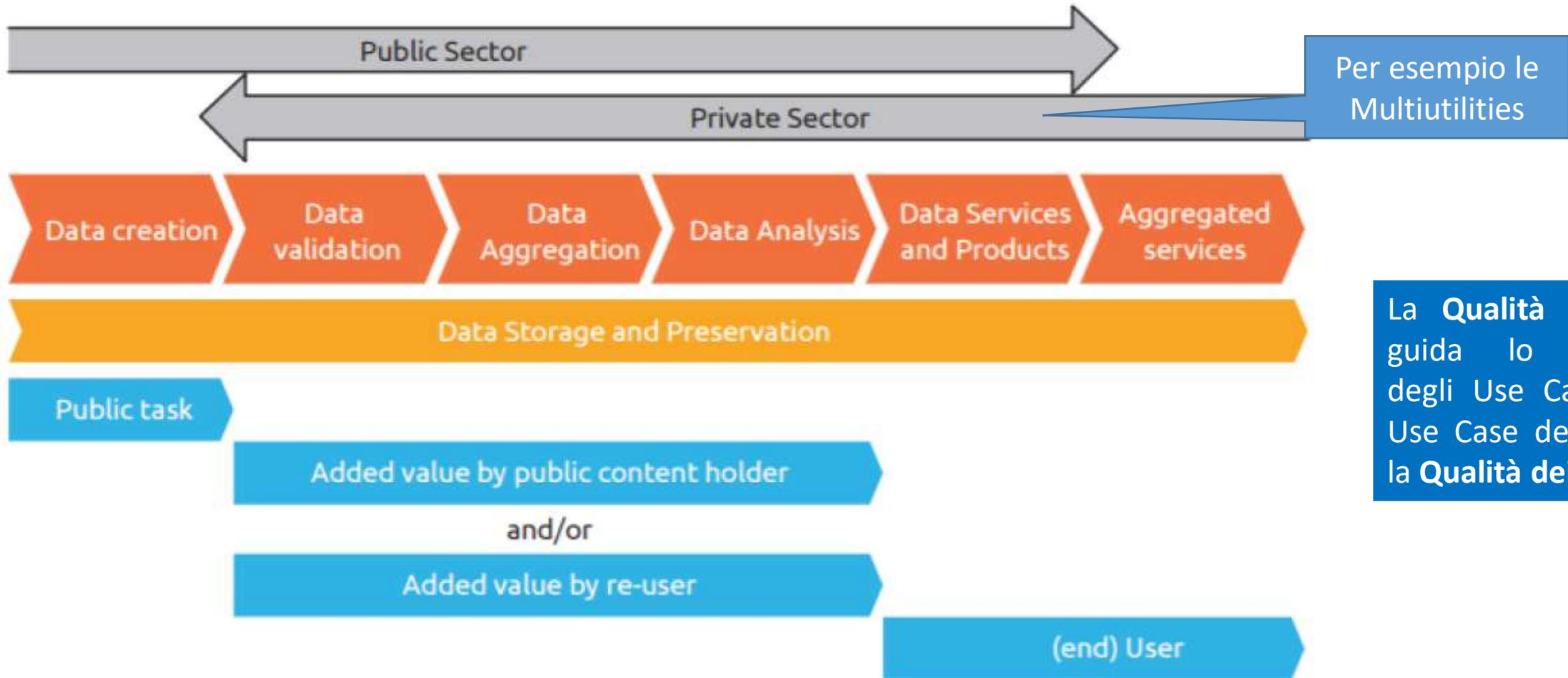


Come definire la Qualità dei Dati?



Fonte: CMMI Institute

La catena del valore degli Open Data

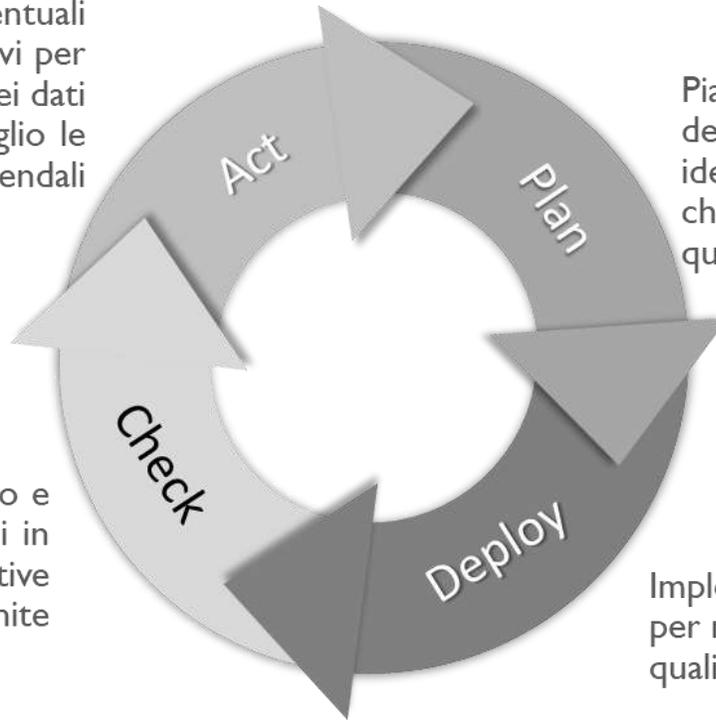


La **Qualità dei Dati** guida lo sviluppo degli Use Case o gli Use Case definiscono la **Qualità dei Dati**?

https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

Data Quality Management Cycle – le attività

Agire per risolvere eventuali problemi identificativi per migliorare la qualità dei dati e soddisfare meglio le aspettative aziendali



Monitoraggio e misurazione dei livelli in relazione alle aspettative aziendali definite

Pianificazione dell'assessment dello stato attuale e identificazione delle metriche chiave per la misurazione della qualità dei dati

Implementazione di processi per misurare e migliorare la qualità dei dati

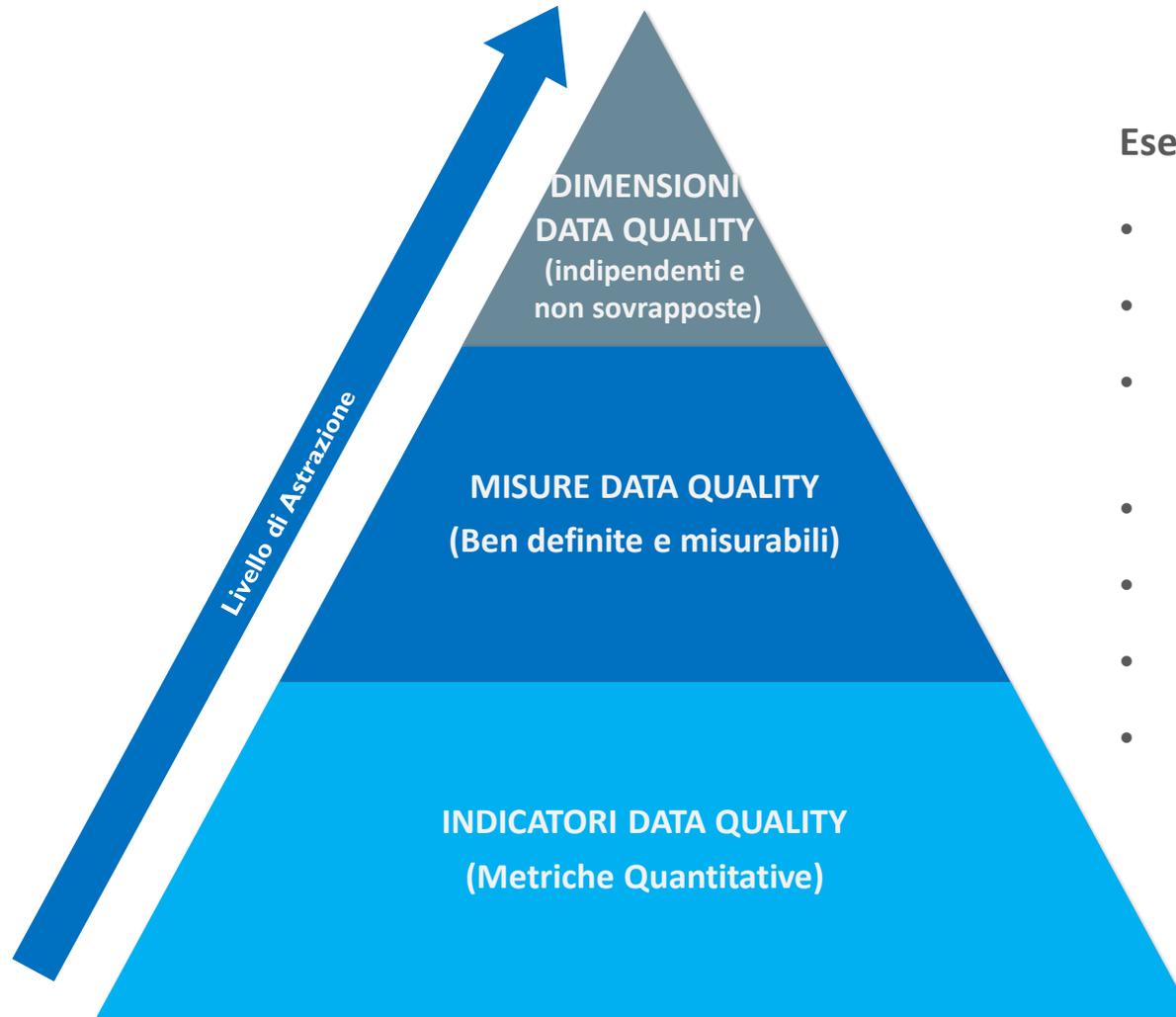
1. Definire **modello di governo** della Data Quality
2. Definire i **requisiti di Data Quality**
3. **Profilare, Analizzare e Valutare** la Data Quality
4. Definire **Data Quality Indicators (DQI)**
5. Definire le regole di **Data Quality**
6. Test & Validazione dei requisiti di Data
7. Impostare e valutare i Data Quality Service Levels
8. (Continuously) Measure & Monitor Data Quality
9. Gestire le Data Quality Issues
10. Pulire e correggere i difetti di Data Quality
11. Definire e implementare procedure di DQM
12. Monitorare l'applicazione delle procedure di DQM e le relative performance

Campi	Dataset				
id_amat	9001	9002	9003	9004	9005
id_via	1064	3377	3377	3377	3377
anagrafica	PIAZZA VENTICINQUE APRILE	VIA GIACOMO MEDICI DEL VASCELLO	VIA G		
gerarchia					
rete	urbana	ciclabile	ciclabile		
tipologia	promiscuo pedoni	ciclabile sede propria	ciclabile		
sede	attraversamento	marciapiede	marciapiede		
marcia	bidirezionale	monodirezionale - concorde	monodirezionale		
norma	attraversamento - generico	ordinanza lavori stradali	ordinanza lavori stradali		
lunghezza	8	7	8		
LONG_X_4326_CENTROID	91.870.305.186.785	924.194.464.907.128	923.9		
LAT_Y_4326_CENTROID	454.809.876.471.498	454.392.715.546.864	454.4		
Location	(45.4809876471498, 9.1870305186785)	(45.4392715546864, 9.24194464907128)	(45.4		

Quali caratteristiche di qualità applicare e su quali campi?

- ### Esempi di Misure della Qualità dei Dati
- **Conformità alle regole (Validità):** il livello per cui il dato soddisfa un set di regole di business (business rules) che governano i valori consentiti per ogni campo
 - **Relationship (Coerenza):** il livello per cui il dato soddisfa un set di regole di business che governano le relazioni fra i vari record
 - **Credibilità (Ragionevolezza):** il livello per cui il dato soddisfa le aspettative in comparazione con altri dati, in rispetto ad altre fonti e nel tempo
 - **Prossimità (Tempestività):** la prossimità fra quando il dato diventa disponibile e quando l'evento che rappresenta è realmente occorso
 - **Puntualità (Tempestività):** la differenza fra quando il dato diventa disponibile e quando sarebbe dovuto essere disponibile (in base alle specifiche)
 - **Copertura (Completezza):** il livello per cui il dato include tutti campi e record rilevanti
 - **Identità (Unicità):** il livello per cui il dato esclude tutti i duplicati

Gerarchia della Data Quality



Esempi di Indicatori della Qualità dei Dati (KQI)

- **DQI-V01**: Percentuale di campi considerati validi
- **DQI-V02**: Percentuale di record considerati validi
- **DQI-B01**: Percentuale di aggregazioni numeriche all'interno della tolleranza
- **DQI-C01**: Percentuale di campi obbligatori popolati
- **DQI-C02**: Percentuali di campi opzionali popolati
- **DQI-C03**: Percentuali di record previsti salvati
- **DQI-U01**: Percentuale di record unici



1 **Definizione Obiettivi**
Definire gli Obiettivi relativi al miglioramento della Qualità dei Dati

2 **Identificazione delle azioni**
• Analisi processi
• Analisi flusso dati
• Reporting & Costi

3 **Registrazione attività**
• Assegnare issue ID
• Prioritizzazione
• Assegnazione al team di lavoro

4 **Analisi delle Cause**
• Analisi processi
• Analisi flusso dati
• Reporting & Costi

5 **Bonifica**
• Prioritizzazione
• Pulizia dati
• Evoluzione sistemi-processi

6 **Sostenibilità**
• Reporting & costi
• Applicazione dei controlli
• Riconciliazione dei dati
• Formazione

Data Governance secondo il DMBOK2



ALCUNI RUOLI STANDARD (definiti da DAMA I e DAMA Italy)

- **Chief Data Officer:** parte dell'Executive Team ed a capo delle attività relative al Data Management dell'azienda. Provvede Leadership per iniziative volte alla data-driven culture. Possibilmente Budget Holder.
- **Data Owner:** solitamente parte del Senior o Middle Management Team a cui fa capo una linea di Business o Dipartimento. Le responsabilità sui Dati di propria competenza sono relative alla definizione dei requisiti di mantenimento e aggiornamento, nonché di accesso all'informazione ed utilizzo.
- **Data Steward:** solitamente SME (Subject Matter Experts) con una conoscenza di tutti i fondamenti di Data Management che operano come «liaison» fra il Business e l'ICT. Spesso non sono full-time su questo ruolo per quanto le attività di Stewardship debbano essere formalizzate e documentate.
- **Data Scientist:** con l'avvento delle nuove tecnologie di BI e metodologie predittive e preventive basate sui concetti introdotti dal Big Data, coloro che storicamente erano conosciuti come Data Engineers stanno vivendo un revamp. Oggi i Data Scientists si occupano principalmente delle attività di Data Mining, utilizzo nuove tecnologie a supporto dei Data Lake e, purtroppo...soprattutto al Data Cleansing!
- **Data Analyst:** solitamente afferenti al Business (PMO o team di Analisti) e si occupano della parte di «analisi contestuale» dei dati, nonché della loro interpretazione e reportistica per il Business.

- ✓ Secondo lo standard ISO, la capacità di creare valore dai dati richiede sia la comprensione delle caratteristiche dei dati che ne determinano la qualità, sia la capacità di misurare, gestire e realizzare report sulla qualità dei dati
- ✓ Le Dimensioni della Qualità dei Dati sono una caratteristica misurabile dei dati.
- ✓ La Gestione della Qualità dei Dati e la Data Governance sono strettamente collegate e devono essere programmi continuativi, non iniziative isolate.
- ✓ Non basta risolvere i problemi di qualità, ma occorre analizzare le cause
- ✓ I dati sono di alta qualità quando sono «Adatti allo scopo» (FIT for Purpose)



**Grazie
per
l'attenzione!**

Domande?



www.agid.gov.it

Franco Francia – Vicepresidente DAMA Italy
vicepresidente@dama-italy.org