



AGID

Agenzia per l'Italia Digitale

FormezPA

FORMAZIONE AGID – FORMEZ SULLA TRANSIZIONE DIGITALE DELLA PA

**Progetto Informazione e formazione per la transizione digitale della PA
nell'ambito del progetto «Italia Login – la casa del cittadino»**

(A valere sul PON Governance e Capacità Istituzionale 2014-2020)



AGID

Agenzia per l'Italia Digitale

FormezPA

Dati aperti dalla teoria alla pratica: la corretta applicazione delle Linee Guida

Dati aperti: qualità e validazione

Domenico Natale

Editor ISO/IEC 25012 e 25024

Membro ISO SC7 dal 1994

Socio onorario UNINFO

QUALITA', ANALISI E VALUTAZIONE DEI DATI

Il webinar approfondisce la fase relativa all'analisi e valutazione della qualità dei dati:

Quali raccomandazioni seguire per i vari formati di dati aperti in modo da garantire buoni livelli di qualità:

- conoscere e applicare il modello di qualità dei dati UNI CEI ISO/IEC 25012

Quali caratteristiche di qualità considerare tra quelle definite negli standard di riferimento e come misurarle:

- sono previste 15 caratteristiche, 2 obbligatorie per legge
- le altre sono selezionabili a scelta secondo il contesto
- l'Agid con la determina 68/2013 propone l'applicazione di quattro caratteristiche inerenti (vedi UNI TS 11725:2018 Linee guida per la misurazione della qualità dei dati)
- sono disponibili 63 misure della qualità tra tutte le caratteristiche

Quale livello di qualità garantire a utenti e ri-utilizzatori:

- il massimo livello raggiungibile dell'intervallo 0 – 1 per ogni caratteristica

DATO e METADATO

DATO: Rappresentazione re-interpretabile dell'informazione in un modo formalizzato idoneo per la comunicazione, l'interpretazione o l'elaborazione

METADATO: dato che descrive un altro dato

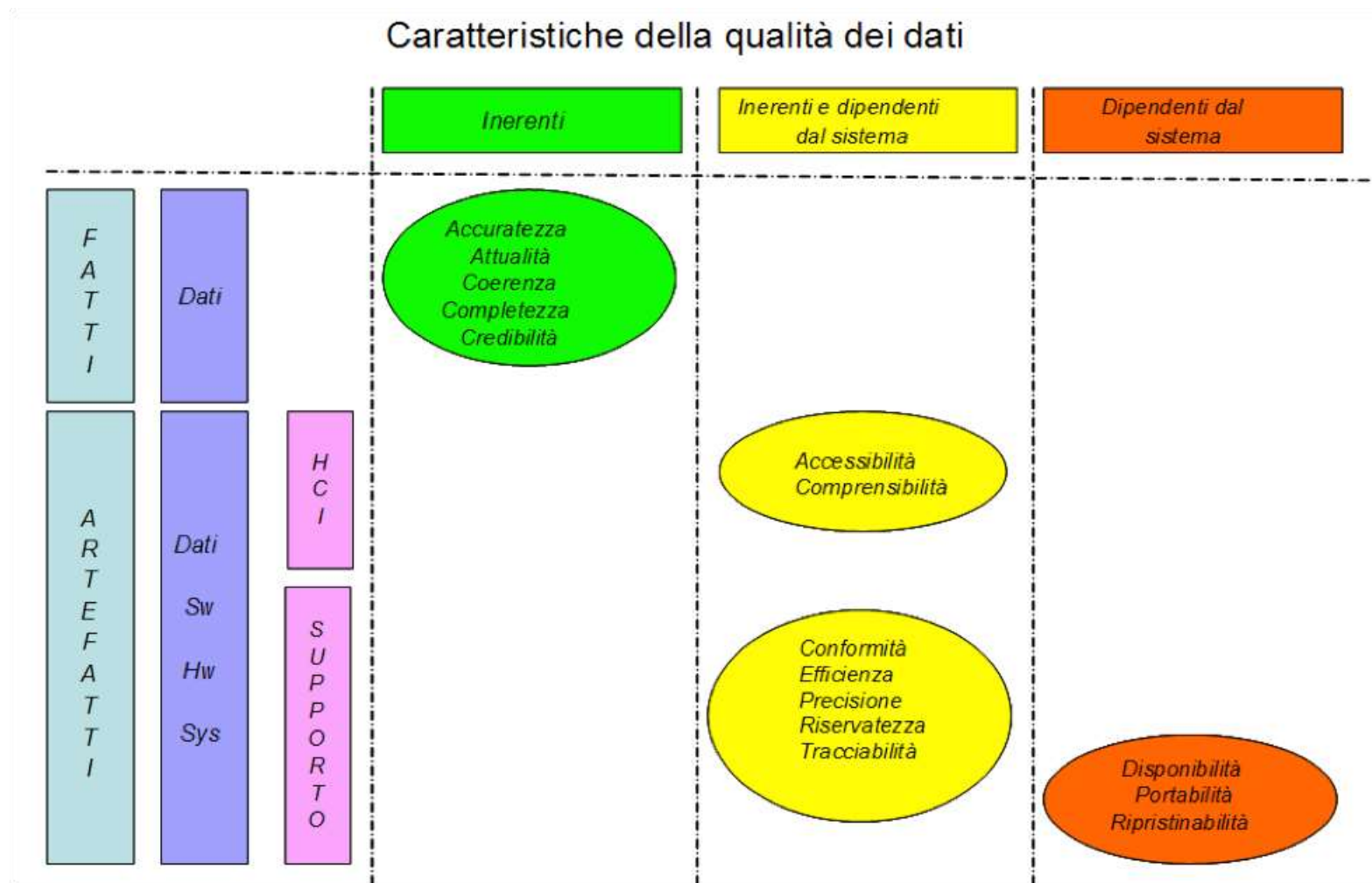
Nota: i dati possono essere elaborati da persone o da mezzi automatici

[Fonte UNI CEI ISO/IEC 25012 e 25024]

Ci si riferisce ai dati conservati con un formato strutturato in un computer;
lo scopo esclude i dati prodotti da dispositivi «embedded or real time sensors» a meno che non siano conservati per successive elaborazioni o finalità storiche

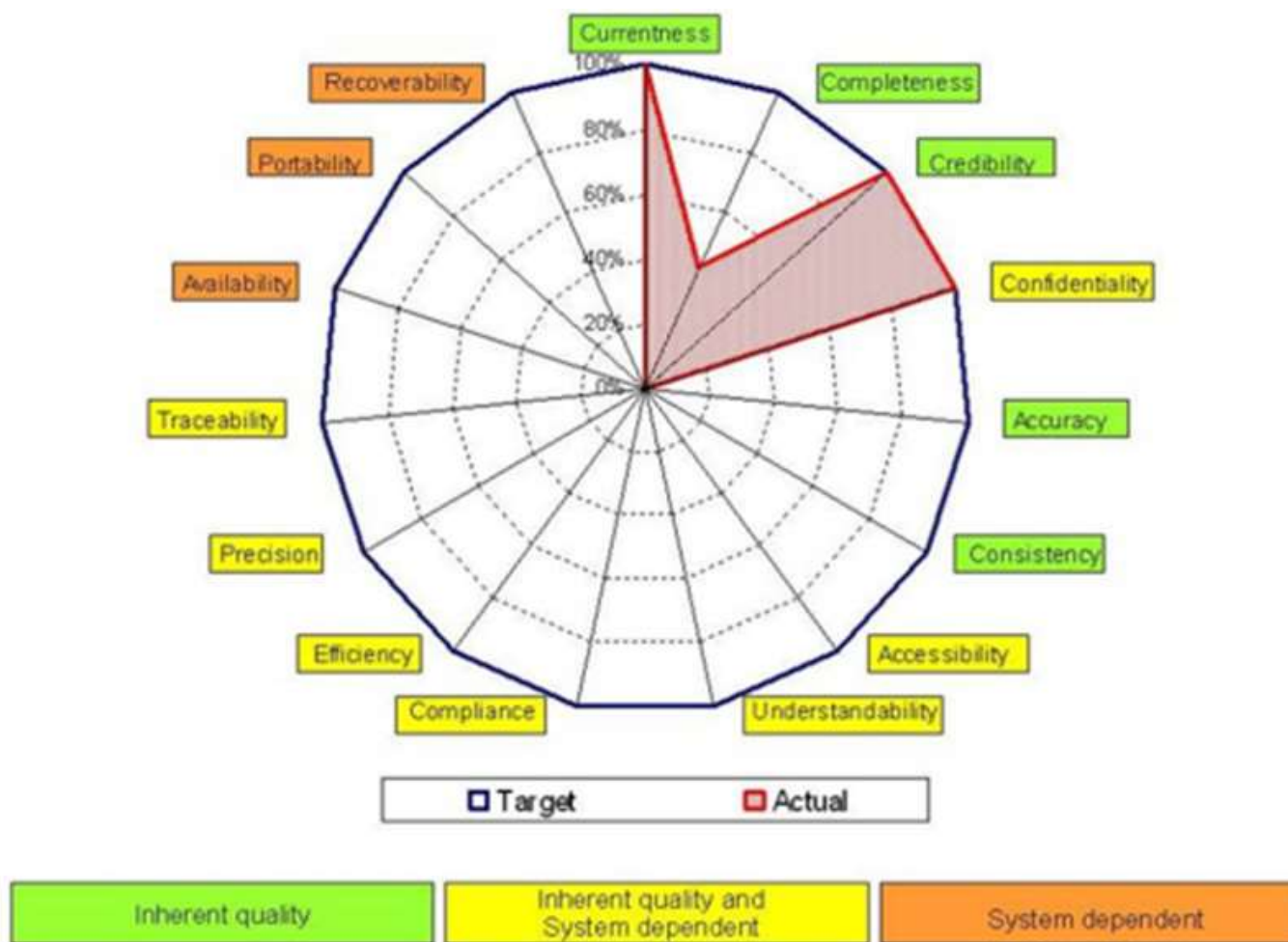
Le caratteristiche di qualità sono applicabili a: stringhe di caratteri, testi, date, numeri, immagini, suoni, ecc.

MODELLO di QUALITA'



Rivista UNI,
Unificazione&Certificazione,
2009, pag. 20

ESEMPI di MISURE



iso25000.it (Dati)

ACCURATEZZA

I dati hanno attributi che rappresentano correttamente il valore vero, cioè reale

Sintattica: i valori dei dati sono uguali o vicini a un set di valori ritenuti sintatticamente corretti (ad esempio quando Maria è memorizzata come Mari)

Es. Misura:

- numero dei dati con valori accurati sul numero totale dei dati

Semantica: i valori dei dati sono uguali o vicini a un set di valori definiti in un dominio considerato semanticamente corretto (ad esempio quando Giovanni è memorizzato come Giovanni se non addirittura come Giove o Giorgio)

Es. Misura:

- numero dei dati semanticamente corretti sul numero totale dei dati per i quali si richiede l'accuratezza semantica

COMPLETEZZA

I dati hanno valori per tutti gli attributi attesi in uno specifico contesto d'uso

Esempio Misure:

- *il numero dei record contenuti in un file sul numero di record attesi*
- *il numero dei campi compilati (non vuoti) in un record sul numero totale dei campi)*
- *numero dei record completamente vuoti sul numero dei record*
- *numero dei Comuni che hanno trasmesso dati, sul numero dei Comuni attesi*

COERENZA

I dati hanno attributi coerenti con altri dati analoghi memorizzati in un altro database o in un altro sistema

Un caso di incoerenza è rappresentato da sinonimi (necessità di un dizionario ufficiale)

Es. Misure:

- numero delle date di assunzione di impiegati risultanti anteriori alla data di nascita sul totale del numero degli impiegati analizzati

- numero dei casi in cui i CF di soggetti memorizzati in un sistema siano diversi dal numero dei CF degli stessi soggetti memorizzati in un altro sistema

CREDIBILITA'

I dati hanno attributi che sono ritenuti credibili dagli utenti

E' incluso il concetto di affidabilità delle fonti, attribuzione, impegno... e della verifica degli outlier

Es. Misura:

- numero delle informazioni i cui contenuti sono validati/certificati da uno specifico processo rispetto al totale delle informazioni

ATTUALITA'

I dati hanno attributi che sono del «giusto tempo»

Es. Misure:

- numero dei dati aggiornati con la frequenza richiesta sul numero totale dei dati in esame

- numero dei dati aggiornati nel tempo utile sul totale dei dati in esame (per esempio il cambio del binario di un treno in partenza)

ACCESSIBILITA'

Accesso ai dati da persone autorizzate e da persone con disabilità che necessitano di tecnologie assistive (caratteristica obbligatoria per la legge 4/2004)

Nota: da fare attenzione ai dati pubblicati su pdf non accessibile (es. da scanner)

Es. Misure:

- numero dei dati rilevanti per i compiti dell'utente che hanno valori accessibili sul totale dei dati per i quali è richiesta accessibilità

- numero dei dispositivi che consentono la visualizzazione di dati accessibili, sul totale dei dispositivi accessibili

COMPLIANCE

Dati con attributi che rispettano standards, convenzioni o regole

Es. Misura:

- numero di dati che hanno valori e/o formati che rispettano standards, convenzioni o regole sul totale dei dati che dovrebbero essere conformi

Nota

Le modalità di attuazione di regole di rappresentazione e uso di dati devono essere conformi alla Legge 4/2004 sull'accessibilità e al GDPR

RISERVATEZZA

I dati hanno attributi che assicurano di essere accessibili e interpretabili da utenti autorizzati

Es. Misure:

- *numero dei dati correttamente criptati e decrittati sul numero del totale dati in esame*
- *numero di accessi conclusi con successo da parte di utenti non autorizzati in uno specifico periodo di tempo sul numero di accessi tentati da utenti non autorizzati*

EFFICIENZA

I dati hanno attributi che consentono di poter essere processati con attesi livelli di performance usando appropriate quantità di risorse

Es. Misure:

- numero dei dati che gli utenti valutano come «usati facilmente» sul numero totale dei dati valutati dagli utenti

- tempo perso dovuto alla rappresentazione dei dati (data format) durante un lavoro sul tempo globale dell'elaborazione

PRECISIONE

I dati hanno attributi che sono esatti o che consentono discriminazione in uno specifico contesto d'uso

Es. Misure:

- numero dei dati con la precisione richiesta sul numero dei dati per i quali il requisito di precisione è definito*
- numero dei dati definiti e percepiti con la precisione richiesta/data format sul numero dei dati la cui precisione di format è richiesta*

TRACCIABILITA'

I dati hanno attributi che consentono un controllo degli accessi ai dati o modifiche

Es. Misure:

- numero dei dati per i quali esiste effettivamente il controllo di tracciabilità sul numero dei dati per i quali è previsto il controllo di tracciabilità*
- numero dei dati per i quali l'accesso di utenti è effettivamente tracciato sul numero dei dati per i quali il controllo utenti è atteso*
- numero dei dati per i quali esiste effettivamente la tracciatura dei valori sul numero dei casi attesi*

COMPRESIBILITA'

I dati hanno attributi che consentono di essere letti e interpretati dagli utenti e sono espressi in linguaggi appropriati, simboli e unità

Es. Misure:

- numero dei dati i cui valori sono rappresentati da simboli conosciuti sul numero dei dati per i quali è richiesta comprensibilità*
- numero dei dati i cui valori sono definiti in un dizionario dati usando un comune vocabolario sul numero dei dati definiti in un dizionario*
- numero dei dati con metadati sul numero dei dati*
- numero dei dati facilmente comprensibili dagli utenti sul numero dei dati*

DISPONIBILITA' (dipendente dal sistema)

I dati hanno attributi che consentono di essere utilizzati dagli utenti e/o applicazioni

Es. Misura:

- numero dei dati disponibili in uno specifico periodo di tempo sul numero dei dati richiesti nello stesso periodo di tempo

Nota: specificare se il periodo di tempo include backup, ripristino o periodo normale

PORTABILITA' (dipendente dal sistema)

I dati hanno attributi che preservano la qualità esistente durante o dopo dopo il trasferimento

Es. Misura:

- numero dei dati la cui qualità è preservata sul numero dei dati trasferiti

RIPRISTINABILITA' (dipendente dal sistema)

I dati hanno attributi che consentono di mantenere e preservare uno specifico livello di operazioni e qualità, anche in caso di eventi di disastro o fallimento

Es. Misure:

- numero dei dati ripristinati con successo e correttezza dal sistema sul numero dei dati per cui è richiesto il ripristino*
- numero dei dati (o file) salvati periodicamente sul numero dei dati per cui è richiesto il backup*

VALIDAZIONE e VALUTAZIONE (Conferma, attraverso la disponibilità di evidenza obiettiva, che i requisiti per uno specifico uso definito, o applicazione, sono stati pienamente adempiuti – ISO/IEC 25000: Guide to SQuaRE) e (determinazione sistematica della misura in cui una entità soddisfa criteri specifici - ISO/IEC 25001 Planning and management)

La valutazione secondo, l'ISO/IEC 25040 della serie 25000, consiste nelle seguenti attività:

- 1) Stabilire il fine della della valutazione
 - Ottenere i requisiti da sottoporre a valutazione
 - Identificare il prodotto e gli oggetti della valutazione
- 2) Selezionare le misure da valutare ed le fasi del ciclo di vita
- 3) Definire il piano di valutazione e i criteri
- 4) Eseguire la valutazione sulla base dei risultati della misurazione in confronto ai requisiti
- 5) Dare feedback all'organizzazione che ha in carico il prodotto
- 5) Produrre un rapporto finale
- 6) Attestato di certificazione di Ente terzo accreditato

SCALA di VALUTAZIONE



RIFERIMENTI

Per l'uso degli standard citati e per approfondimenti si raccomanda l'uso degli standard originali disponibili sul sito dell'UNI www.uni.com

UNI CEI ISO/IEC 25012:2014 Ingegneria del software - Requisiti di qualità e valutazione del prodotto software (SQuaRE) - Modello di qualità dei dati

UNI CEI ISO/IEC 25024:2016 Ingegneria del software e di sistema - Requisiti e valutazione della qualità dei sistemi e del software (SQuaRE) - Misurazione della qualità dei dati

UNI TS 11725:2018 Ingegneria del software e di sistema - Linee guida per la misurazione della qualità dei dati

DATI APERTI: QUALITA', VALIDAZIONE, VALUTAZIONE

Domenico Natale

Chairman Commissione UNINFO UNI/CT 504 Software Engineering

Per altre informazioni sulla qualità dei dati e sulla serie 25000
è possibile consultare www.iso25000.it

www.agid.gov.it

Domenico Natale

Chairman Commissione UNINFO UNI/CT 504 Software Engineering

dnatale51@gmail.com