



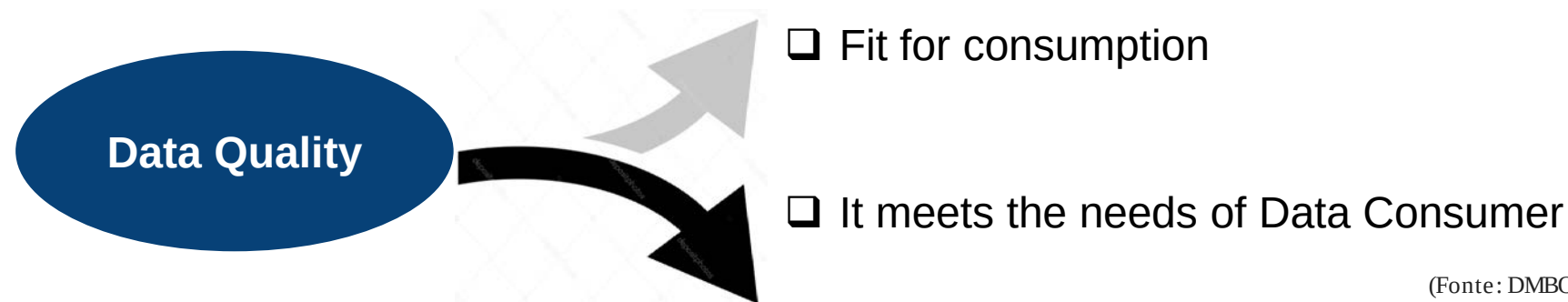
Data Engineering Essentials: Governare il Dato per Creare Valore.

Hands-on: misurare la Data Quality

Ing. Fabio Gatti
Ordine degli Ingegneri della Provincia di Roma

Roma, 01 Luglio 25

DQ: review principi fondamentali



(Fonte: DMBOK–Ch 13 Data Quality)



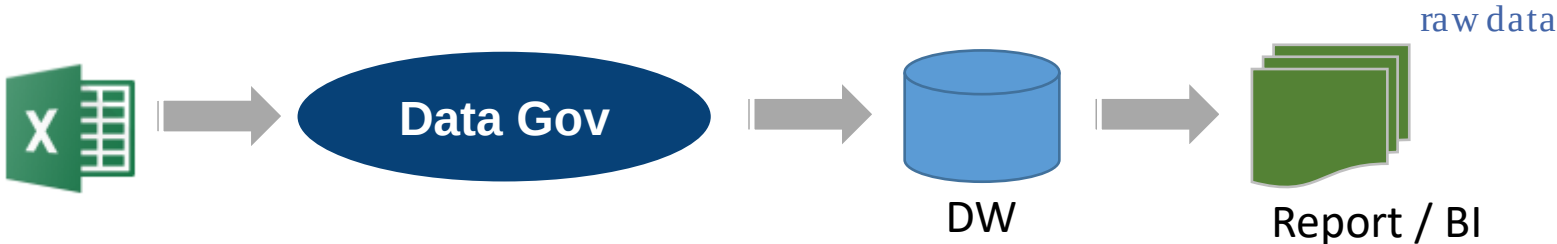
- ☐ DQ dimensions: caratteristiche **misurabili** per valutare la Qualità dei Dati
- ☐ Valutarle nel **loro complesso**
- ☐ Integrare con QGates
- ☐ Rapportarle ai **business** needs



Use case#1: Source spreadsheet

Scenario

L' unità operativa X fornisce alla funzione DG un foglio di calcolo per Data Ingestion e costruzione report periodico



Data Issue

Violazione Schema

Variazione struttura

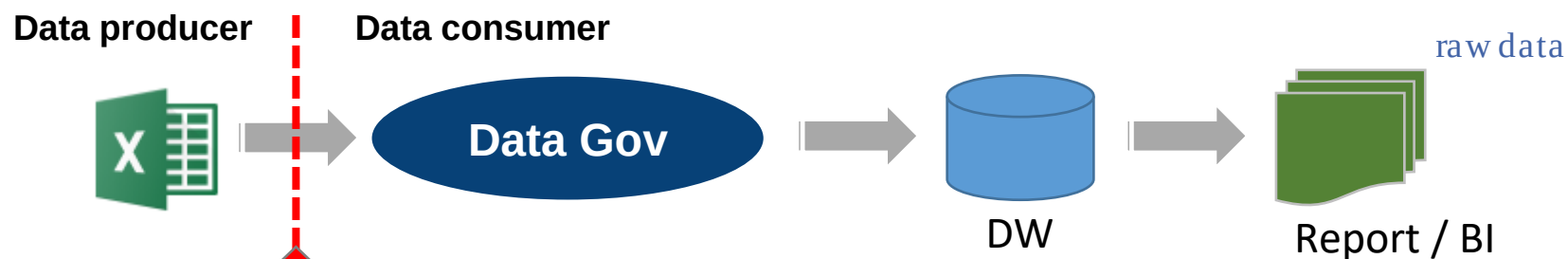
Modifica nomi colonne / ordinamento

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	eliminate missing data from a sample														
2															
3		Poverty	White	Crime	Doctors	University	Income			Poverty	White	Crime	Doctors	University	Income
4	Alabama	15.7	71.0	448	218.7	22.9	42,666		Alabama	15.7	71.02718	448	218.1734	22	42666
5	Alaska	8.4	70.6	661	228.5	27.3	68,460		Alaska	8.4	70.62319	661.2	228.4503	27.3	68460
6	Arizona	14.7	86.5	483	209.7	25.1	50,958		Arizona	14.7	86.5057	482.7	209.6666	25.1	50958
7	Arkansas	17.3	80.8		203.4	18.8	38,815		California	13.3	76.64005	522.6	268.6906	29.6	61021
8	California	13.3	76.6	523	268.7	29.6	61,021		Connecticut	9.3	84.27865	256	376.3753	35.6	68595
9	Colorado		89.7	348	259.7	35.6	56,993		Delaware	10	74.26606	689.2	250.9418	27.5	57989
10	Connecticut	9.3	84.3	256	376.4	35.6	68,595		Florida	13.2	79.81107	722.6	247.8636	25.8	47778
11	Delaware	10.0	74.3	689	250.9	27.5	57,989		Georgia	14.7	65.38772	493.2	217.4457	27.5	50861
12	Florida	13.2	79.8	723	247.9	25.8	47,778		Hawaii	9.1	29.66733	272.8	316.9829	29.1	67214
13	Georgia	14.7	65.4	493	217.4	27.5	50,861		Illinois	12.2	79.13311	533.2	280.1539	29.9	56235
14	Hawaii	9.1	29.7	273	317.0	29.1	67,214		Iowa	11.5	94.17046	294.7	189.2832	24.3	48980

Use case#1: Invio spreadsheet

Remediation

Enforcement rapporto Cliente / Fornitore e istituzione di un Data Quality Gate tramite Data Contract



Data Contract

```
dataContractSpecification: 0.9.3
id: urn:datacontract:orders-latest
info:
  title: Orders Latest
  version: 1.0.0
models:
  orders:
    type: table
    fields:
      order_id:
        type: text
        format: uuid
```

*datacontract.yaml
or odcs.yaml*

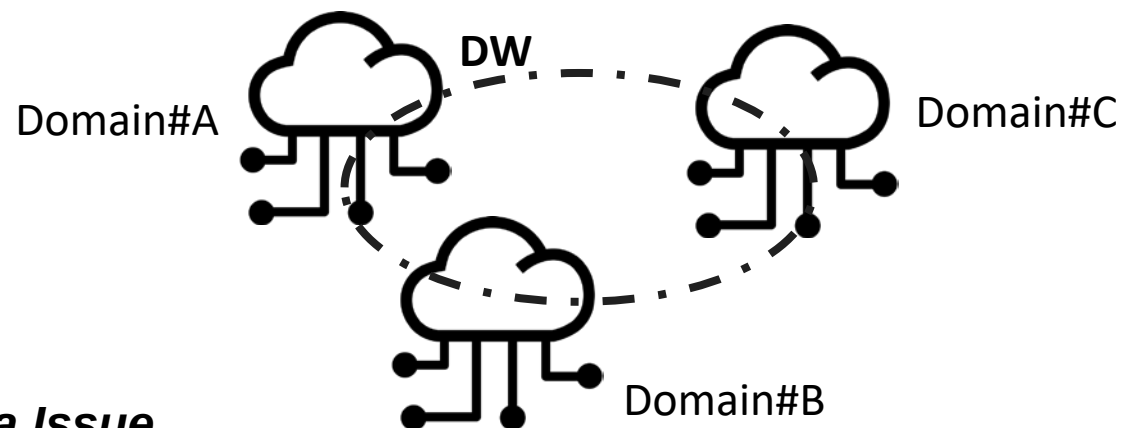
RC

*Elaborazione manuali
Change in data structure
Weak lineage*

Use case#2: Multi Domain DW

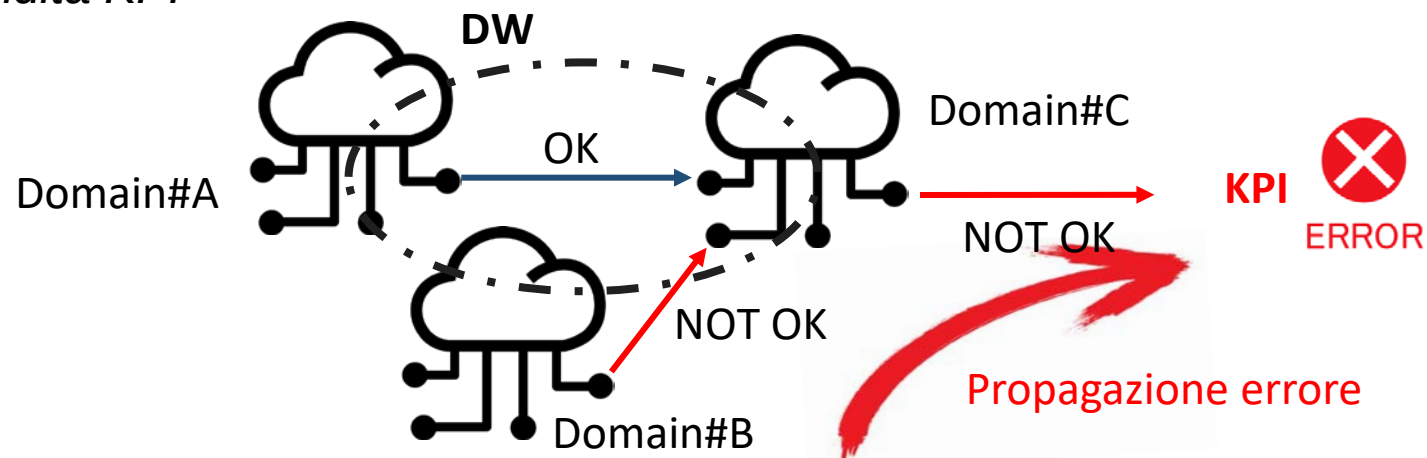
Scenario

DW alimentato da dati provenienti da diversi Data Domain e con KPI dipendenti



Data Issue

Inconsistenza / Validità KPI

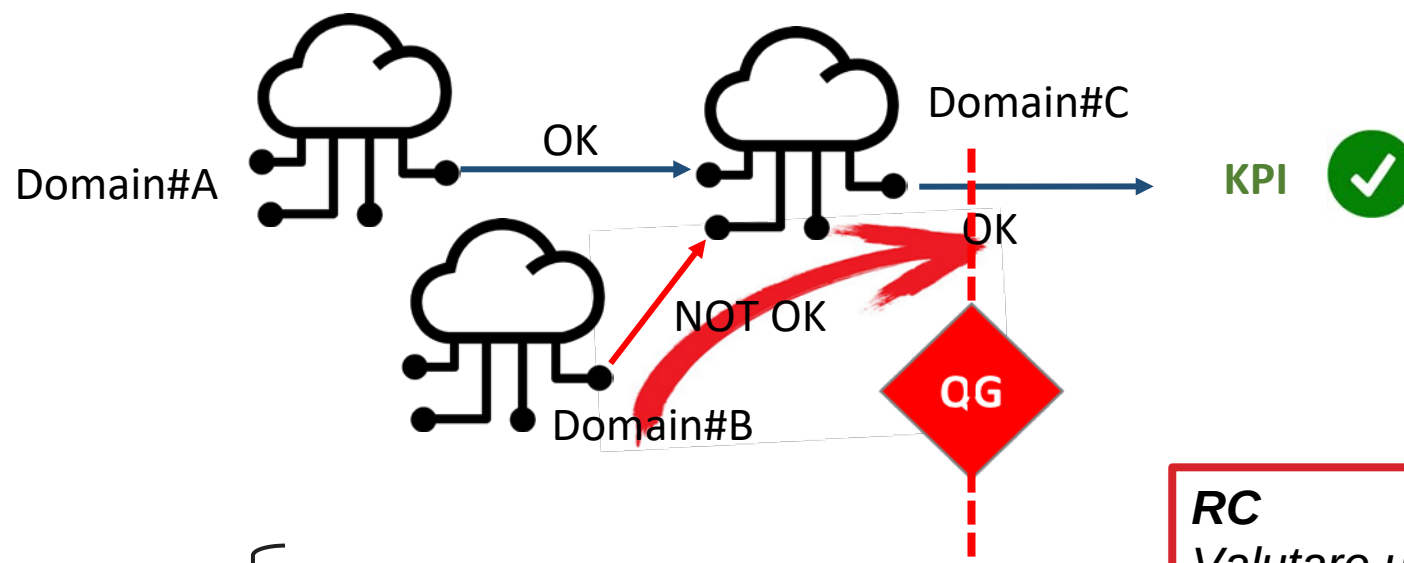




Use case#2: Multi Domain DW

Remediation

Trasmettere Metadato relativo a DQ e utilizzare la metrica per valutare se il dato e' Fit for Consumption



"KPI": {
 "data" : [#,#,..]
 "meta" : {
 "DQ#1" : #
 "DQ#2" : #
 }
}

RC

*Valutare un Dato con i propri Metadati
Concetto di DQ Inter Domain*



Use case#3: ERP

Scenario

DW alimentato da dati proveniente da ERP (es. SAP) con campi parziali

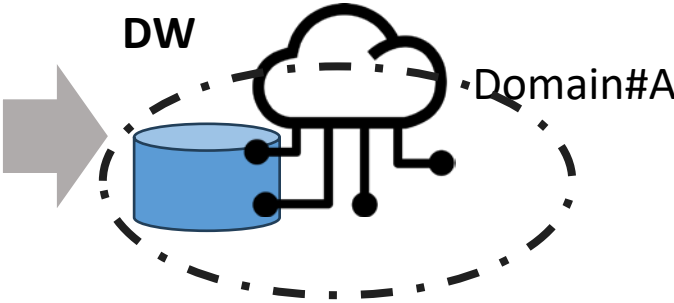
ERP

Create PM Notification: Malfunction Report

Notification: %00000000001 M2 Test

Notific. Status: OSNO

Order:



Data Issue

Completeness / Validity KPI

	student_id	lname	fname	course_id	grade
4	1	Salard	Jean	20	89
5	2	Le Bert	Antoi...	40	80
6	2	Le Bert	Antoi...	10	90
7	2	Le Bert	Antoi...	30	61
8	2	Le Bert	Antoi...	20	50
9	3	Legend...	Pierre	10	45
10	3	Legend...	Pierre	20	84
11	3	Legend...	Pierre	30	67
12	3	Legend...	Pierre	40	67
13	4	Turcotte	Cedric	10	67
14	4	Turcotte	Cedric	30	33
15	4	Turcotte	Cedric	40	77

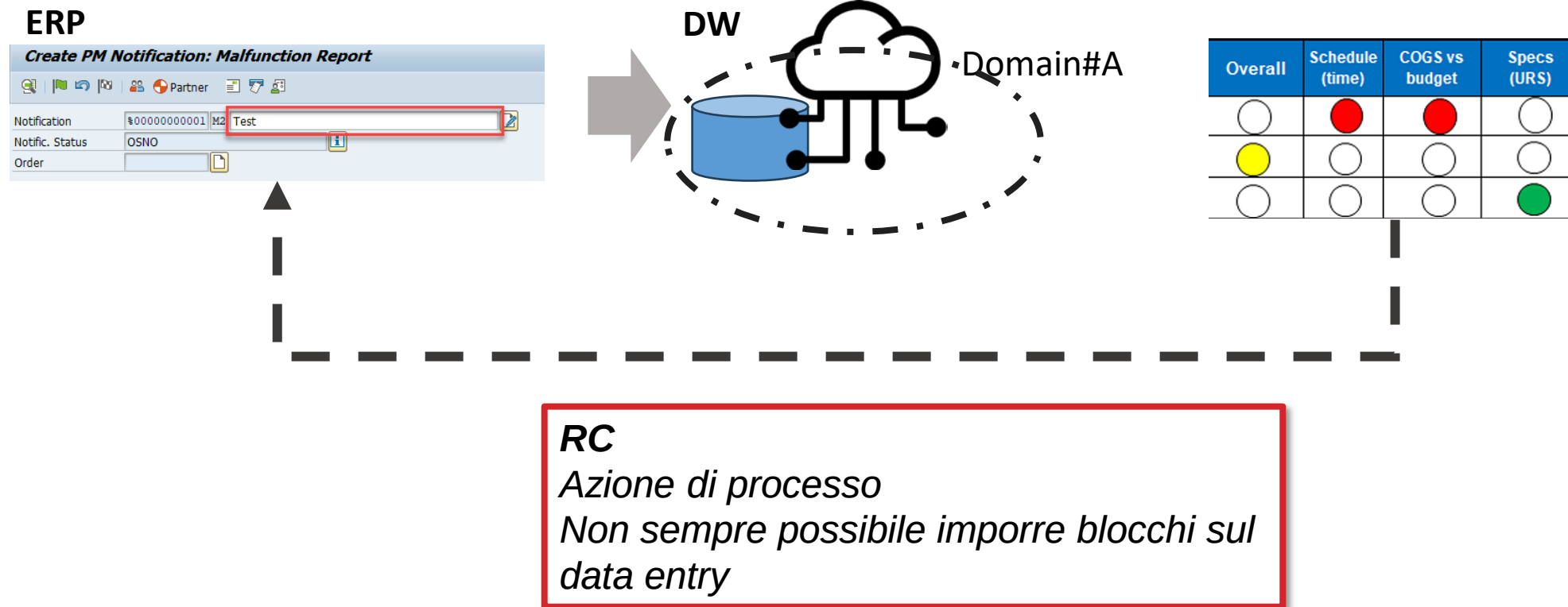
NOT OK

KPI ERROR

Use case#3: ERP

Remediation

Valutare DQ del KPI per stabilire la validita' dell'indicatore

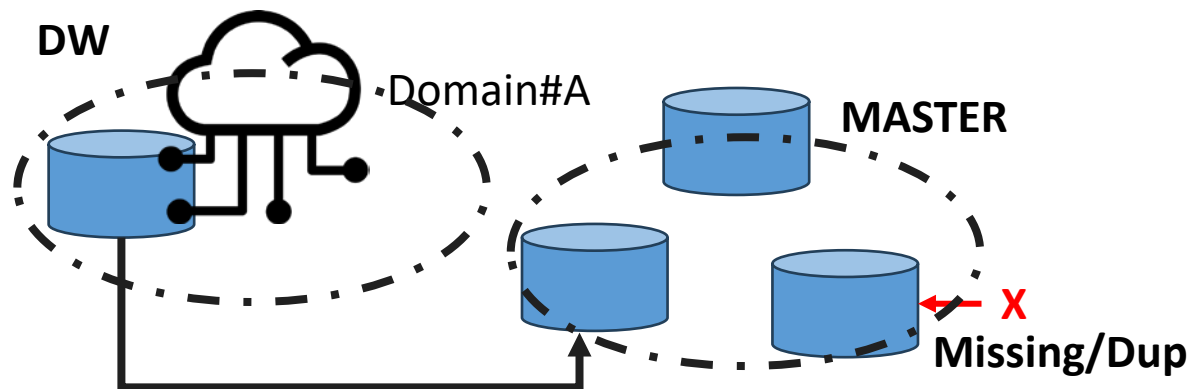




Use case#4: Master Data Management

Scenario

DW collegato (join) a dati Master con dati mancanti o duplicati (es CRM, anagrafica cliente)



Data Issue

Completeness / Validity KPI

	student_id	lname	fname	course_id	grade
4	1	Salard	Jean	20	89
5	2	Le Bert	Antoi...	40	80
6	2	Le Bert	Antoi...	10	90
7	2	Le Bert	Antoi...	30	61
8	2	Le Bert	Antoi...	20	50
9	3	Legend...	Pierre	10	45
10	3	Legend...	Pierre	20	84
11	3	Legend...	Pierre	30	67
12	3	Legend...	Pierre	40	67
13	4	Turcotte	Cedric	10	67
14	4	Turcotte	Cedric	30	33
15	4	Turcotte	Cedric	40	77

NOT OK

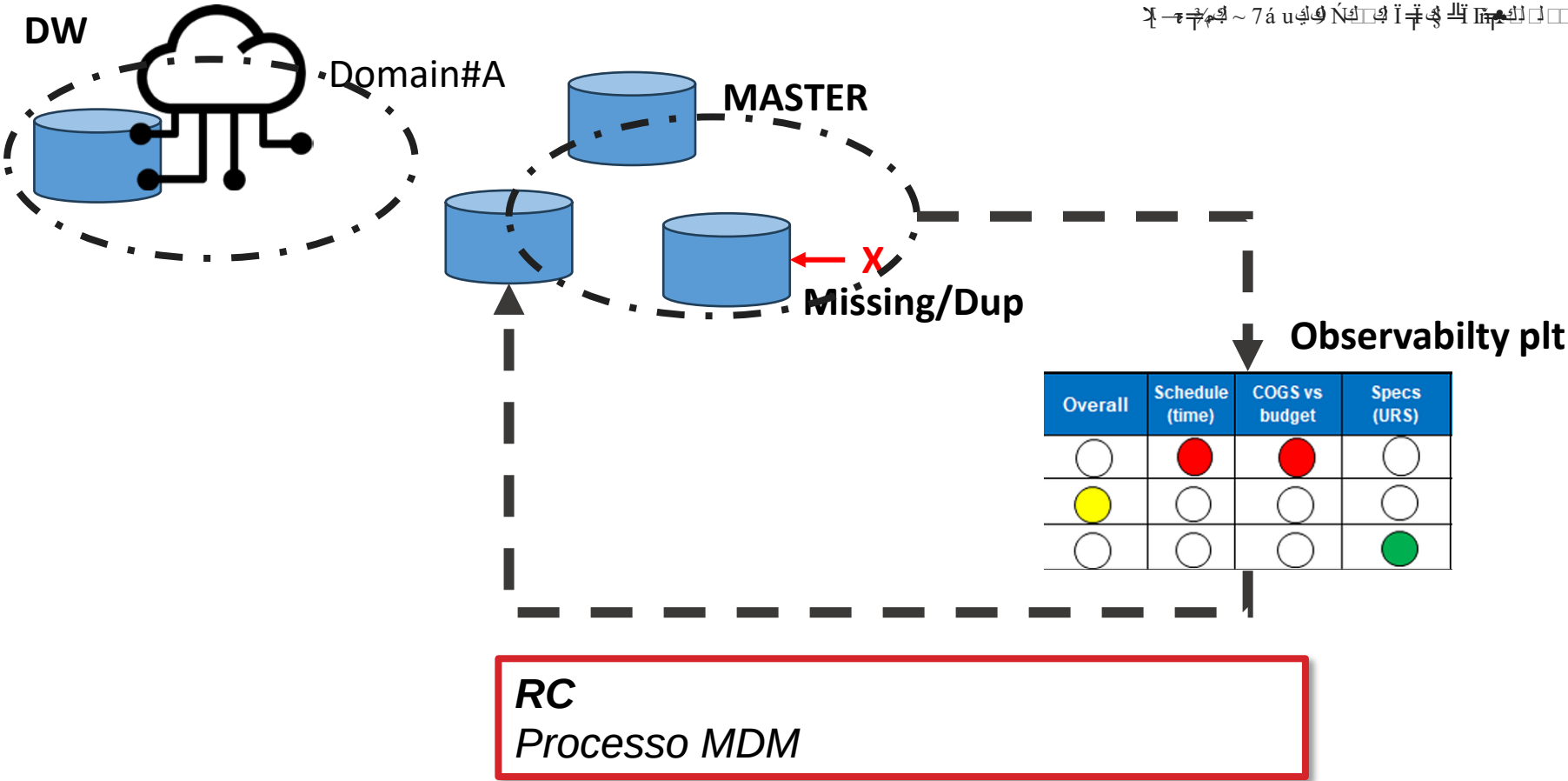
KPI 
ERROR



Use case#4: Master Data Management

Remediation

I Master Data sono dati critici by definition. Applicare processi di Master Data Management





Misura quantitativa

- *Selezione delle aree di miglioramento*
- *Indice Fit for Consumption*

Focus sui right data

Maggiore evidenza sulle aree critiche



pag. 11

Verso una cultura del dato

Business-Driven

I Data Project e la Data Culture in azienda a supporto delle iniziative di Business

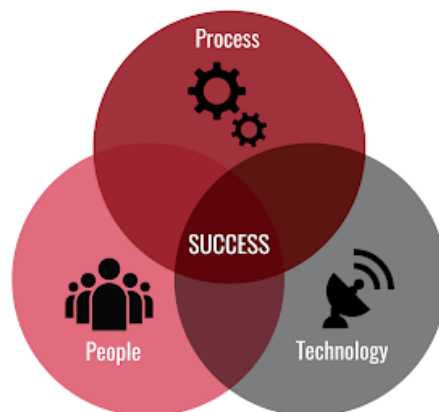
Framework based

Sfruttiamo i framework consolidati



I pillar in priorità

People, processes, tools





GRAZIE PER L'ATTENZIONE

ing. Fabio Gatti



fabiogattibox@gmail.com



<https://www.linkedin.com/in/fabiogatti1969/>