



Data Engineering Essentials: Governare il Dato per Creare Valore.

DataChain: dalla qualità alla tracciabilità del dato.

Ing. Fabio Gatti

Ordine degli Ingegneri della Provincia di Roma

Roma, 01 Luglio 25



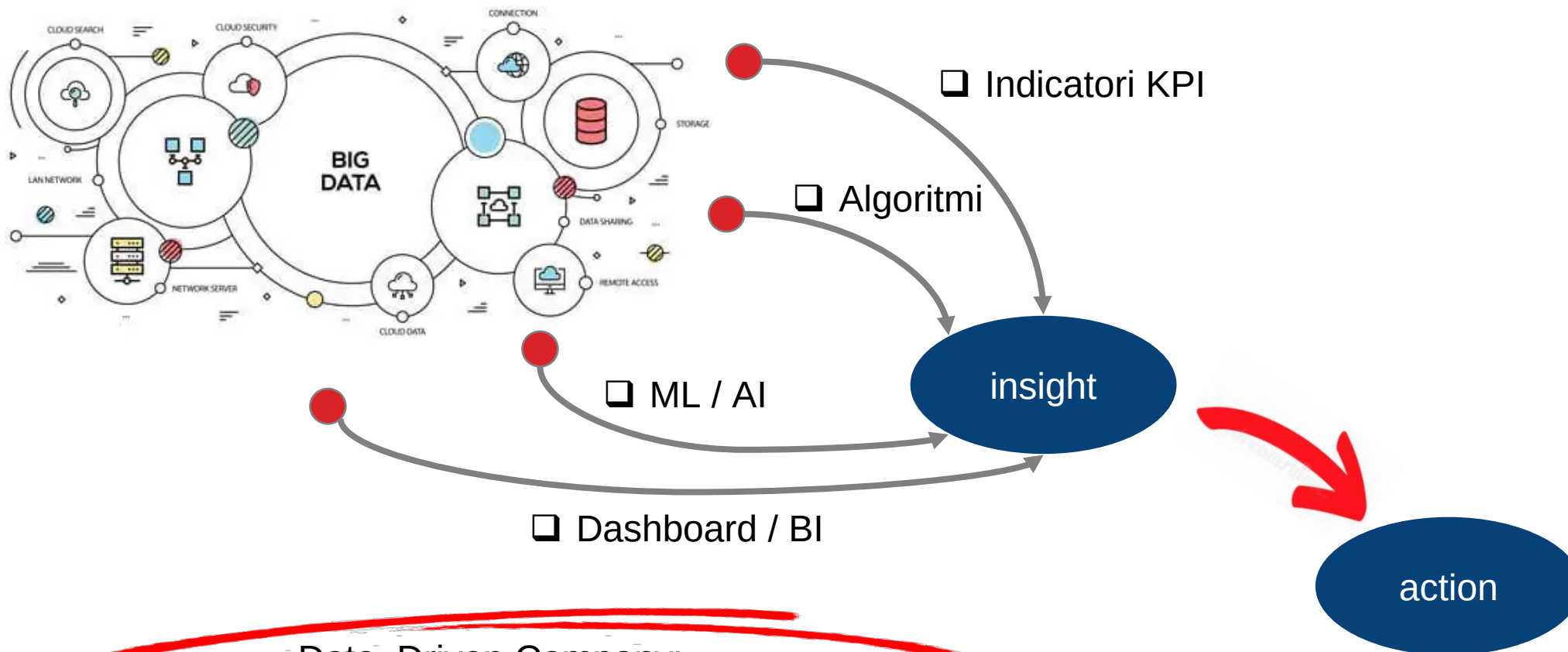
...decisioni Aziendali

Il Board della azienda X analizza il report PEAR contenente gli indicatori di processo KPI e li valuta per prendere alcune decisioni strategiche. Il report e' stato preparato dalle singole funzioni senza alcun controllo e auditing sul dato.

L'azienda Y utilizza un insieme di indicatori per valutare una opportunità di business. Gli indicatori mostrano un aumento nella difettosità del prodotto e un aumento nei rientri dal campo. I responsabili di prodotto decidono di non inserire il prodotto in portafoglio.

L'azienda Z sfrutta un algoritmo di ML per la predizione della efficienza della linea di produzione e utilizza i risultato per decidere un piano di acquisto di nuovi macchinari

Data-Driven Company

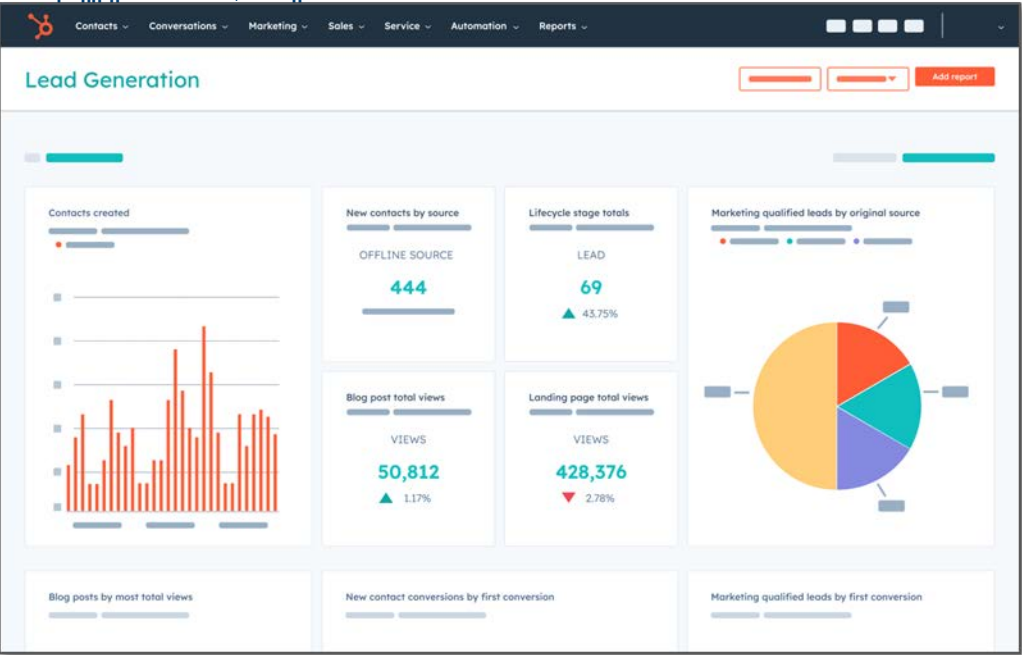


Data-Driven Company:
le decisioni vengono prese sulla base dei dati



Come appaiono i report

Et 7/4/10 13/49 I=IN -I=7K



çE • ¼-¾ ¾/¾

APPENDIX C – PROCESS EFFECTIVENESS ASSESSMENT REPORT (PEAR)

Auditing Company Name		Auditing Company Logo	
Organization:		Site:	DOB:
PEAR Number:	Audit Report Number:	Issue Date:	
Process Name:			
Process details, including associated process interfaces:			
Applicable 5100/5110/5120 clause(s):			
Organization's method for determining process effectiveness:			
Auditor observations and comments supporting process effectiveness determination:			
Statement of Effectiveness Level: The audited process is: ☐ 1. Not implemented; planned results are not achieved. ☐ 2. Implemented; planned results are not achieved and appropriate actions not taken. ☐ 3. Implemented; planned results are not achieved, but appropriate actions being taken. ☐ 4. Implemented; planned results are achieved.			
Auditor Name(s):		Auditor Representative Acknowledgement Name:	
Signature(s):		Signature(s):	

Et 7/4/10 13/49 I=IN -I=7K



¿İçinde Nereye Gidilecek?

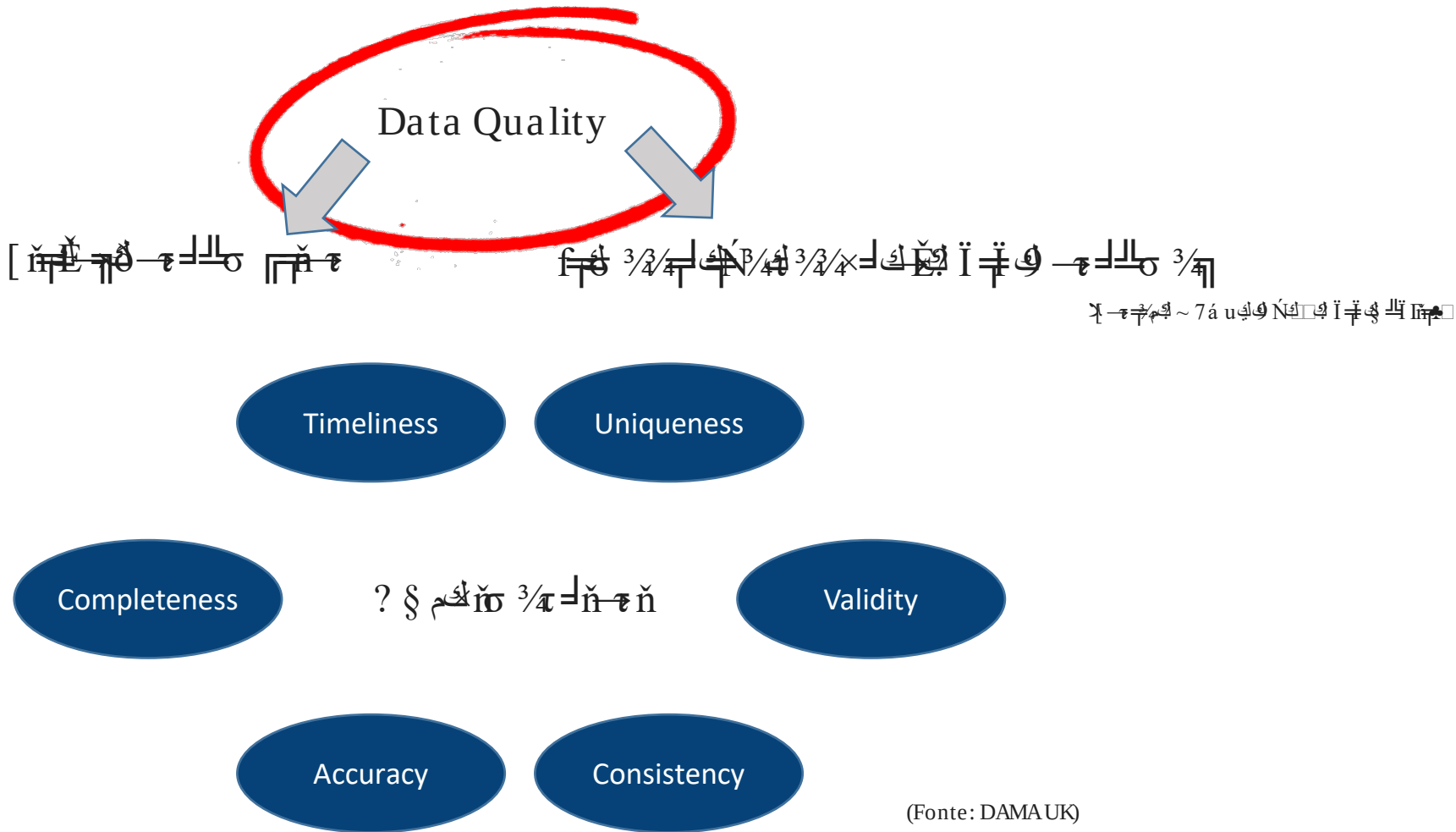


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Eliminate missing data from a sample														
2															
3		Poverty	White	Crime	Doctors	University	Income			Poverty	White	Crime	Doctors	University	Income
4	Alabama	15.7	71.0	448	218.2	22.0	42,666		Alabama	15.7	71.02718	448	218.1734	22	42666
5	Alaska	8.4	70.6	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	482	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	68599
9	Colorado		89.7	348	259.7	35.6	56,993		Delaware	10	74.26606	689.2	250.9418	27.5	57988
10	Connecticut	9.3	84.3	256	376.4	35.6	68,595		Florida	13.2	79.81107	722.6	247.8636	25.8	47777
11	Delaware	10.0	74.3	689	250.9	27.5	57,989		Georgia	14.7	65.38772	493.2	217.4457	27.5	50866
12	Florida	13.2	79.8	723	247.9	25.8	47,178		Hawaii	9.1	29.66733	272.8	316.9829	29.1	67211
13	Georgia	14.7	65.4	493	217.4	27.5	50,861		Illinois	12.2	79.13311	533.2	280.1539	29.9	56233
14	Hawaii	9.1	29.7	273	317.0	29.1	67,214		Iowa	11.5	94.17046	294.7	189.2832	24.3	48988
15	Idaho		94.6	239	168.8	24.0	47,576		Kansas	11.3	88.70372	452.7	222.5477	29.6	50177
16	Illinois	12.2	79.1	533	280.2	29.9	56,235		Kentucky	17.3	89.90433	295	232.3485	19.7	41533
17	Indiana	14.1		216.9			47,966		Louisiana	17.3	64.83954	729.5	262.6615	20.3	43733
18	Iowa	11.5	94.2	235	189.3	24.3	48,940		Maine	12.3	90.38985	118	278.3948	25.4	46588
19	Kansas	11.3	88.7	453	222.5	29.6	50,117		#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
20	Kentucky	17.3	89.9	295	232.3	19.7	41,533		#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
21	Louisiana	17.3	64.8	730	262.7	20.3	43,733		#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
22	Maine	12.3	96.4	118	278.4	25.4	46,581		#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
23															
24									Number of rows without missing data				15		





La Qualità del dato: fondamenti



Data Quality: come iniziare

- ❑ Framework: DAMA, DGI, EDM, AIM, SAM, ISO8000 ...
- ❑ Processi: Definizione obiettivi, Assessment, Adozione di Standard, Piano implementativo, definizione Ruoli e Responsabilita'
- ❑ Working Group: IAQG, Kaggle, ...
- ❑ Tecnologia: ETL, Pandas, OpenMetadata,...



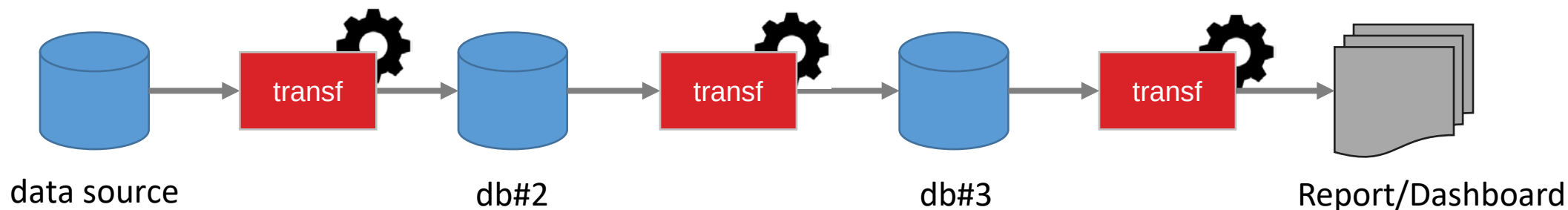
In che ordine?

1. People

2. Process

3. Tool

Il salto di paradigma: dalla qualità alla tracciabilità



Perché la qualità non basta

- ☐ Compliance (GDPR, BCBS239)
- ☐ Auditing
- ☐ Change Management
- ☐ Transparency

Casi concreti in cui la tracciabilità fa la differenza

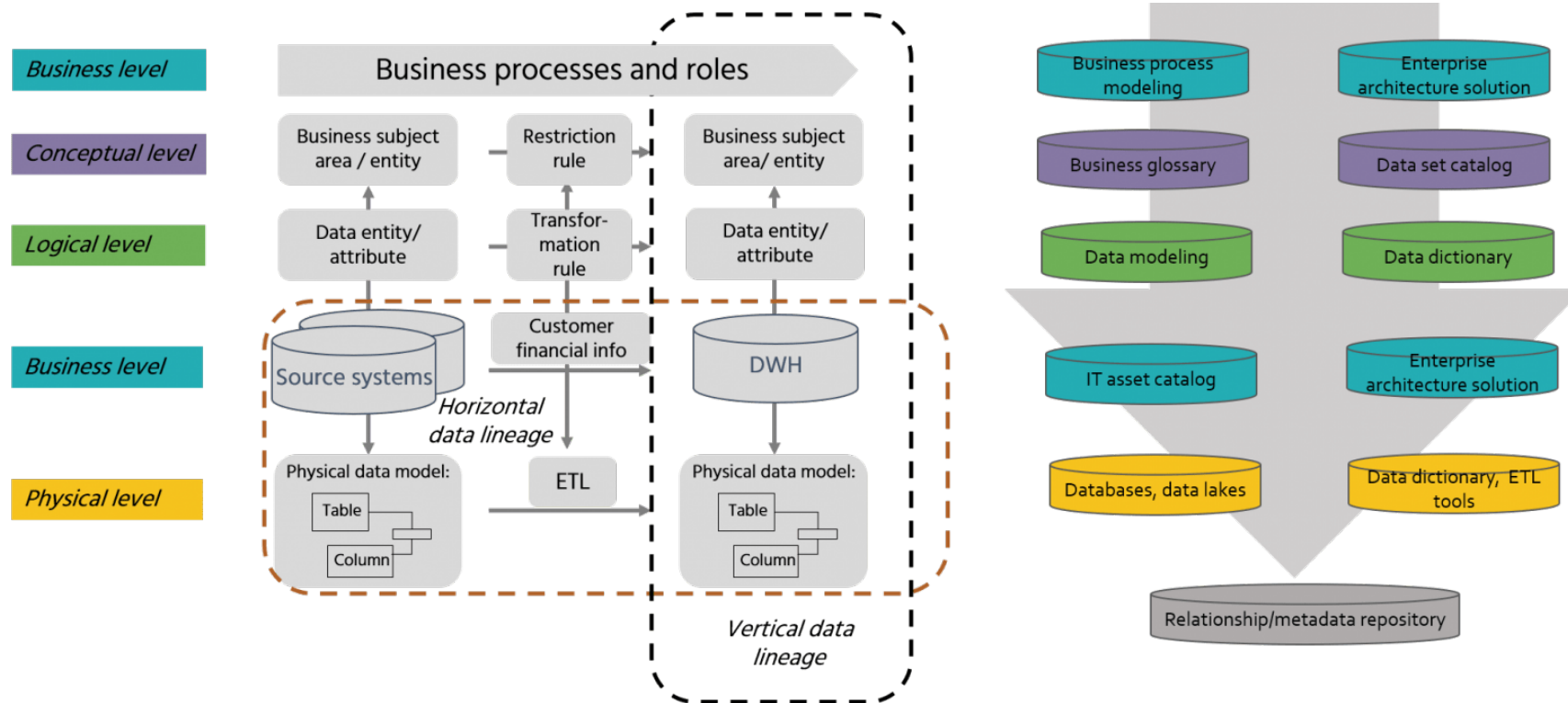
Audit,
Contenziosi,
AI explainability



La tracciabilità del dato: principi e architetture

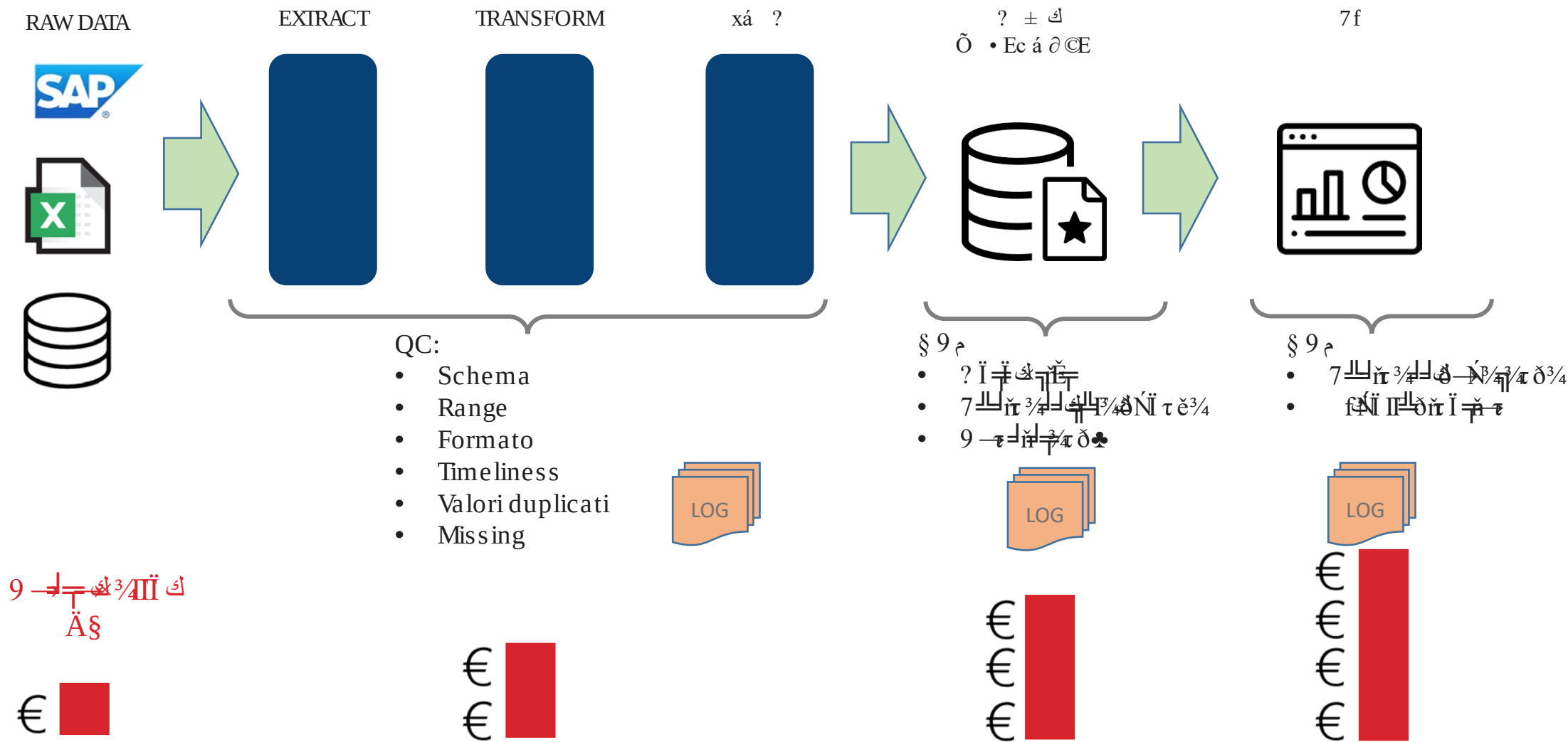
Data Lineage is a description of the pathway from the data source to their current location and the alterations made to the data along the pathway.

(DAMA Dictionary of DM, 2nd Ed)





Datachain: controllo Qualità e lineage

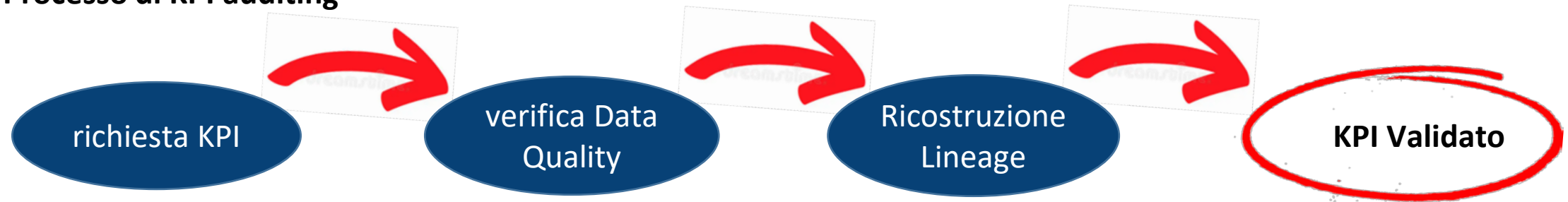




Data Governance in azione: modello operativo



Processo di KPI auditing



Takeaway

DATA LINEAGE CON CONTROLLO QUALITÀ



- ❑ La qualità del dato e' **necessaria** ma **non sufficiente**
- ❑ La **tracciabilità** rende il dato auditabile e affidabile
- ❑ Serve un modello organizzativo e tecnologico **coerente**

**“In God we trust.
All others must
bring data”.**

W. Edwards Deming





GRAZIE PER L'ATTENZIONE

ing. Fabio Gatti



fabiogattibox@gmail.com



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