

Carlo Batini

Università di Milano Bicocca

Dipartimento di Informatica, Sistemistica e Comunicazione

batini@disco.unimib.it

La Qualità delle Informazioni

Cosa sono i *dati* e le *informazioni*?

Un **dato** e' una discriminazione tra stati fisici delle cose (es. la colonnina di un termometro, i valori 36,5 38,9...), che puo' o meno trasmettere una **informazione** su un fenomeno o evento a una persona.

La trasmette o no, a seconda della **conoscenza** a priori che quella persona ha del fenomeno o evento.



Il valore **38,5** della colonnina del mercurio non fornisce nessuna informazione a una persona che non ha il concetto di temperatura (es. un bambino). Se la persona:

- conosce il concetto di **temperatura** in gradi **Celsius**, e
 - si fida della **precisione** del termometro,
- allora arriva alla conclusione: **"Hai la febbre!"**.

L'informazione come rappresentazione della realtà



Strutture linguistiche
e/o basate sui sensi



Processo di rappresentazione



Ho visto un bellissimo
tramonto in montagna,
era una giornata con
poche nuvole....



In quanti modi si puo' rappresentare Portofino

Place	Country	Population	Main economic activity
Portofino	Italy	7.000	Tourism



Dati strutturati

Immagine

Mappa

Portofino

Testo

Dear Laure, I try to describe the wonderful harbour of Portofino as I have seen this morning a boat is going in, other boats are along the wharf. Small pretty buildings and villas are looking on to the harbour.

Forza lavoro in Francia



DEPARTMENTS	QUANTITIES (000)				PROPORTION	
	A	I	S	TOTAL	A	I
1 AIN	67	43	40	150	45	38
2 AISNE	96	71	66	193	29	37
3 ALLIER	65	46	57	167	39	37
4 Hautes ALPES	15	6	12	33	43	34
5 Hautes ALPES	16	0	13	37	44	31
6 ALPES Maritimes	31	61	122	214	14	39
7 ARDECHES	48	32	25	105	46	47
8 ARDENNES	25	55	35	115	22	47
9 ARIEGES	33	17	14	64	52	26
10 AUBE	28	46	35	112	25	43
11 AUDE	30	20	32	102	46	19
12 AVEYRON	70	32	39	151	54	24
13 BOUCHES DU RH.	42	145	226	412	10	35
14 CALVADOS	70	56	69	194	36	28
15 CANTAL	45	15	20	70	55	16
16 CHARENTE	65	38	38	140	47	26
17 CHARENTE M.	79	39	65	183	43	21
18 CHER	43	41	35	120	36	34
19 CORREZE	64	23	29	116	55	20
21 COTE D'OR	43	41	59	143	30	20
22 COTES DU NORD	131	35	62	228	58	15
23 CREUSE	38	13	17	68	65	15
24 DORDOGNE	104	34	41	179	59	10
25 DOUBS	35	87	39	142	25	47
26 DROME	46	36	35	119	39	31
27 EURE	48	52	45	145	33	36
28 EURE & LOIR	44	27	38	110	41	28
29 FINISTERE	164	76	89	329	50	30
30 GARD	40	51	52	144	28	36
31 HAUTE GARONNE	64	87	84	235	30	31
32 GERS	63	19	16	98	71	11
33 GERONDE	115	107	179	392	30	27
34 HERAULT	62	40	71	173	36	23
35 ILLE & V.	137	60	82	279	49	21
36 INDE	54	30	32	116	46	28
37 INDE & L.	61	41	55	157	39	26
38 ISERE	66	136	78	282	24	46
39 JURA	39	34	27	100	39	34
40 LANDES	70	25	20	125	57	20
41 LOIR & CHER	51	27	30	108	47	25
42 LOIRE	56	160	82	298	19	54
43 Haute LOIRE	52	23	22	97	54	24
44 LOIRE INF.	101	108	105	314	32	34
45 LOIRET	51	51	54	156	32	33
46 LOT	41	20	15	77	61	15
47 LOT & GAR.	70	24	30	124	57	19
48 LOZERE	22	5	7	34	64	15
49 MAINE & L.	104	65	55	224	44	28
50 MANCHE	115	42	36	214	54	20
51 MARNE	44	37	67	148	28	34
52 Haute MARNE	25	26	26	77	31	35
53 MAYENNE	74	23	28	125	59	19
54 MEURTHE & M.	23	127	91	241	9	33
55 MEUSE	24	31	27	82	30	36
56 MORRHAN	132	47	59	238	55	20
57 MOSELLE	35	173	94	302	12	57
58 NIÈVRE	34	27	33	94	36	28
59 NORD	61	483	296	860	9	56
60 OISE	40	69	55	164	24	43
61 ORNE	65	30	35	130	50	23
62 P. D. C.	94	242	137	473	20	51
63 PUY DE DOME	60	79	63	202	36	30
64 Hautes PYRENEES	60	49	62	171	42	35
65 Hautes PYRENEES	37	27	28	92	40	38
66 PYRENEES ORIENT.	35	20	33	88	40	39
67 BAS-RHIN	76	122	114	312	24	38
68 Haute-RHIN	40	131	74	245	17	51
69 RHONE	44	815	194	1033	10	47
70 Haute SAONE	34	32	23	89	38	36
71 SAONE & L.	64	77	62	203	41	33
72 SARTHE	87	45	58	190	46	24
73 SAVOIE	44	38	35	117	38	39
74 Haute SAVOIE	32	42	45	120	37	30
75 PARIS	8	875	940	1823	0	38
76 SEINE	8	874	880	1762	1	51
77 SEINE INF.	75	132	174	401	19	38
78 SEINE & M.	37	72	76	185	20	39
79 SEINE & O.	40	336	306	730	6	45
80 SEINE-ET-M.	71	29	33	133	53	23
81 SOMME	57	66	61	184	31	36
82 TARN	55	47	33	135	41	35
83 TARN & G.	44	15	18	77	58	11
84 TARN	35	30	31	96	36	37
85 VAUCLUSE	40	30	41	111	36	38
86 VANDIEU	110	26	43	186	58	20
87 Vienne	55	29	38	122	47	28
88 Haute Vienne	33	47	45	125	41	30
89 Vienne	56	66	49	171	33	34

I. Per dipartimento

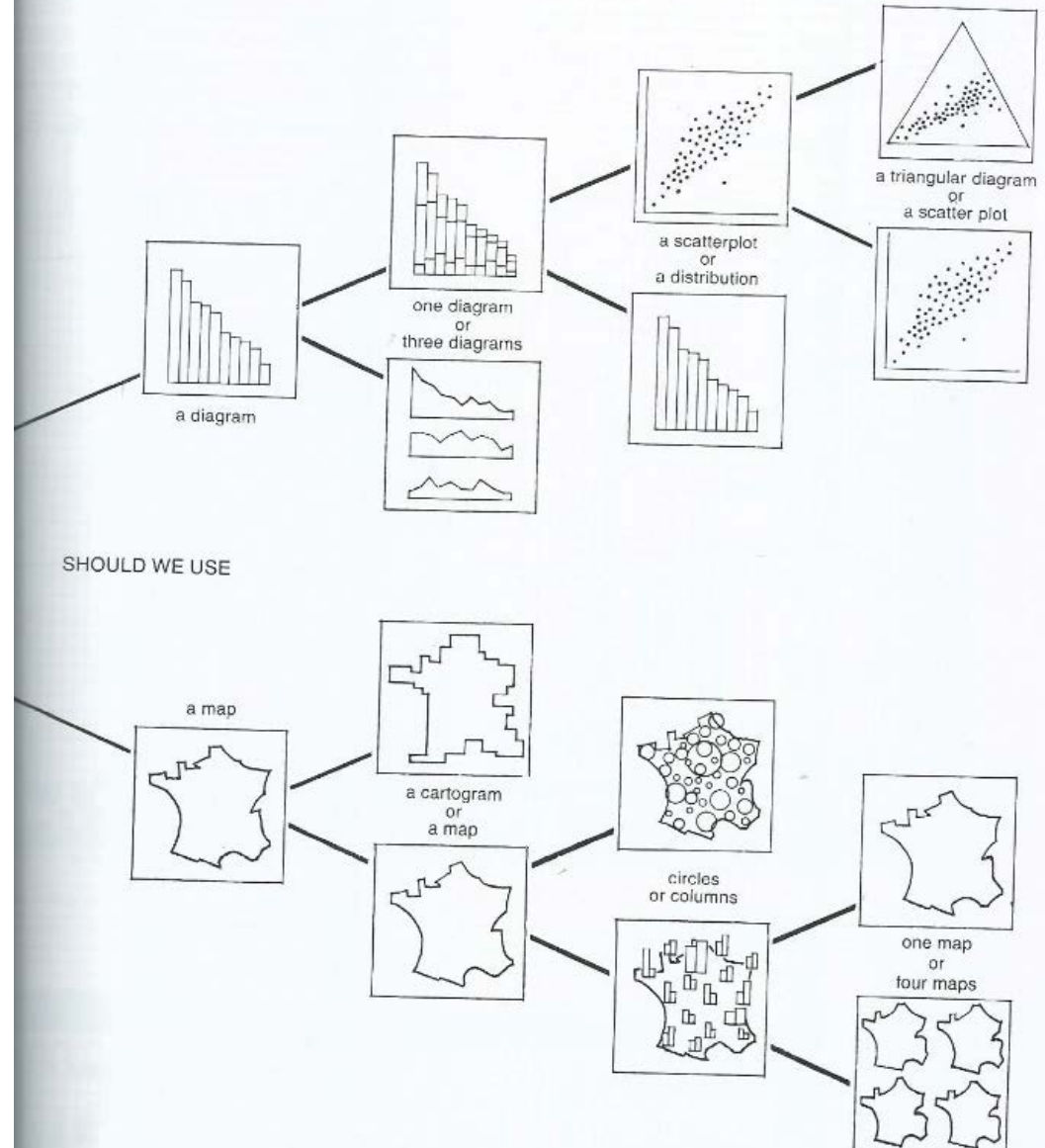
1. Agricoltura
2. Industria
3. Servizi

II. Percentuale sul totale

1. Agricoltura
2. Industria
3. Servizi

Cento differenti grafici per la stessa informazione...

DEPARTMENTS	QUANTITIES (000)			PROPORTION	
	I	II	Total	I	II
1 AIN	89	43	132	45	22
2 AISNE	18	11	29	29	17
3 ALLIER	15	48	63	39	27
4 ALPE ALPES	15	8	23	43	34
5 ALPE HES	18	8	26	44	31
6 ALPES MONT	31	81	112	24	29
7 ANDECH	48	39	87	45	31
8 ARDENNES	35	53	88	32	41
9 ARIEGE	33	17	50	66	34
10 AUDE	28	48	76	12	43
11 AUDE	50	20	70	102	40
12 AVEYRON	70	32	102	14	34
13 BORDOIS DU RH.	42	143	285	10	39
14 CALVADOS	70	55	125	86	38
15 CANTAL	45	13	58	10	18
16 CHARENTE	65	36	101	47	38
17 CHARENTE MME	70	39	109	183	43
18 CHER	45	41	86	120	30
19 CORREZE	84	25	109	115	53
20 COTE D'OR	43	41	84	143	30
21 COTES DU NORD	101	25	126	228	58
22 CREUSE	38	13	51	88	19
23 DORDOGNE	104	84	188	179	38
24 DOUBS	35	47	82	142	25
25 DROME	45	30	75	119	39
26 EURE	48	59	107	145	35
27 EURE & LOIR	44	27	71	110	41
28 FINISTERE	104	10	114	320	50
29 GARD	40	51	91	144	28
30 GERS	64	21	85	216	30
31 HAUTE GARONNE	63	10	73	88	71
32 HERAULT	115	197	312	302	31
33 ILE A V.	42	40	82	179	36
34 ILE A V.	137	80	217	279	49
35 INDRE	34	30	64	116	40
36 INDRE & L.	41	41	82	107	39
37 ISERE	48	158	206	282	24
38 JURA	39	34	73	100	39
39 LANDER	70	35	105	128	57
40 LOIR & CHER	51	27	78	108	47
41 LOIRE	36	160	196	288	19
42 LOIRE	52	23	75	87	54
43 LOIRE INF.	101	168	269	314	32
44 LOIRET	51	51	102	156	32
45 LOT	42	10	52	67	13
46 LOT & GAR.	70	24	94	124	57
47 LOZERE	22	5	27	34	64
48 MAINE & L.	104	40	144	224	44
49 MANCHE	118	42	160	214	34
50 MARNE	44	57	101	168	30
51 MARNE	25	28	53	81	31
52 MAYENNE	74	23	97	125	59
53 MEURTHE & M.	23	127	150	242	9
54 MEUSE	24	31	55	82	36
55 MORRHAN	132	47	179	238	55
56 MOSELLE	36	173	209	302	12
57 NIEVRE	34	17	51	94	38
58 NORD	81	483	564	860	9
59 OISE	40	59	99	164	24
60 ORNE	65	39	104	130	50
61 P. D. C.	84	242	326	478	20
62 PUY DE DOME	80	79	159	222	38
63 PYRENEES	80	49	129	191	43
64 PYRENEES ORIENT.	37	27	64	92	40
65 BAS-RHIN	35	20	55	86	40
66 BAS-RHIN	76	122	198	312	24
67 BAS-RHIN	40	121	161	235	17
68 RHONE	44	215	259	433	10
69 SAONE	34	32	66	89	38
70 SAONE & L.	94	71	165	233	41
71 SARTHE	87	43	130	190	46
72 SAVOIE	44	38	82	117	38
73 SAVOIE	32	42	74	130	37
74 PARIS	8	375	383	1117	0
75 SEINE	8	374	382	1132	1
76 SEINE INF.	75	152	227	401	18
77 SEINE & M.	37	72	109	165	20
78 SEINE & O.	46	338	384	730	6
79 SEINE-ET-LOIRE	71	39	110	138	31
80 SOMME	57	68	125	186	31
81 TARN	55	47	102	135	41
82 TARN & G.	44	13	57	75	17
83 VAR	39	60	99	104	30
84 VAUCLUSE	40	39	79	111	36
85 VANDER	118	78	196	181	39
86 VANDER	80	28	108	128	47
87 VANDER	84	47	131	148	51



Ogni giorno incontriamo tanti casi
di scarsa qualità della informazione

Israelitico, truffa da 15 milioni Indagato Mastrapasqua (ex Inps)

Al centro dell'inchiesta depistaggi, cartelle cliniche falsificate e fatture gonfiate. Una visita da 250 euro spacciata per un intervento da duemila. Perquisiti l'ospedale, la Regione e due Asl

di Lavinia Di Gianvito e Ilaria Sacchettoni



ROMA - Una visita in ambulatorio per un alluce valgo, pagata dalla Regione Lazio 250 euro, contrabbandata per un intervento complesso per la quale l'Israelitico avrebbe chiesto anche duemila euro. E l'alterazione non avrebbe riguardato solo le prestazioni ortopediche, ma anche quelle geriatriche e oncologiche, fino a derubare la sanità di almeno 15 milioni di euro. Sarebbe questa l'entità del raggio - ma i calcoli sono ancora in corso - per la quale questa mattina sono scattate le perquisizioni dei carabinieri del Nas nella struttura sanitaria e le acquisizioni di documenti nelle sedi della Regione e delle Asl RmA e RmD. Dieci gli indagati per truffa al servizio sanitario nazionale e falso: il direttore generale Antonio Mastrapasqua, ex presidente dell'Inps (accusato solo di falso), due funzionari della Regione e sette dipendenti dell'ospedale, tra cui il direttore sanitario Luigi Antonio Spinelli e il vice direttore amministrativo Tiziana D'Agostino.

Corriere
della Sera
29/9/2014

Economist
15/10/2015

China's economy

Rate cut shows that even China's government doesn't believe its own data

Oct 23rd 2015, 18:20 BY S.R.



264



23

IF GROWTH is so strong, why did the Chinese central bank feel the need to cut interest rates again? It is only reasonable to ask this after the People's Bank of China cut rates on October 23rd, the sixth time in a year, in the same week that the government said the economy grew 6.9% year-on-year in the third quarter, ahead of market expectations. The answer lies in the question: growth is not as strong as official data suggest, and the government knows that.

Nominal growth has slowed to 6.2%, the weakest since 1999, translating directly into weak cash flow for companies that already face heavy debt burdens. The industrial sector has been especially hard hit, with the country's northern rustbelt on the brink of a recession. And more than a percentage point of overall growth this year has stemmed from activity in the financial sector, a contribution that will fall away quickly in the wake of this summer's stockmarket collapse. Many analysts reckon real growth is closer to 5-6%, well shy of the government's 7% target. More than three years' worth of producer price deflation tilt risks to the downside.

Il Giornale
24/10/
2015

Centinaia di dirigenti e funzionari dell'Inps sono al centro di un'indagine della Procura di Nocera Inferiore (Salerno) sull'erogazione di premi di produttività nel biennio 2012-2013.



Per raccogliere elementi utili all'inchiesta, carabinieri del reparto territoriale di Nocera Inferiore, con la collaborazione di quelli territorialmente competenti e del Ris di Roma, hanno eseguito a Roma un decreto di perquisizione presso la Direzione generale dell'Inps e in tutti i capoluoghi di regione un ordine di esibizione di documenti presso le direzioni regionali dello stesso istituto. L'Inps, in una nota, ha assicurato la "piena collaborazione con la magistratura e le forze dell'ordine", mentre il presidente dell'Anticorruzione, Raffaele Cantone, ha detto di ritenere che si tratti, come per altre vicende, di uno "spaccato inquietante dello stato della Pubblica amministrazione".

Secondo l'ipotesi investigativa, gli incentivi di produttività sarebbero stati concessi sulla base di ispezioni "gonfiate": nel database dell'Istituto sarebbero stati inseriti dati falsi, maggiorati rispetto a quelli reali, relativamente ai risultati delle ispezioni effettuate nelle aziende e ai rapporti di lavoro annullati. In questo modo sarebbero stati alterati i risultati delle "performance", cioè i parametri degli obiettivi indicati dalla direzione generale per la corresponsione degli incentivi. I dati attuali sono ancora parziali ma solo per gli anni 2012 e

Lo scandalo dei ticket fantasma “Nel Lazio truffa da 100 milioni”

Repub-
blica
5/11/2015

RICEVUTA n. 11012010
Data 11/01/2010
Importo 50.00
Paga di spesa al...
Totale 50.00
Pagato € 50.00

Nicola Zingaretti, commissario di governo per la Sanità regionale, «di presentare denuncia contro quanti, abusando della buona fede dei cittadini e nostra, hanno intascato soldi dai primi e dalla Regione concorrendo a dilapidare le risorse destinate alla salute di tutti».

TOTALE 50.00
Da Pagare € 50.00

Importo evaso: € 388,39
Interessi: € 34,33
Spese amm.: € 25,00
Importo dovuto: € 447,72

I DOCUMENTI
La fattura pagata dal paziente (50 euro). A destra l'accertamento inviato dalla Regione: 388,39 euro per prestazioni mai eseguite

REGIONE LAZIO
Certificazione stato pratica
Data certificazione: 18/11/2015
Situazione familiare
Anno 2009
Tipo famiglia: 2
Importo complessivo: € 22.850,00
Prestazioni fruite in esenzione
Importo evaso: € 388,39
Interessi: € 34,33
Spese amm.: € 25,00
Importo dovuto: € 447,72

Il Lazio ha avviato da una settimana una maxi-operazione per recuperare l'evasione del ticket dovuto per visite, analisi ed esami diagnostici. Sono già 25 mila i cittadini che, ricevuta una lettera di "avviso bonario", (le "cartelle pazze" per dirla con l'opposizione di centrodestra che contesta l'operazione), han-

Corriere della Sera 14/11/2015

Inconveniente in aeroporto

Linate, l'umidità fa scattare l'allerta sicurezza

Il volo Milano-Napoli AZ1285 ha fermato il rullaggio in pista su ordine della torre di controllo. Ore 10 di ieri mattina, aeroporto di Linate.

Un allarme sicurezza — quasi subito rientrato — ha preoccupato i passeggeri in partenza dallo scalo milanese. Dal velivolo sono partiti telefonate e cinguettii su Twitter. Sulla pista sono entrati i vigili del fuoco e i responsabili delle procedure di emergenza. È stato ispezionato l'aereo: tutto a posto, un falso allarme.

Ma la causa del ritardo resta comunque anomala. L'umidità accumulata sulla pista ha «alterato» la sensibilità dei segnali di sicurezza. Il sistema informatico ha percepito — sbagliando — una presenza estranea.

© RIPRODUZIONE RISERVATA

Un esempio di scarsa qualità: il 3/3/2013 sul sito on-line di Repubblica..

27ª Giornata — 03/03/2013

Catania 2 • 4 Inter

MARCHESE (19° 1T) BERGESSIO (7° 1T) PALACIO (48° 2T) PALACIO (47° 2T) PALACIO (25° 2T) ALVAREZ (7° 2T)

CATANIA 90

INTER

TABELLINO

Catania	Inter
21 ANDUJAR	1 HANDANOVIC
5 ROLIN	40 JUAN JESUS
12 MARCHESE	31 PEREIRA
3 SPOLLI	26 CHIVU
22 ALVAREZ	11 ALVAREZ
27 BIAGIANTI	14 GUARIN
13 IZCO	4 ZANETTI
10 LODI	22 KUZMANOVIC
19 CASTRO	21 GARGANO
17 GOMEZ	7 SCHELOTTO

COMMENTO IN DIRETTA di Pasquale Notargiacomo

Primo Tempo **Secondo Tempo**

49° FINISCE LA PARTITA! Vince l'Inter 3-2

48° IL GOL. Cambiasso in area, vicino alla linea di fondo, serve un assist perfetto per il rasoterra in area di Palacio che completa la rimonta

47° INTER IN VANTAGGIO!!!! 3-2

46° Destro di Almiron da fuori area: potente ma impreciso

45° Saranno quattro i minuti di recupero

44° L'INTER SFIORA IL TERZO GOL! Diagonale riavvicinato di Cambiasso in area, fuori di

DIRETTE

Calcio - Serie A

03.03 Torino - Palermo	0 - 0	12:30
03.03 Bologna - Cagliari	2 - 0	15:00
03.03 Catania - Inter	2 - 2	15:00
03.03 Fiorentina - Chievo	2 - 1	15:00
03.03 Pescara - Udinese	0 - 1	15:00
03.03 Sampdoria - Parma	1 - 0	15:00
03.03 Siena - Atalanta	0 - 2	15:00
03.03 Roma - Genoa	0 - 0	20:45

Legenda: ⌚ In programma 👁 Diretta 🛑 Finita

RISULTATI

Serie A **Marcatori**

Di cosa
non mi occupo
in questo
seminario...



Cosa è la qualità della informazione?
La intendiamo tutti nello stesso modo?

Quale di queste due immagini e' di migliore qualità?



Queste sono due foto di Marte.....

Istintivamente, quale è di maggiore qualità?



www.hoax-slayer.com



astrobiology.nasa.gov

La prima immagine è stata scaricata da un blog amatoriale, la seconda dal sito della Nasa.

Le vostre valutazioni sono probabilmente basate sul vostro “modello di Marte”. Ora che sapete da quali siti provengono, pensate di cambiare opinione?

Tanti approcci alla qualità

- La qualità della informazione è **oggettiva**
- La qualità della informazione è **soggettiva**
- La qualità **dipende** dallo **scopo** per cui la si usa

La qualità oggettiva

Esempio di (scarsa) qualità: Film, Registi, Remakes

Ci sono altri errori?

Id	Title	Director	Year	#Remakes	LastRemakeYear
1	Casablanca	Weir	1942	3	1940
2	Dead Poets Society	Curtiz	1989	0	NULL
3	Rman Holiday	Wylder	1953	0	NULL
4	Sabrina	NULL	1964	0	1985

Inaccurato

Incompleto

Inconsistente

La qualità soggettiva

La stessa esibizione della danzatrice
e' *percepita* diversamente.....



Esempio di percezione soggettiva



Scarsa qualità



Stessa qualità?

La qualità soggettiva di un libro

Recensioni clienti



1

5,0 su 5 stelle



Semiology of Graphics: Diagrams, Networks, Maps

da Jacques Bertin

Formato: Copertina rigida | [Cambia](#)

Prezzo: 68,37 € Prime

★★★★★ **Amazing book**

Da [Amazon Customer](#) il 1 luglio 2013

Formato: Copertina rigida | **Acquisto verificato**

Jacques Bertin's book provides an exhaustive and perceptive analysis of the semiology of diagrams and maps. I highly recommend it to anyone interested in this topic.

► [Commento](#) | Ti è stata utile questa recensione? [Segnala un abuso](#)

Qualità che dipende
dallo scopo:
esempio personale

- **Operazione di
protesi all'anca**
- **Dolori**
- **Radiografia →**



Qualità che dipende dall'uso: esempio personale

- **Radiografia**

- Esito: si vede una linea più chiara
- sospetta frattura dell'osso della gamba
in corrispondenza dell' attacco della protesi

- Parere dell'esperto radiologo: al 90% non è
frattura, sembra più una vena, ma per la certezza
serve la **risonanza magnetica**, è più precisa...

- **Risonanza** → Non c'è frattura

Quante dimensioni ha la qualità?

Id	Title	Director	Year	#Remakes	LastRemakeYear
1	Casablanca	Weir	1942	3	1940
2	Dead Poets Society	Curtiz	1989	0	NULL
3	Rman Holiday	Wylder	1953	0	NULL
4	Sabrina	NULL	1964	0	1985

Inaccurato

Incompleto

Inconsistente

Inconsistente

Tante!!!

Acronym	Data Quality Dimension
TDQM	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of manipulation, Value added, Free of error, Interpretability, Objectivity, Relevance, Reputation, Security, Timeliness, Understandability
DWQ	Correctness, Completeness, Minimality, Traceability, Interpretability, Metadata Evolution, Accessibility (System, Transactional, Security), Usefulness (Interpretability, Timeliness (Currency, Volatility), Responsiveness, Completeness, Credibility, Accuracy, Consistency, Interpretability
TQDM	Inherent dimensions: Definition conformance (consistency), Completeness, Business rules conformance, Accuracy (to surrogate source), Accuracy (to reality), Precision, Nonduplication, Equivalence of redundant data, Concurrency of redundant data, Pragmatic dimensions: accessibility, timeliness, contextual clarity, Derivation integrity, Usability, Rightness (fact completeness), cost.
AIMQ	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of operation, Freedom from errors, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, Understandability
CIHI	Dimensions: Accuracy, Timeliness Comparability, Usability, Relevance Characteristics: Over-coverage, Under-coverage, Simple/correlated response variance, Reliability, Collection and capture, Unit/Item non response, Edit and imputation, Processing, Estimation, Timeliness, Comprehensiveness, Integration, Standardization, Equivalence, Linkage ability, Product/Historical comparability, Accessibility, Documentation, Interpretability, Adaptability, Value.
DQA	Accessibility, Appropriate amount of data, Believability, Completeness, Freedom from errors, Consistency, Concise Representation, Relevance, Ease of manipulation, Interpretability, Objectivity, Reputation, Security, Timelines, Understandability, Value added.
IQM	Accessibility, Consistency, Timeliness, Conciseness, Maintainability, Currency, Applicability, Convenience, Speed, Comprehensiveness Clarity, Accuracy, Traceability, Security, Correctness, Interactivity.
ISTAT	Accuracy, Completeness, Consistency
AMEQ	Consistent representation, Interpretability, Case of understanding, Concise representation, Timeliness, Completeness Value added, Relevance, Appropriateness, Meaningfulness, Lack of confusion, Arrangement, Readable, Reasonability, Precision, Reliability, freedom from bias, Data Deficiency, Design Deficiency, Operation, Deficiencies, Accuracy, Cost, Objectivity, Believability, Reputation, Accessibility, Correctness, Unambiguity, Consistency
COLDQ (Loshin)	Schema: Clarity of definition, Comprehensiveness, Flexibility, Robustness, Essentialness, Attribute granularity, Precision of domains, Homogeneity, Identifiability, Obtainability, Relevance, Simplicity/Complexity, Semantic consistency, Syntactic consistency. Data: Accuracy, Null Values, Completeness, Consistency, Currency, Timeliness, Agreement of Usage, Stewardship, Ubiquity, Presentation: Appropriateness, Correct Intepretation, Flexibility, Format precision, Portability, Consistency, Use of storage, Information policy: Accessiiblity, Metadata, Privacy, Security, Redundancy, Cost.
DaQuinCis	Accuracy, Completeness, Consistency, Currency
QAFD	Syntactic/Semantic accuracy, Internal/External consistency, Completeness, Currency, Uniqueness.
CDQ	Accuracy, Completeness, Consistency, Currency, Timeliness, Completability, Reputation, Accessibility, Cost.

L'informazione come rappresentazione della realtà



**Strutture linguistiche
e/o basate sui sensi**



Processo di rappresentazione

**La qualità dipende
dal tipo di
informazione
(testo, immagine,
ecc.)**

Ho visto un bellissimo
tramonto in montagna
era una giornata con
poche nuvole....

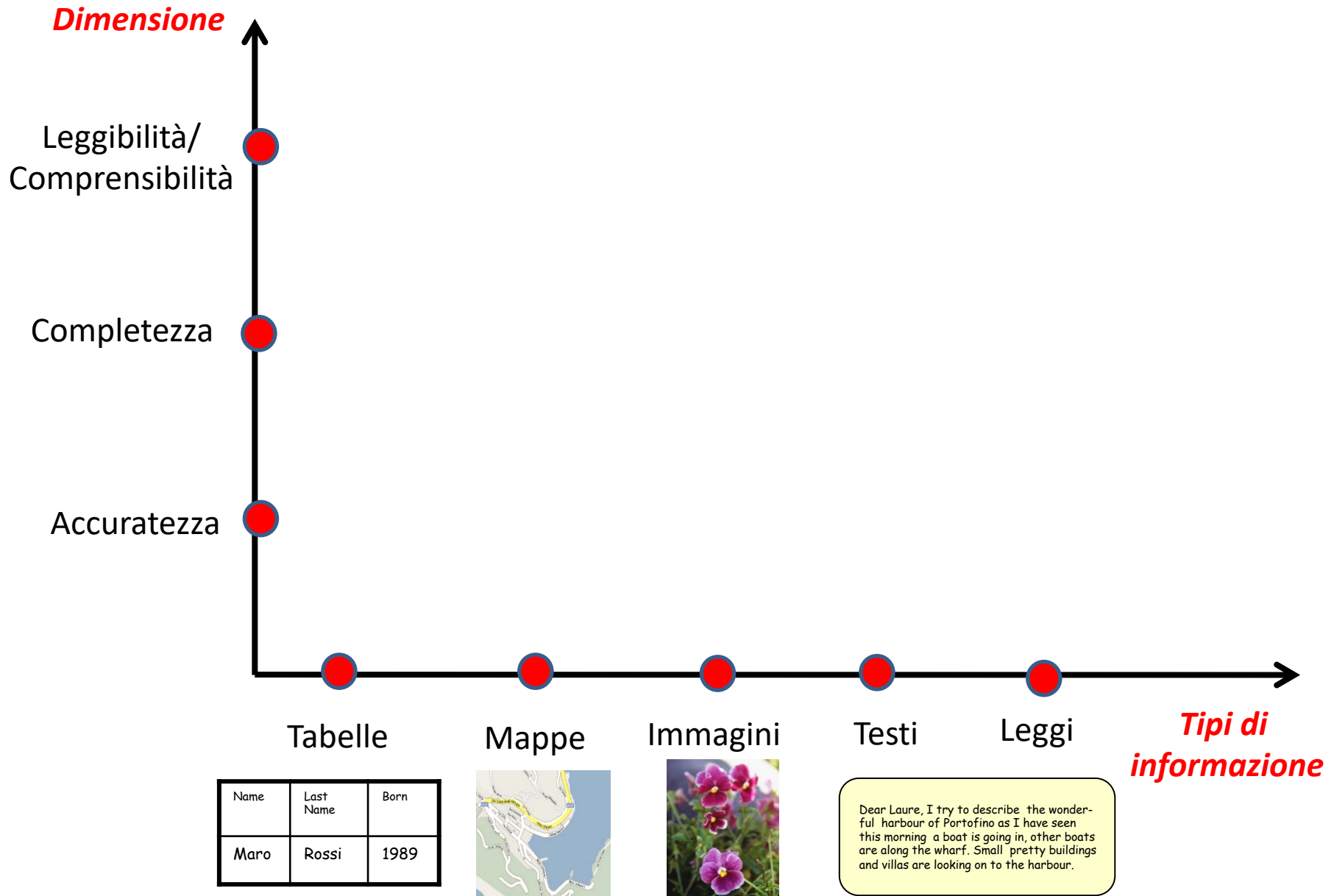


La qualità ha tante **dimensioni**

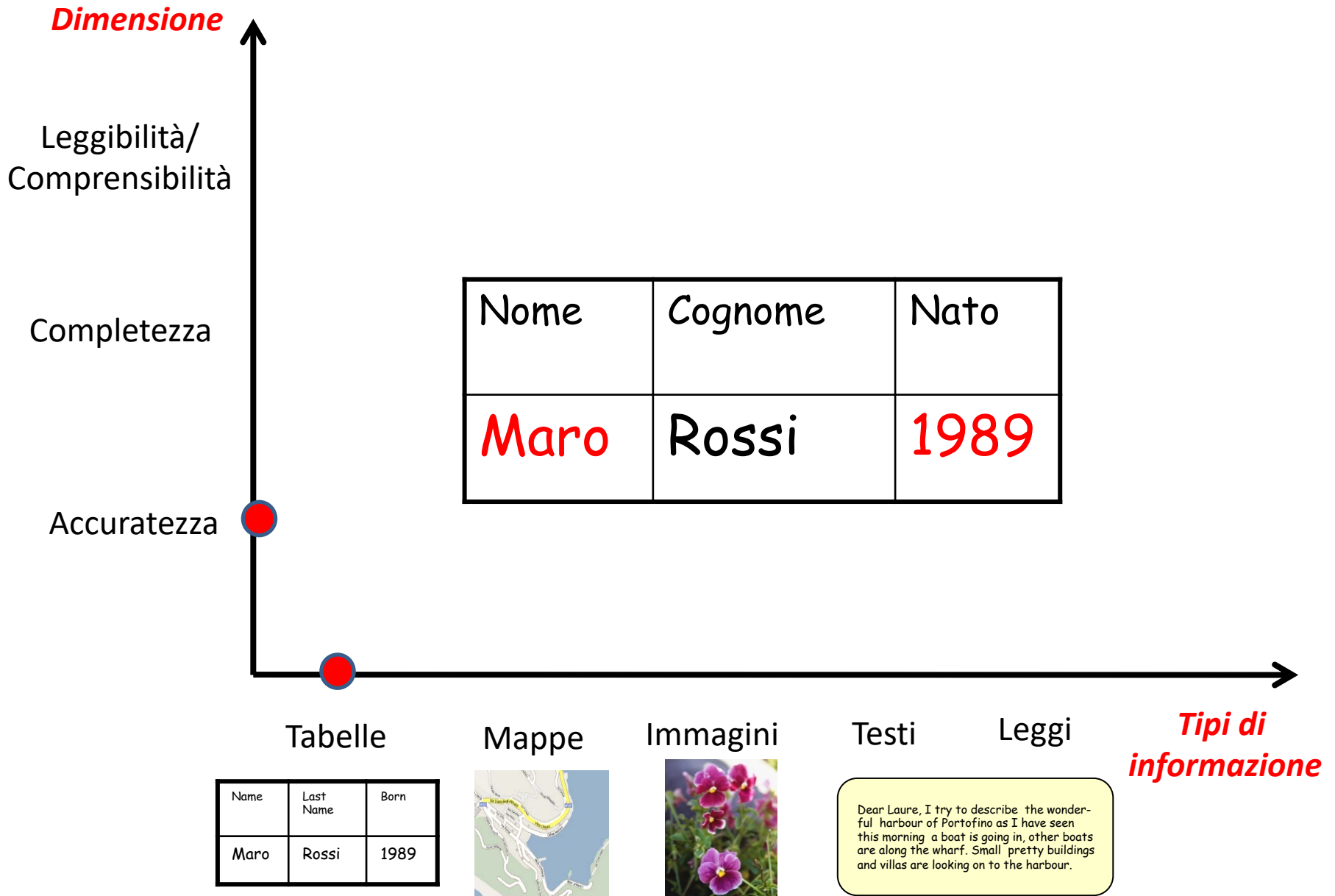
&

La qualità dipende dal
tipo della informazione

Dimensioni e tipi di informazione



Accuratezza delle tabelle



Nomi Italiani Maschili e Femminili - Lettera M



Ring billigt med Comviq

Under hundraen med Fastpris Mini. Betala endast 95 kr/mån!



Quanti sono i nomi di battesimo italiani? Quasi impossibile dirlo ma sembra siano **oltre 10 mila**! Ovviamente molti di essi sono davvero più unici che rari. Ecco un elenco di circa 1700 nomi italiani (in **grassetto** i più comuni e diffusi), suddivisi tra maschili e femminili: alcuni potranno sembrare strani...pensa a quelli che non sono inclusi in questo elenco!

Nomi che iniziano con la lettera:

A B C D E F G I L M NO PQ R S TUV WZ

Maria Rossi e'
nata nel 1949



Nome	Cognome	Nato
Maro	Rossi	1989

Maschili

Macario	Maccabeo
Maffeo	Maggiorino
Magno	Maiorico
Malco	Mamante
Mancio	Manetto
Manfredo	Manilio
Manlio	Mansueto
Manuele	Marcello
Marciano	Marco
Mariano	Marino
Mario	Marolo

Femminili

Macaria	Maddalena
Mafalda	Magda
Maida	Manuela
Mara	Marana
Marcella	Mareta
Margherita	Maria
Marianna	Marica
Mariella	Marilena
Marina	Marinella
Marinetta	Marisa
Marita	Marta

Accuratezza lessicografica per attributi composti di stringhe di lettere



Nome	Cognome	Nato
Maro	Rossi	1989

Corrispondenza
lessicografica

Corrispondenza
lessicografica

Nomi nelle
anagrafi italiane

.....

Marco

Cesira

Maria

Mario

.....

Accuratezza semantica

Maria Rossi e'
nata nel 1949



Nome	Cognome	Nato
Maro	Rossi	1989

Corrispondenza
lessicografica

Corrispondenza
semantica

Corrispondenza
lessicografica

Nomi nelle
anagrafi italiane

.....

Marco

Cesira

Maria

Mario

.....

Accuratezza lessicografica e semantica nel caso di attributi numerici

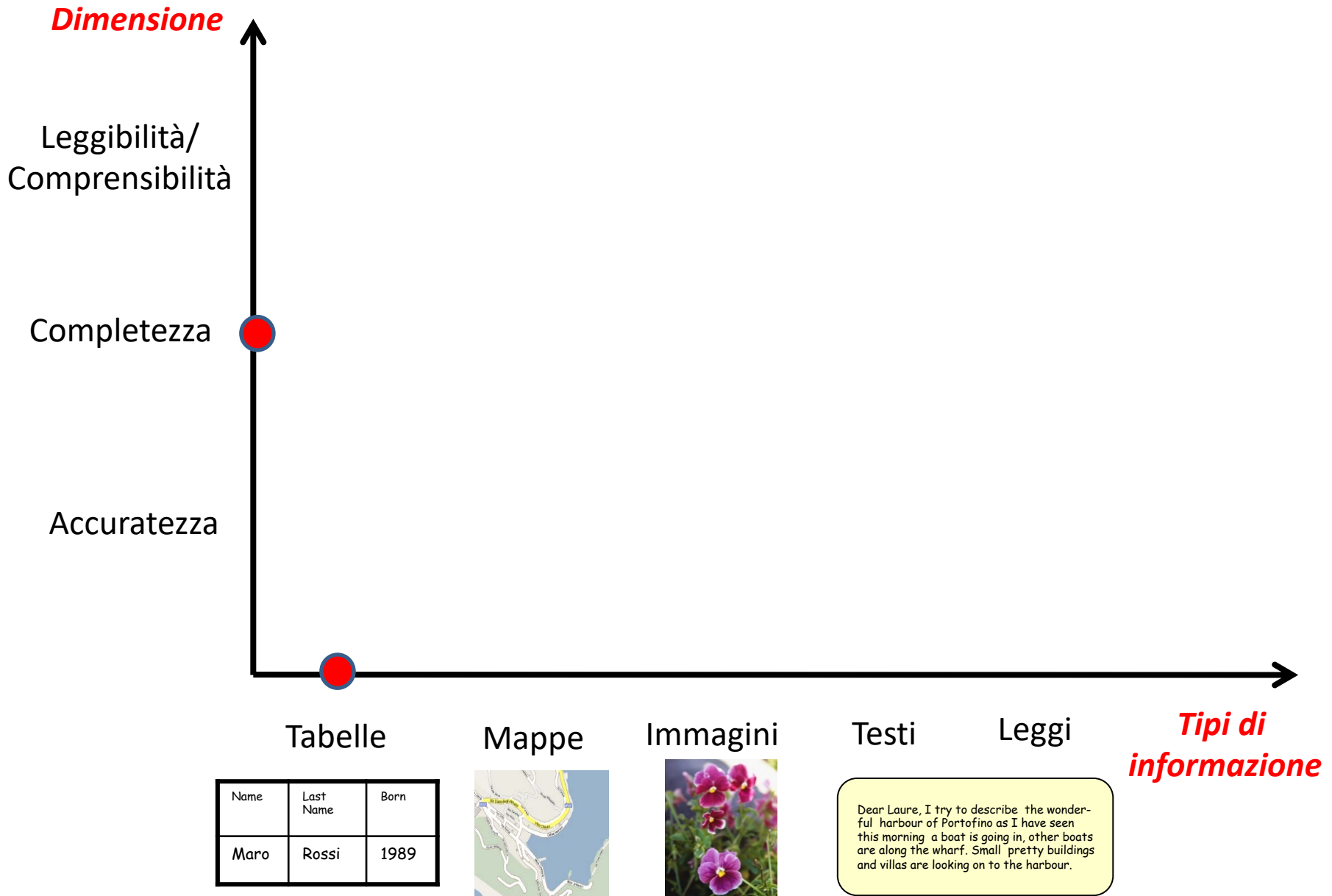
Maria Rossi e'
nata nel 1949



Nome	Cognome	Nato
Maro	Rossi	1989

?

Completezza delle tabelle



Completezza di riga

Id	Title	Director	Year	#Remakes	LastRemakeYear
1	Casablanca	Weir	1942	3	1940
2	Dead Poets Society	Curtiz	1989	0	NULL
3	Rman Holiday	Wylder	1953	0	NULL
4	Sabrina	NULL	1964	0	1985

Incompleto

Completezza di tabella

Id	Title	Director	Year	#Remakes	LastRemakeYear
1	Casablanca	Weir	1942	3	1940
2	Dead Poets Society	Curtiz	1989	0	NULL
3	Rman Holiday	Wylder	1953	0	NULL
4	Sabrina	NULL	1964	0	1985

Incompleto

I valori nulli significano tutti la stessa cosa?

Citta'	Indirizzo prefettura
Roma	Via Manzoni
Firenze	Null
Tivoli	Null
Chiasso	Null

Almeno tre significati differenti...

Citta'	Indirizzo prefettura
Roma	Via Manzoni
Firenze	Null → Non lo sappiamo
Tivoli	
Chiasso	

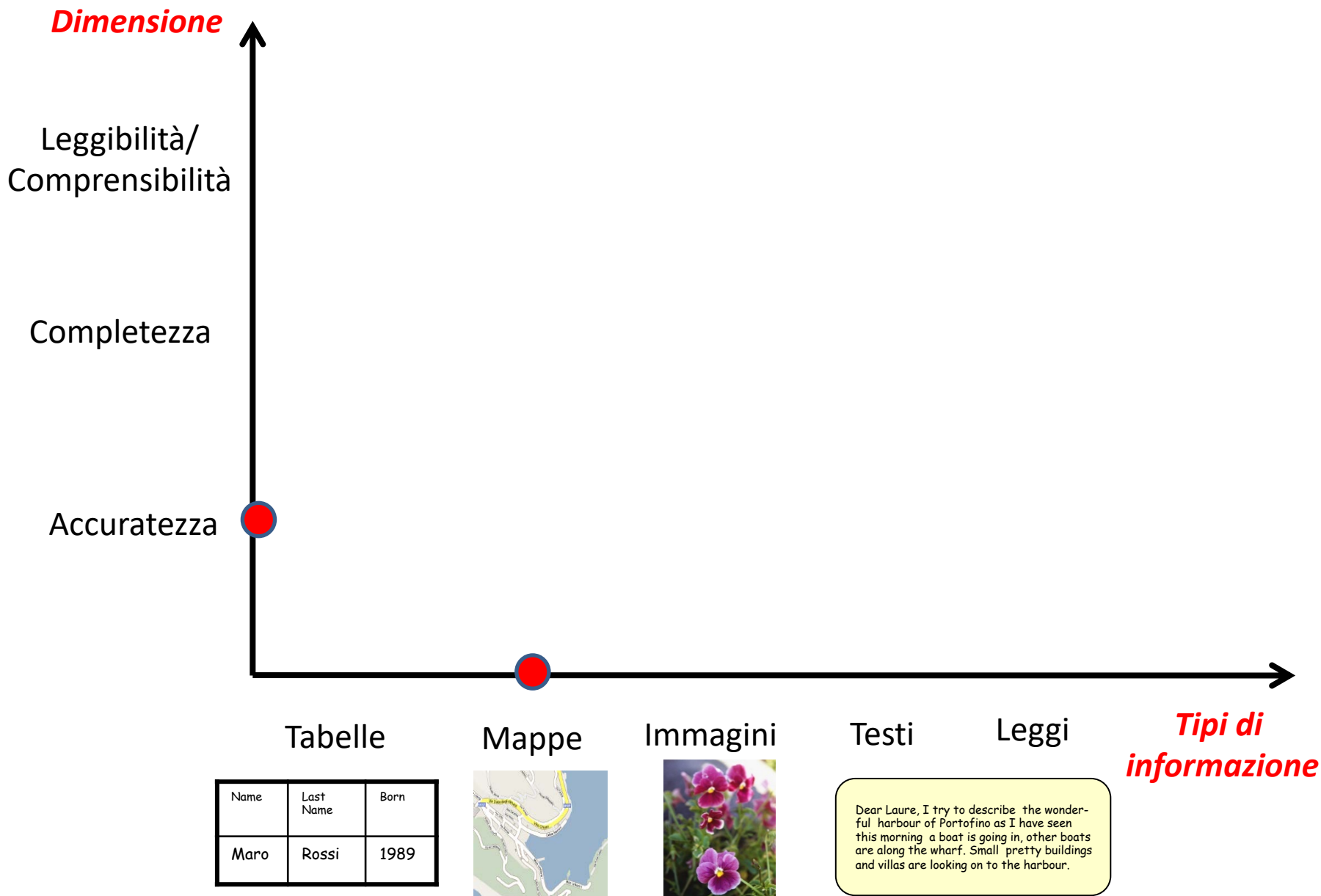
Almeno tre...

Citta'	Indirizzo prefettura
Roma	Via Manzoni
Firenze	Non lo sappiamo
Tivoli	Null → Non esiste
Chiasso	

Almeno tre...

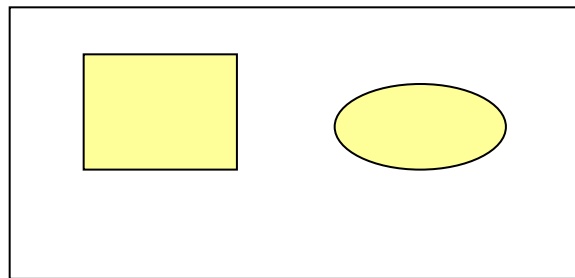
Citta'	Indirizzo prefettura
Roma	Via Manzoni
Firenze	Non lo sappiamo
Tivoli	Non esiste
Chiasso	Null → Non si sa nulla

Dimensioni e tipi di informazione

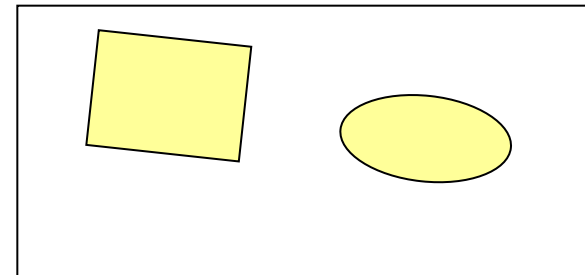
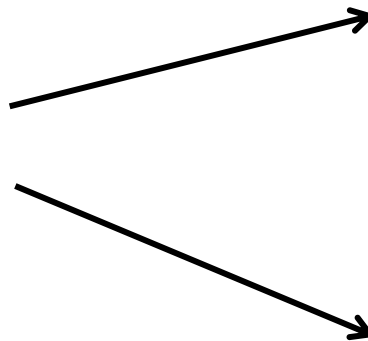


Accuratezza per le mappe

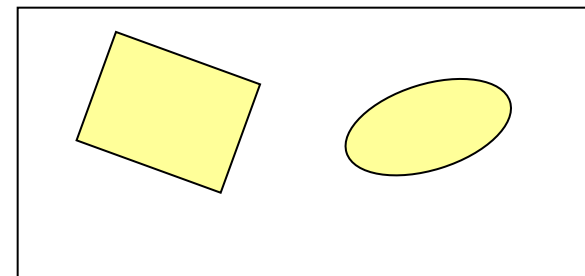
La accuratezza o *fedeltà geometrica* e' la differenza tra la forma di un oggetto o insieme di oggetti geometrici come sequenza ordinata di punti in una mappa, e la sequenza dei corrispondenti punti nel mondo reale.



Geometria nel mondo reale

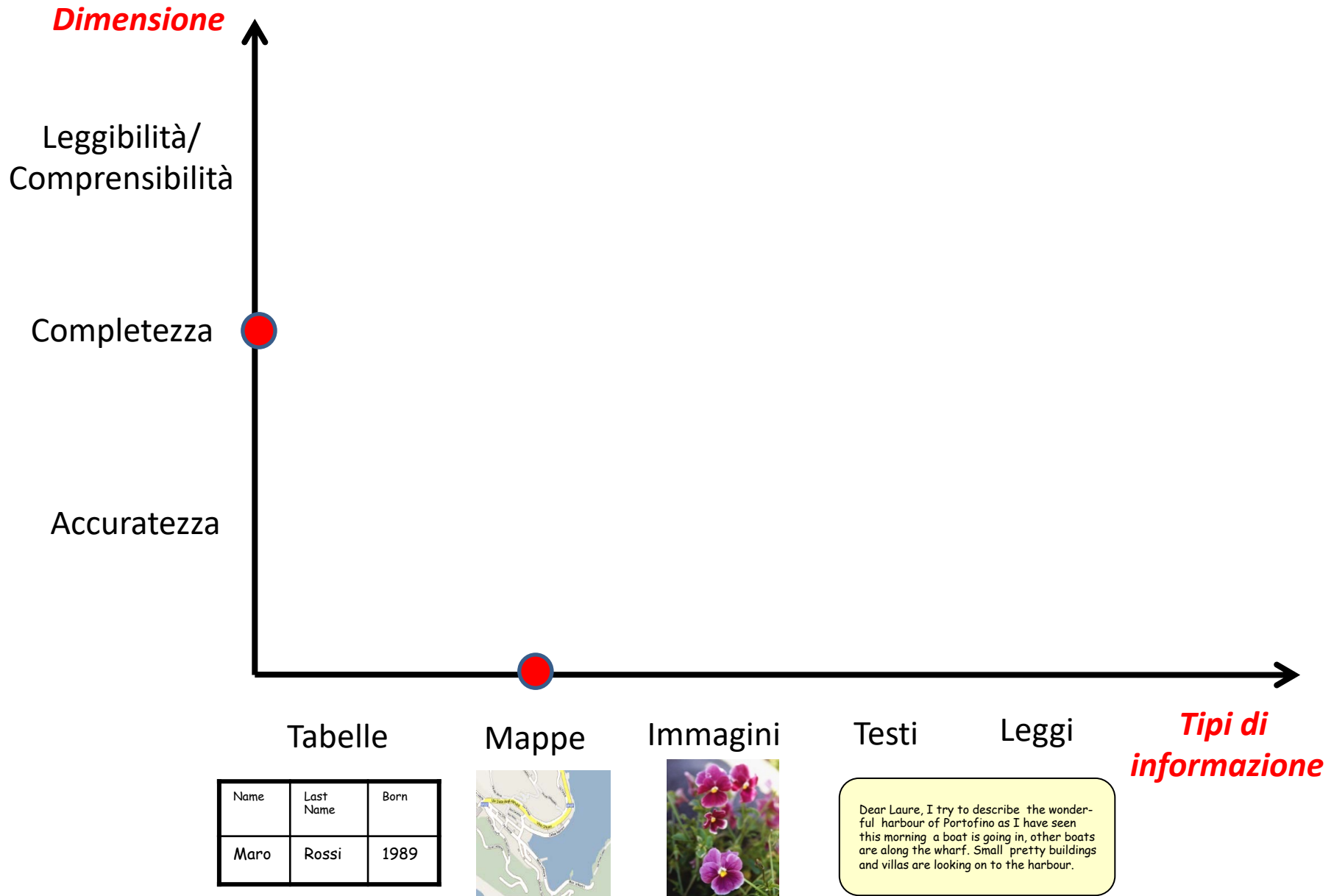


Alta fedeltà

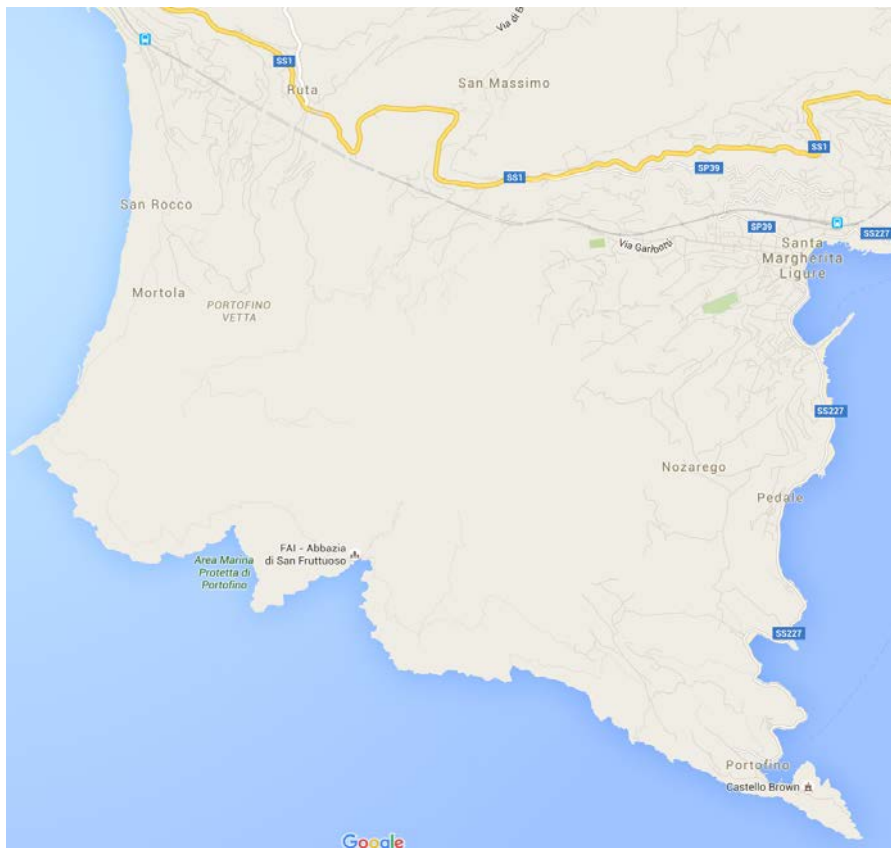


Bassa fedeltà

Dimensioni e tipi di informazione



I sentieri del monte di Portofino: discussione tra mio figlio e me...

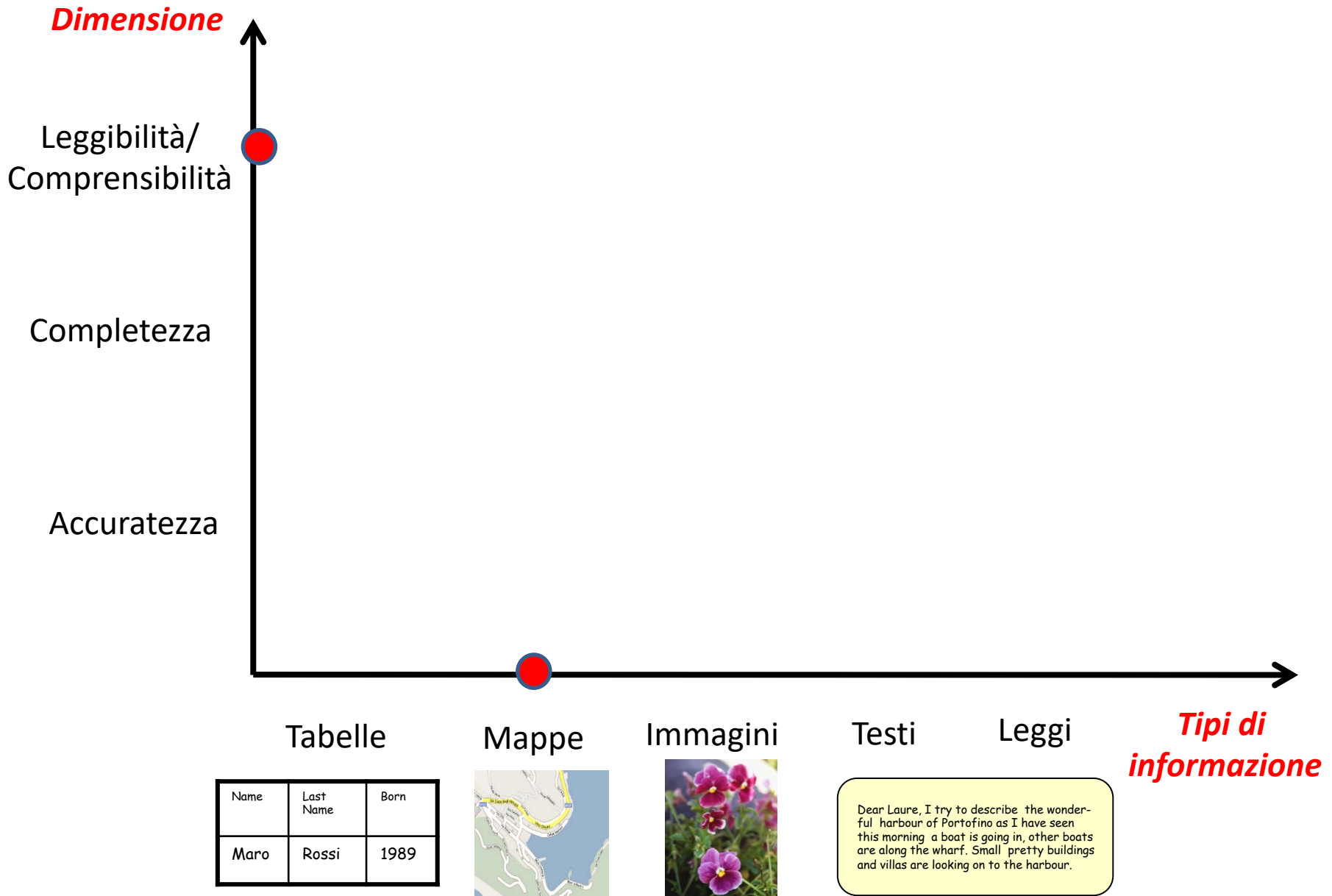


<https://maps.google.it/>



www.portofinotrek.com/trek/6-mappa

Dimensioni e tipi di informazione



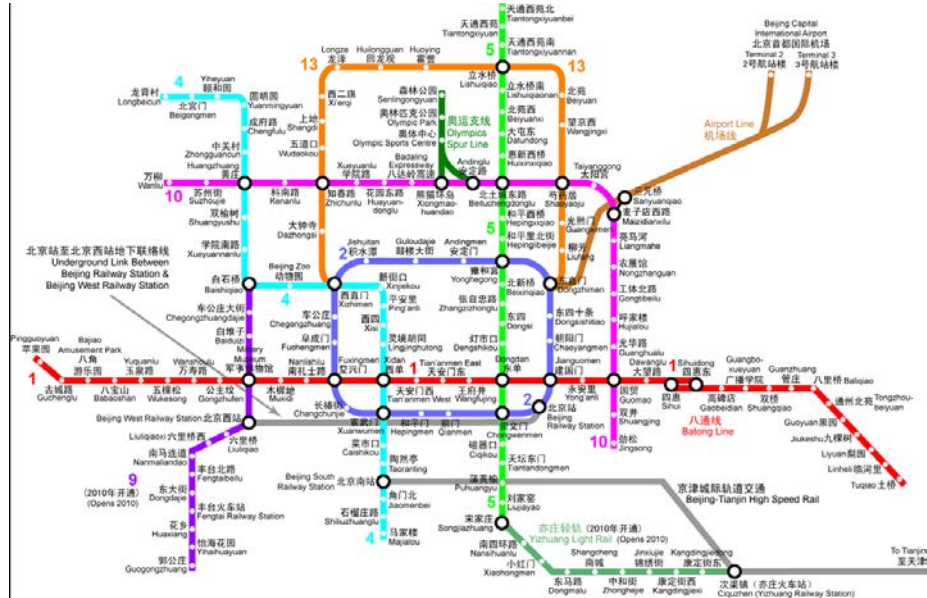
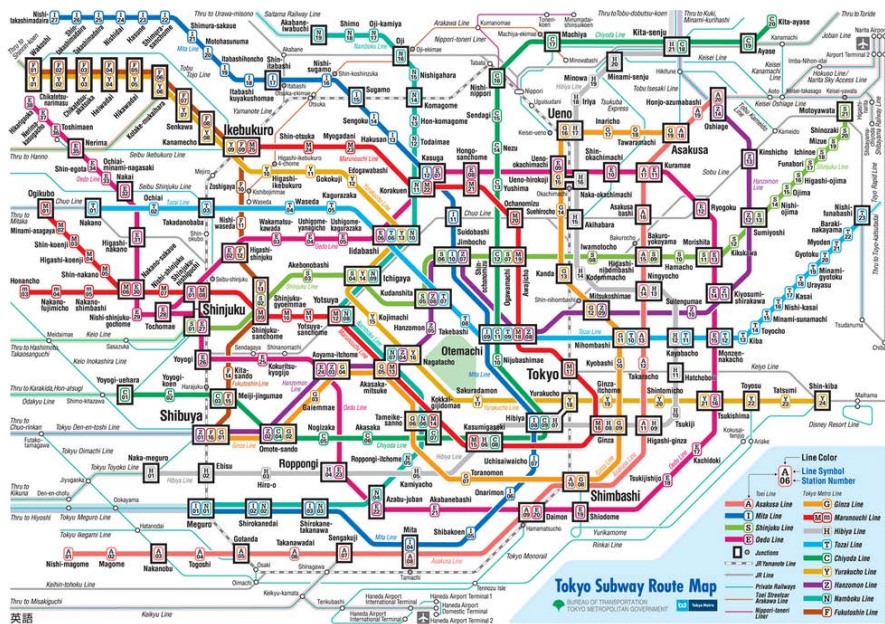
Subway maps di Londra e New York notate delle differenze?



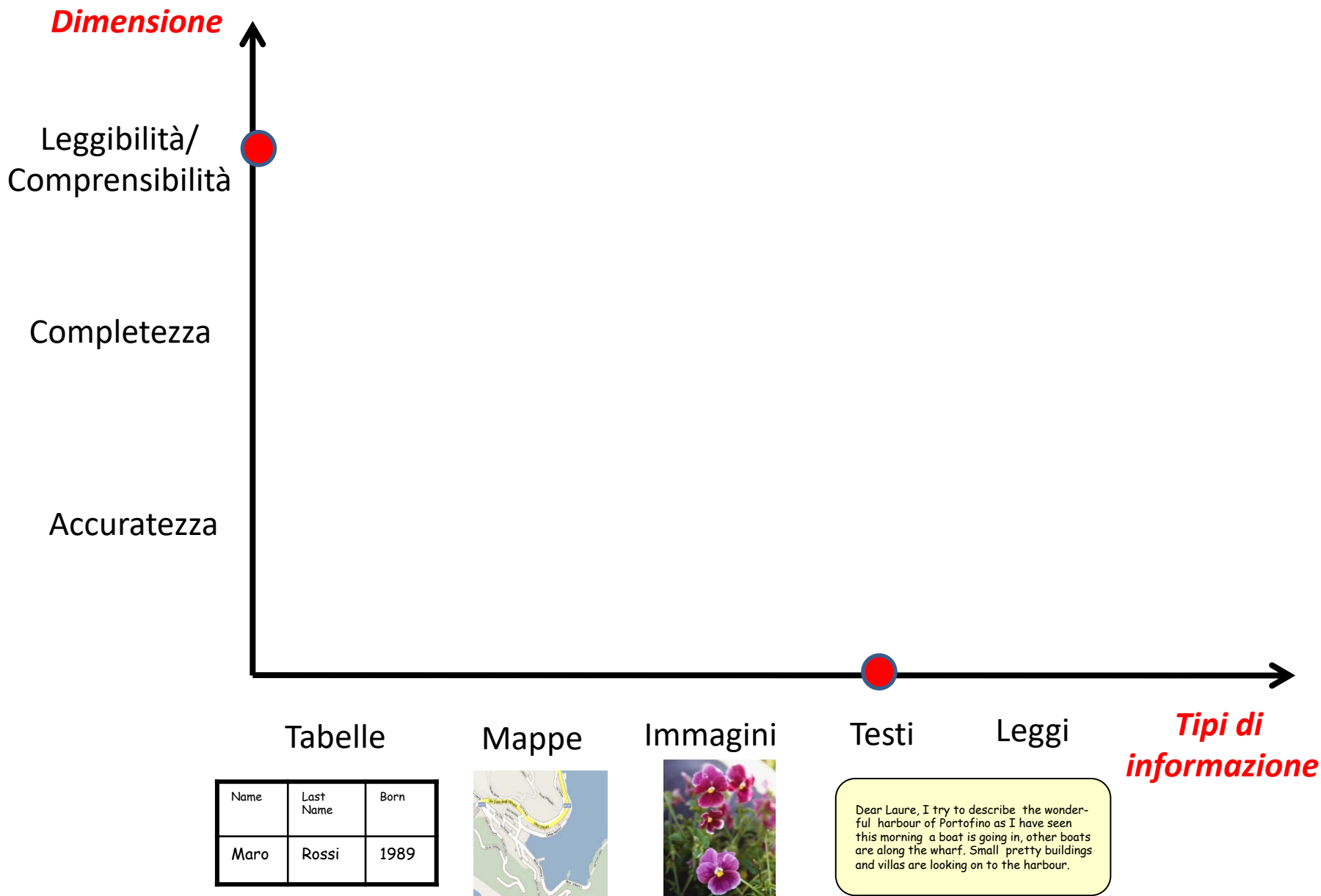
E tra New York e Lione?



E tra Tokio e Pechino?



Dimensioni e tipi di informazione

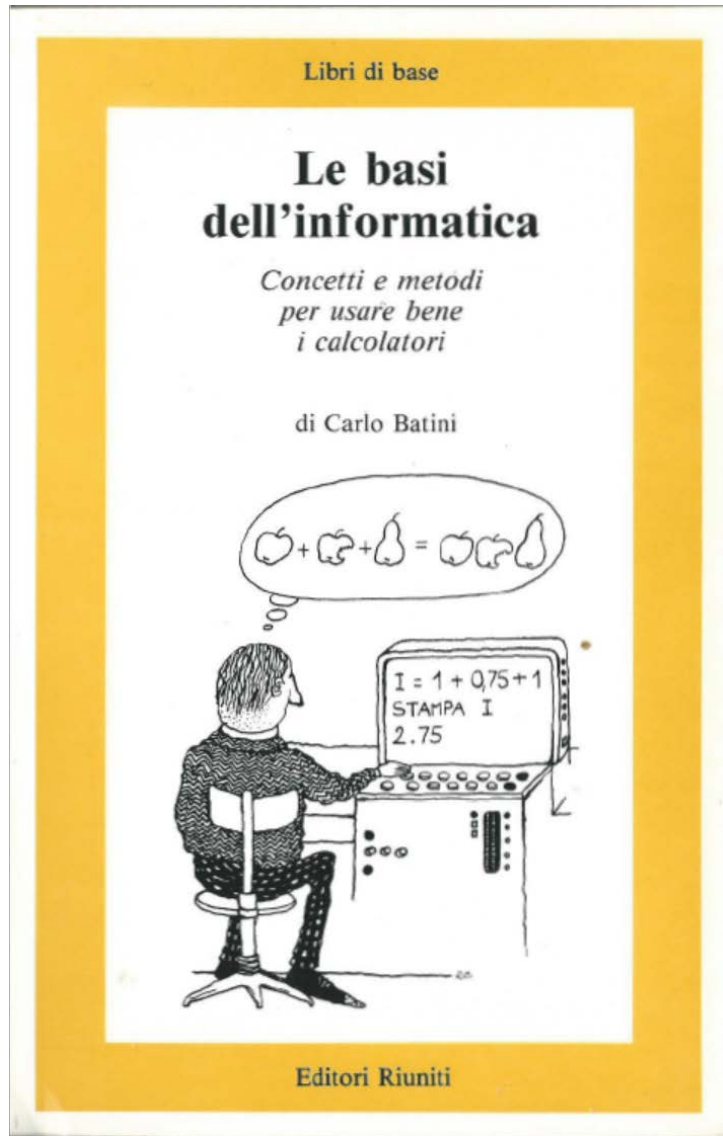


Un testo completamente senza struttura..

-I saw them not long ago I love flowers Id love to have the whole place swimming in roses God of heaven theres nothing like nature the wild mountains then the sea and the waves rushing then the beautiful country with the fields of oats and wheat and all kinds of things and all the fine cattle going about that would do your heart good to see rivers and lakes and flowers all sorts of shapes and smells and colours springing up even out of the ditches primroses and violets nature it is as for them saying theres no God I wouldnt give a snap of my two fingers for all their learning why dont they go and create something I often asked him atheists or whatever they call themselves go and wash the cobbles off themselves first then they go howling for the priest and they dying and why why because theyre afraid of hell on account of their bad conscience ah yes I know them well who was the first person in the universe before there was anybody that made it all who ah that they dont know neither do I so there you are they might as well try to stop the sun from rising tomorrow the sun shines for you he said the day we were lying among the rhododendrons on Howth head in the grey tweed suit and his straw hat the day I got him to propose to me yes first I gave him the bit of seedcake out of my mouth and it was leapyear like now yes 16 years ago my God after that long kiss I near lost my breath yes he said I was a flower of the mountain yes so we are flowers all a womans body yes that was one true thing he said in his life and the sun shines for you today yes that was why I liked him because I saw he understood or felt what a woman is and I knew I could always get round him and I gave him all the pleasure I could leading him on till he asked me to say yes and I wouldnt answer first only looked out over the sea and the sky I was thinking of so many things he didnt know of Mulvey and Mr Stanhope and Hester and father and old captain Groves and the sailors playing all birds fly and I say stoop and washing up dishes they called it on the pier and the sentry in front of the governors house with the thing round his white helmet poor devil half roasted and the Spanish girls laughing in their shawls and their tall combs and the auctions in the morning the Greeks and the jews and the Arabs and the devil knows who else from all the ends of Europe and Duke street and the fowl market all clucking outside Larby Sharons and the poor donkeys slipping half asleep and the vague fellows in the cloaks asleep in the shade on the steps and the big wheels of the carts of the bulls and the old castle thousands of years old yes and those handsome Moors all in white and turbans like kings asking you to sit down in their little bit of a shop and Ronda with the old windows of the posadas 2 glancing eyes a lattice hid for her lover to kiss the iron and the wineshops half open at night and the castanets and the night we missed the boat at Algeciras the watchman going about serene with his lamp and O that awful deepdown torrent O and the sea the sea crimson sometimes like fire and the glorious sunsets and the figtrees in the Alameda gardens yes and all the queer little streets and the pink and blue and yellow houses and the rosegardens and the jessamine and geraniums and cactuses and Gibraltar as a girl where I was a Flower of the mountain yes when I put the rose in my hair like the Andalusian girls used or shall I wear a red yes and how he kissed me under the Moorish wall and I thought well as well him as another and then I asked him with my eyes to ask again yes and then he asked me would I yes to say yes my mountain flower and first I put my arms around him yes and drew him down to me so he could feel my breasts all perfume yes and his heart was going like mad and yes I said yes I will Yes.

- Trieste-Zurich-Paris 1914-1921

Leggibilità/accessibilità culturale



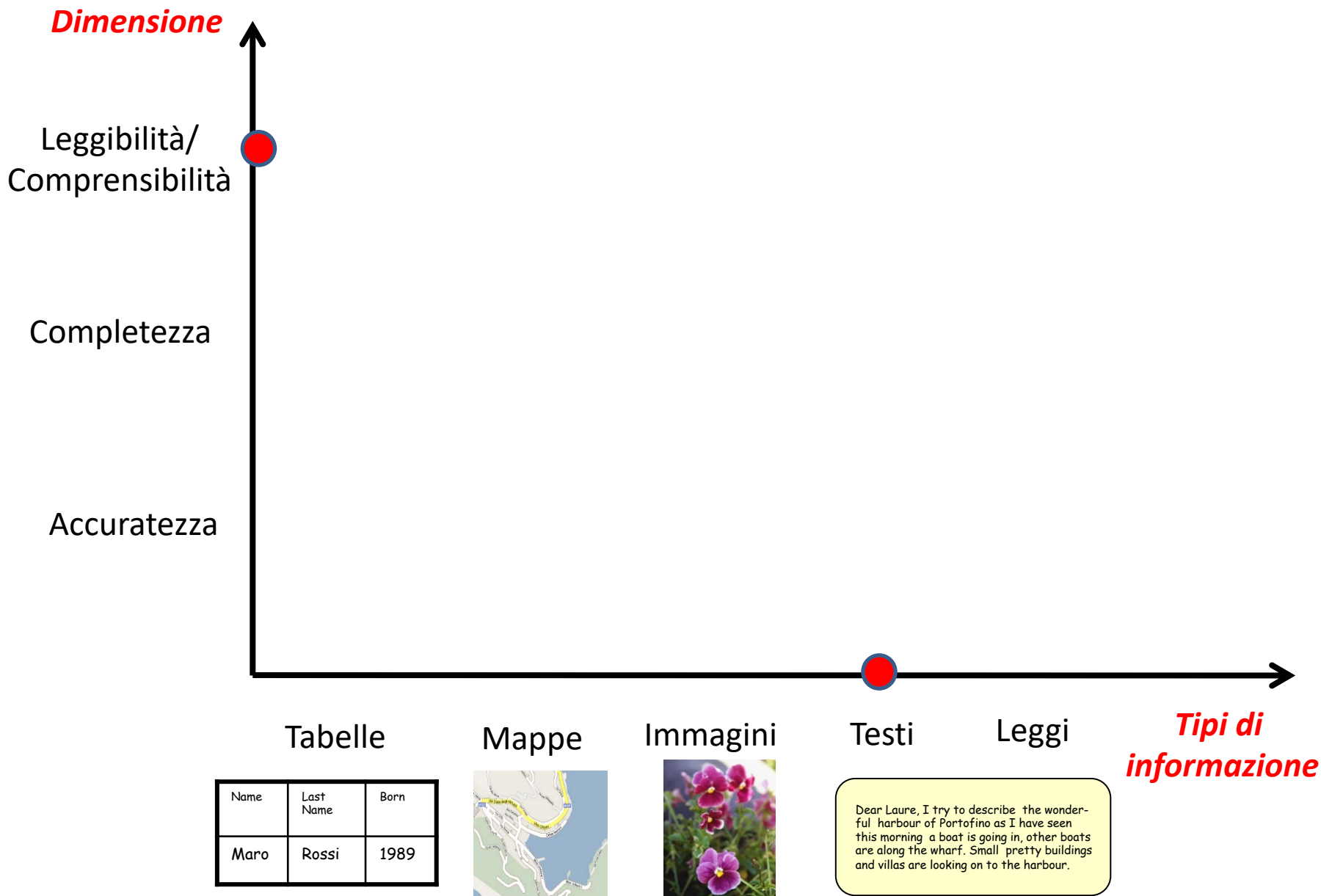
Il libro iniziava con ..

«Quotidianamente ciascuno di noi ha la necessità di usare e produrre un grande numero di informazioni.»

Mi fu cambiato in ...

«**Ciascuno di noi ogni giorno** ha la necessità di usare e produrre un grande numero di informazioni.»

Dimensioni e tipi di informazione



Comprensibilità

Esempio: in molti dei suoi gialli, Agatha Christie introduce un preamolo in cui elenca i protagonisti principali.

Characters in "The Mysterious Affair at Styles"

Captain Hastings, the narrator, on sick leave from the Western Front.

Hercule Poirot, a famous Belgian detective exiled in England; Hastings' old friend

Chief Inspector Japp of Scotland Yard

Emily Inglethorp, mistress of Styles, a wealthy old woman

Alfred Inglethorp, her much younger new husband

John Cavendish, her elder stepson

Mary Cavendish, John's wife

Lawrence Cavendish, John's younger brother

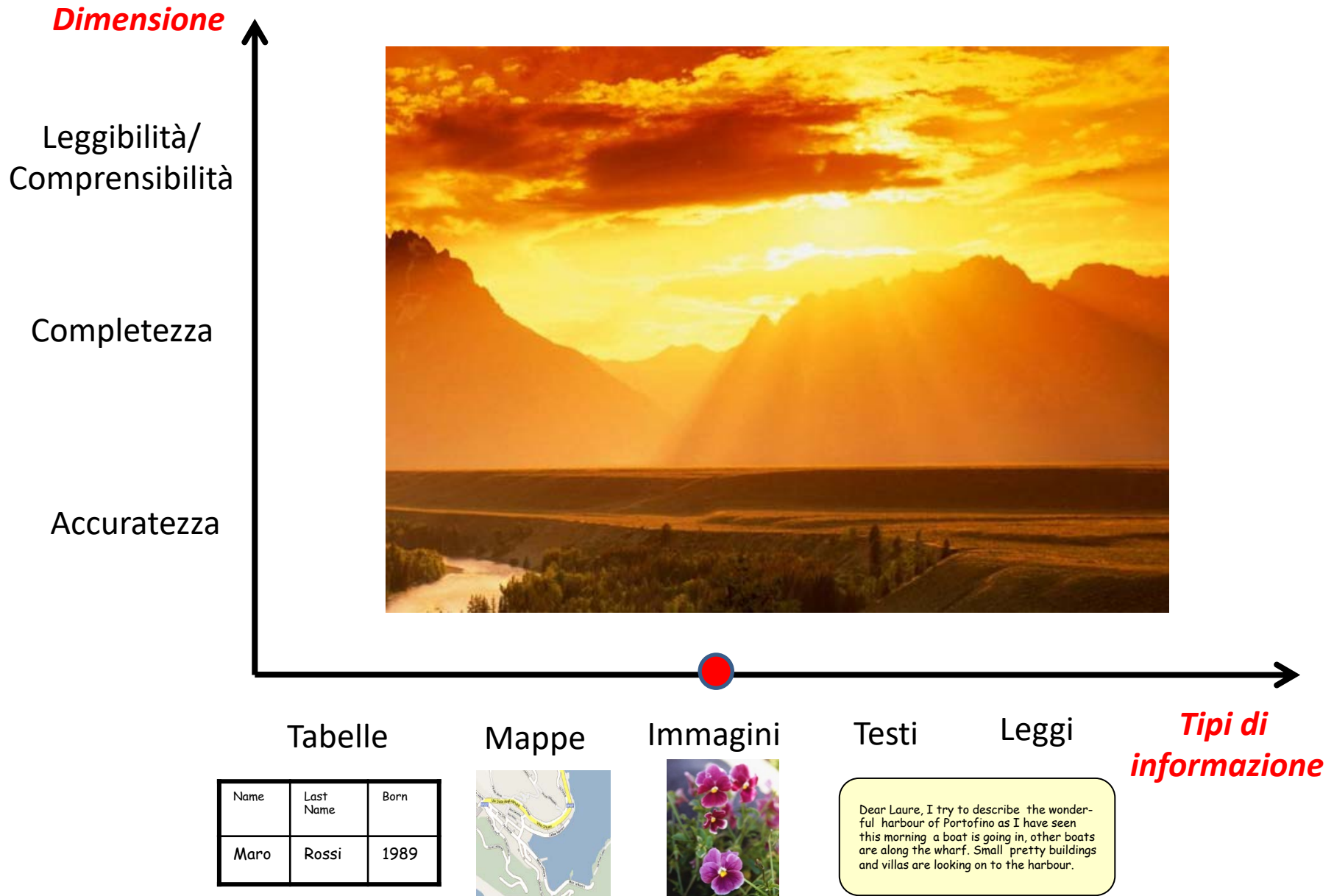
Evelyn Howard, Mrs. Inglethorp's companion

Cynthia Murdoch, the beautiful, orphaned daughter of a friend of the family

Dr. Bauerstein, a suspicious toxicologist

Questo e' un esempio di informazione che il lettore può usare come conoscenza a priori per migliorare la comprensione del testo

Dimensioni e tipi di informazione



Prima idea: le caratteristiche elementari della immagine: **profondita' di colore e risoluzione**

Image Type	Black and White Photographs	Color Photographs	Works of Art on Paper	Microfilm
Master	Scan Type: 8-bit grayscale Resolution: 600 DPI Format: Uncompressed TIFF	Scan Type: 24 bit color Resolution: 400 DPI Format: Uncompressed TIFF	Scan Type: 8-bit Grayscale/ 24-bit color Resolution: 400 DPI Format: Uncompressed TIFF	Scan Type: Bitonal Resolution: 600 DPI (down back to original size) Format: Uncompressed TIFF
Access	Type: 8-bit grayscale Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 8-bit grayscale/24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 Pixels	Type: 8-bit grayscale/24-bit color Format: JPEG Compression: Medium Spatial Resolution: Resize to 1024 x 768 pixels
Thumbnail	4-bit grayscale Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	8-bit color Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	4-bit grayscale/8-bit color Format: GIF (or JPEG) Spatial Resolution: Resize to 150 - 200 pixels across the long dimension 72 DPI	4-bit grayscale/8-bit color Format: GIF Spatial Resolution: Resize to 150-200 pixels across the long dimension 72 DPI

Number of bit
per pixel

Resolution

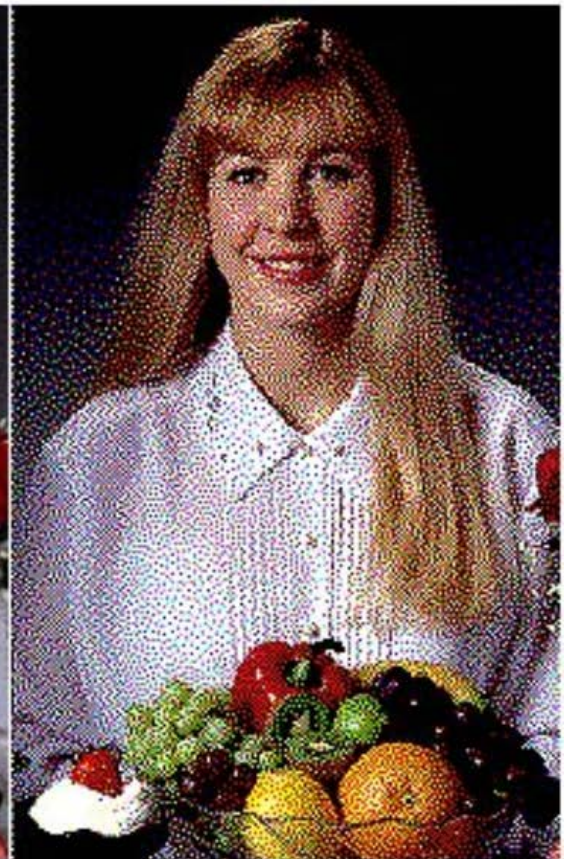
Ancora su profondità del colore



16.7 Million
Colors



256
Colors



16
Colors

Queste immagini hanno stessa profondità di colore e risoluzione: hanno anche la stessa qualità?



Seconda idea: le –nesses
(in inglese)

Nitidezza / Contrasto

Il grado in cui i contorni e le superfici della immagine si modificano rapidamente, con alto dettaglio.



Luminosità / Brillantezza

Correttezza e piacevolezza
nella riproduzione dei toni dei colori



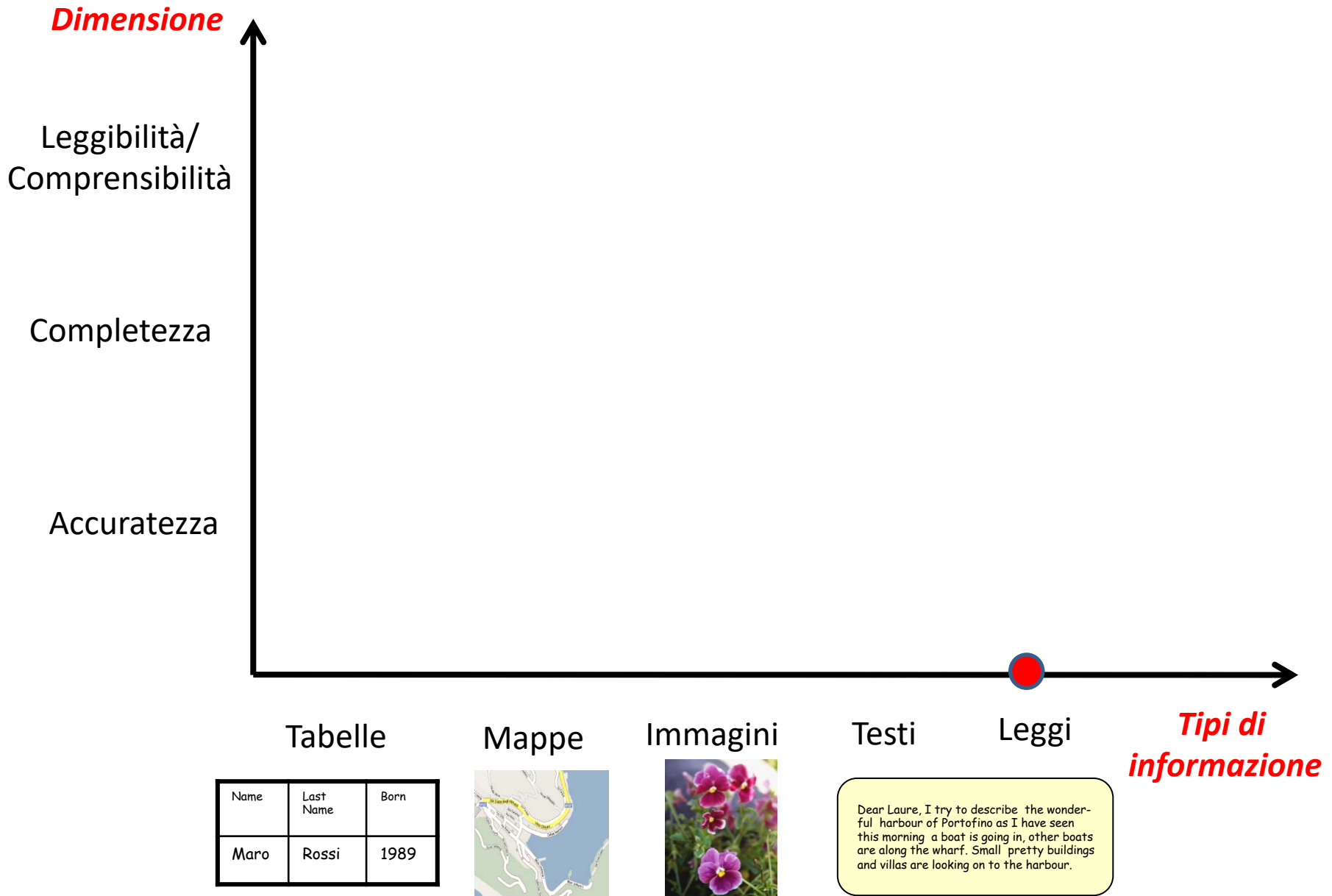
Altri indici sintetici: il **mean square error (MSE)** rispetto a una immagine di riferimento

L' MSE non ha buona correlazione con le misure soggettive di qualità delle immagini



Fig. 3. Evaluation of “Lena” images distorted by different means. (a) Original “Lena” image, 512×512 , 8bits/pixel; (b) Contrast stretched image, MSE = 225, Q = 0.9372; (c) Gaussian noise contaminated image, MSE = 225, Q = 0.3891; (d) Impulsive noise contaminated image, MSE = 225, Q = 0.6494; (e) Blurred image, MSE = 225, Q = 0.3461; (f) JPEG compressed image, MSE = 215, Q = 0.2876.

Dimensioni e tipi di informazione



Esempi di **accuratezza** **nei riferimenti**

Riferimento implicito per la
legge e gli articoli

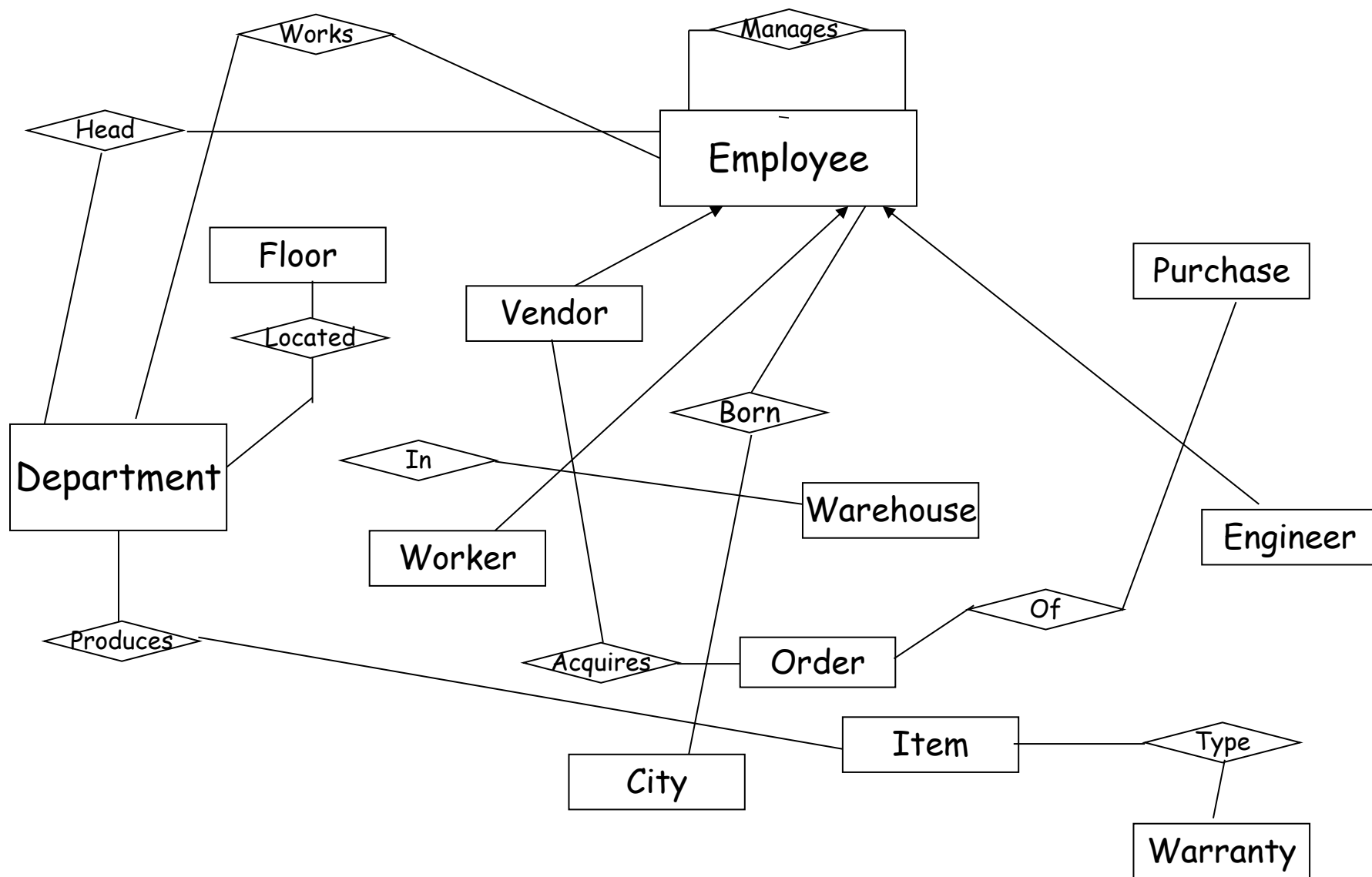
- Articolo 1 – Questa legge cancella tutte le leggi precedenti in tema di contrasto alla evasione fiscale
- Articolo 1 – Questa legge sostituisce la Legge 320/2007 in tema di contrasto alla evasione fiscale
- Articolo 1 - Questa legge sostituisce la Legge 320/2007, intero Art. 1 e Art 7 nei commi 3 e 5.

Riferimento esplicito per la
legge, implicito per gli articoli

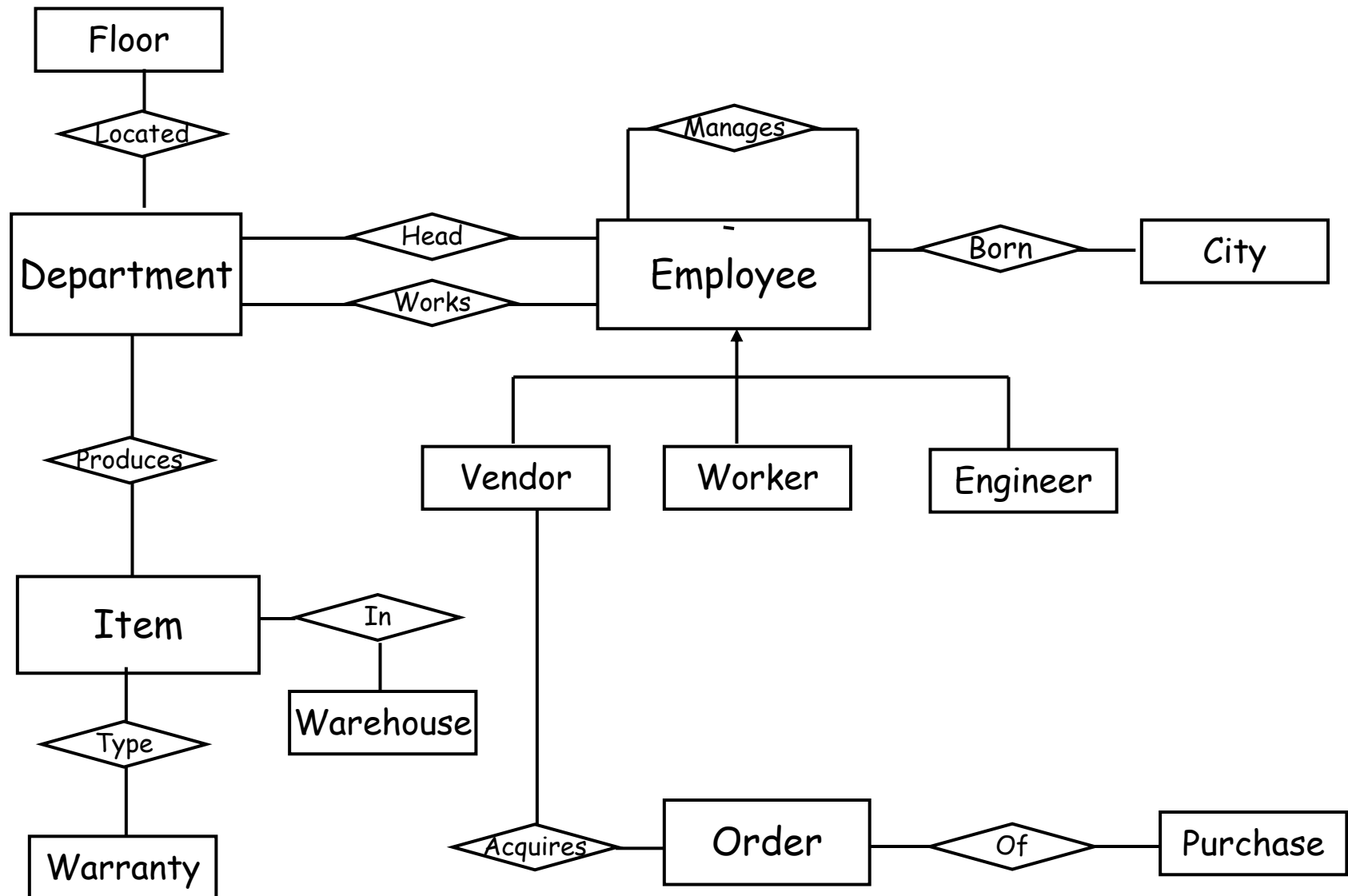
Riferimento esplicito
per la legge e gli articoli

La qualita' dipende dalla cultura

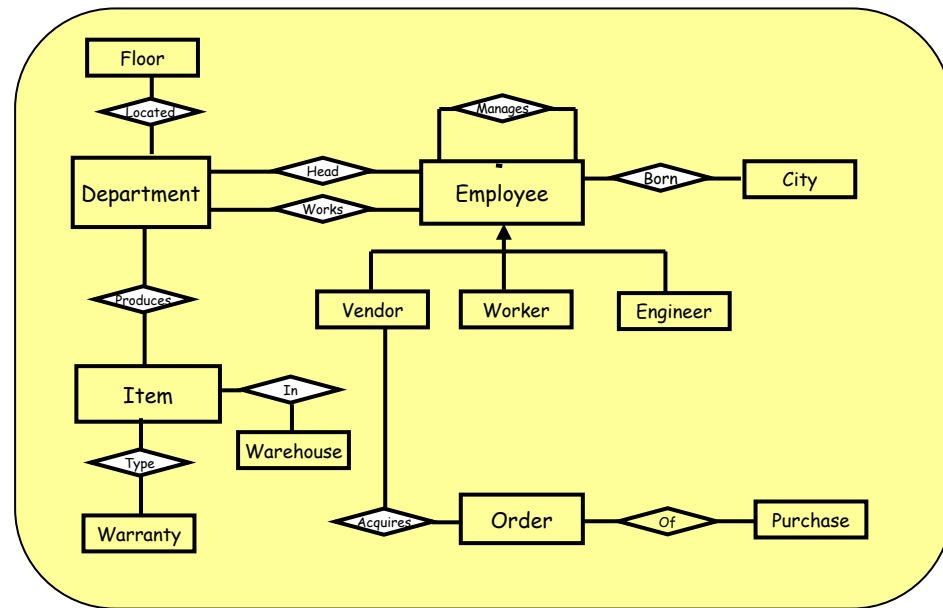
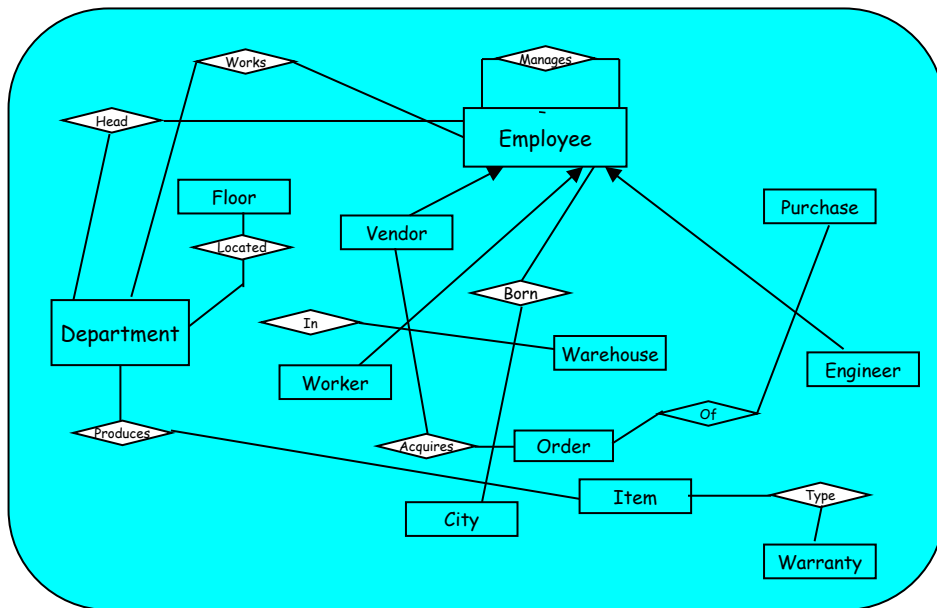
Uno schema dati «Spaghetti style»



Lo stesso schema piu' ordinato



Quale diagramma «vi piace» di piu'?



Qualità ed estetica.....

Minimizza
i piegamenti

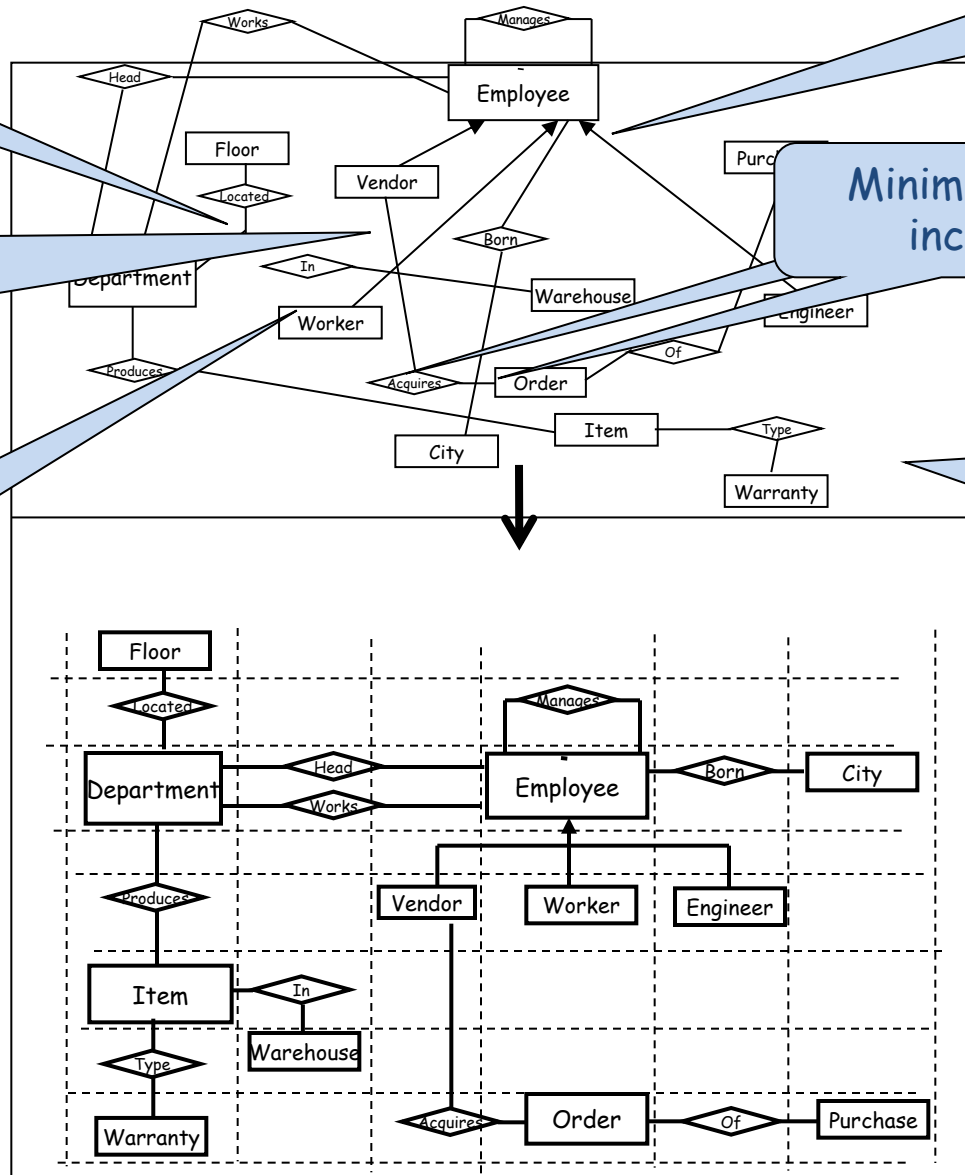
Metti il
concetto più
importante nel
mezzo

Usa solo linee
orizzontali o
verticali

Metti vicini i
concetti simili

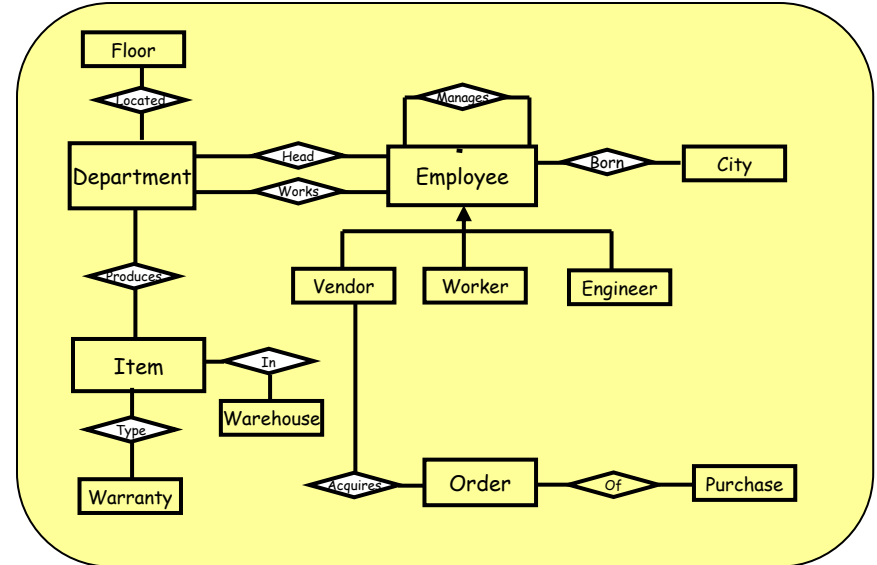
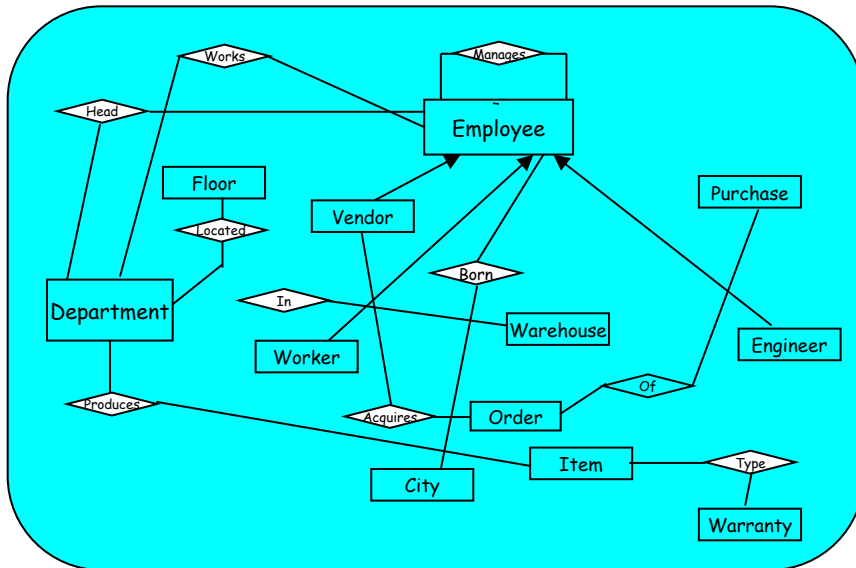
Minimizza gli
incroci

Disponi il
disegno in una
griglia



Esperienza in Cina, 1985

Domanda ai professori cinesi:
quale diagramma vi piace di piu'?



***Risposta: quello a sinistra,
a noi piace la asimmetria e il movimento***

La qualità come tradeoff di dimensioni

Fedeltà vs Qualità Complessiva Percepita



Fedeltà vs Utilità



Efficienza vs Qualità complessiva percepita



188 Kb



63 Kb



33 Kb



16 Kb



5 Kb



2 Kb

I grafici: gioia (leggibilità)
e dolori (imprecisione)
della nostra percezione
della informazione

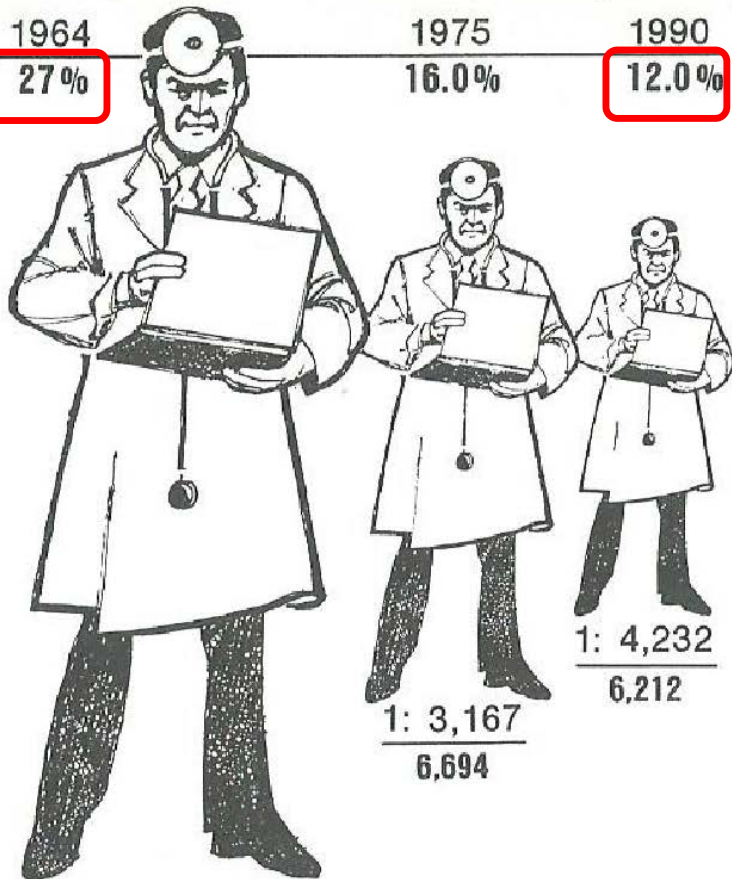
Un grafo e' meglio di 1.000 parole...
E' sempre vero?

Quante bugie nei grafici.... - 1

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964 27% 1975 16.0% 1990 12.0%



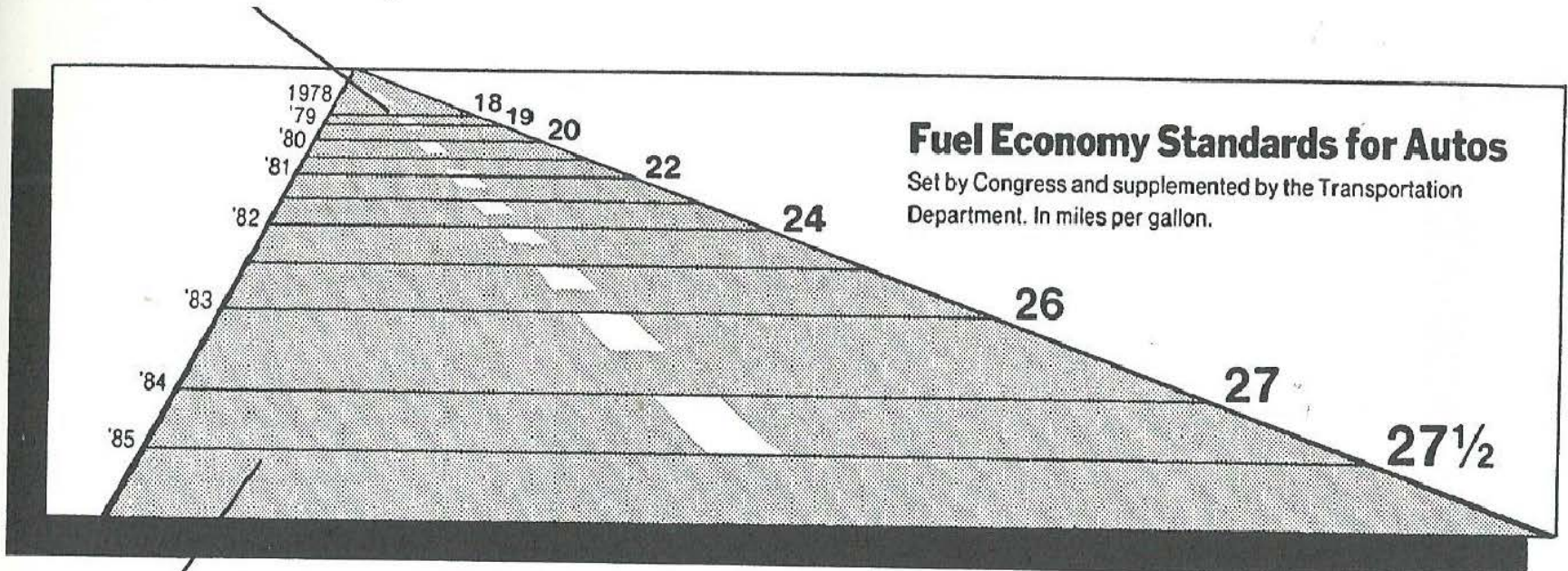
Il lie factor

*Misura quanto e' bugiardo il grafico =
misura dell'effetto che si vuole
rendere nel grafico / misura
dell'effetto nei dati*

Lie factor = 2.8

Quante bugie nei grafici.... - 2

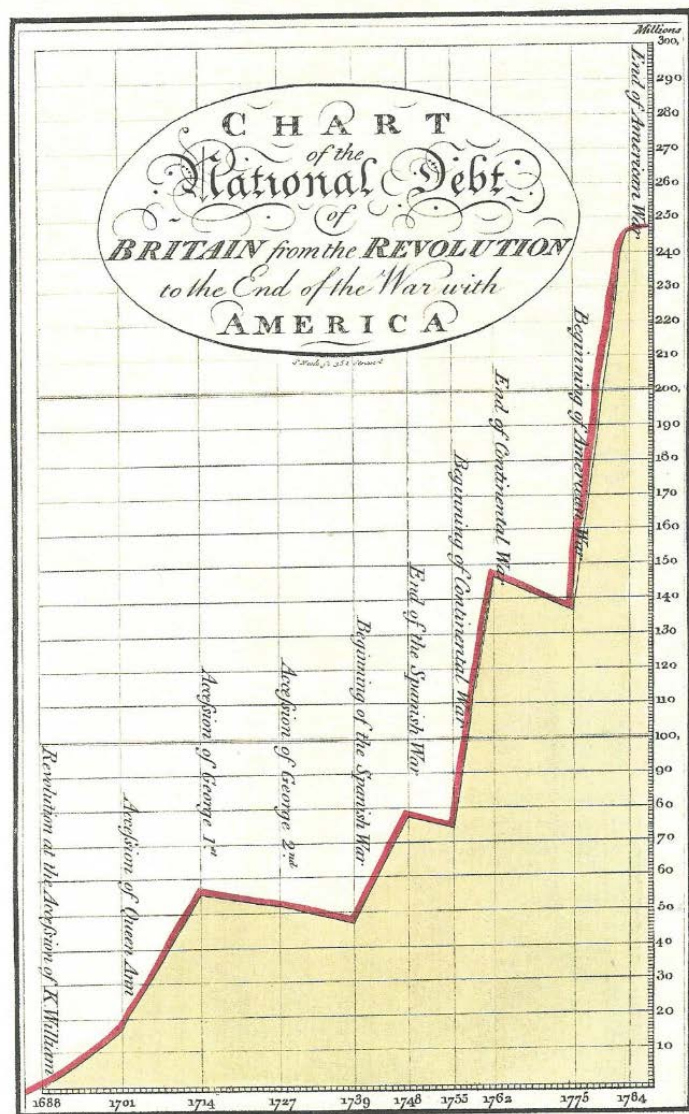
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



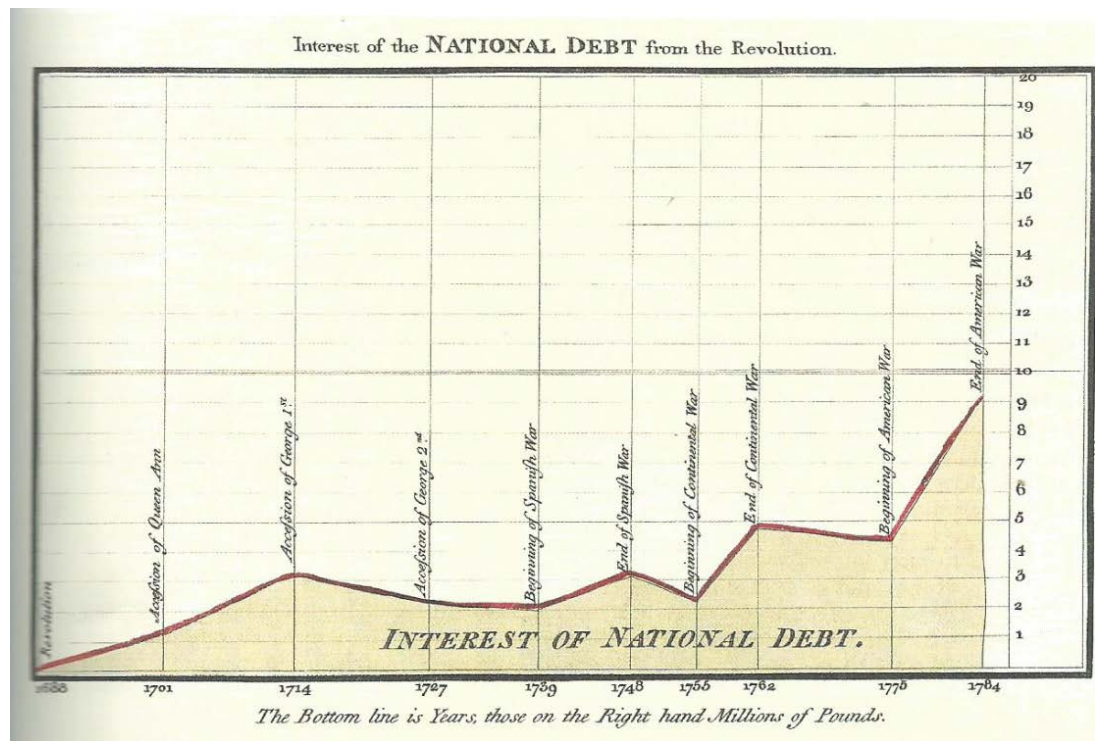
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Lie factor = 14.8

Quante bugie nei grafici.... - 3

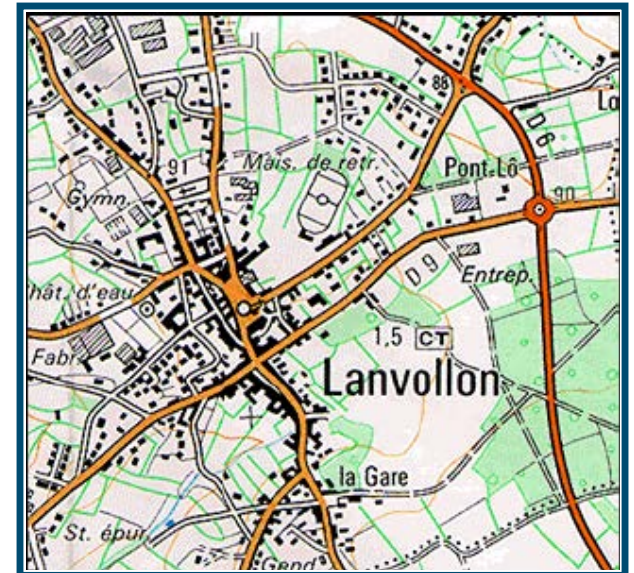
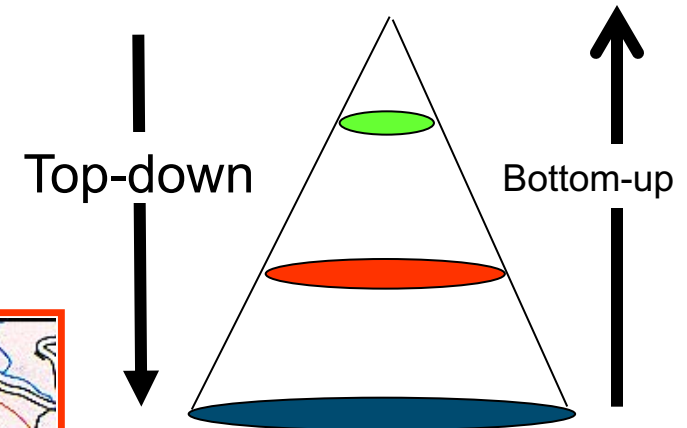
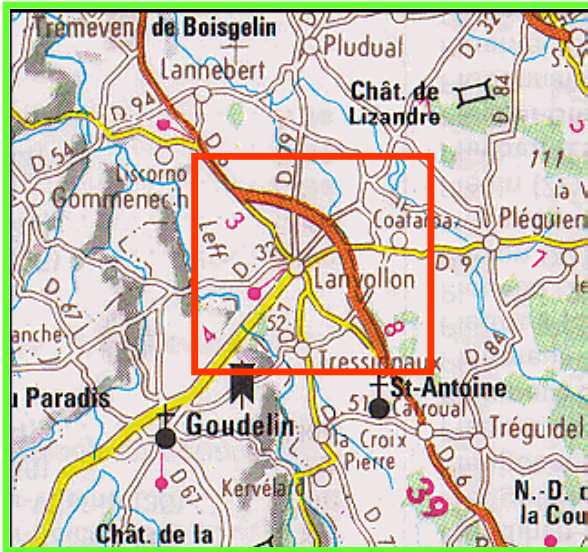


The Divisions at the Bottom are Years, & those on the Right hand Money.



Qualità e astrazioni

Per le mappe l'astrazione e' spesso indispensabile



Generalizzazione e Simbolizzazione

Operazioni di generalizzazione cartografica

Operator	Before	After
(a) Smoothing Reduce angularity of the map object.		
(b) Collapse Reduce dimensionality of map object (area to point, linear polygon to line).		
(c) Displacement Small movement of map objects in order to minimise overlap.		
(d) Enhancement Emphasize characteristics of map feature and meet minimum legibility requirements.		
(e) Typification Replacement of a group of map features with a prototypical subset.		
(f) Text Placement Non overlapping unambiguous placement of text.		
(g) Symbolization Change of symbology according to theme (pictorial, iconic), or reduce space required for symbol.		

Riduzione vs astrazione



Riduzione fotografica



Generalizzazione/Astrazione

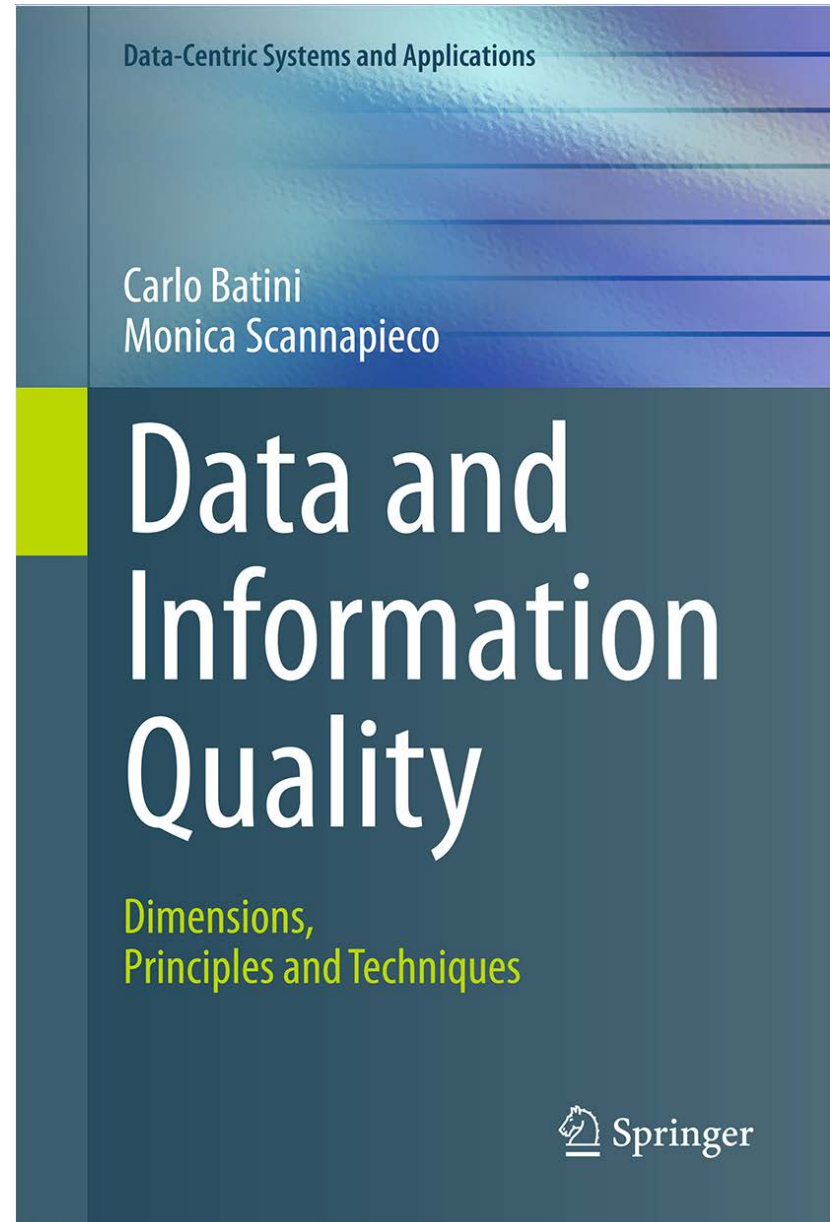


Bugie e verità

tradeoff tra Accuratezza and Leggibilità

The process is one of compromise reflecting the long held view among cartographers that making maps involves telling small lies in order to tell the truth!

Un libro
per approfondire



Resti

Attività' basate sui processi

A Database is Like a Lake

To Clean Up The Lake, One Must First
Eliminate The Sources Of Pollutant

