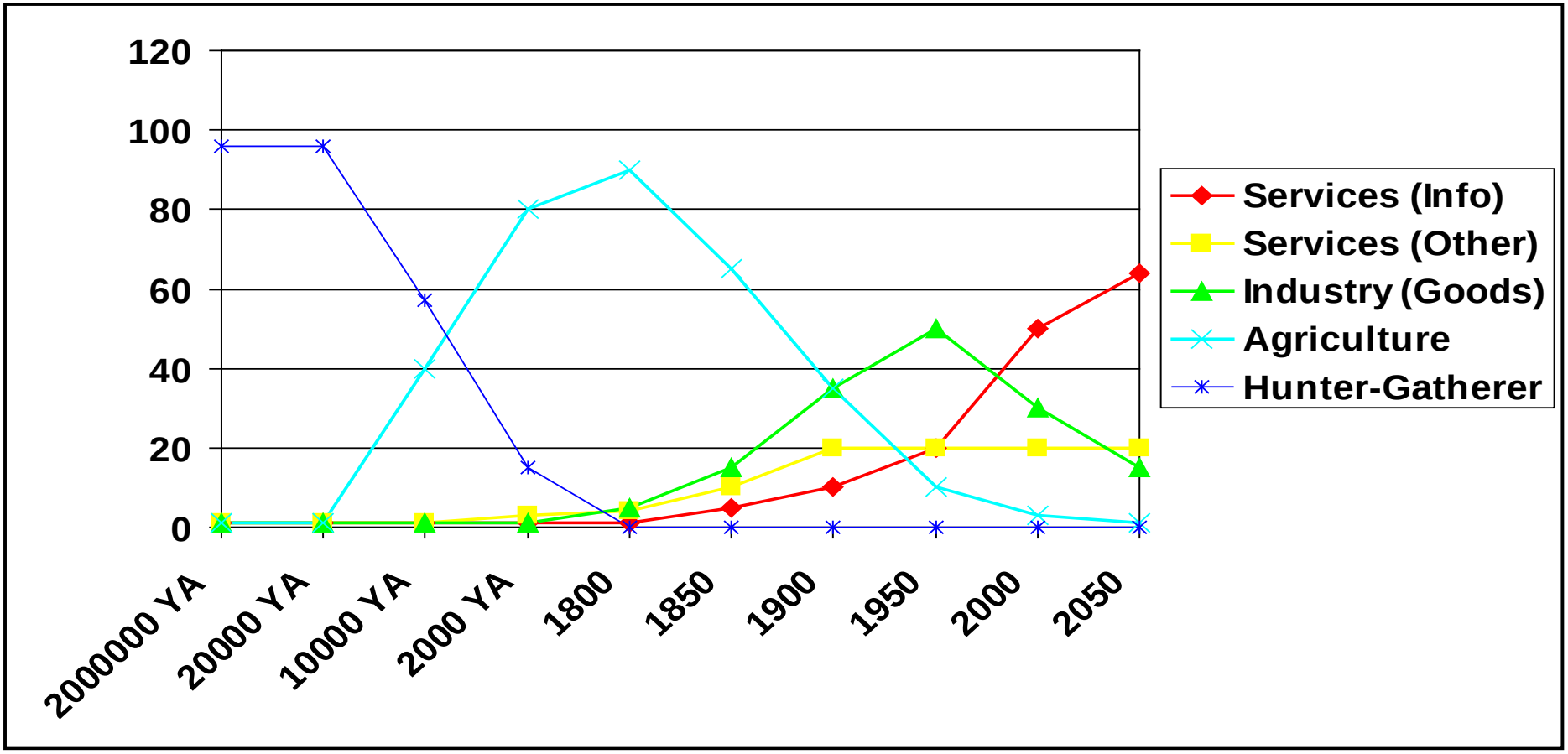
C. Batini & M. Scannapieco
Data and Information Quality Book Figures

Chapter 11: Information Quality in Use

Evolution of U.S. Labor Percentages by Sector in the last 2.000.000 years



Estimations based on Porat, M. (1977) Info Economy: Definitions and Measurement

Sales transactions example from [211]

ID	Date	Customer code	Product code	Quantity	Price	Amount
1	June 7, 2015	C	X	20	€ 5.000	€ 100.000
2	June 7, 2015	B	Y	3	€ 1.000	€ 3.000
3	June 8, 2015	A	Y	1	€ 1.000	€ 1.000
4	June 8, 2015	B	Z	5	€ 3.000	€ 15.000

a. Illustrative Sale Transaction Dataset

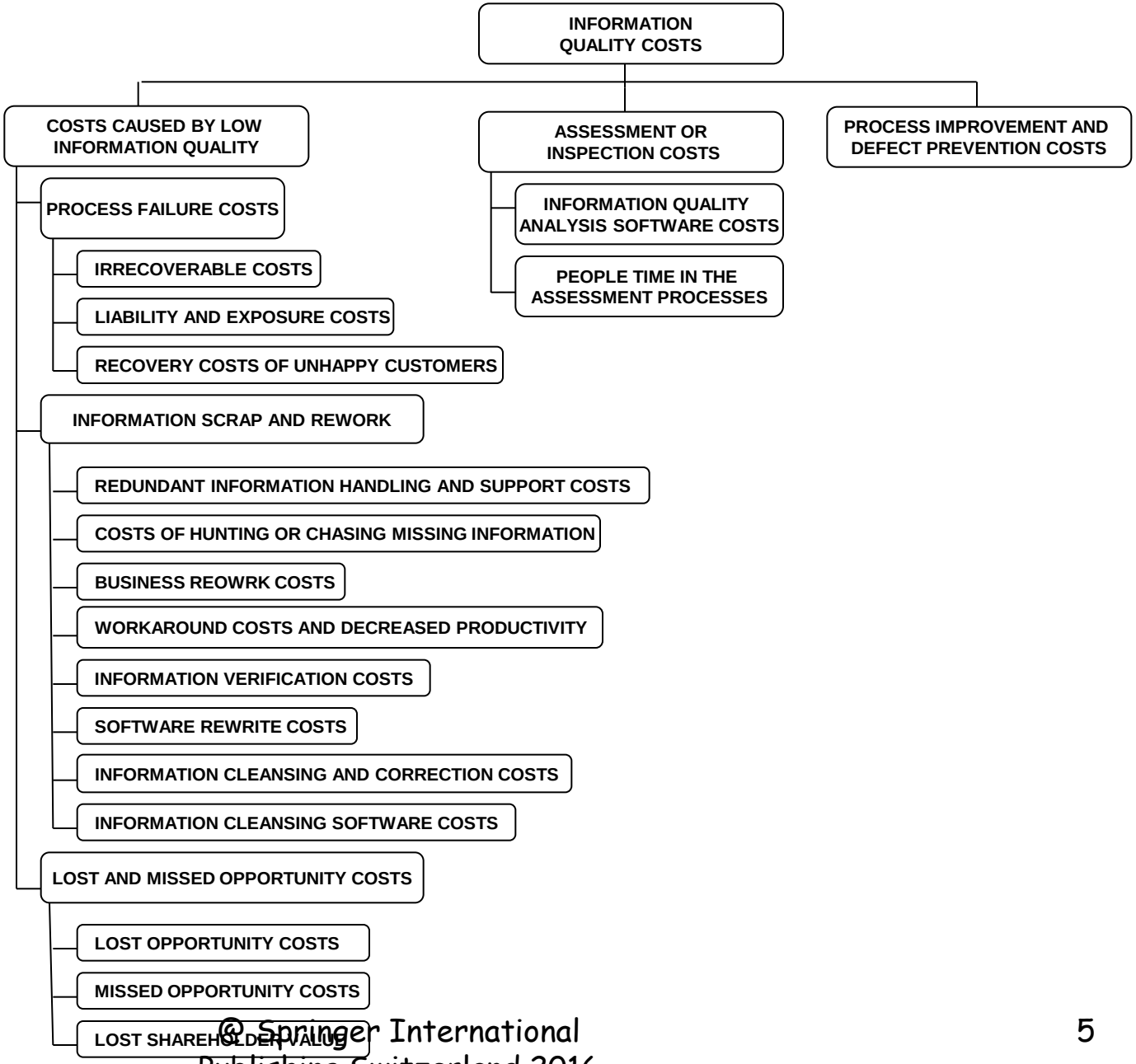
ID	Date	Customer code	Product code	Quantity	Price	Amount
1	June 7, 2015	C	X	20	€ 5.000	€ 100.000
2	June 7, 2015	B	Y	3	€ 1.000	€ 3.000
3	June 8, 2015	A	Y	1	€ 1.000	€ 1.000
4	June 8, 2015	B	Z	5	€ 3.000	€ 15.000

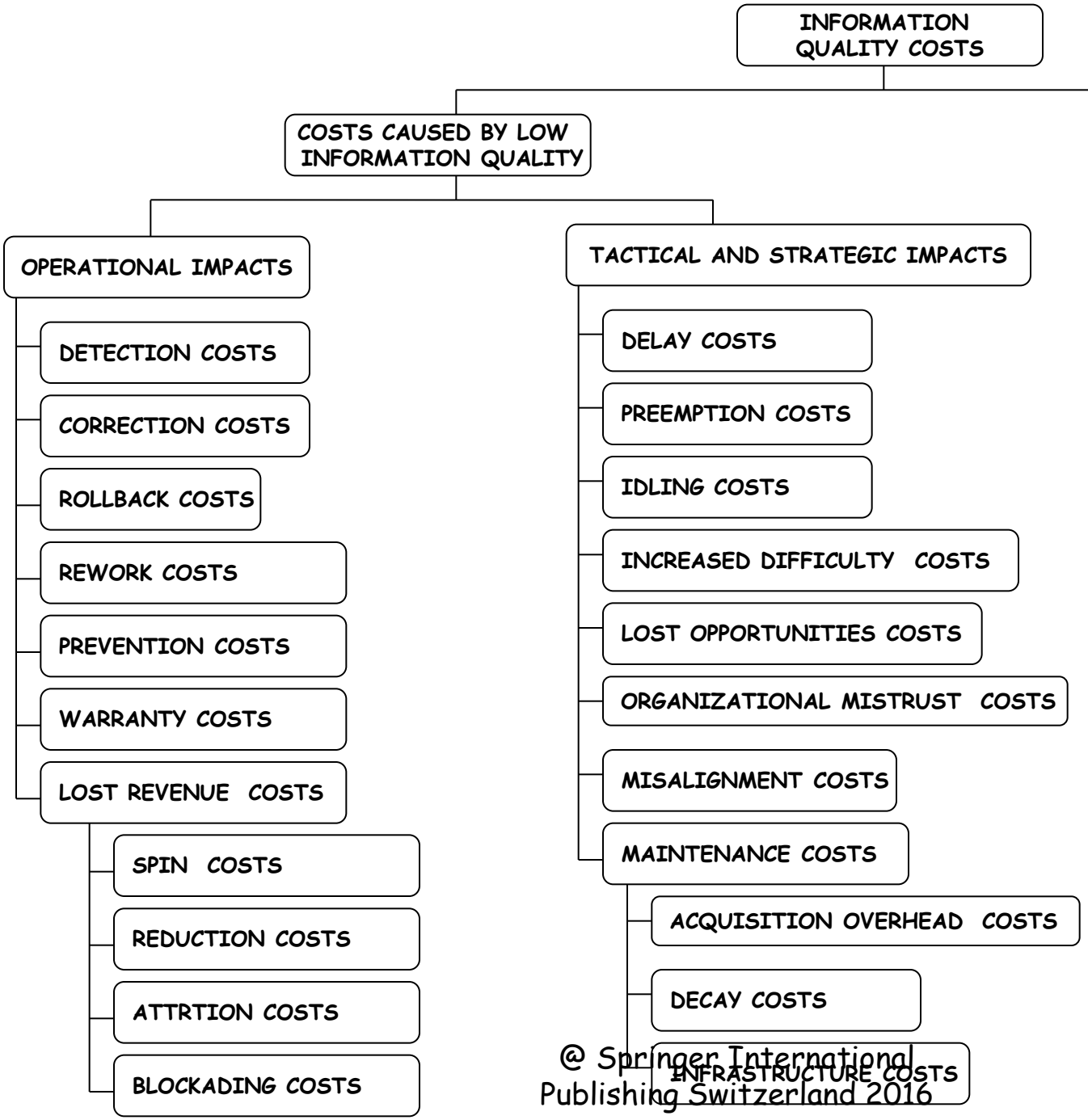
b. Dataset actually delivered

Alumni profile example from [212]

ID	Gender	Marital Status	Income Level	Record Complete (Absolute)	Record Complete (Grade)	Last Update	Recent Update	Up-to-date rank	Inclination	Amount
A	Male	Married	Medium	1	1	2015	1	1	1	200
B	Female	Married	NULL	0	0.667	2012	0	0.47	1	800
C	NULL	Single	NULL	0	0.333	2013	0	0.78	0	0
D	NULL	NULL	NULL	0	0	2005	0	0.08	0	0
								Total	2	1.000

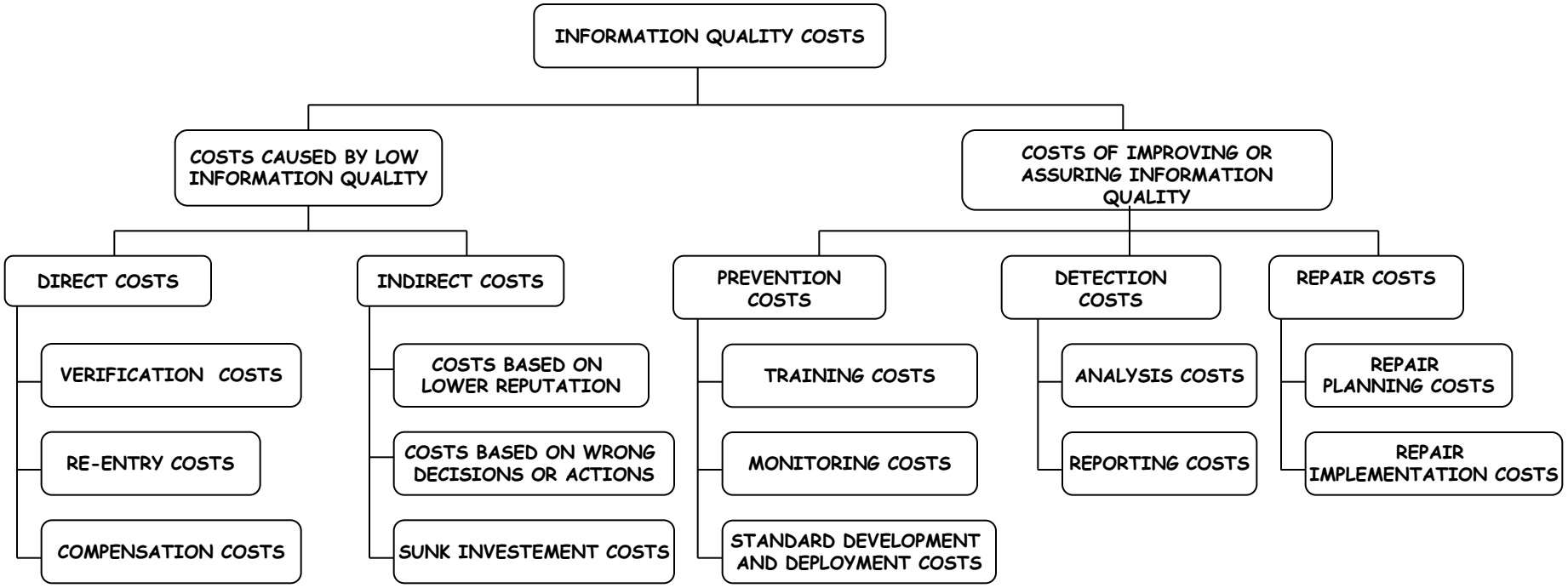
The English classification



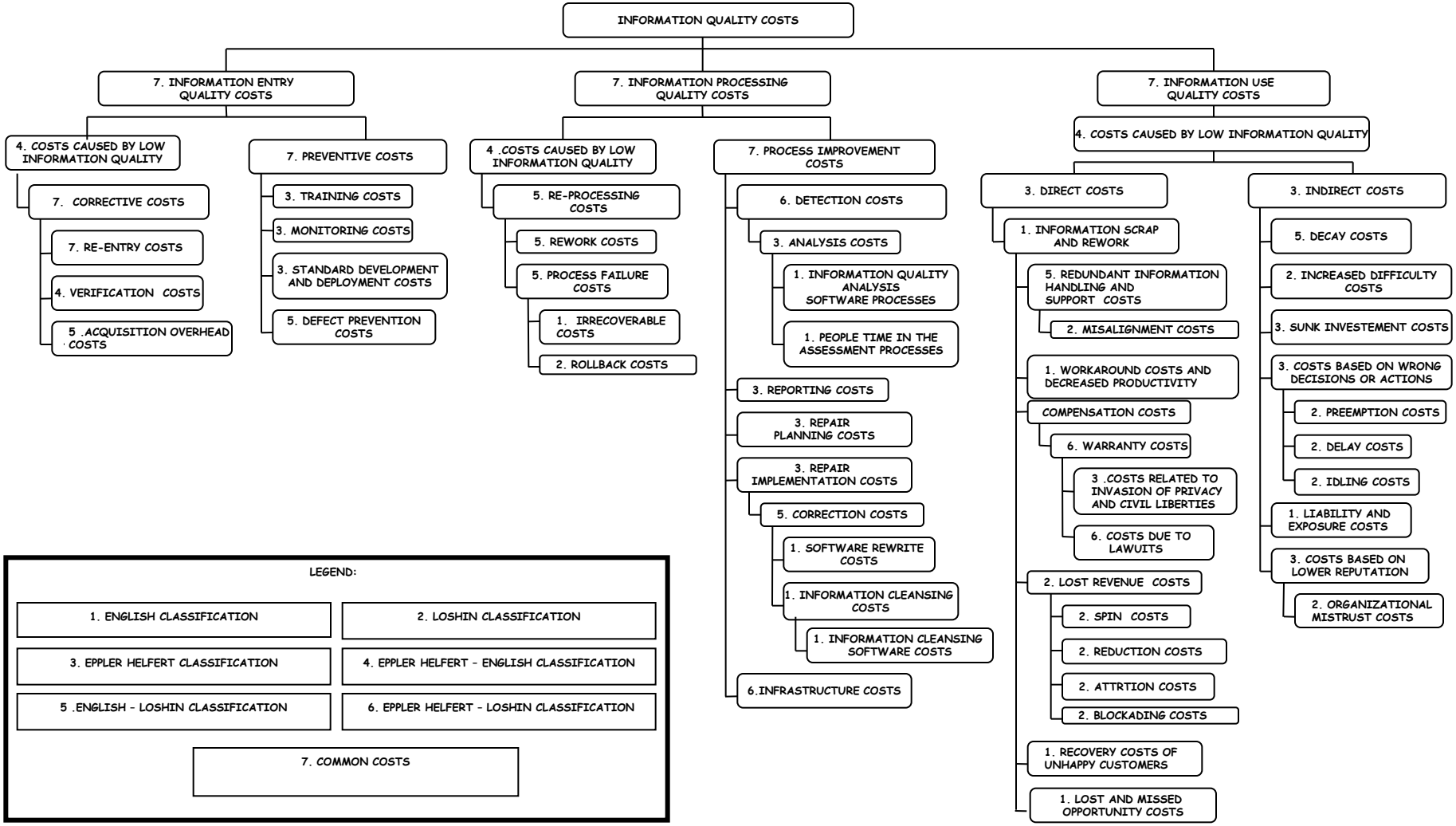


The Loshin classification

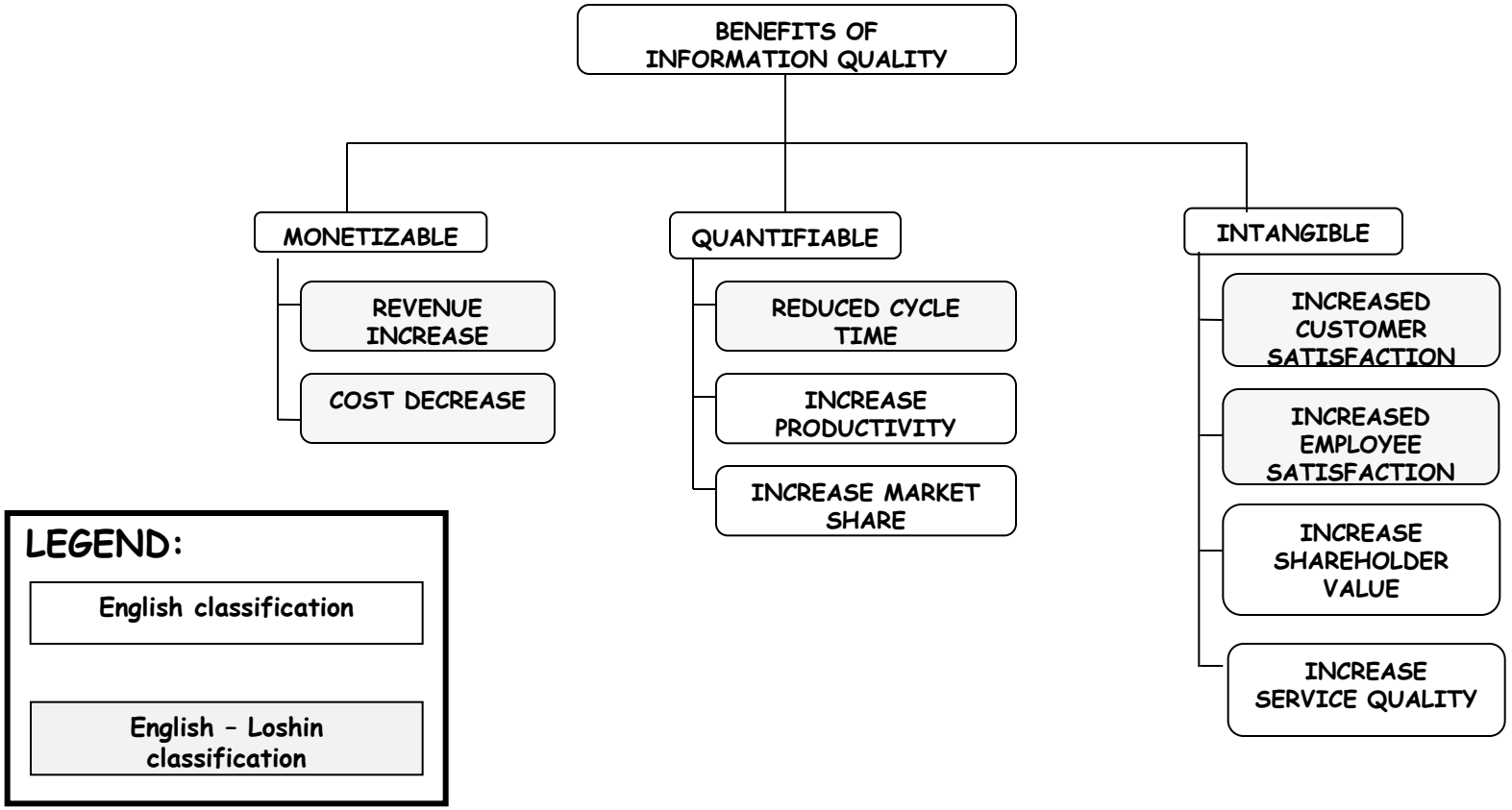
The EpplerHelfert classification



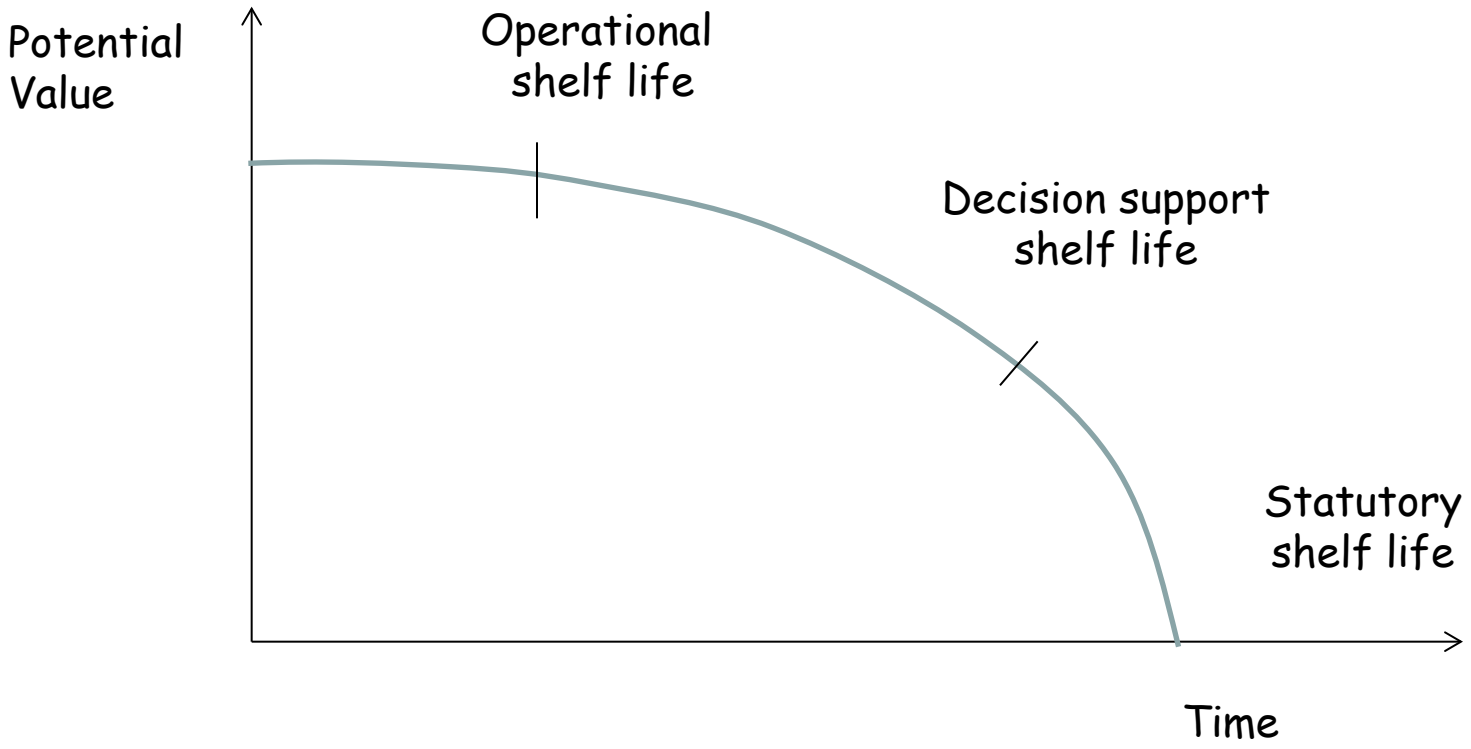
A comparative classification for costs



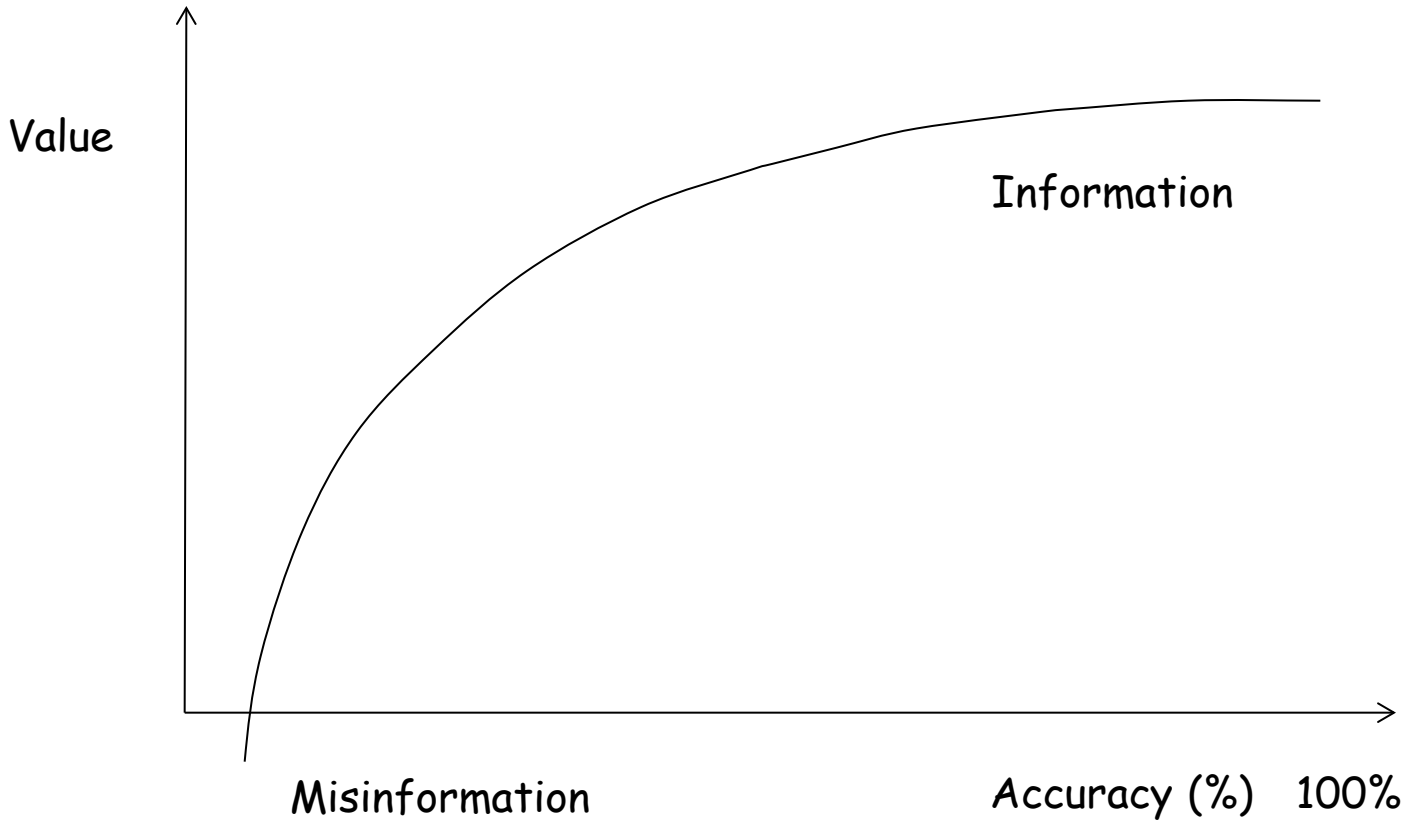
A comparative classification for benefits



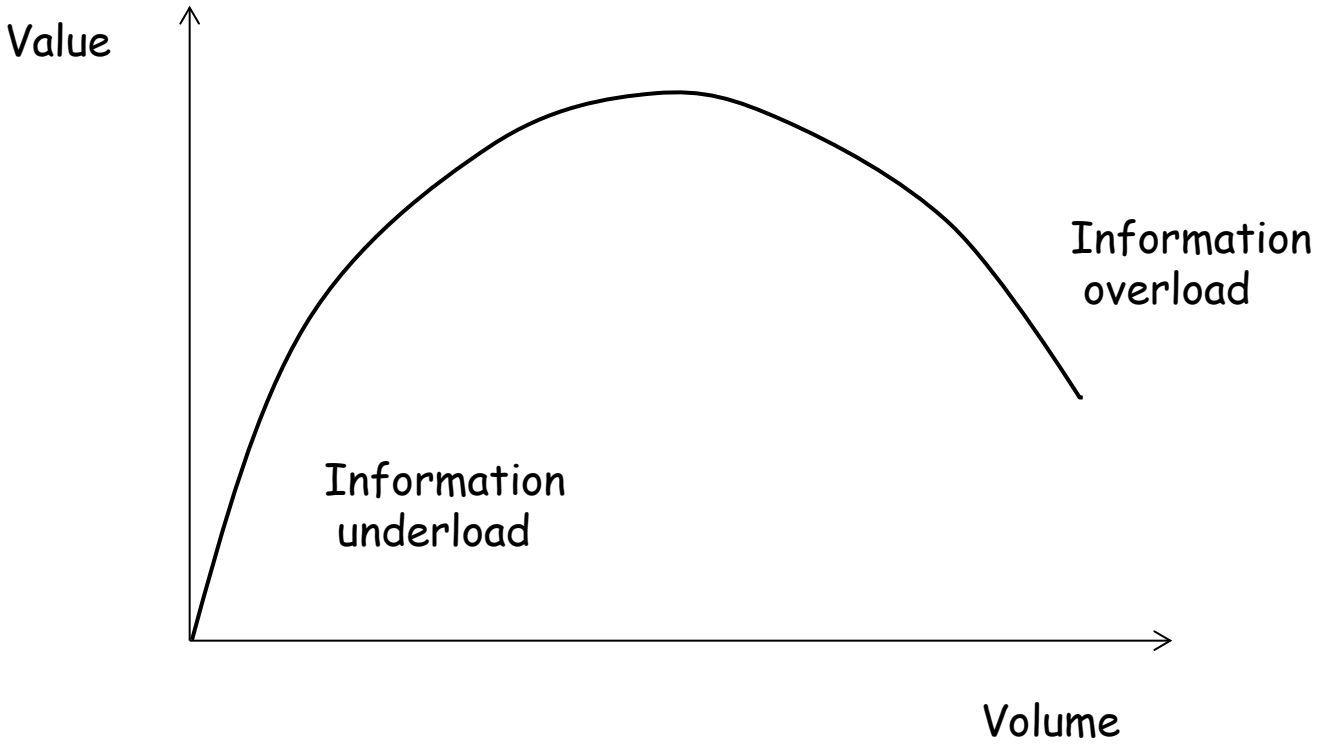
Law 3: Information is perishable, from [449]



Law 4: The value of information increases with accuracy, from [449]



Law 5: More is not necessarily better, from [449]



Integer programming formulation proposed in [34]

Value of Project L = $\sum_{\text{All I}}$ Weight(I) $\sum_{\text{ALL J}}$ $\sum_{\text{All K}}$ Utility(I,J,K;L)

Maximize: Total Value from all projects

$$\sum_{\text{All L}} X(L) * \text{Value}(L)$$

Resource Constraint: $\sum_{L} X(L) * \text{Cost}(L) \leq \text{Budget}$

Exclusiveness Constraint: $X(P(1)) + X(P(2)) + \ldots + X(P(S)) \leq 1$

Interaction Constraint: $X(P(1)) + X(P(2)) + X(P(3)) \leq 1$

Integer Constraints: 1 if project L is selected; 0 otherwise

$$X(L) = \begin{cases} 0 \\ 1 \end{cases}$$

The effect of relation completeness on utility, cost, and net-benefit

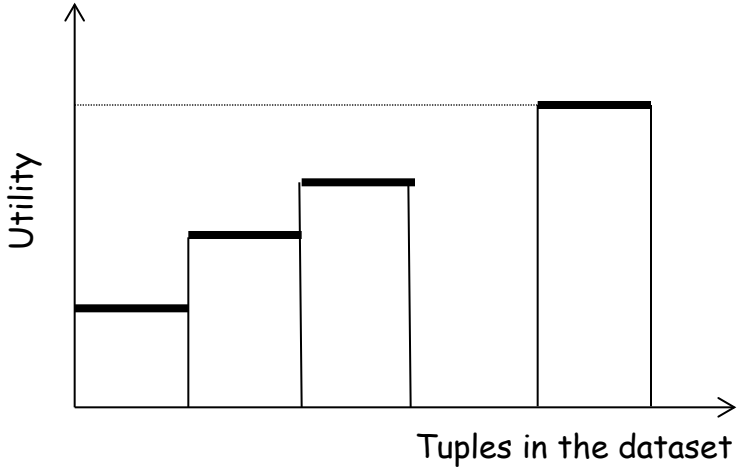


Figure a

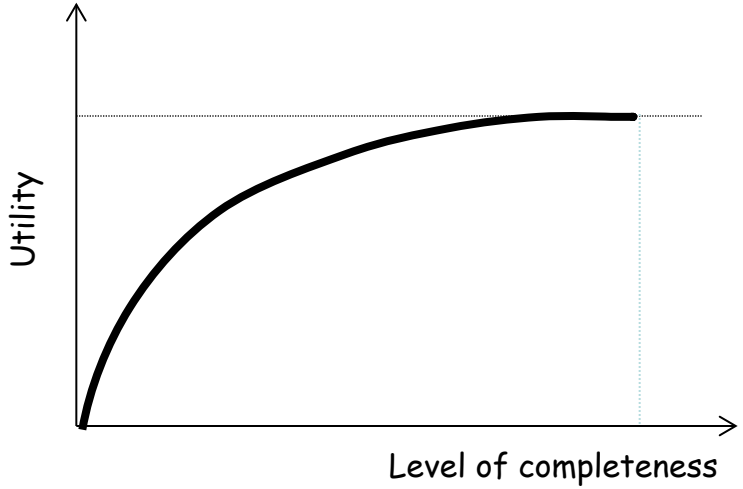


Figure b

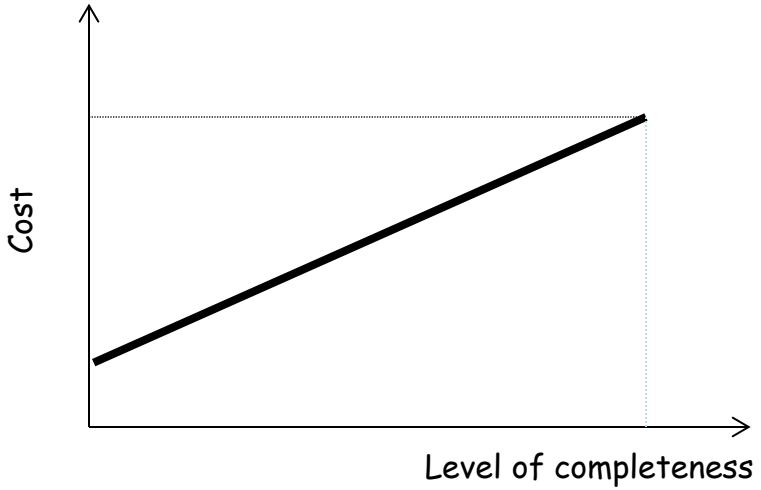


Figure c

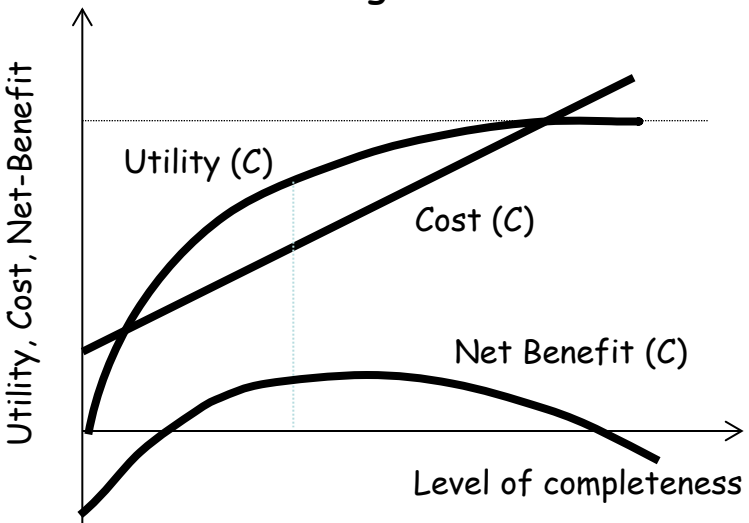
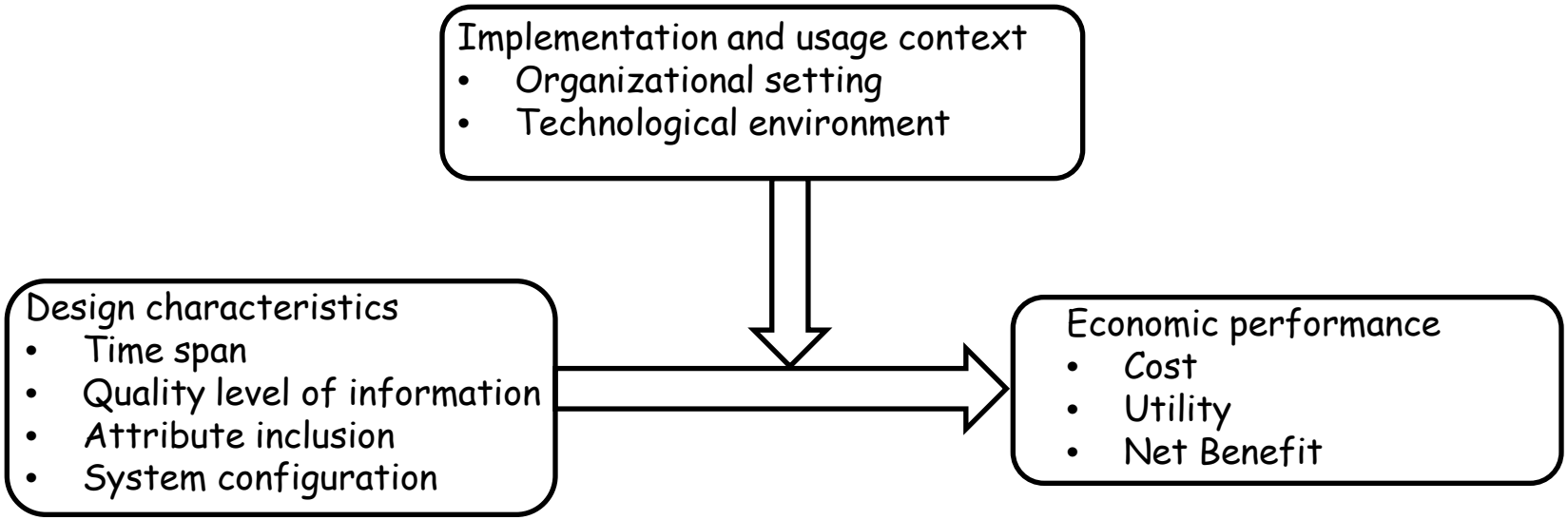
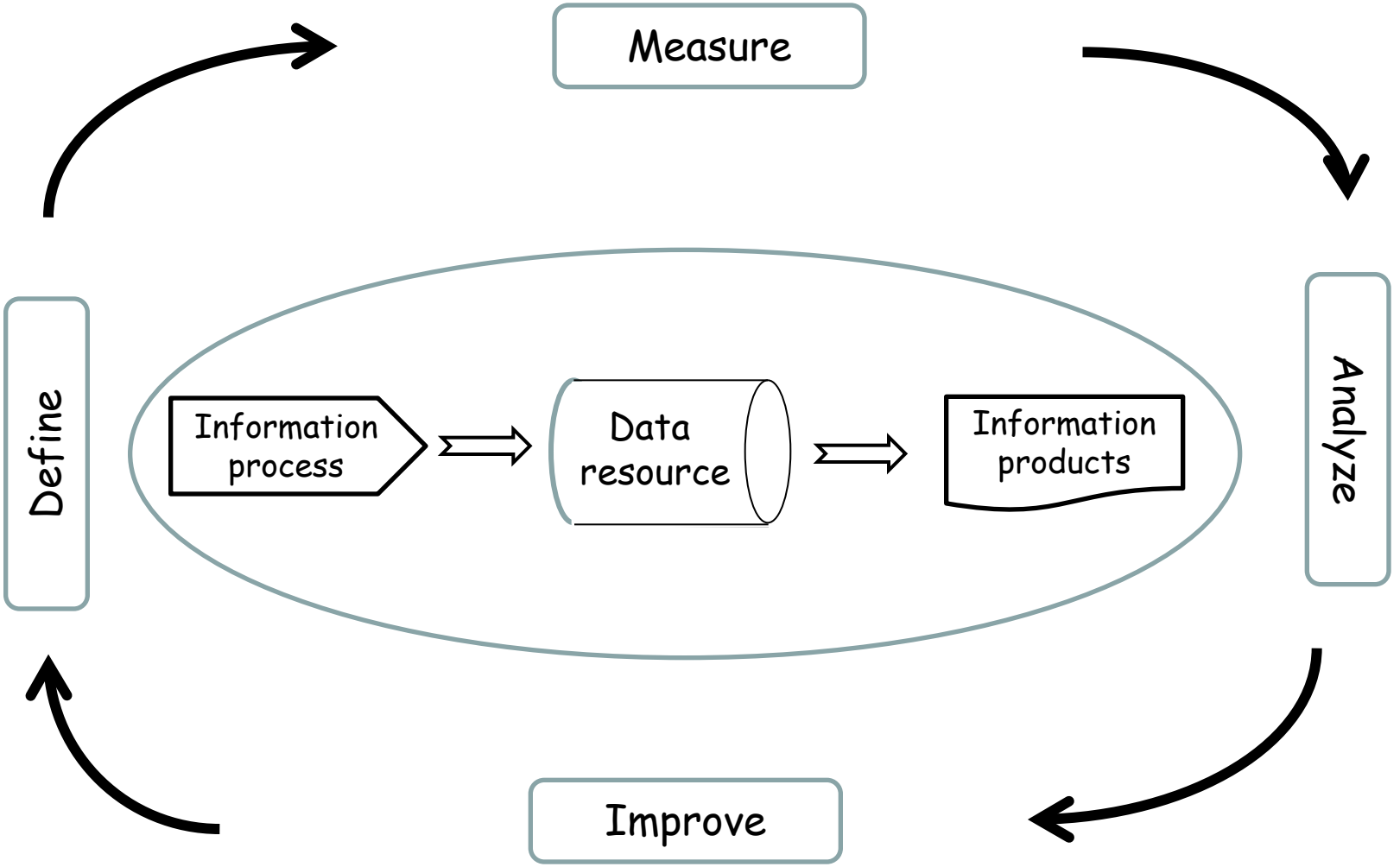


Figure d

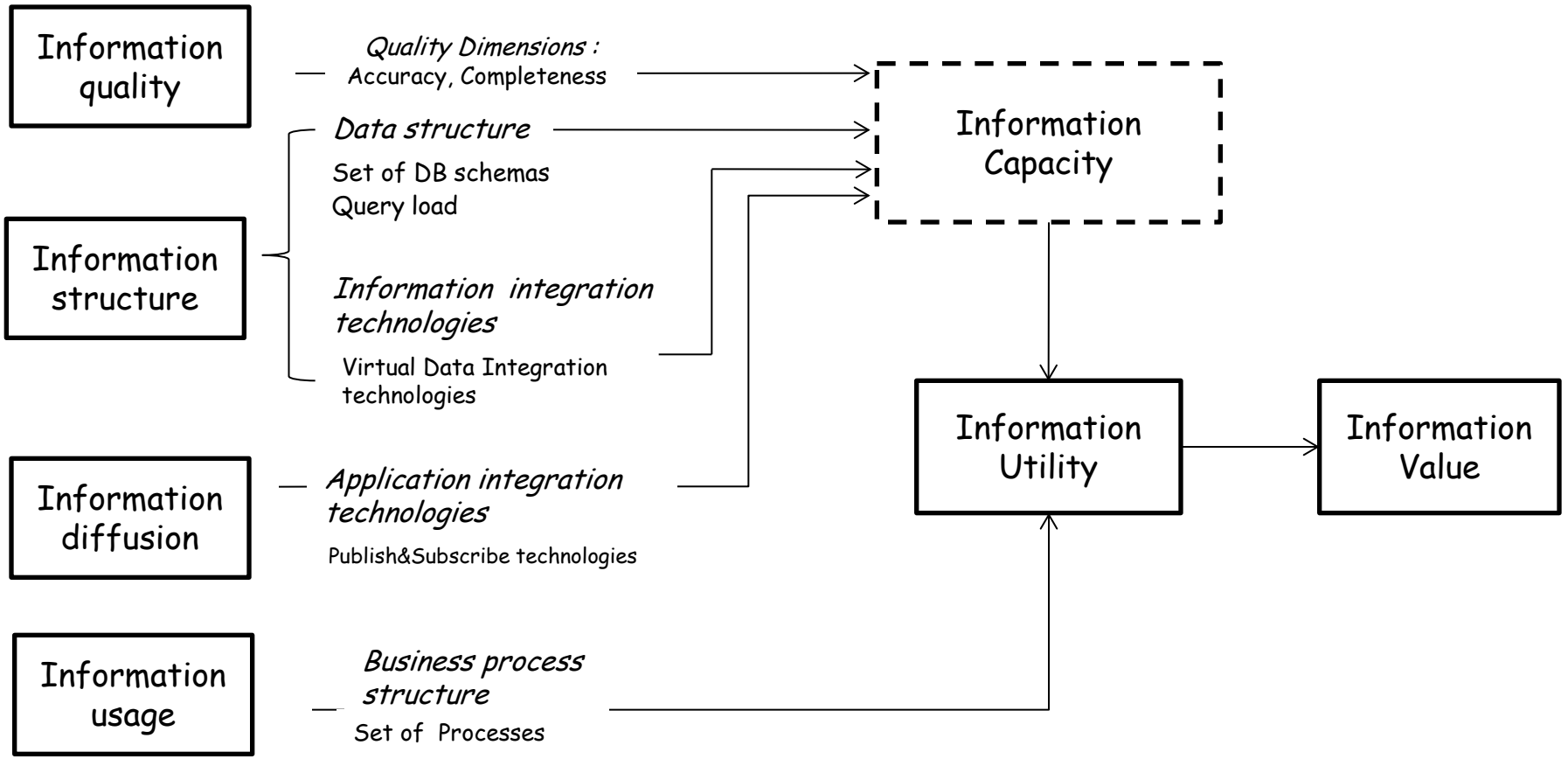
Net-benefit maximization framework in [216]



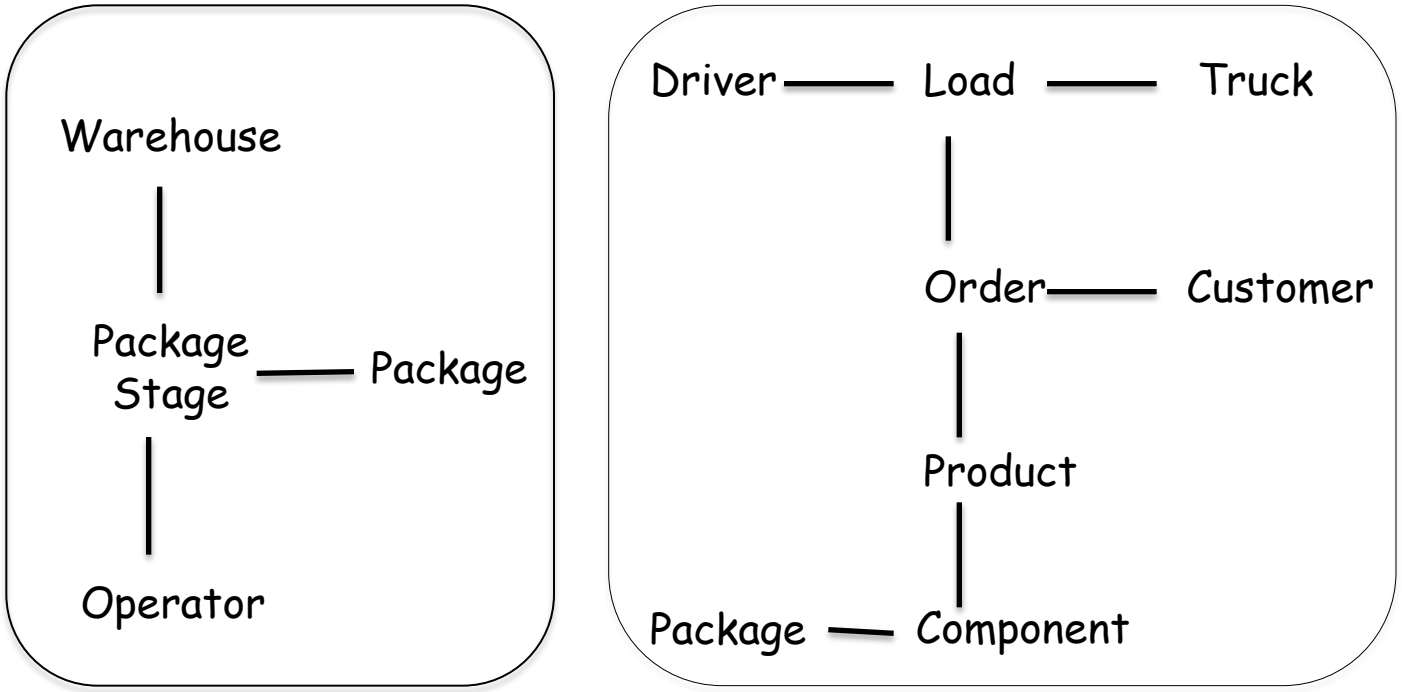
A framework for the assessment/improvement IQ life cycle



A unified model of information quality, capacity, utility and value



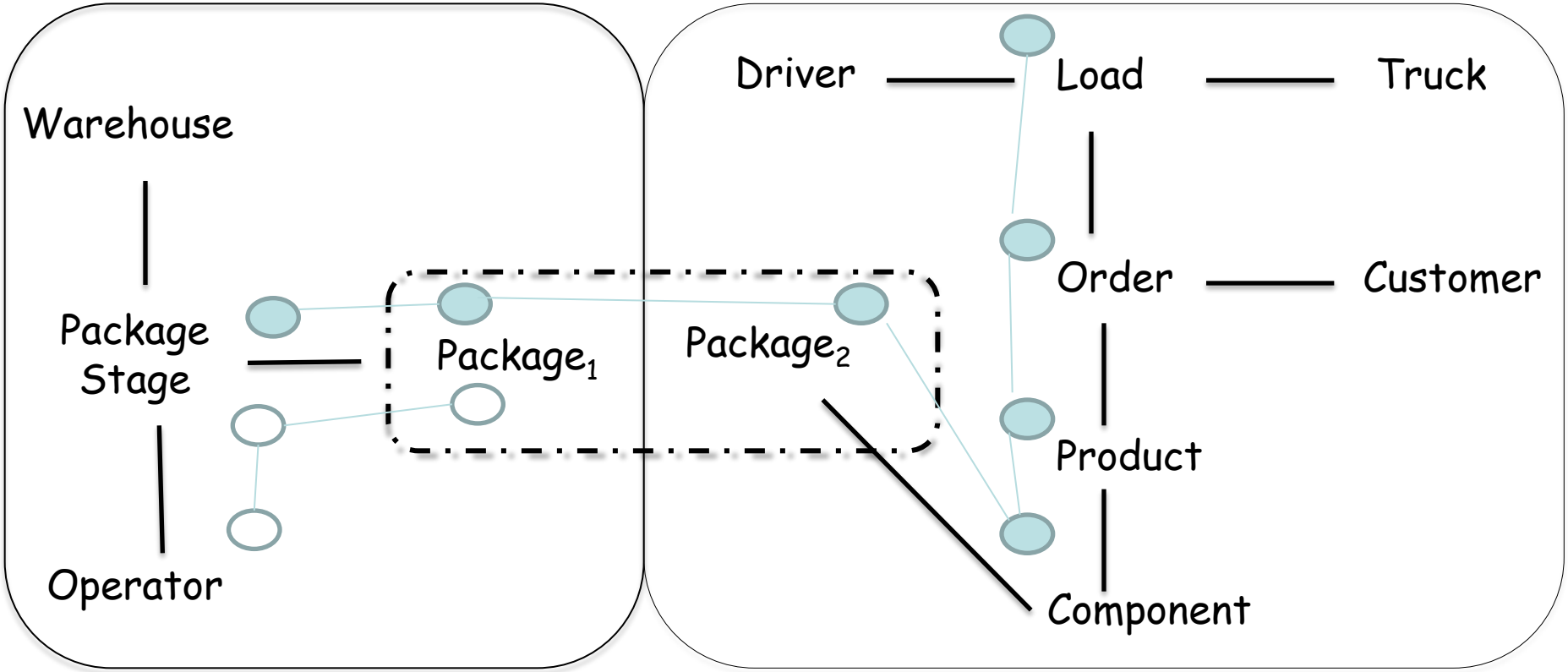
Two distinct databases of a furniture company



DB1

DB2

Integrated schema and new queries that can be performed on it



a. Integrated schema

Q1. Progress of a package whose order is scheduled for a given load

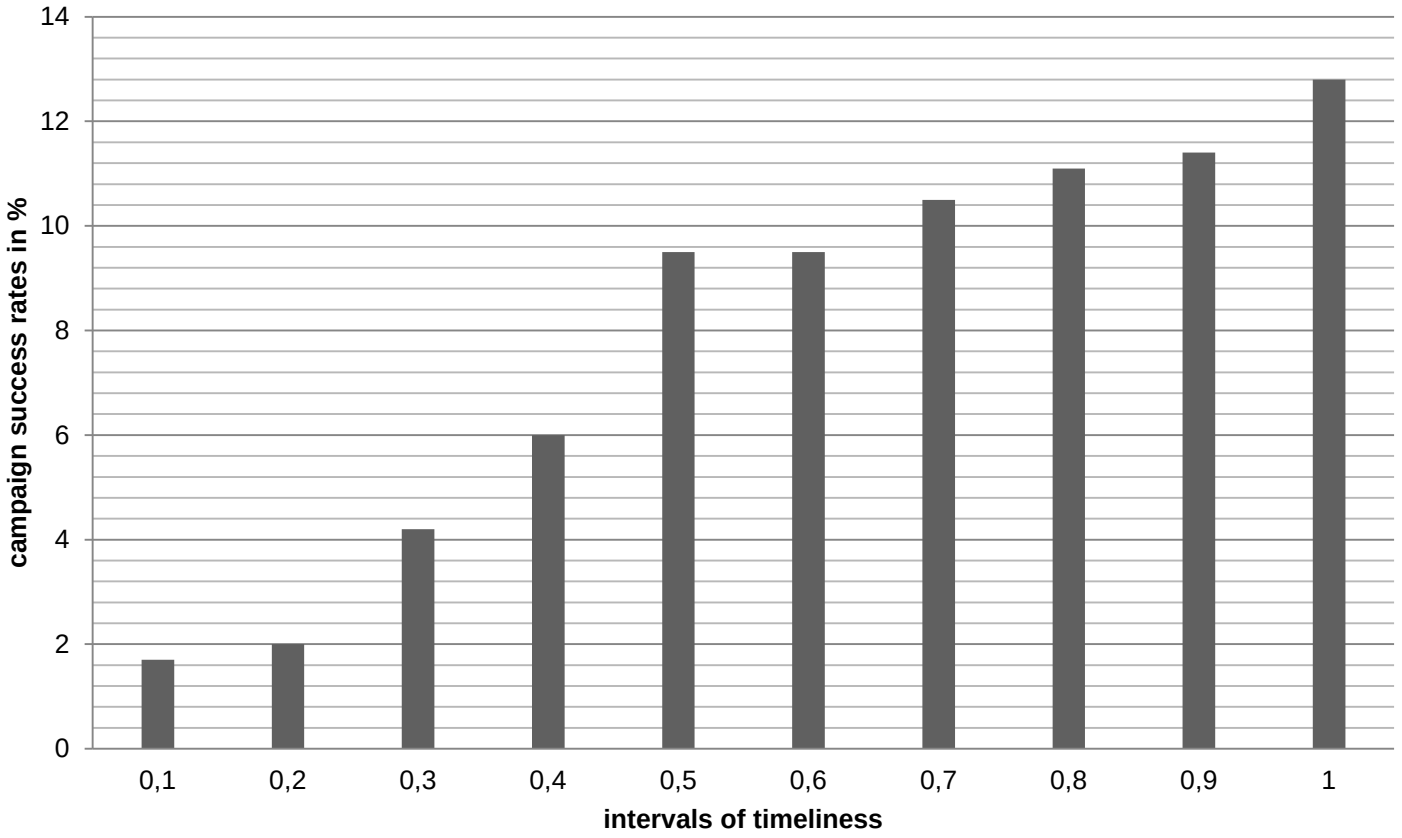
Q2. Packages managed by a given operator

b. Queries with higher utility

Evaluation of relevance and timeliness for the attributes in the table

$Attribute_i$	Surname	First Name	Address	Current Tariff
$relevance_i$	0.9	0.2	0.9	1.0
decline(Ai) [1/year]	0.02	0.0	0.1	0.4
age(Ai) [year]	0.5	0.5	2	0.5
$Q_{timeliness}$ (Ai)	0.99	1.00	0.82	0.82

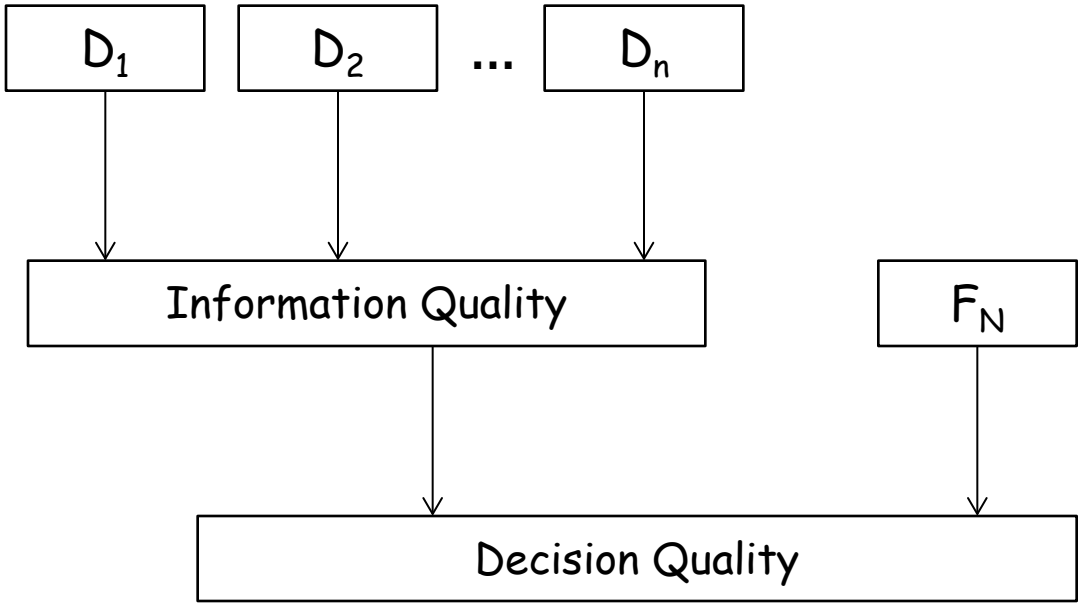
Success rate of a former campaign



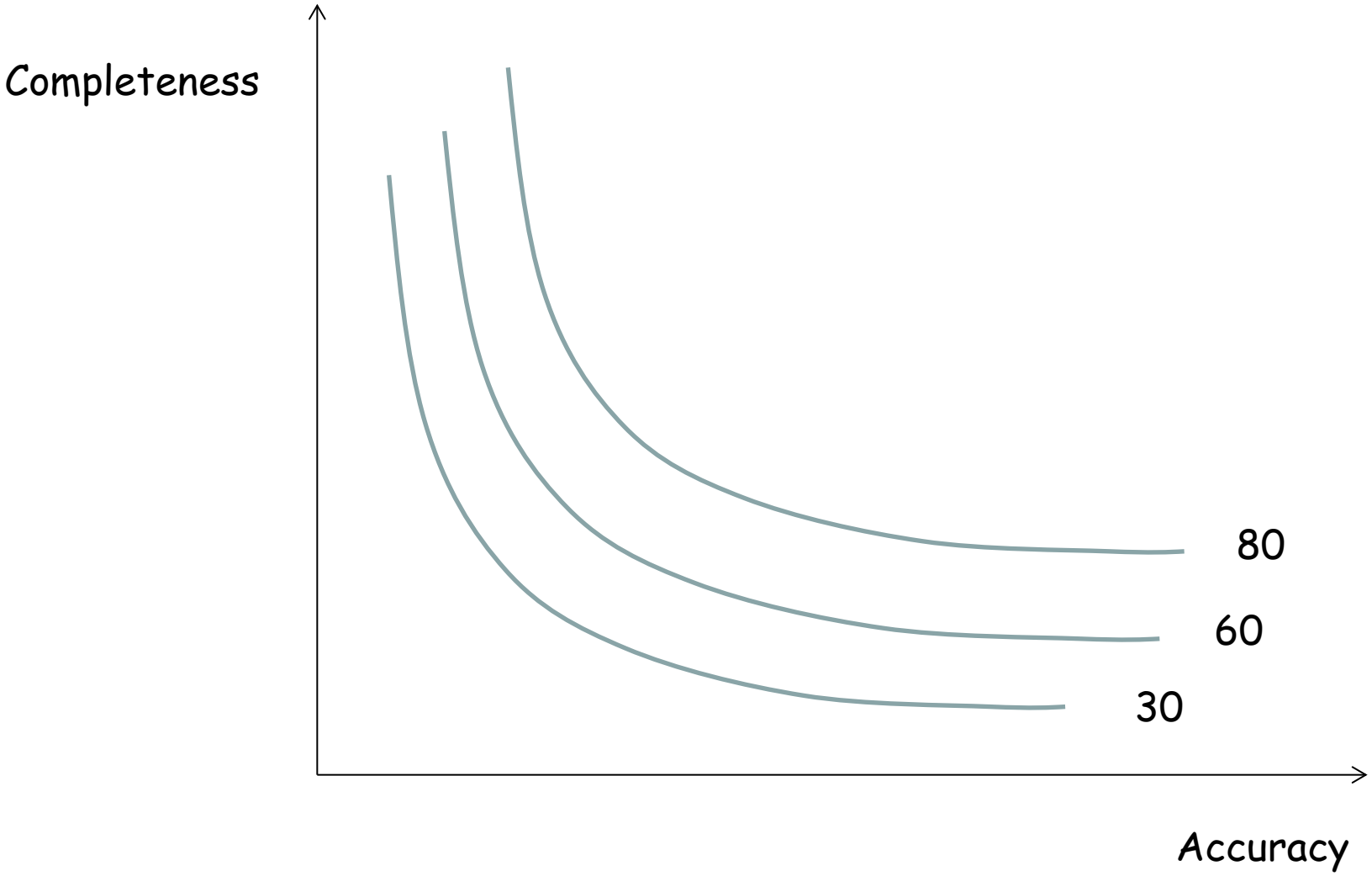
Main papers addressing the relationship between information quality and decision making

Paper	Independent variable	Measured as	Dependent variable	Modeled as	Domain
Jarvenpaa 1985	IQ dimensions	- Intepretation accuracy - Measurement validity - Consistency	Decision performance	- Display format - Task complexity	Managerial decision
Gonzales 1997	IQ dimension	Clarity of the animation	Decision quality	% of correct answers	- Rental decision - Fluidynamics problem
Ahituv 1998	IQ dimension	Completness	Decision efficiency	Number of enemy aircrafts hits	Reaction to an hostile air attack
Raghunathan 1999	IQ dimension	Accuracy	Decision quality	- Closeness of beleif output - Probability of output	
Chengalur-Smith 1999	- IQ metrics - Experience - Time	Reliability of information	Decision making outcome	Choice of best apartment	- Apartment selection - Restaurant site selection
Fisher 2003	Metadata on IQ	Present/ not present	Decision making outcome	- Complacency - Consensu - Consistency	- Apartment selection - Job transfer
Jung 2005	- IQ category - IQ dimensions	- Contextual quality - Completness/ Relevance/Aggregation	Decision quality	# of correct answers	Restaurant site selection
Ge 2006	IQ dimensions	- Accuracy - Completness	Decision quality → Decision effectiveness	% of right decisions	Investment decision
Shankarana-rayan 2006	- Metadata on data processing - Quality assessment	Accuracy, completness, currency, consistencym, relevance	Decision making outcome	Perceived usefullness	Allocation of advertising budget

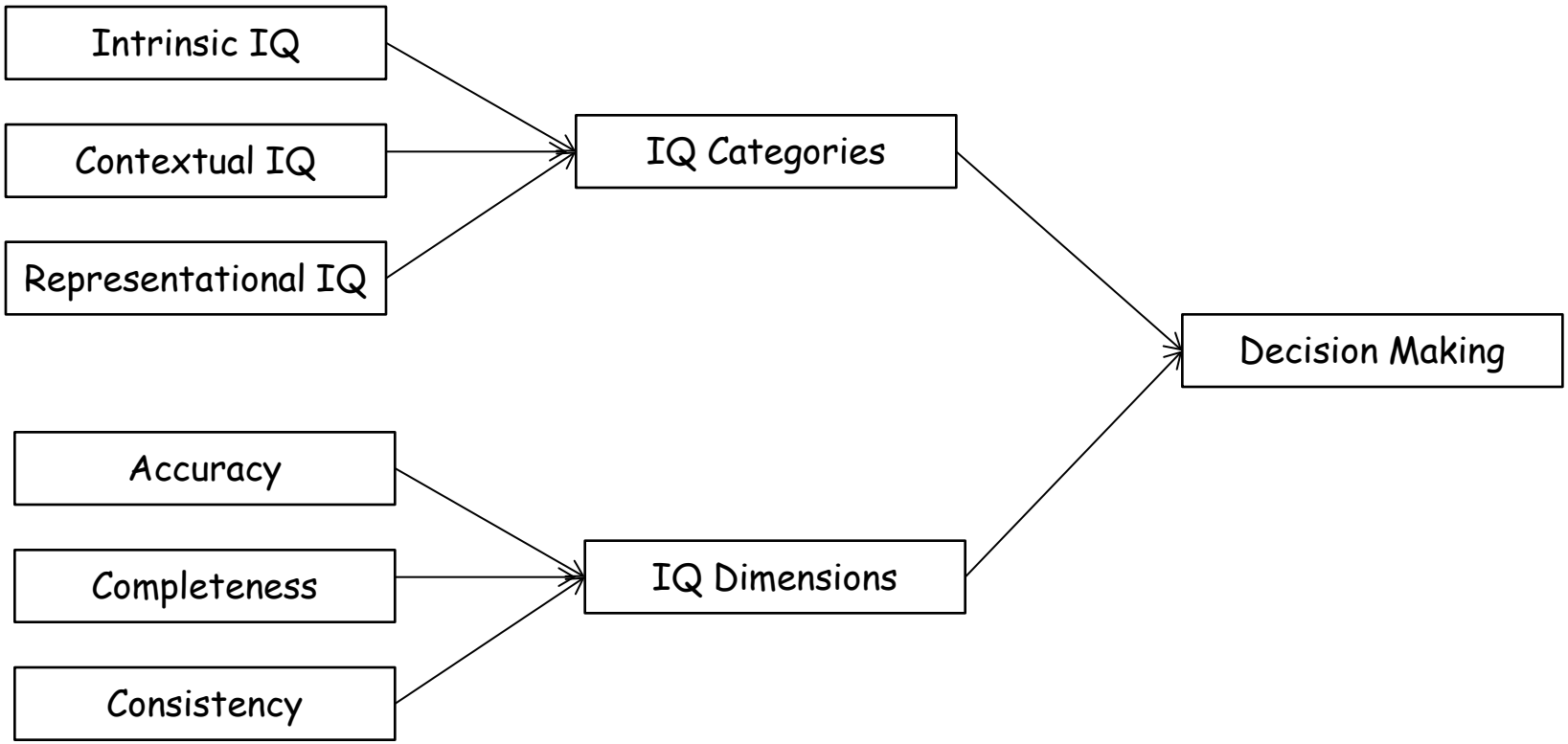
General approach to factors influencing decision quality proposed in [260]



Decision quality contours as a function of completeness and accuracy



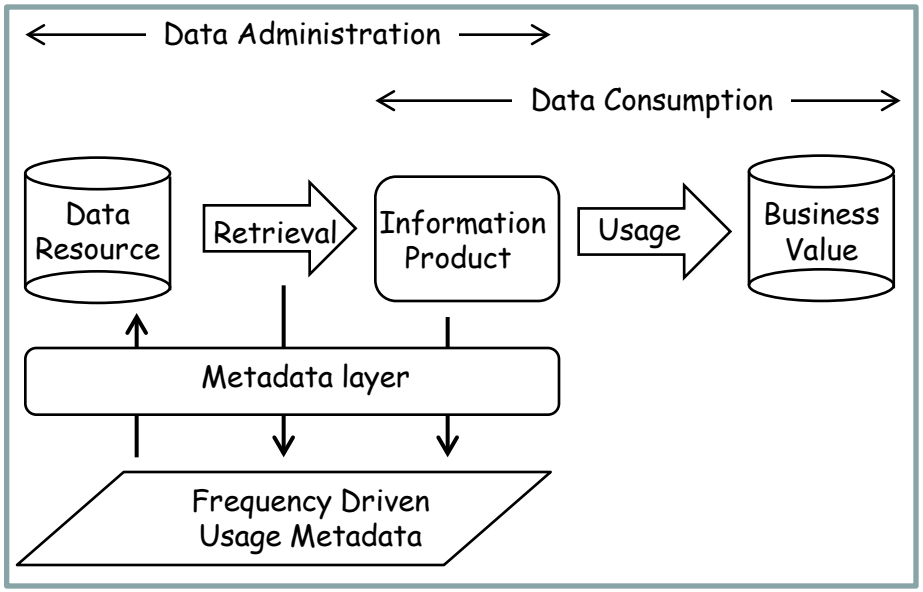
Model proposed in [256]



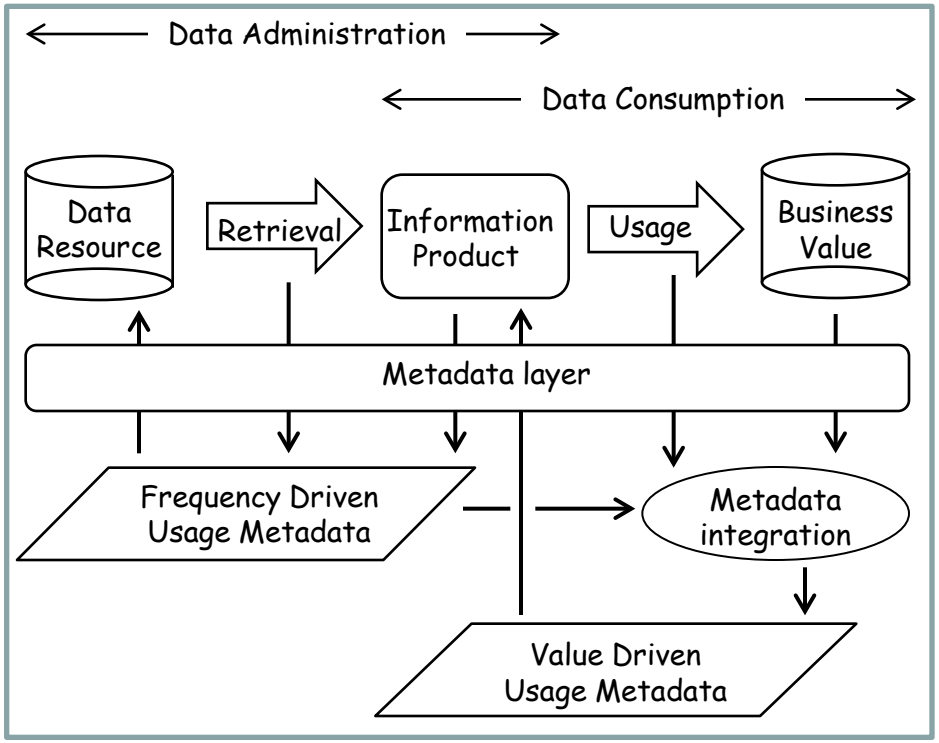
Scenarios proposed in [256] and [258]

	Order Complexity	Objective	Optimal Decision
Scenario 1	•One identical brand of beer over 10 weeks •One decision in each week	Minimize inventory	Zero inventory
Scenario 2	•Order 10 different brands of beer •Make one decision for each brand	Minimize total costs	Minimal total costs

Frequency driven vs value driven usage metadata



a. Frequency Driven Usage Metadata



b. Value Driven Usage Metadata

Example from [375] and frequency driven usage metadata

Customers

#	Customer	Gender	Income	Children	Status	Frequency
1	James	Male	High	0	Single	1
2	Sarah	Female	Low	1	Married	2
3	Isaac	Male	Medium	2	Married	1
4	Rebecca	Female	Low	0	Single	1
5	Jacob	Male	Medium	3	Married	1
6	Lea	Female	High	2	Married	3
7	Rachel	Female	Low	4	Single	0
Frequency		3	1	2	1	

Queries

WHERE Condition	Attributes Used	Tuples Retrieved
Gender = "Male" and Children > 0	Gender, Children	[3], [5]
Gender = "Female" and Children < 3	Gender, Children	[2], [4], [6]
Gender = "Female" and Status = "Married"	Gender, Status	[2], [6]
Income = "High"	Income	[1], [6]

Value driven usage metadata from [375]

Customers

#	Customer	Gender	Income	Children	Status	Value
1	James	Male	High	0	Single	1
2	Sarah	Female	Low	1	Married	2
3	Isaac	Male	Medium	2	Married	1
4	Rebecca	Female	Low	0	Single	1
5	Jacob	Male	Medium	3	Married	1
6	Lea	Female	High	2	Married	3
7	Rachel	Female	Low	4	Single	0
……						
Value		515	2.000	60	500	

Queries

WHERE Condition	Attributes Used	Tuples Retrieved	Total Value
Gender = "Male" and Children > 0	Gender, Children	[3], [5]	100
Gender = "Female" and Children < 3	Gender, Children	[2], [4], [6]	30
Gender = "Female" and Status = "Married"	Gender, Status	[2], [6]	1000
Income = "High"	Income	[1], [6]	2000