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

Chapter 1
Introduction to Information Quality

A relation Movies
with data quality problems

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

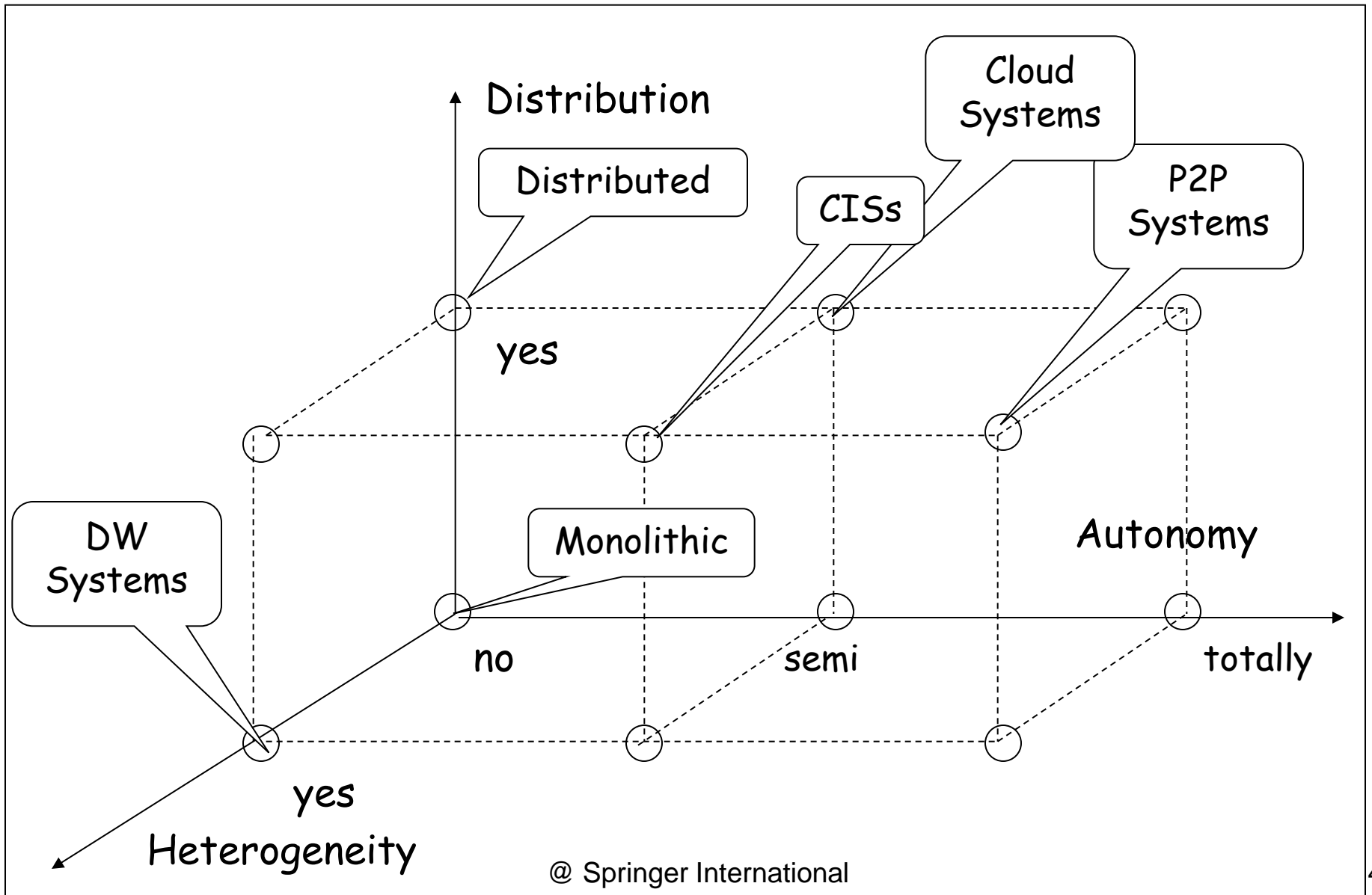# Two images of the same flower with intuitively different image quality levels
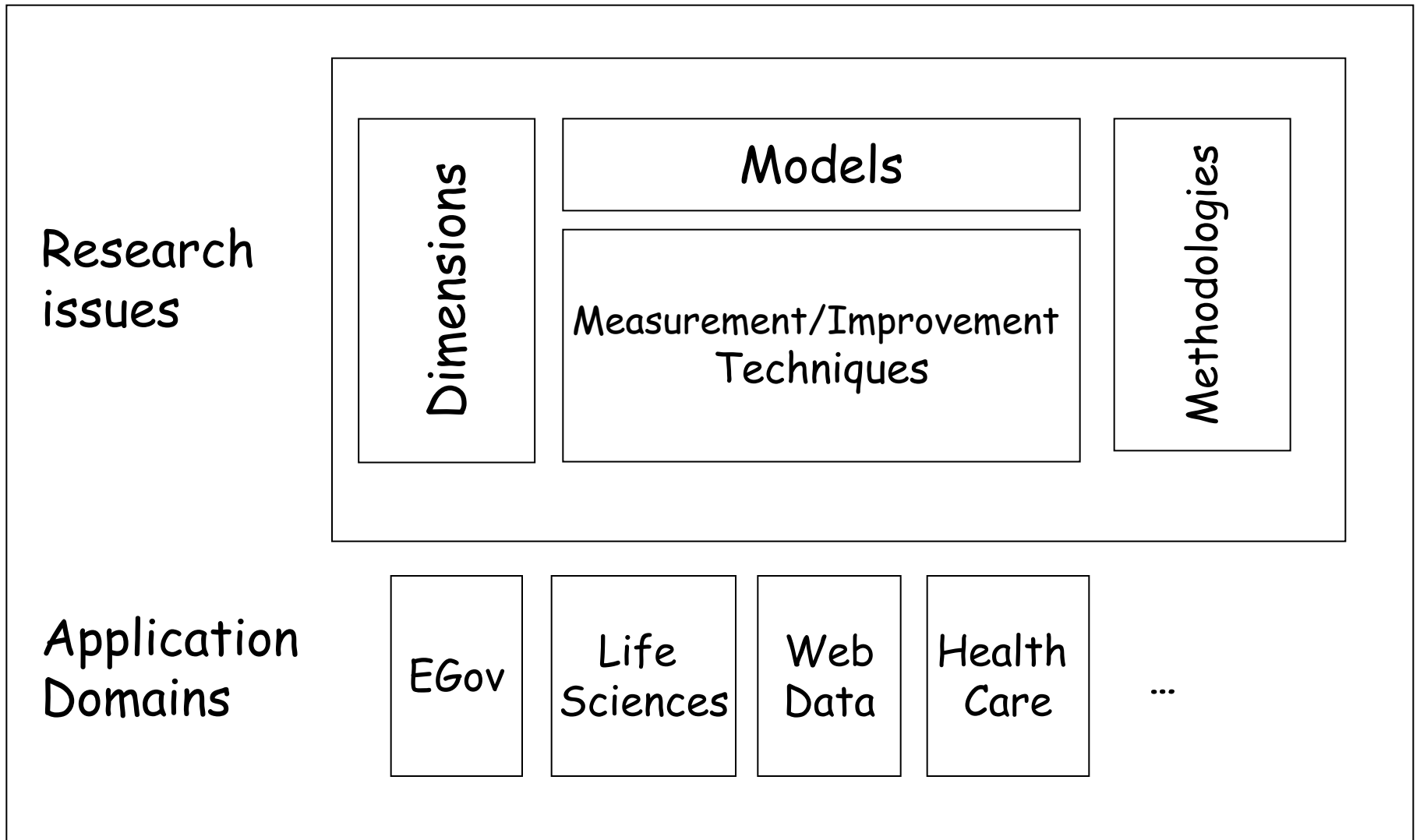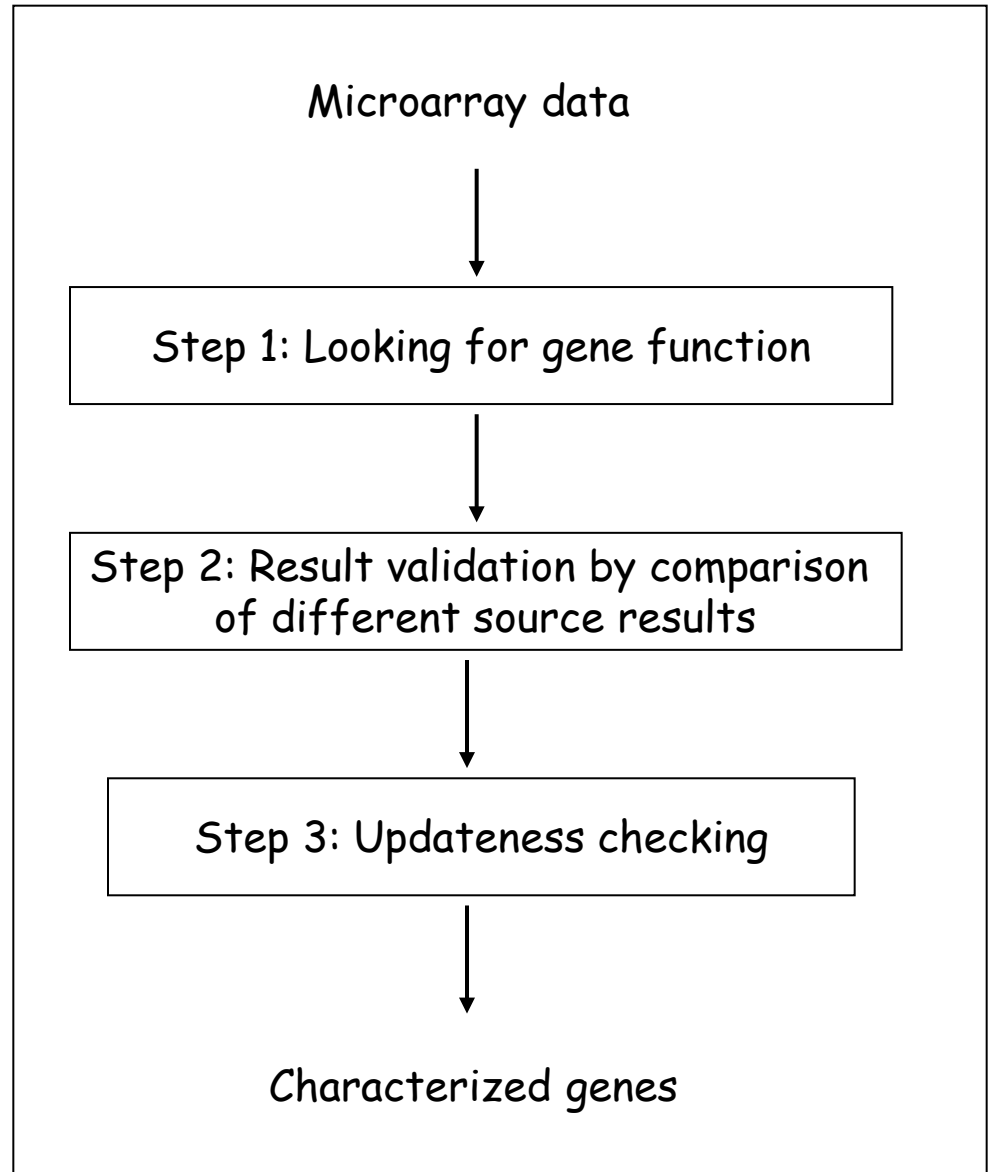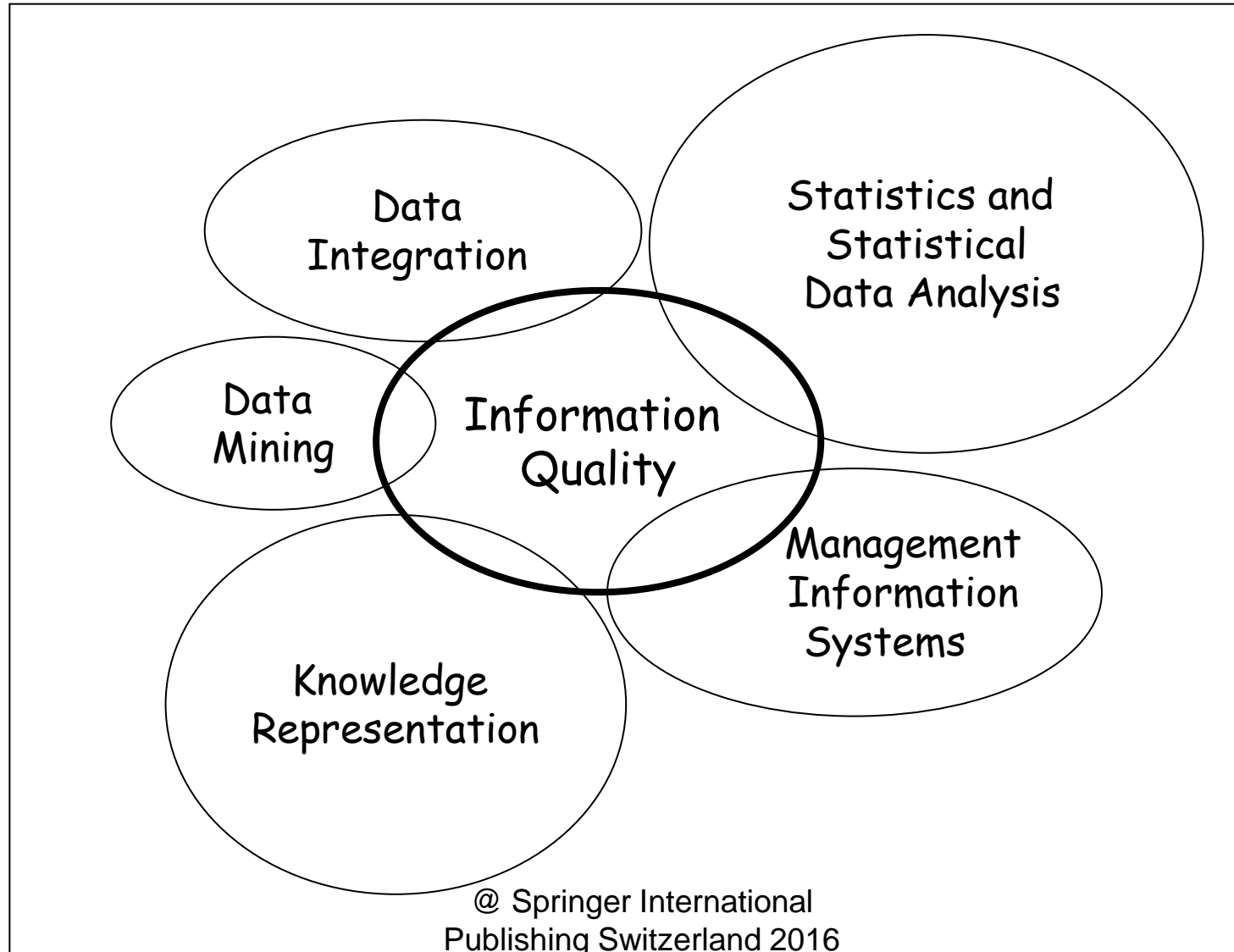

Bad quality

Good quality

3

Types of information systems

Main issues in information quality

Research issues

Dimensions	Models	Methodologies
	Measurement/Improvement Techniques	

Application Domains

| EGov | Life Sciences | Web Data | Health Care | ... |

Example of biological information analysis process

Microarray data

↓

Step 1: Looking for gene function

↓

Step 2: Result validation by comparison of different source results

↓

Step 3: Updateness checking

↓

Characterized genes

Research areas related to information quality

Data Integration

Statistics and Statistical Data Analysis

Data Mining

Information Quality

Management Information Systems

Knowledge Representation

ISO standard data quality dimensions

DQ characteristic	Definition (all definitions except for completeness and accessibility begin with: the degree to which data has attributes that…")
Correctness	correctly represent the true value of the intended attribute of a concept or event in a specific context of use
Completeness	subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use
Consistency	are free from contradiction and are coherent with other data in a specific context of use
Credibility	are regarded as true and believable by users in specific context of use.
Currentness	are of the right age in a specific context of use
Accessibility	data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability
Compliance	adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use
Confidentiality	ensure that it is only accessible and interpretable by authorized users in a specific context of use
Efficiency	can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use
Precision	are exact or that provide discrimination in a specific context of use
Traceability	provide an audit trail of access to the data and of any changes made to the data in a specific context of use
Understandability	enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use
Availability	enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use
Portability	enable it to be installed, replaced or moved from one system to another preserving the existing quality in a specific context of use
Recoverability	enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use