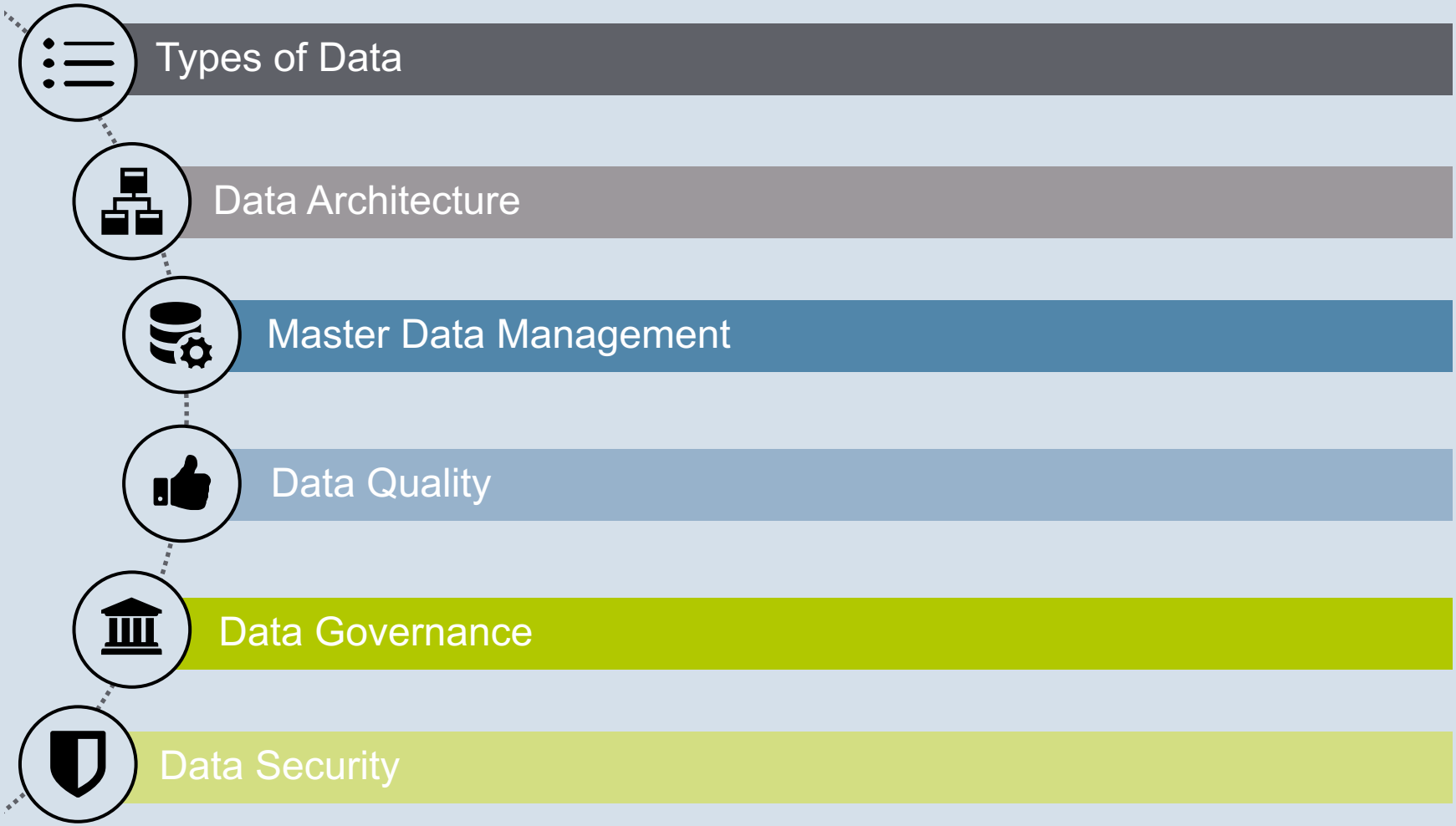
MANNHEIM BUSINESS SCHOOL

MANAGING BIG DATA

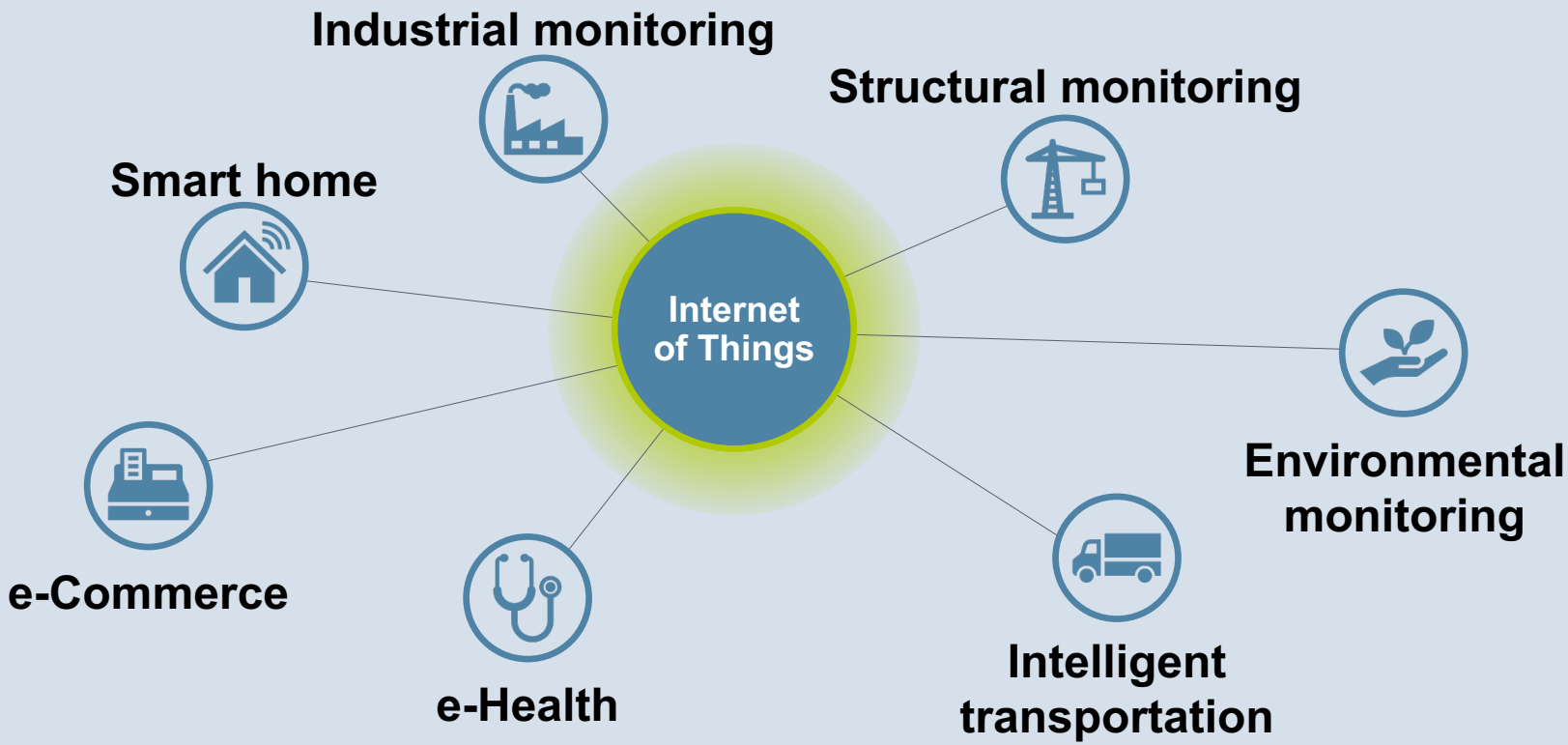
PROF. DR. FLORIAN STAHL

Managing Big Data



Due to the digitalization of business processes and peoples' daily life, humanity produces 2,300 million gigabytes of data every day

Machine-generated Data

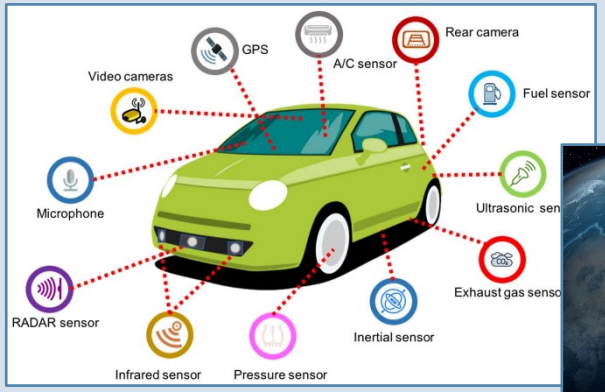


Machine data is the largest source of Big Data!

Machine-generated Data

More data with more complex relationships

… in Real Time and At Scale **… to Manage, to Govern and to Analyze**



Sensors

Satellite Imaging

Video Monitoring

People-generated Data

Mobile Data

Social Media Data

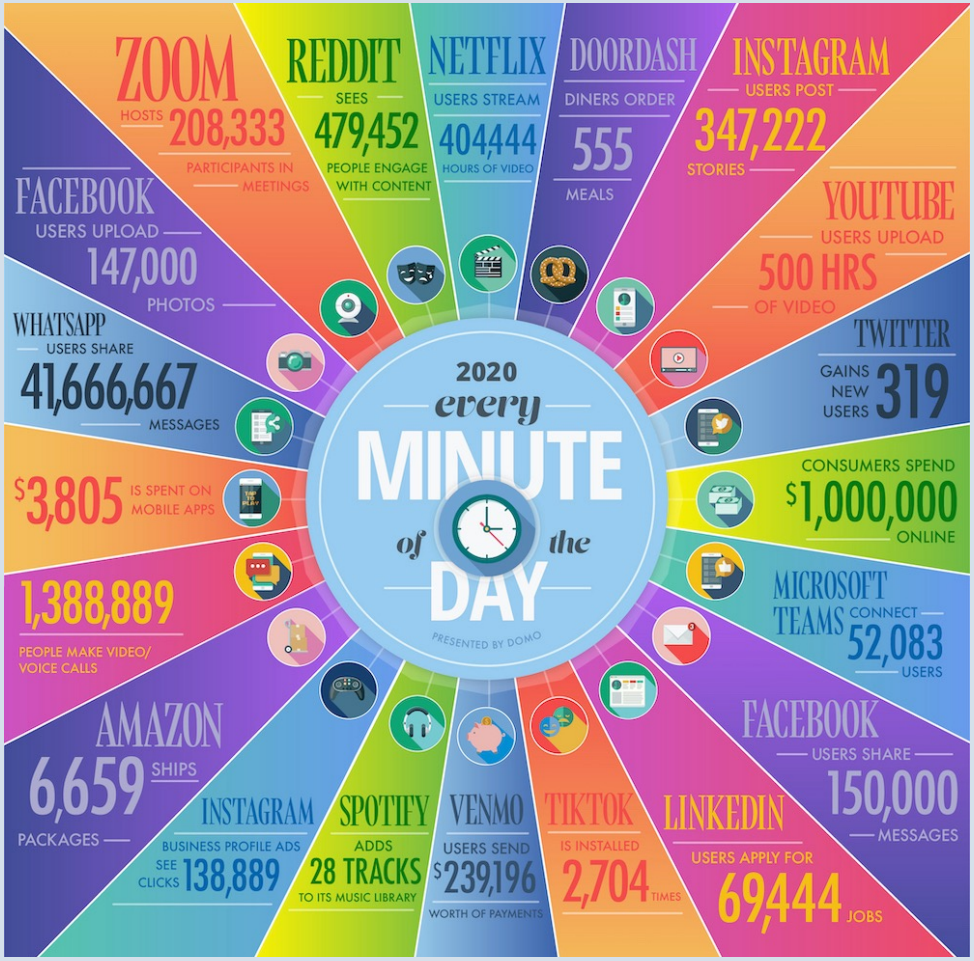
CRM / Loyalty Data

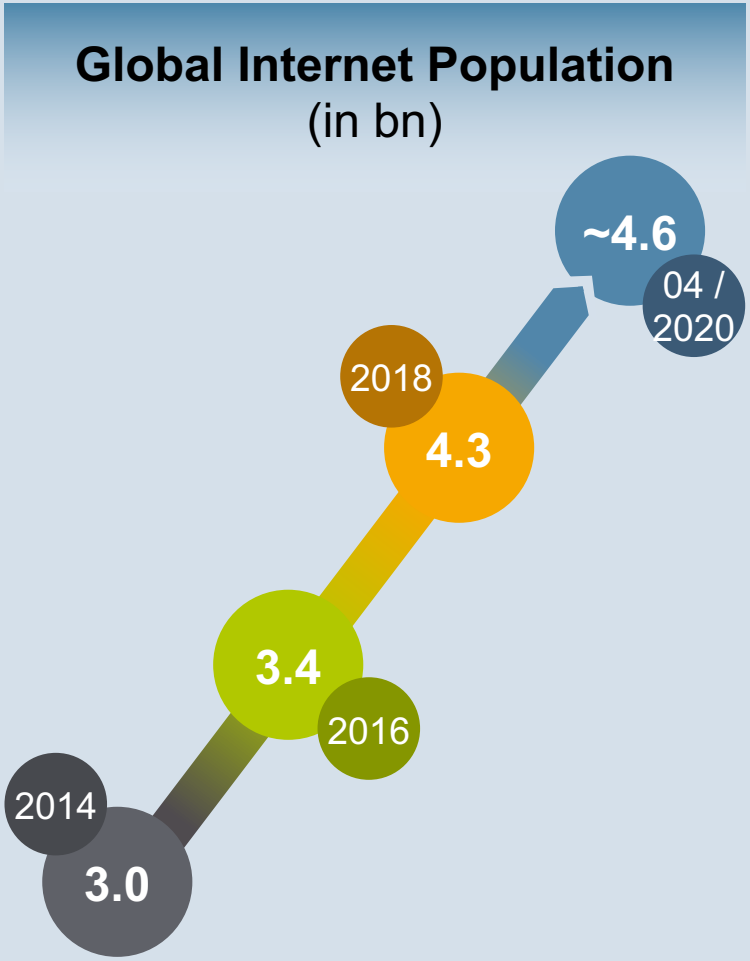
Geo Location Data

Online Search / Weblog Data

E-commerce Data

People-generated Data – Data NEVER Sleeps

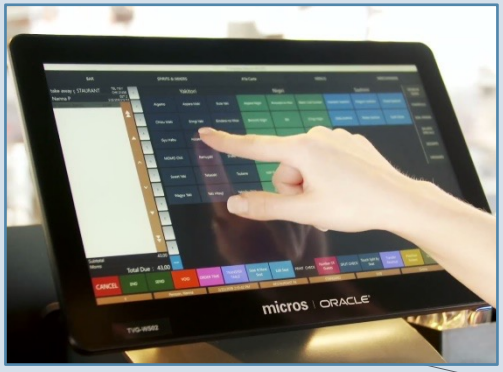




Source: https://web-assets.domo.com/blog/wp-content/uploads/2020/08/20-data-never-sleeps-8-final-01-Resize.jpg

Organization-generated Data

Sales Transaction Data

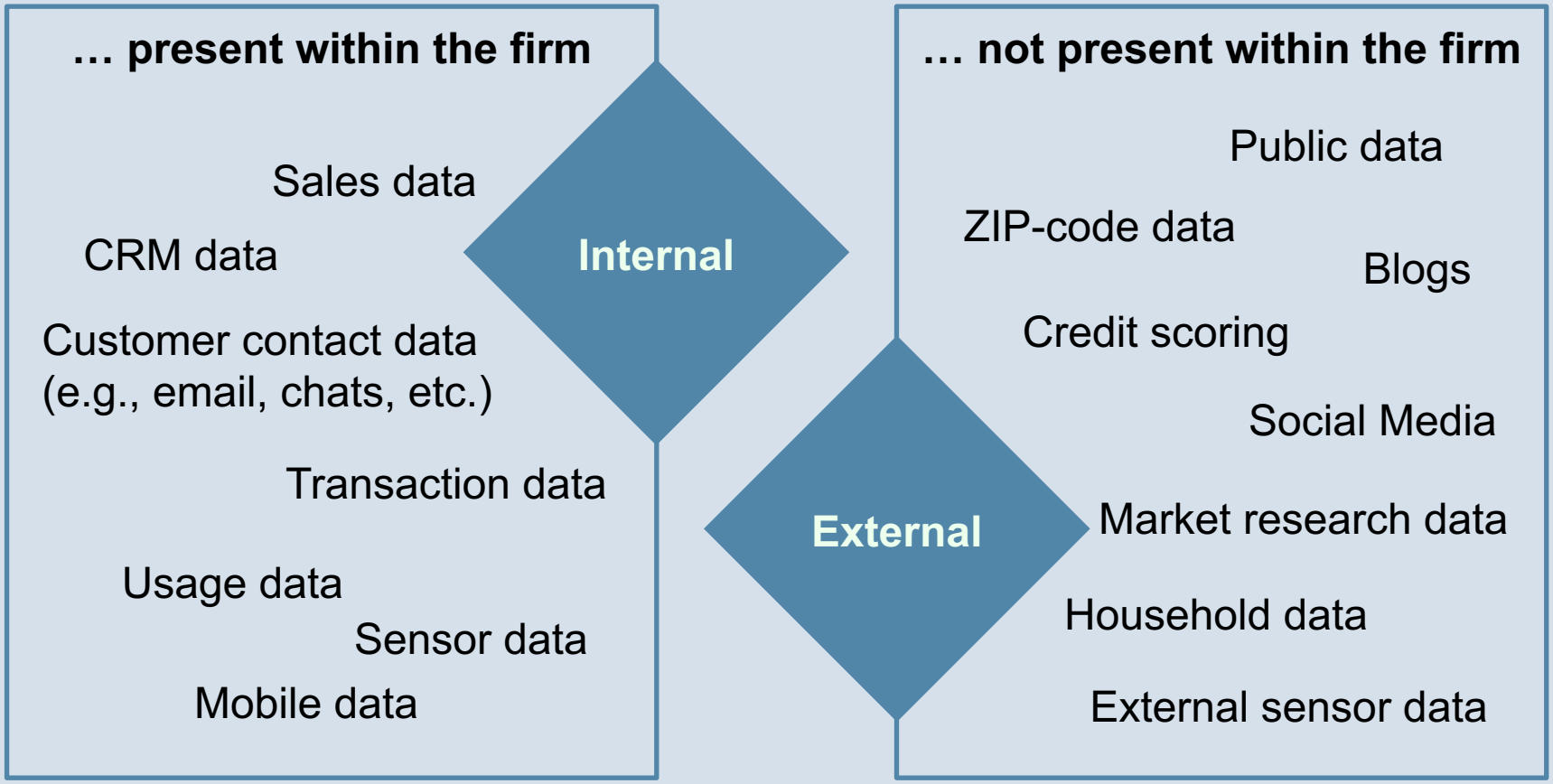


Operations / Logistics Data

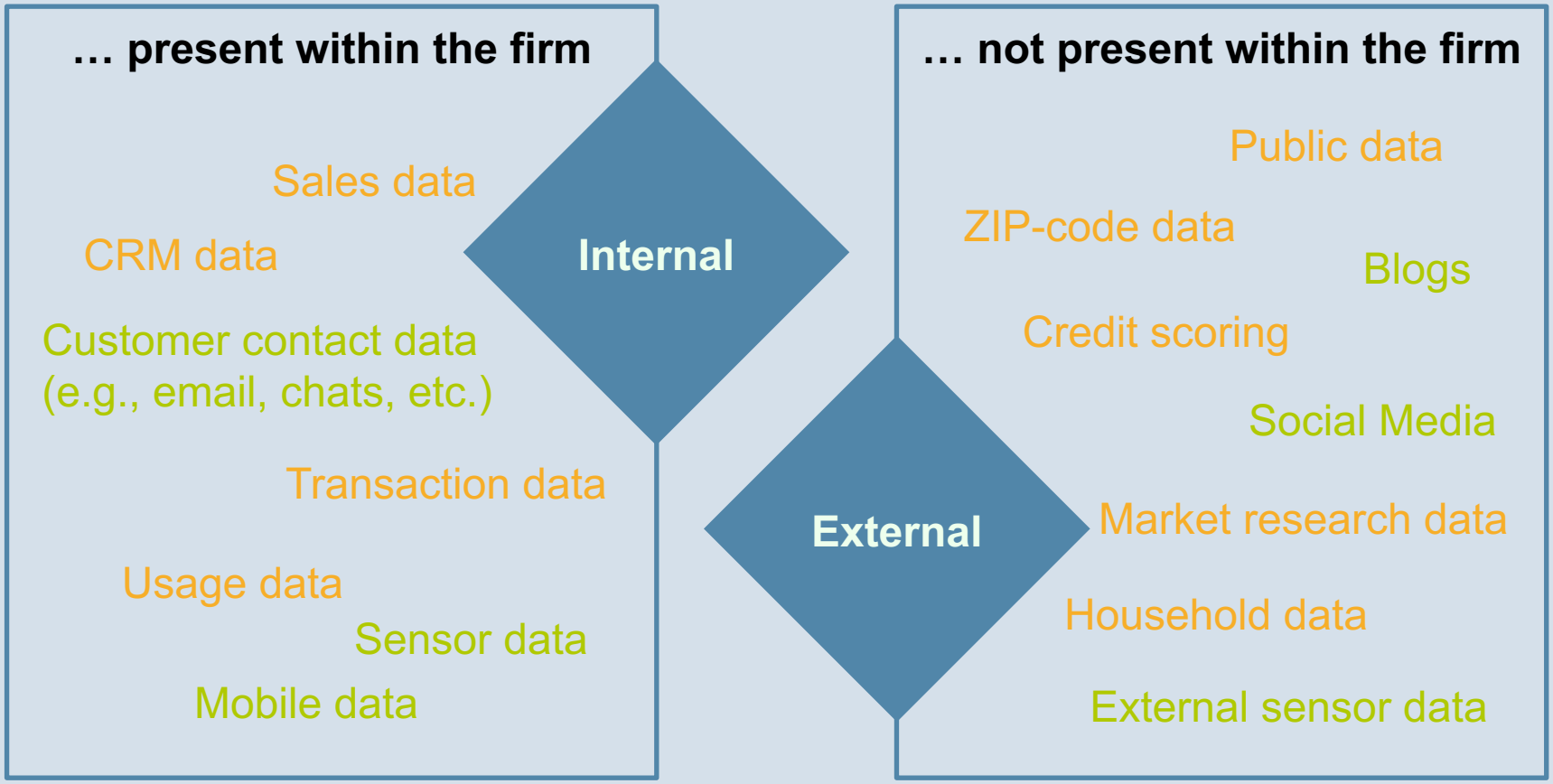
Banking / Financial Data

Government-generated Data

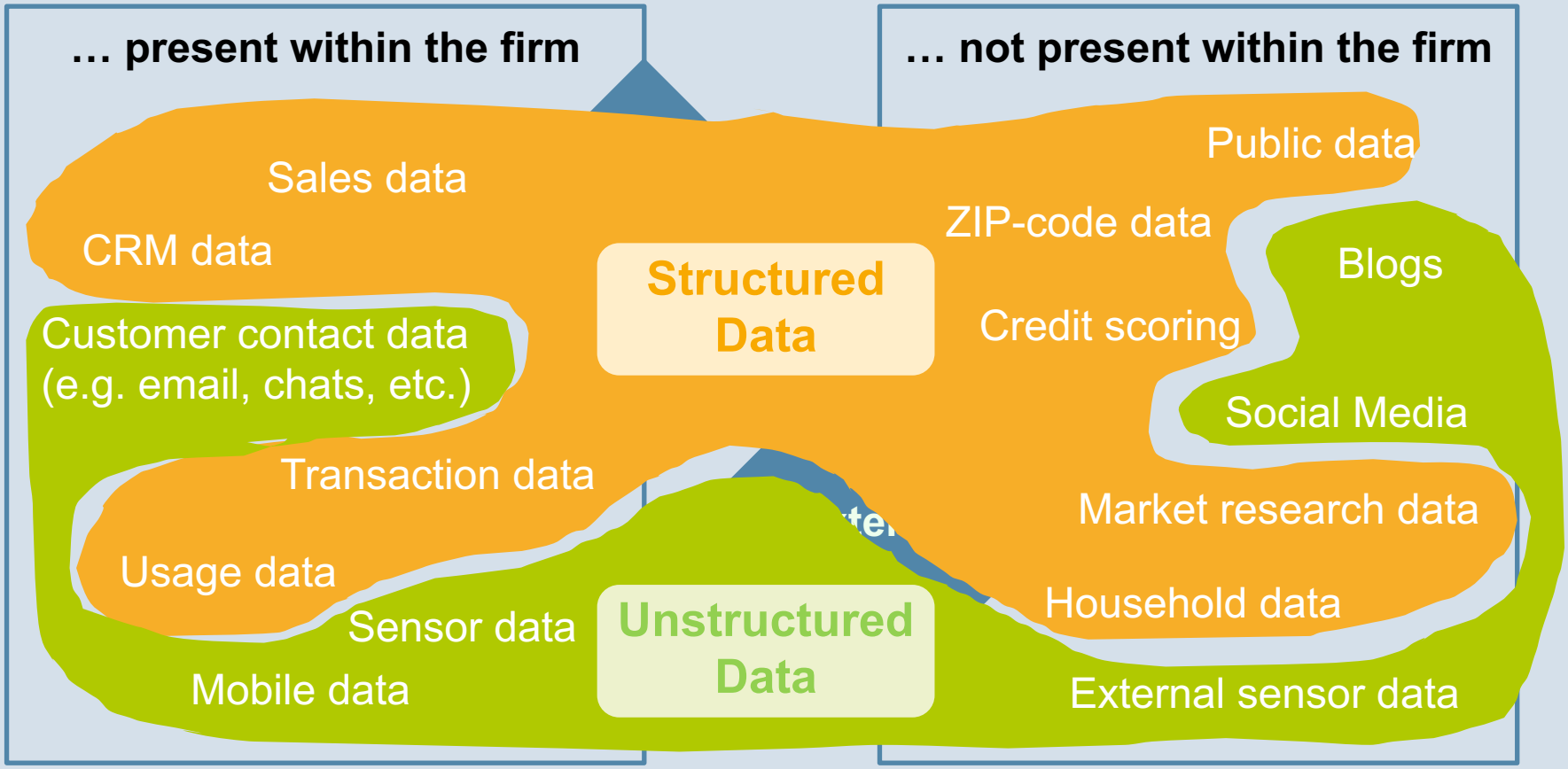
Sources of Data: Internal vs. External



Sources of Data: Internal vs. External



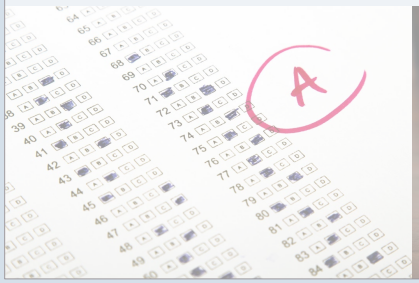
Nature of Data: Structured vs. Unstructured



Types of Data: Quantitative vs. Qualitative

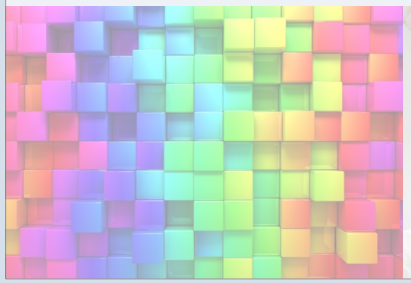
Quantitative Data

- Expressed as a number
- Can be quantified




Qualitative Data

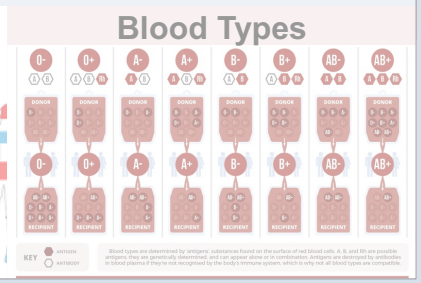
- Sorted by category not number
- Can't be measured (but documented)



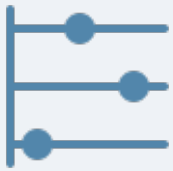

Types of Data: Nominal vs. Ordinal

Nominal Data

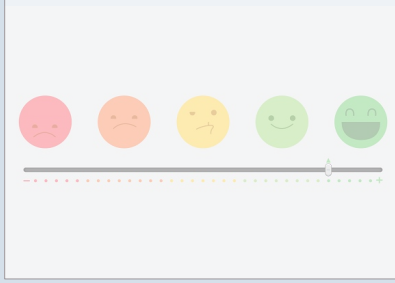
- Used for labelling variables
- Does not attach any quantitative value

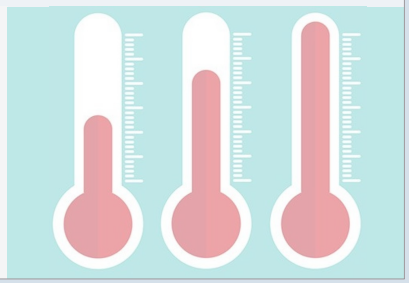


Ordinal Data



- Is placed in some kind of order (usually by position on a scale)





Types of Data: Discrete vs Continuous

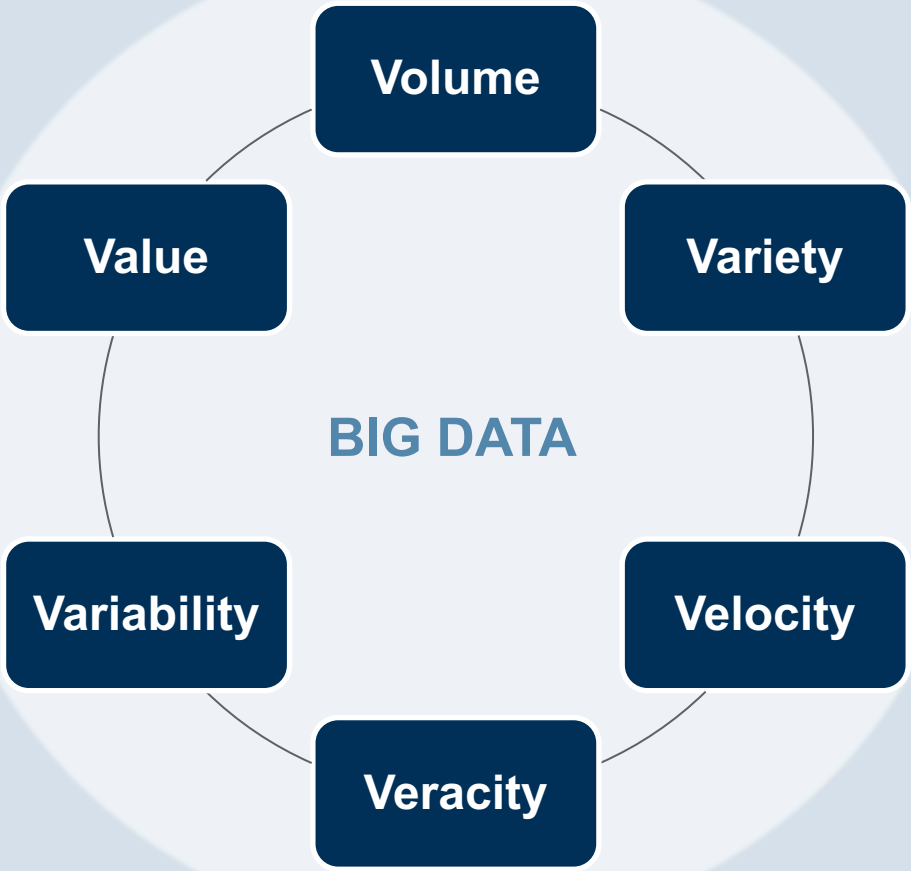
Discrete Data

- Only involves integers
- Cannot be subdivided into parts

Continuous Data

- Can be measured on a scale/continuum
- Can take on any numeric value

Characterizing Dimensions of Big Data



Characterizing Dimensions of Big Data

Volume
- Magnitude of data
- Immediacy with which data is available on a real-time basis

Variety
- Structural variety within the dataset
- Structured vs. unstructured

Velocity
- Rate at which data is generated
- Speed on which data should be analyzed and exploited

Veracity
- Unreliability inherent in some sources of data

Variability
- Variation in data flow rates and sources
- Cause for need to connect, match, cleanse and transform data

Value
- Relatively 'low value density'
- High value can only be obtained by analyzing large volumes of data

Analyzing the Variety of Big Data

Text analytics

e.g., statistical methods, computational linguistics, and machine learning

Audio analytics

e.g., transcript-based or phonetic-based approach

Predictive analytics

e.g., moving averages or linear regression

Video analytics

e.g., automatic video indexing

Social media analytics

e.g., content-based or structure-based

Small and Wide Data vs. Big Data

Big Data

- Big data key characteristics are volume, variety, velocity and variability
- As such, big data is limited to building bigger picture ideas

Small Data

"Data sets sourced within individual organizations or based on individual problem-solving examples."

- ⊕ More human and company specific
- ⊕ Valuable decision making

Wide Data

"Data sets combining a variety of small and large, unstructured and structured data."

- ⊕ Specific and direct inferences on organizational decision making process

In combination, Small and Wide Data allows more in-depth insights than Big Data.

Leveraging Small and Wide Data