Introduction to Data Exploration

Introduction to Data Exploration

K. Selcuk Candan, Ph.D

Professor of Computer Science and Engineering

Arizona State University

Ira A. Fulton Schools of Engineering

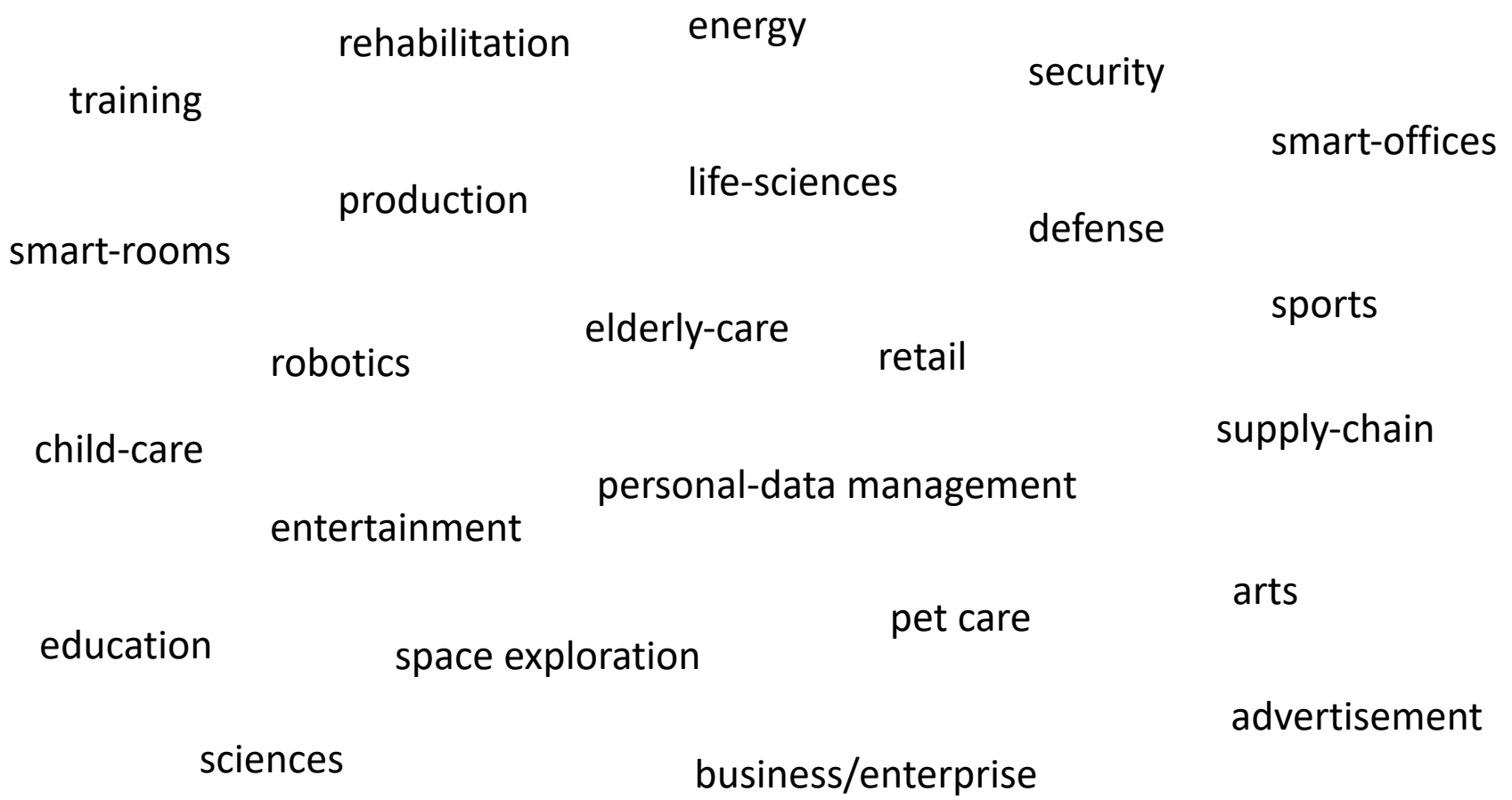
Arizona State University

Objectives

Objective

Explain what are the data challenges

We are living in a data-rich world…



How can we make SENSE from the REAL WORLD data

Data Challenges

INS

(**I**)mprecision
(**N**)oise
(**S**)parsity

3Vs

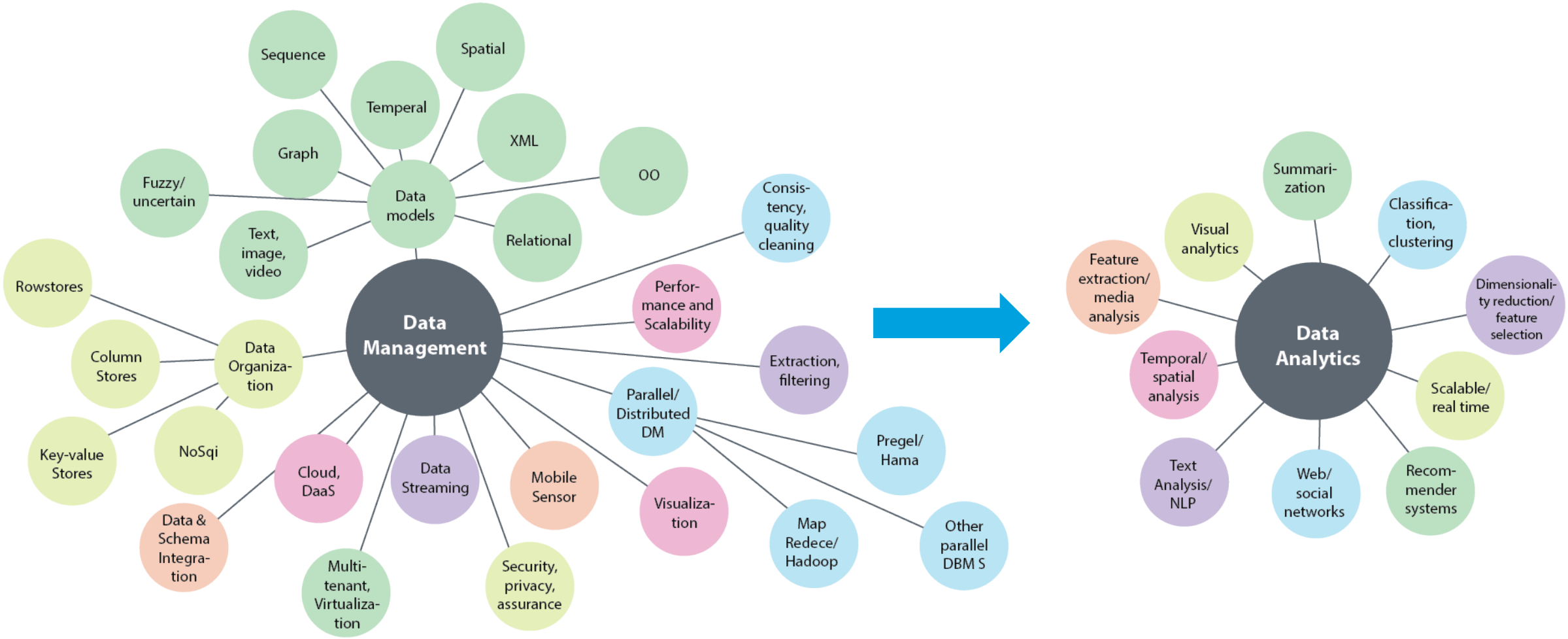
(**V**)olume
(**V**)Elocity
(**V**)ariety

HMLE

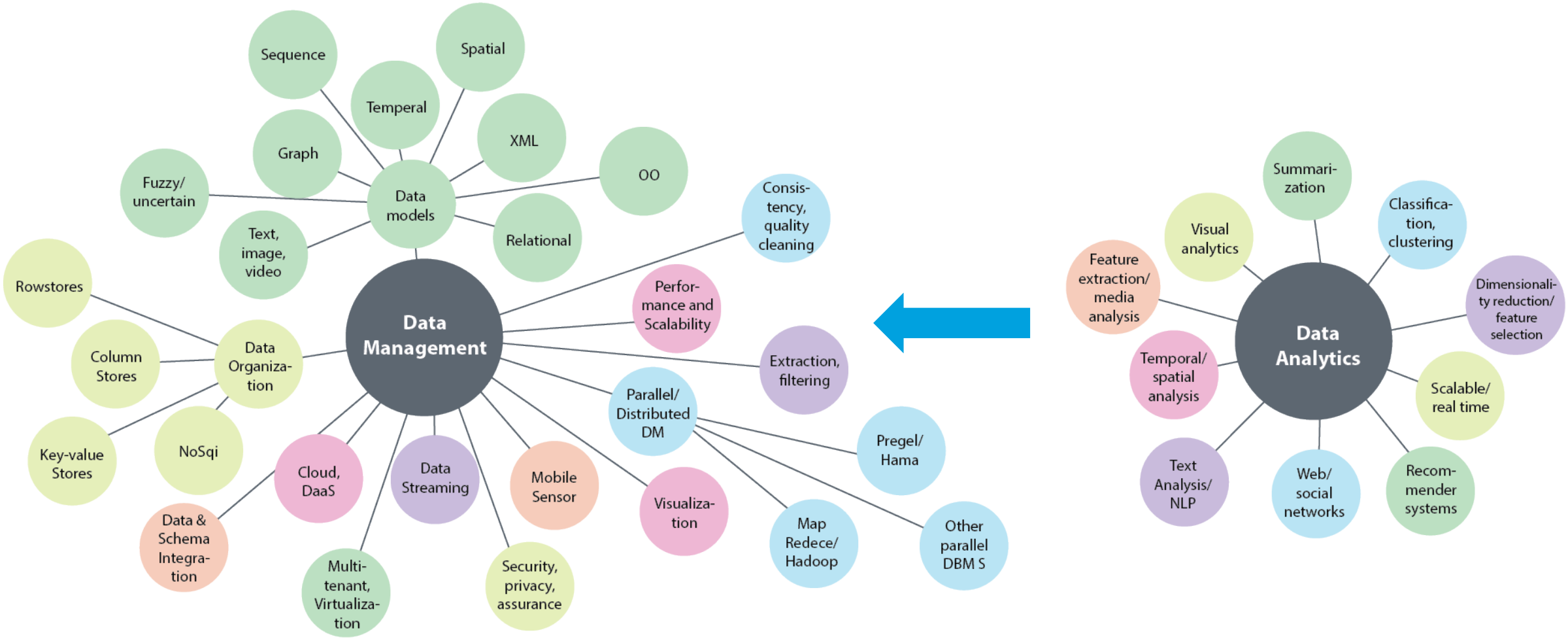
(**H**)igh-dimensional
(**M**)ulti-modal
Inter-(**L**)inked
(**E**)volving

Human Challenges: for many applications the final consumer is **HUMAN**

Data management/mining techniques for supporting scalable, real-time, analysis and exploration



Most data in the real world are imprecise, multi-modal, and subjective



Therefore, data exploration systems need to support both…

…effective data manipulation

- **Filtering**: Projection, selection
- **Integration**: Join, nearest neighbor joins
- **Set operations**: Union and Intersection

…effective data analysis/retrieval

- **Feature extraction**
- **Similarity search**
  (top-k, range, skyline, nearest-neighbor)
- **Clustering**, partitioning
- **Aggregation**, summarization
- **Classification**, latent analysis
- **Preference-driven** retrieval