

DATA SCIENCE

R.Loka vigneswar reddy
21AK5A0321

DATA SCIENCE VOCABULARY



WHAT IS DATA SCIENCE?

- “**Data science**, also known as **data-driven science**, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining.”

WHAT IS DATA SCIENCE?

- “**Data science**, also known as **data-driven science**, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining.”
- “Data science intends to analyze and understand actual phenomena with ‘data’. In other words, the aim of data science is to reveal the features or the hidden structure of complicated natural, human, and social phenomena with data from a different point of view from the established or traditional theory and method.”

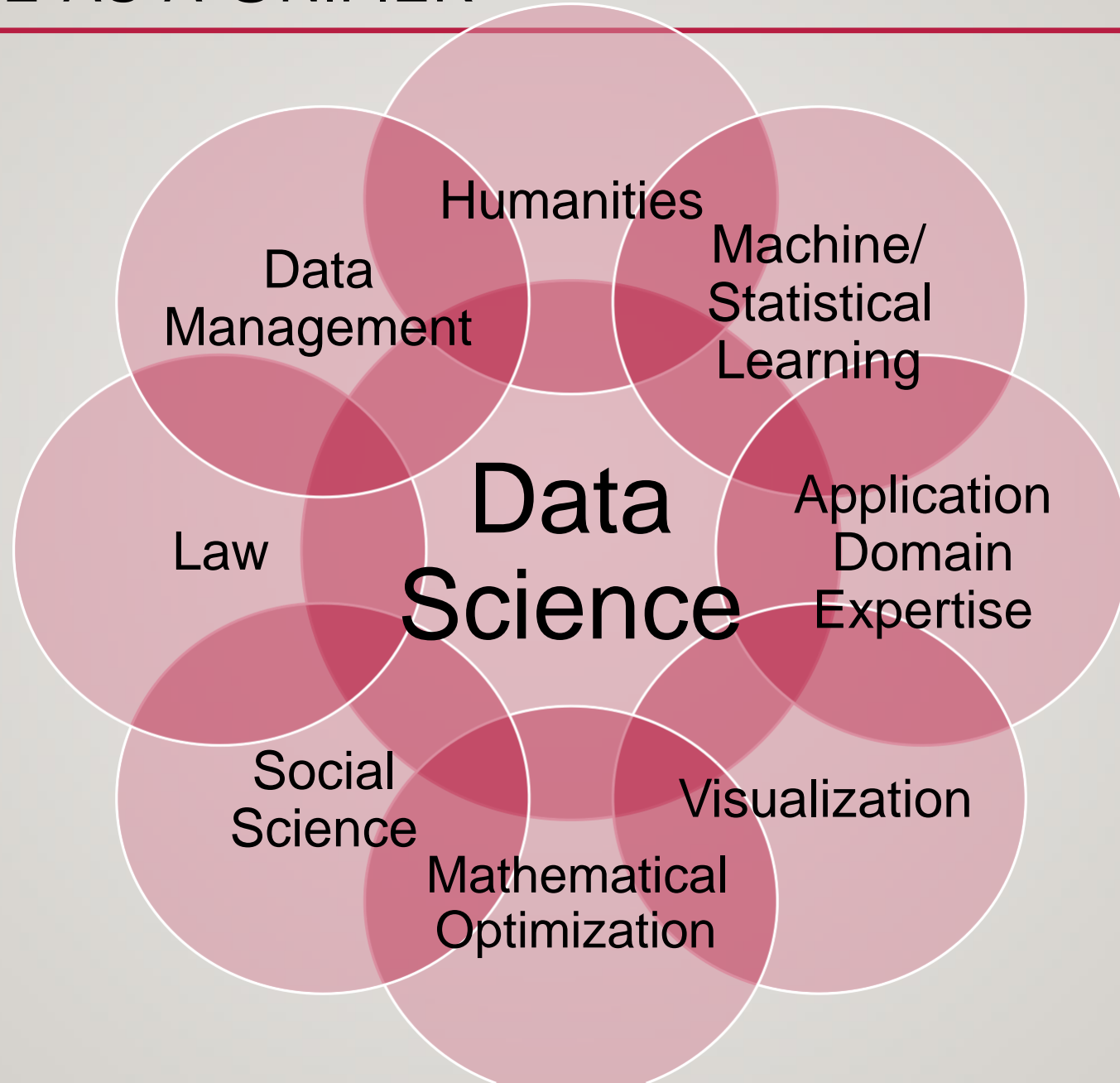
WHAT IS DATA SCIENCE?

- Fourth paradigm
 - “... change of all sciences moving from observational, to theoretical, to computational and now to the 4th Paradigm – Data-Intensive Scientific Discovery”

WHAT IS IMPORTANT?

Need to solve a real problem using data...
No applications, no data science.

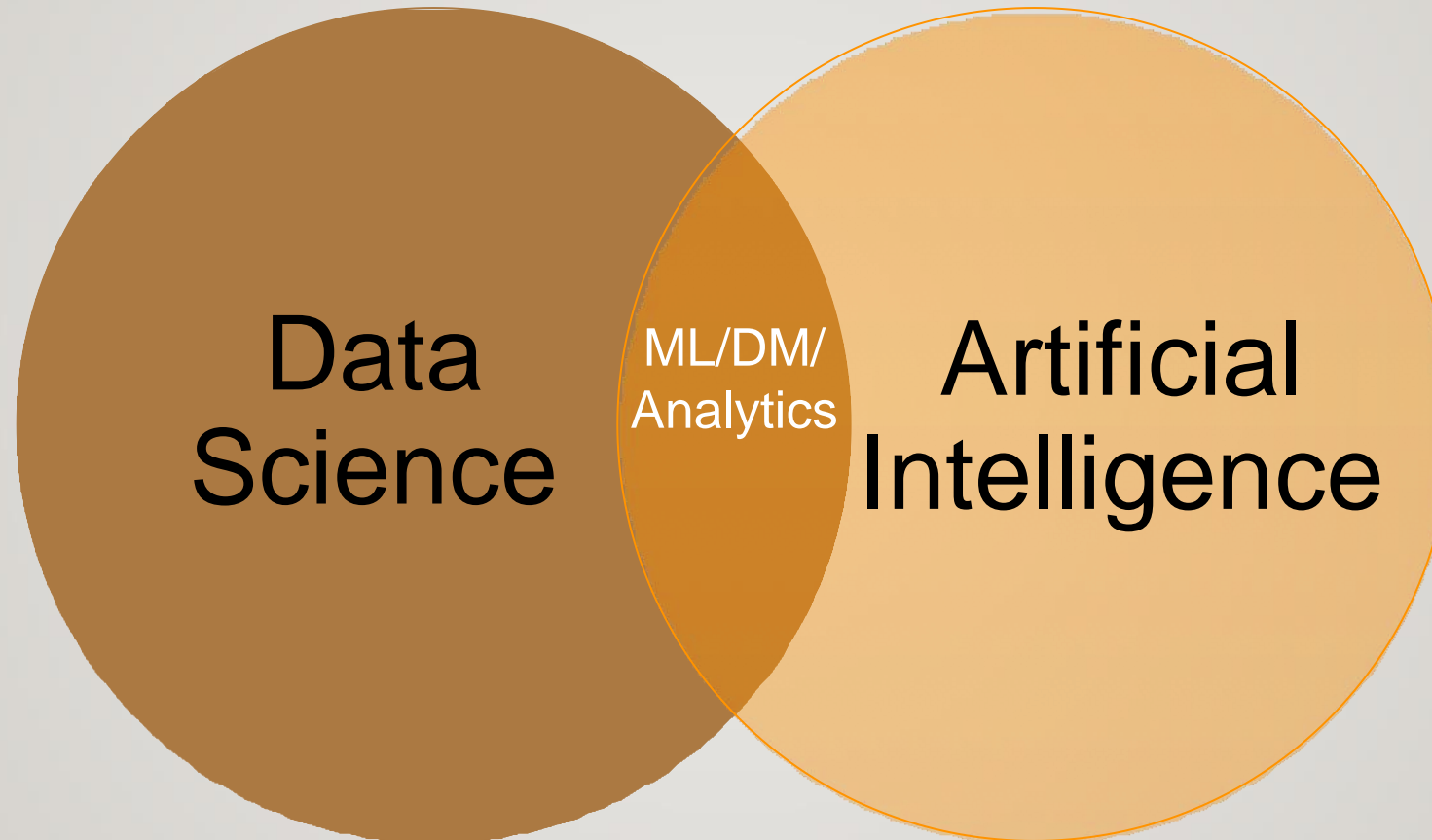
DATA SCIENCE AS A UNIFIER



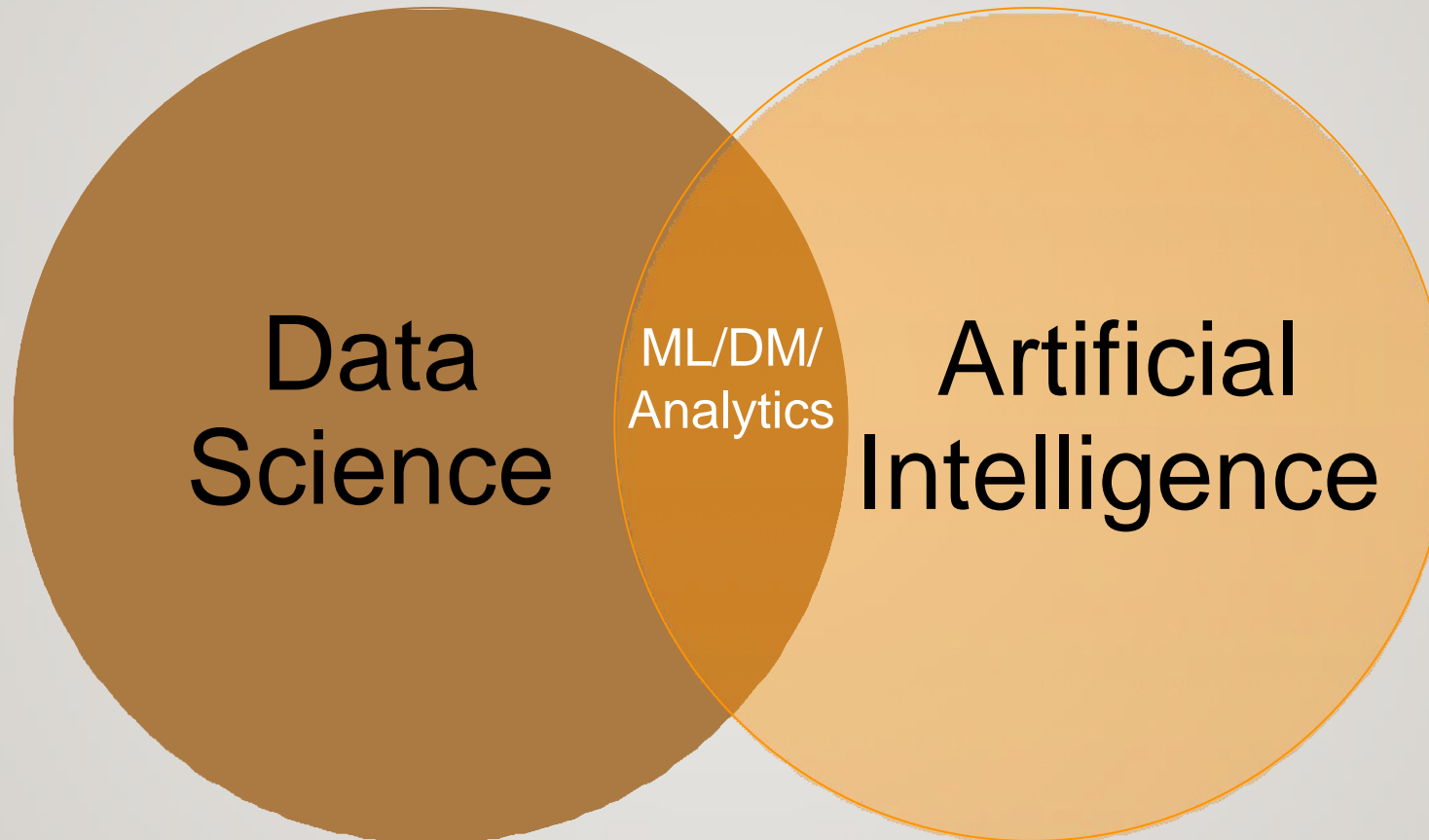
DATA SCIENCE AND BIG DATA

- They are not the “same thing”
- Big data = crude oil
 - Big data is about extracting “crude oil”, transporting it in “mega tankers”, siphoning it through “pipelines”, and storing it in “massive silos”
- Data science is about refining the “crude oil”

DATA SCIENCE AND ARTIFICIAL INTELLIGENCE



DATA SCIENCE AND ARTIFICIAL INTELLIGENCE



**“Data science produces insights.
Machine learning produces predictions”**

DATA SCIENCE APPLICATION EXAMPLES

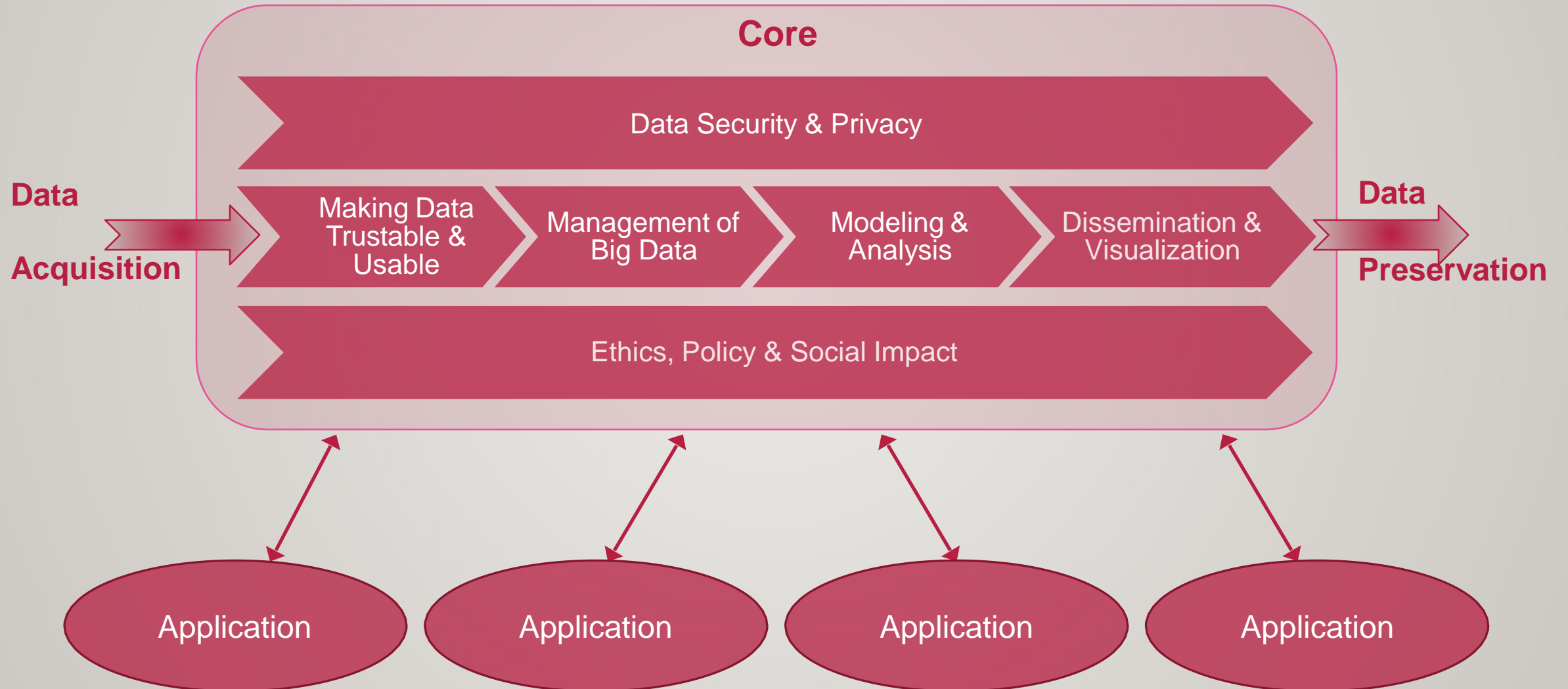
- Fraud detection
 - Investigate fraud patterns in past data
 - Early detection is important
 - Before damage propagates
 - Harder than late detection
 - Precision is important
 - False positive and false negative are both bad
 - Real-time analytics

DATA SCIENCE APPLICATION EXAMPLES

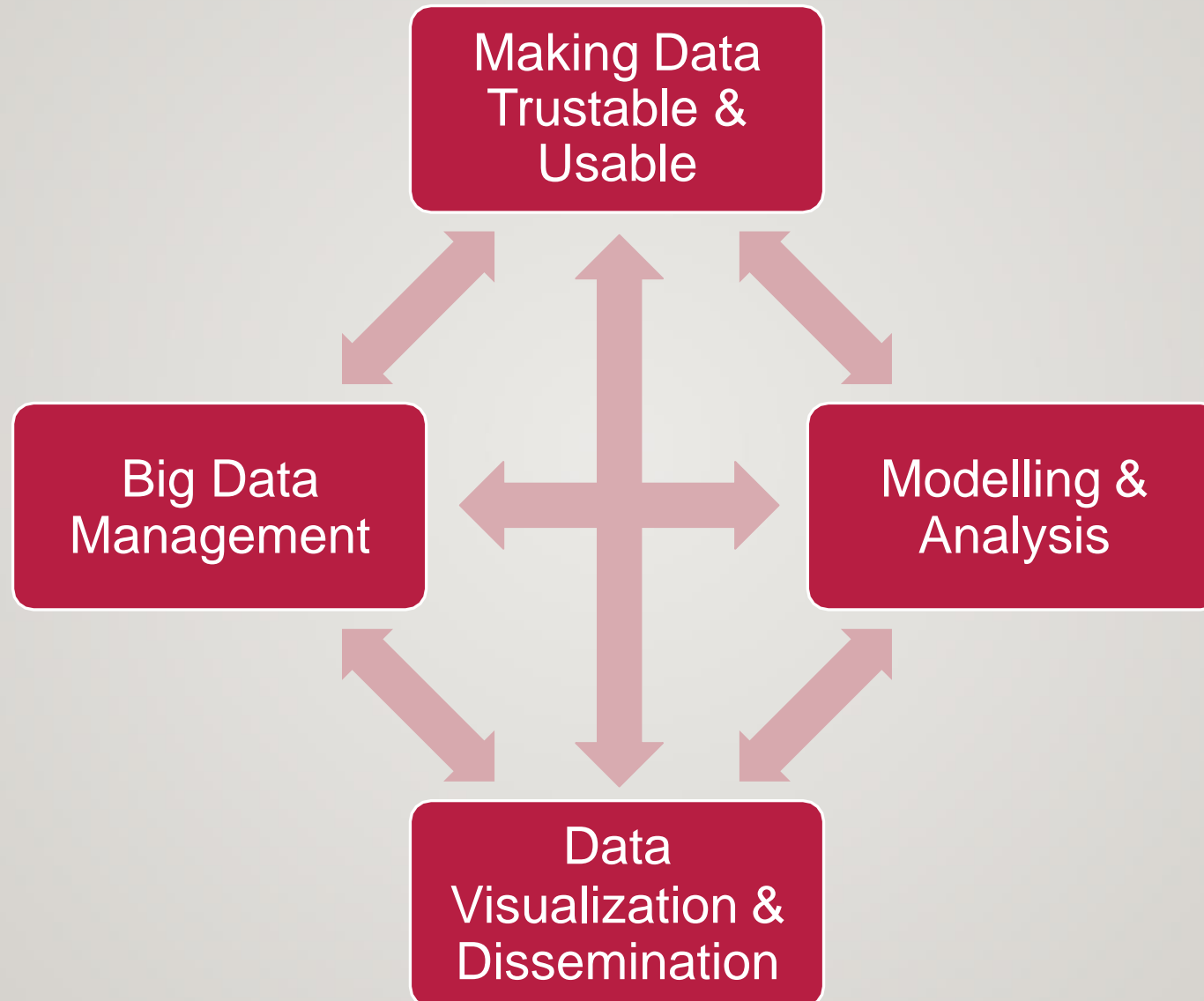
- “Smart cities”
 - Not well-defined



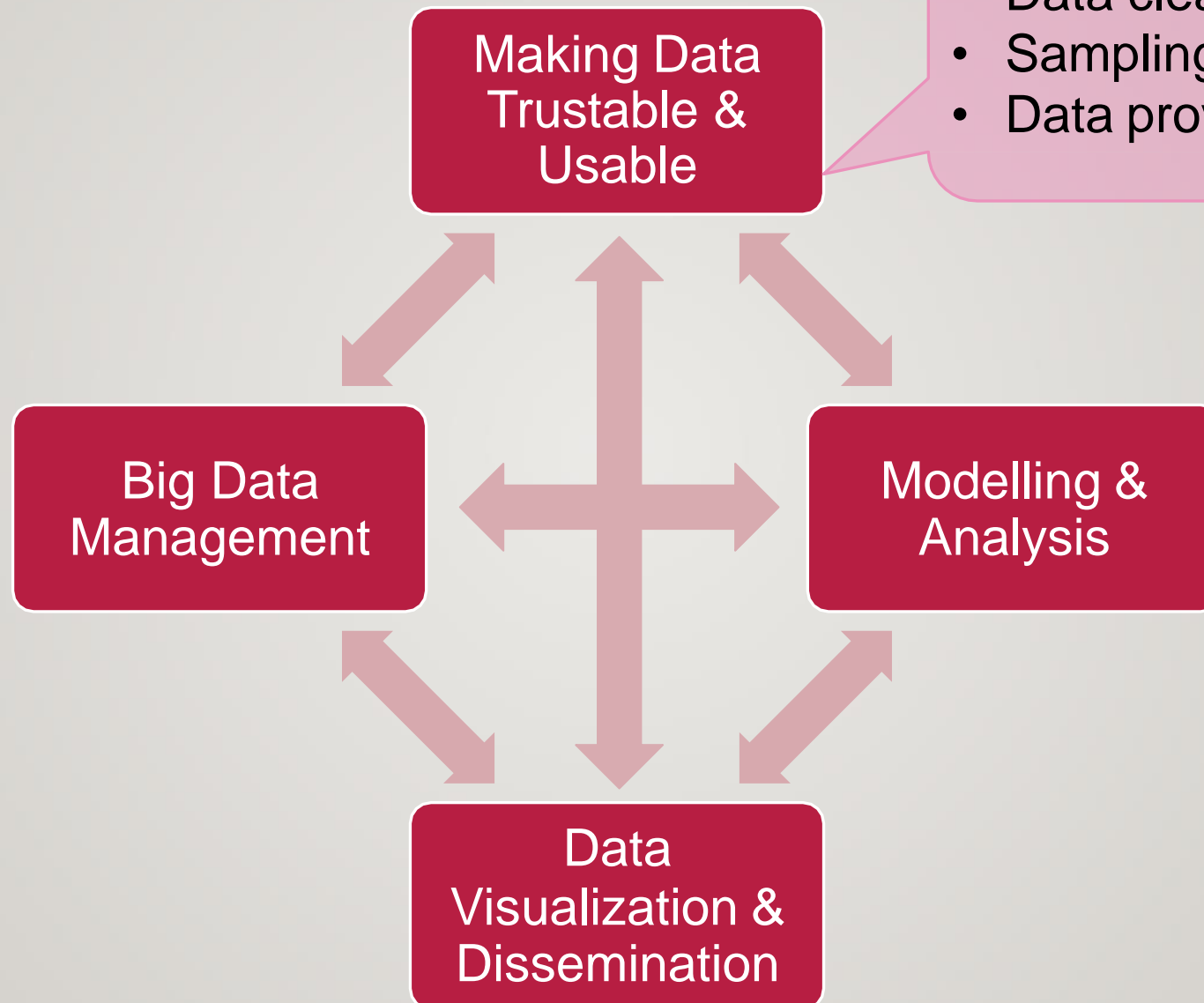
HOLISTIC APPROACH TO DATA SCIENCE



CORE RESEARCH ISSUES & INTERACTIONS



CORE RESEARCH ISSUES & INTERACTIONS



CORE RESEARCH ISSUES & INTERACTIONS

- Data lakes
- Batch & online access
- Platforms

Making Data
Trustable &
Usable

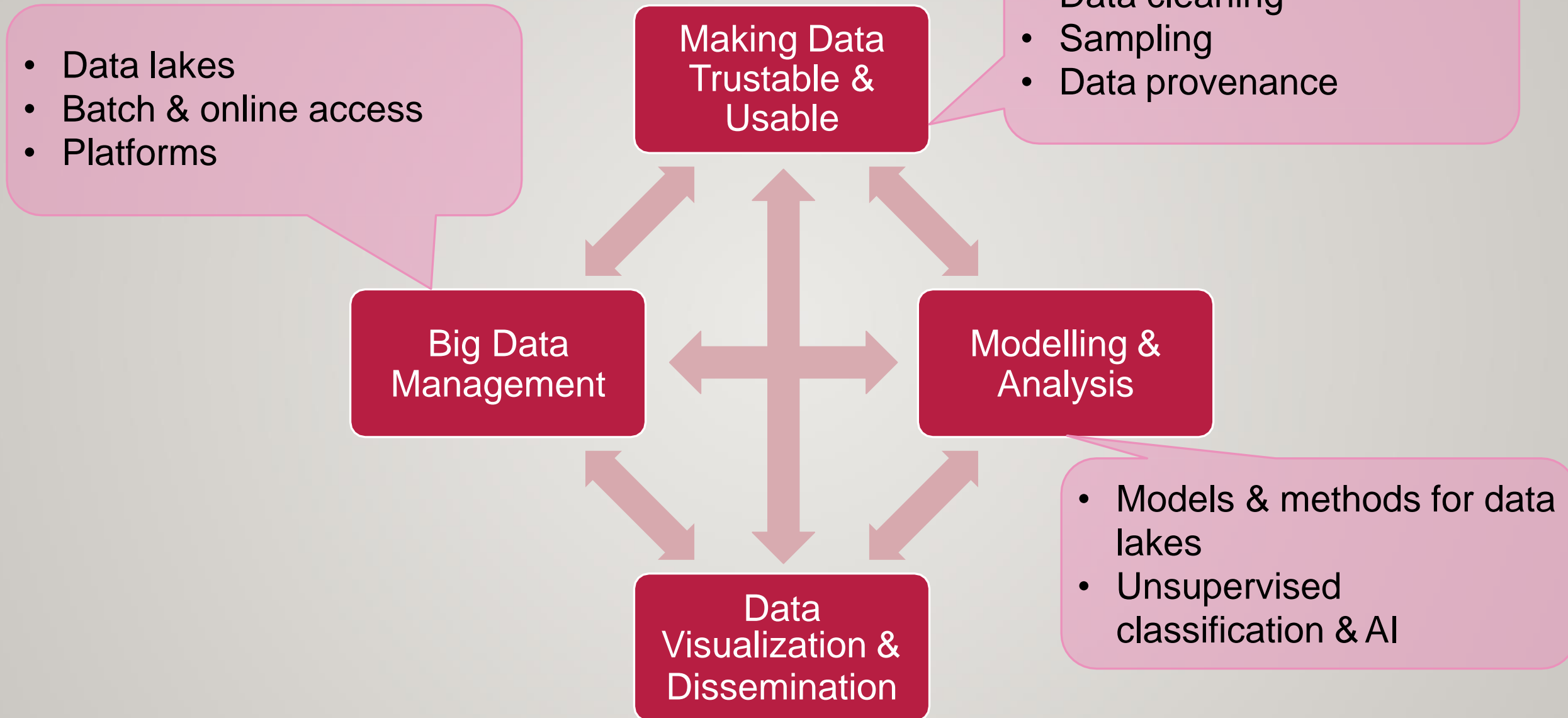
- Data cleaning
- Sampling
- Data provenance

Big Data
Management

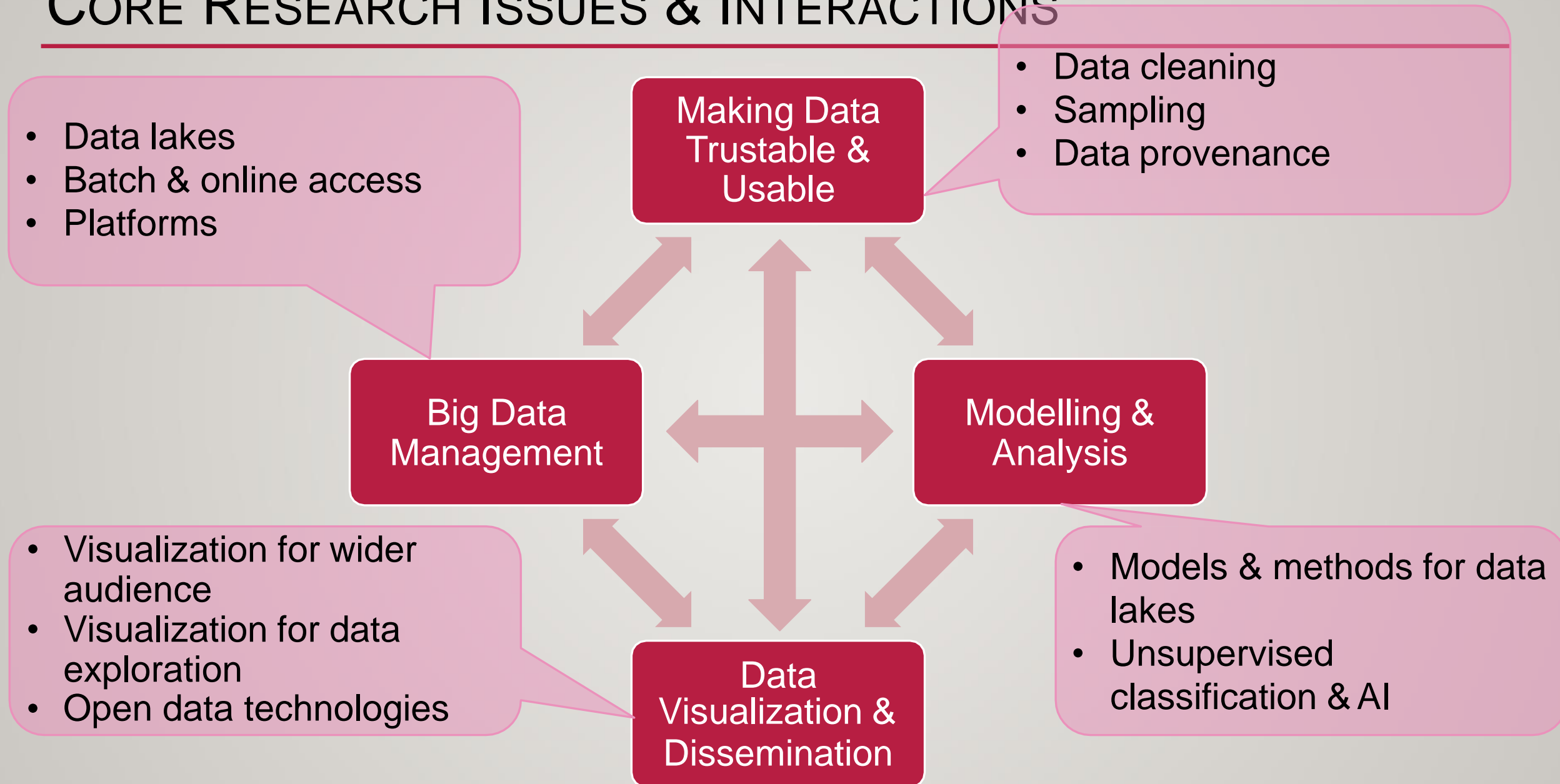
Modelling &
Analysis

Data
Visualization &
Dissemination

CORE RESEARCH ISSUES & INTERACTIONS



CORE RESEARCH ISSUES & INTERACTIONS



THAK YOU FOR
GIVING THIS
OPPORTUNITY

I WILL IMPROVE MY KNOWLADGE BY THIS
PROJECT