AmazonFresh: Changing the Nature of Urban Logistics

Brandeis Marshall, Purdue University
brandeis@purdue.edu
http://web.ics.purdue.edu/~brandeis
@csdoctorsister

AmazonFresh (https://fresh.amazon.com/)

- Free same-day and early morning delivery on orders over $35 of more than 500,000 Amazon items, including fresh grocery and local products.

- Place your order by 10am and have it by dinner, or by 10pm and have it by breakfast.

- Amazon Prime instance
  - ✔ Free Two-Day Shipping on millions of items
  - ✔ No minimum order size
  - ✔ Unlimited instant streaming of thousands of movies and TV shows with Prime Instant Video
  - ✔ Read free books each month through Kindle First and the Kindle Owners' Lending Library

AmazonFresh (https://fresh.amazon.com/)

* AmazonFresh and Prime Fresh Membership Delivery Options

    * Doorstep Delivery: orders are placed at your home in sealed, temperature-controlled tote bags.

    * Attended Delivery: these tote bags are delivered in person to someone that must be home to receive them, e.g., required for sensitive purchases such as alcohol.

* Amazon Prime Air (http://www.amazon.com/b?node=8037720011)

    * 30-minutes or less delivery to customers' hands using unmanned aerial vehicles

    * Amazon R&D released initial details in December 2013

    * Expected Prime Air launch is 2015, currently working with FAA on UAV regulations and standards

AmazonFresh (https://fresh.amazon.com/)

- AmazonFresh Feedback & Customer Reviews: http://www.yelp.com/biz/amazon-fresh-los-angeles

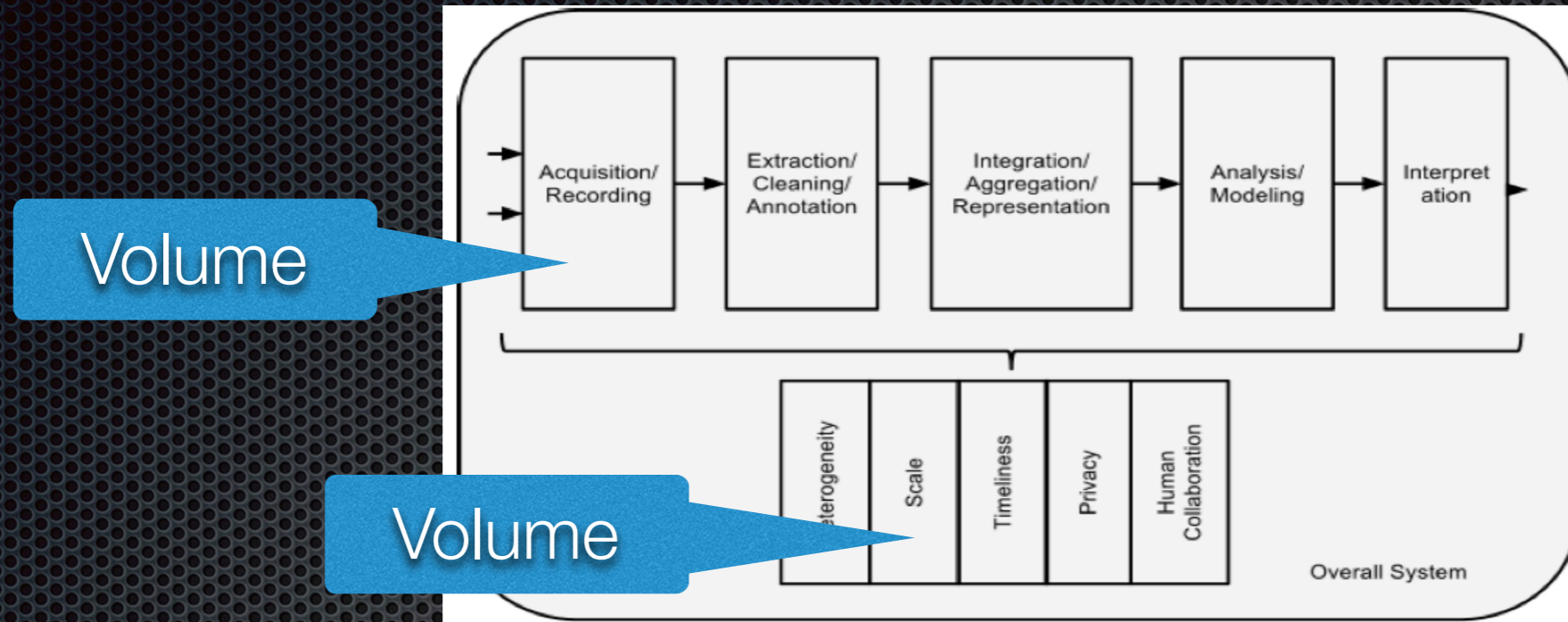# The Business of Data

Big Data Analysis Pipeline (CRA)
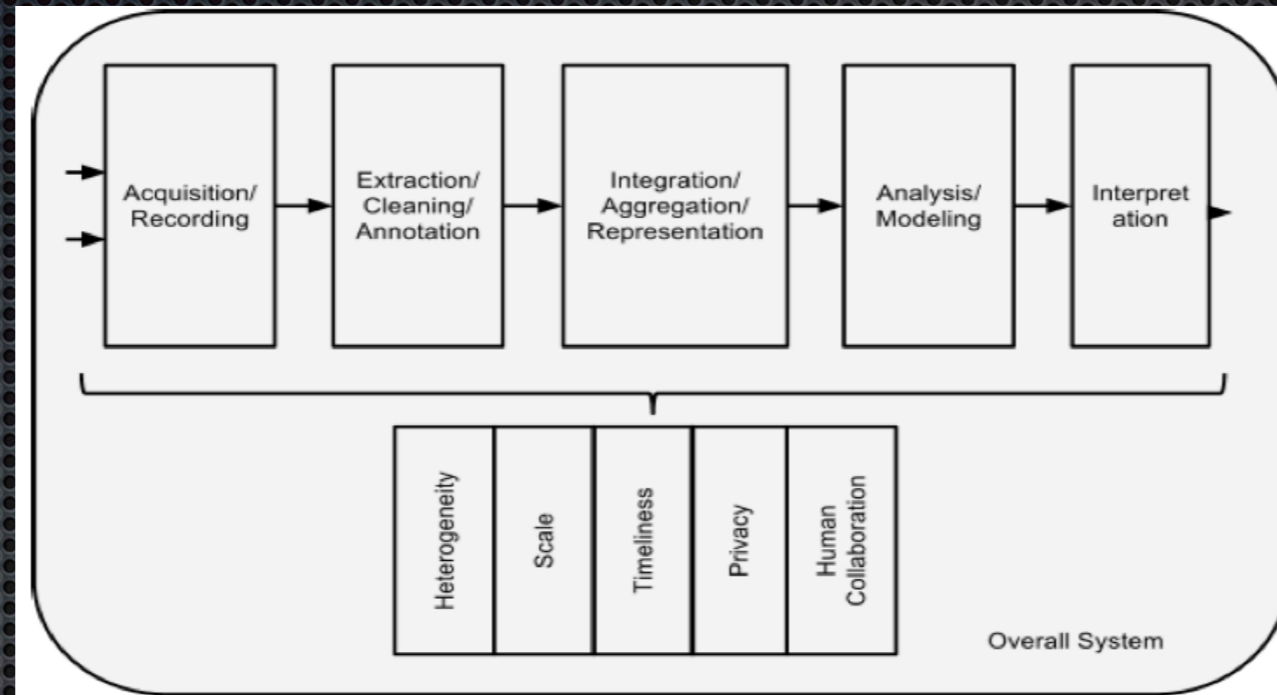
- Big Data V's

  - **The Classics:** Volume+Variety+Velocity (circa 2001, Gartner Inc.)

  - **The Fourth:** Viability/Veracity (circa 2012)

  - **The Fifth:** Value (circa late 2012/early 2013)

The Business of Data

Big Data Analysis Pipeline (CRA)

Volume

Volume

- **Volume: Every click, google search, text, tweet, review, phone call, transaction, purchase…must be gathered and stored**
  - Example: Amazon Cyber Monday 2013 sales = 426 items purchased/sec (36.8 million items ordered worldwide)
  - Example: Amazon Prime Services membership = >1M in 3rd week of December 2013

- One machine is obviously not enough…

  - We need distributed processing of large data sets across clusters of computers using simple programming models

  - Apache Hadoop: open-source software for scalable, reliable, distributed computing

  - Hadoop Enterprise Vendors: Cloudera, HortonWorks

The Business of Data

Big Data Analysis Pipeline (CRA)
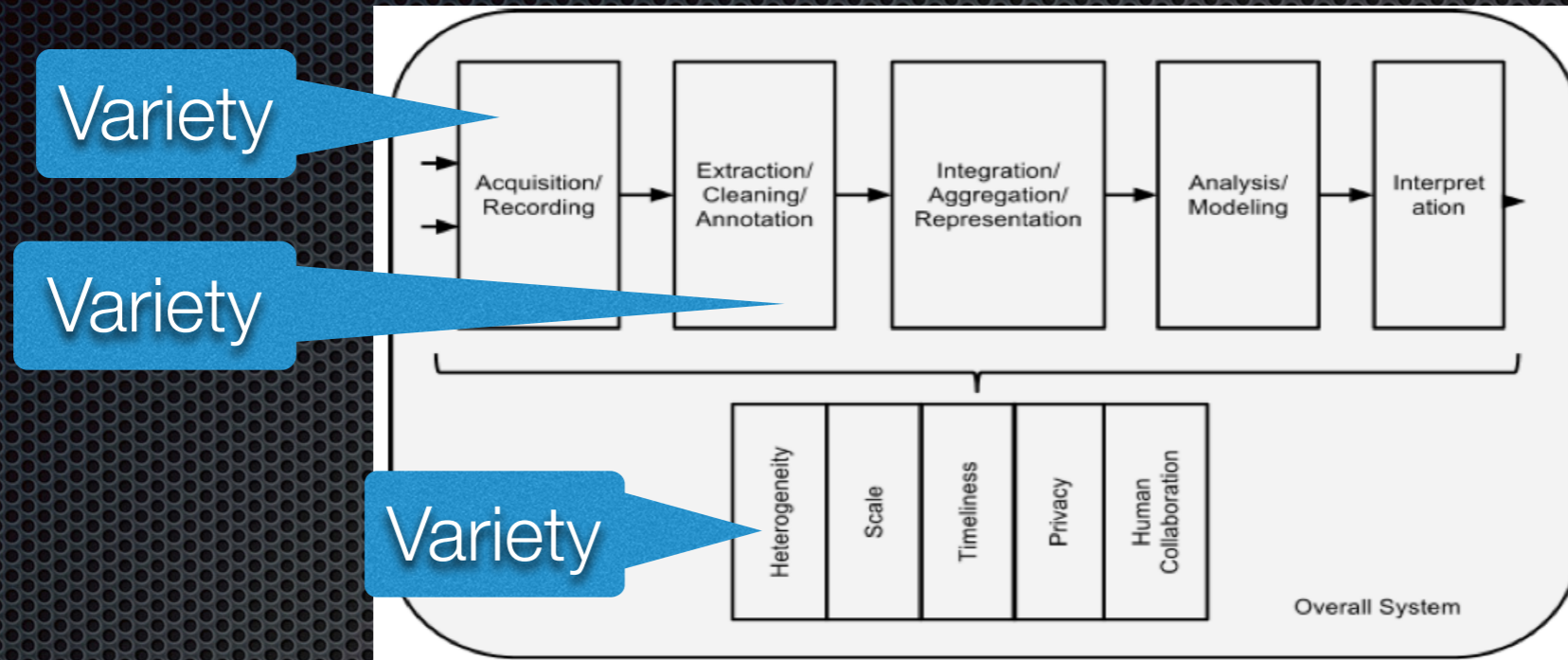
- **Variety: Bringing together the heterogeneous data streams from structured to unstructured sources**

  - Example: geolocations, social networking

- Amazon Kinesis: real-time processing of streaming data at massive scale

  - collect and process hundreds of terabytes of data per hour

  - from hundreds of thousands of sources (web site click-streams, marketing and financial information, manufacturing instrumentation and social media, and operational logs and metering data)

The Business of Data

Big Data Analysis Pipeline (CRA)
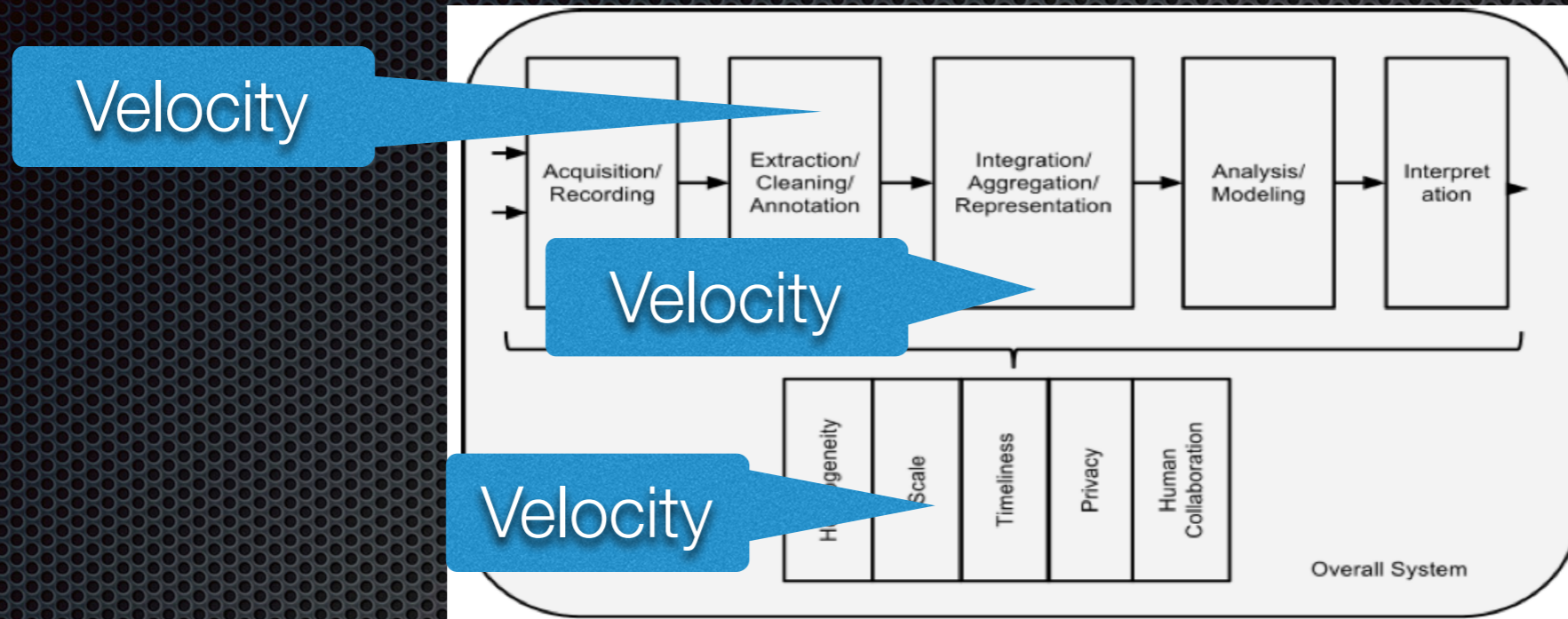
- **Velocity: running at the speed of data**
  - Example: Amazon Cyber Monday 2013 sales = 426 items purchased/sec (36.8 million items ordered worldwide)

- Amazon Web Services (EC2/S3): Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) aka public cloud services

- Databases: Oracle (RDBMS), Cassandra (distributed DBMS)

- KNIME: visually create data flows, selectively execute some or all analysis steps, and later inspect the results, models, and interactive view

The Business of Data

Big Data Analysis Pipeline (CRA)
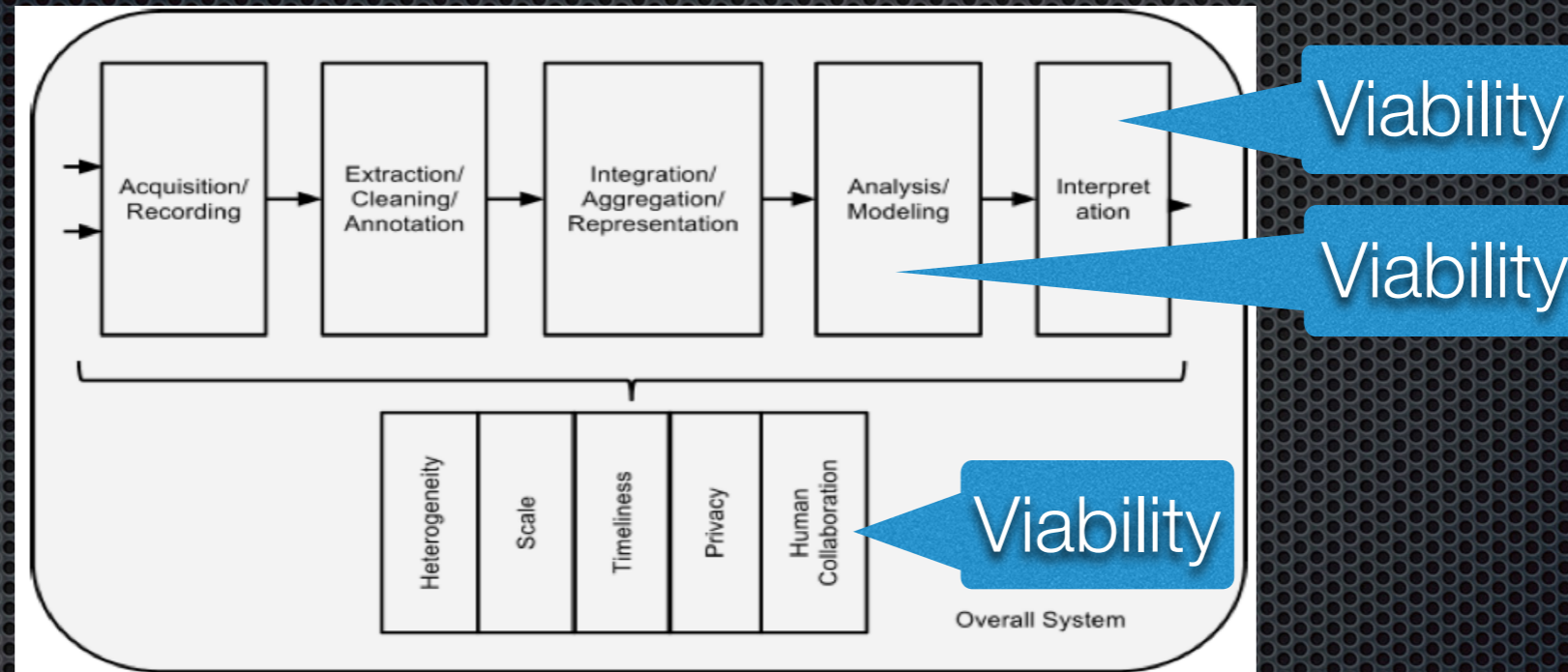
- **Viability: identifying relevant factors & predicting outcomes important to the business**
  - Example: How do geolocation, product availability, time of day, purchasing history, age, family size, credit limit, and vehicle type all converge to predict a consumer's propensity to buy?

- R, WEKA 3: a suite of machine learning algorithms designed for data mining tasks

- Predictive Model Markup Language: statistical and data mining models standard

The Business of Data

Big Data Analysis Pipeline (CRA)
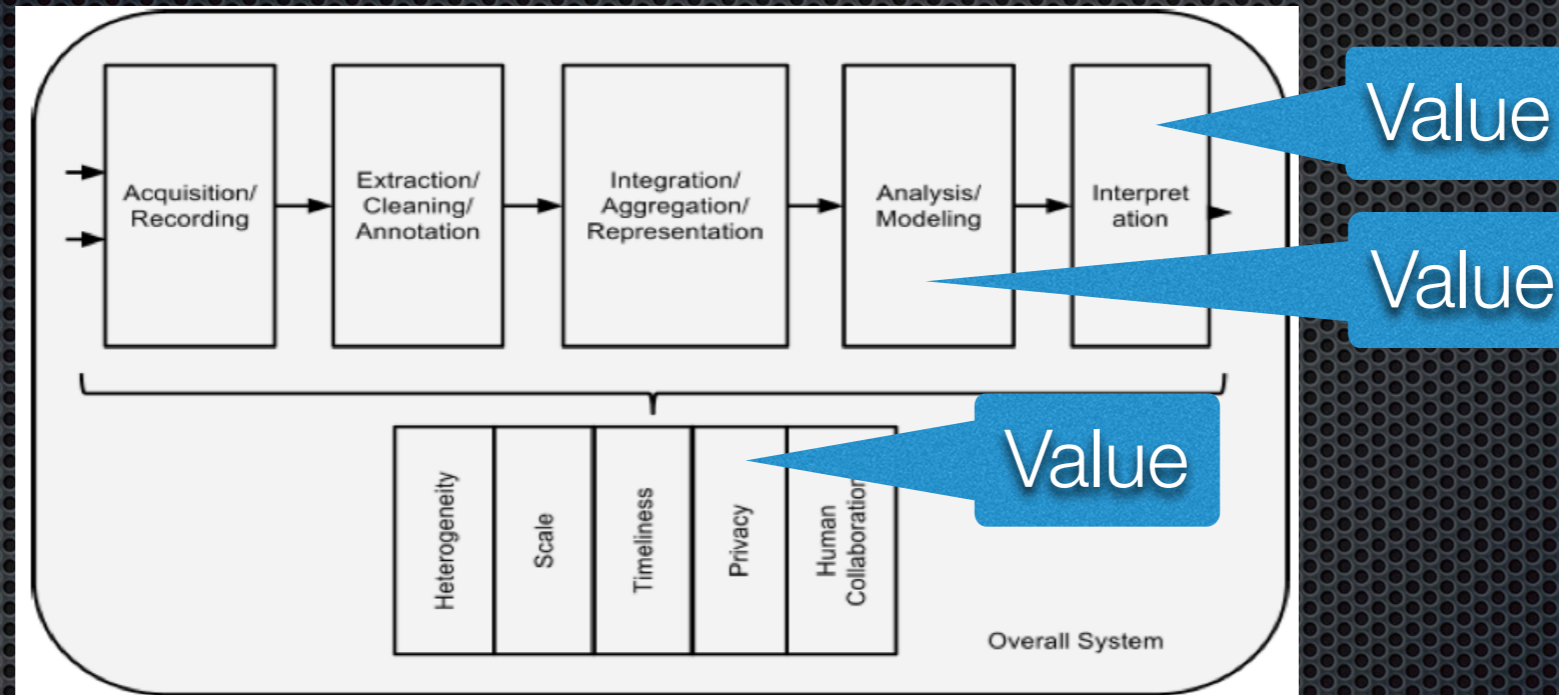
- **Value: uncover subtle interactions, e.g., combinations of additional variables or long-tail**
  - Example: predicting what customers in <region> with <educational background> who own <make, model, year> automobile and <X> credit score show a statistically significant likelihood to <make some transaction> in the next <Y> days.

- Predictive Analytics: Relationship + Persona + Intent
  - R, WEKA 3: a suite of machine learning algorithms designed for data mining tasks
  - Predictive Model Markup Language: statistical and data mining models standard

Thank you for your time and attention

Brandeis Marshall, Purdue University
brandeis@purdue.edu
http://web.ics.purdue.edu/~brandeis
@csdoctorsister