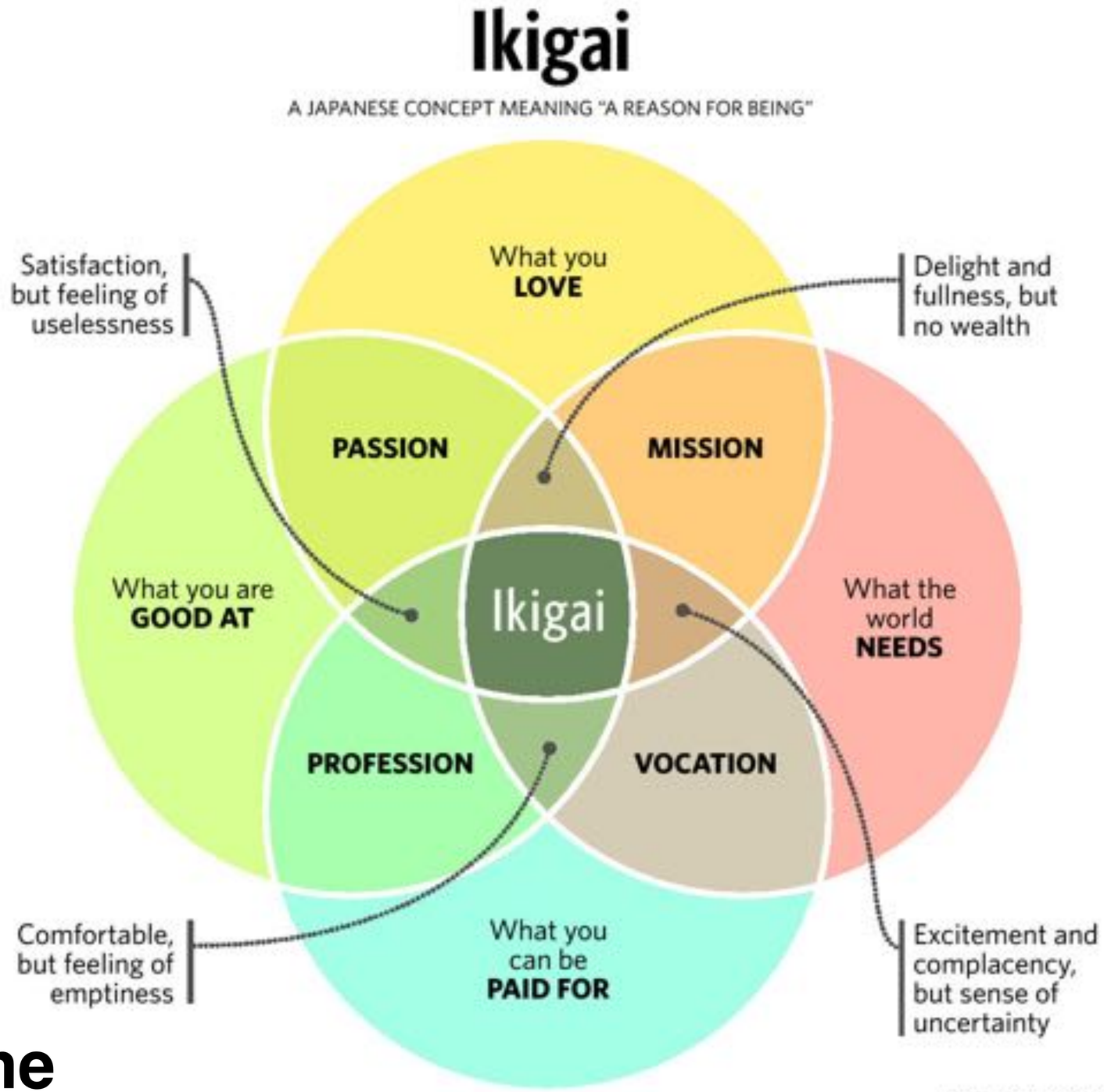
Corporate Data Science

Angela Bassa (@angebassa)

csv,conf 2017

vs. Indie Data Science

Good Thing There's > 7 Billion Of Us



Source: Dreamstime

"Data is useless without context"

@thegrugq

A Little Context, Then

It’s Not Personal

(And It’s Not Academic Either);

It’s A Business.

Agenda

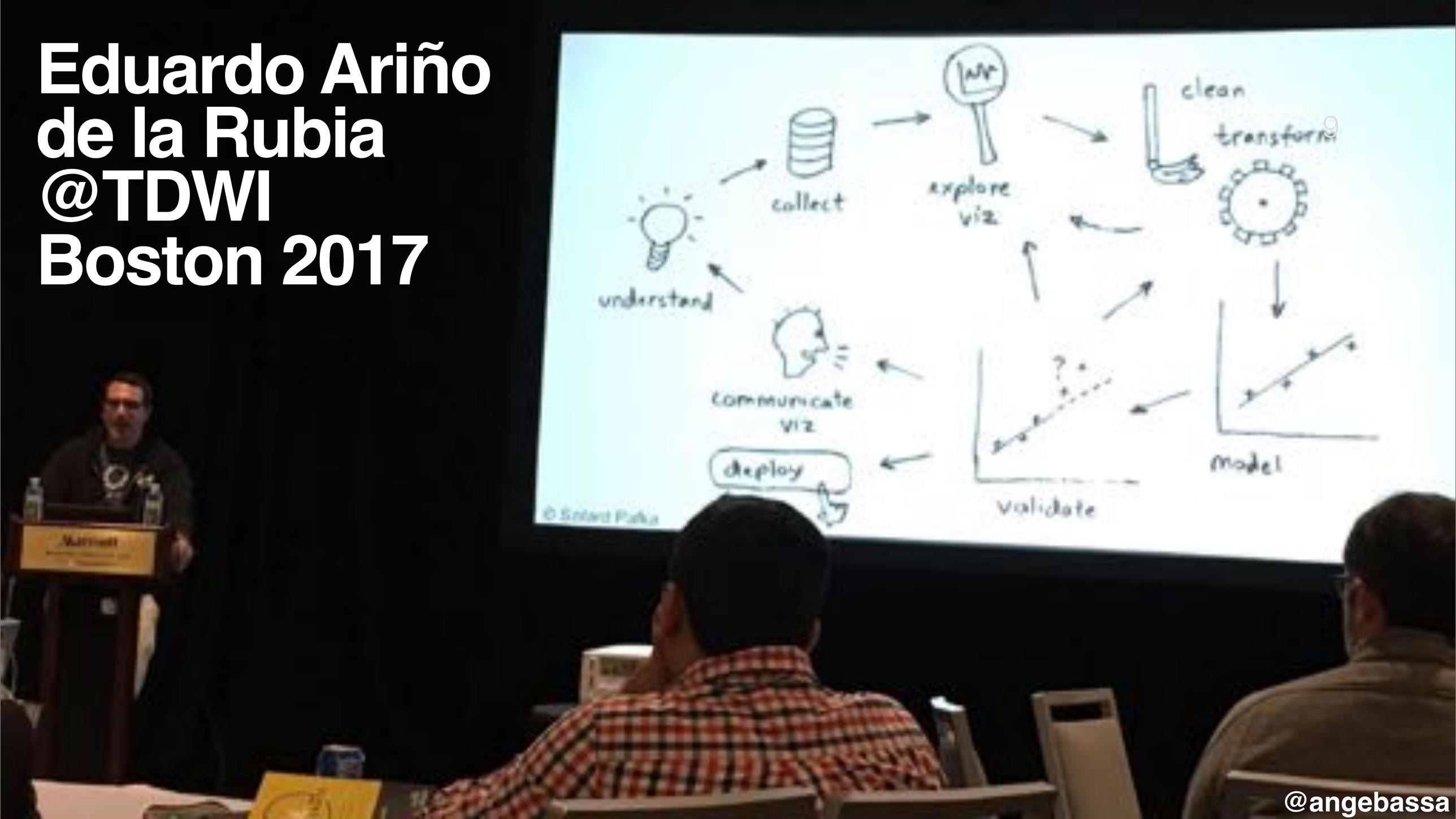
Data Sciencing In Industry

Building Effective Data Science Teams

Proper Care And Maintenance Of Data Teams

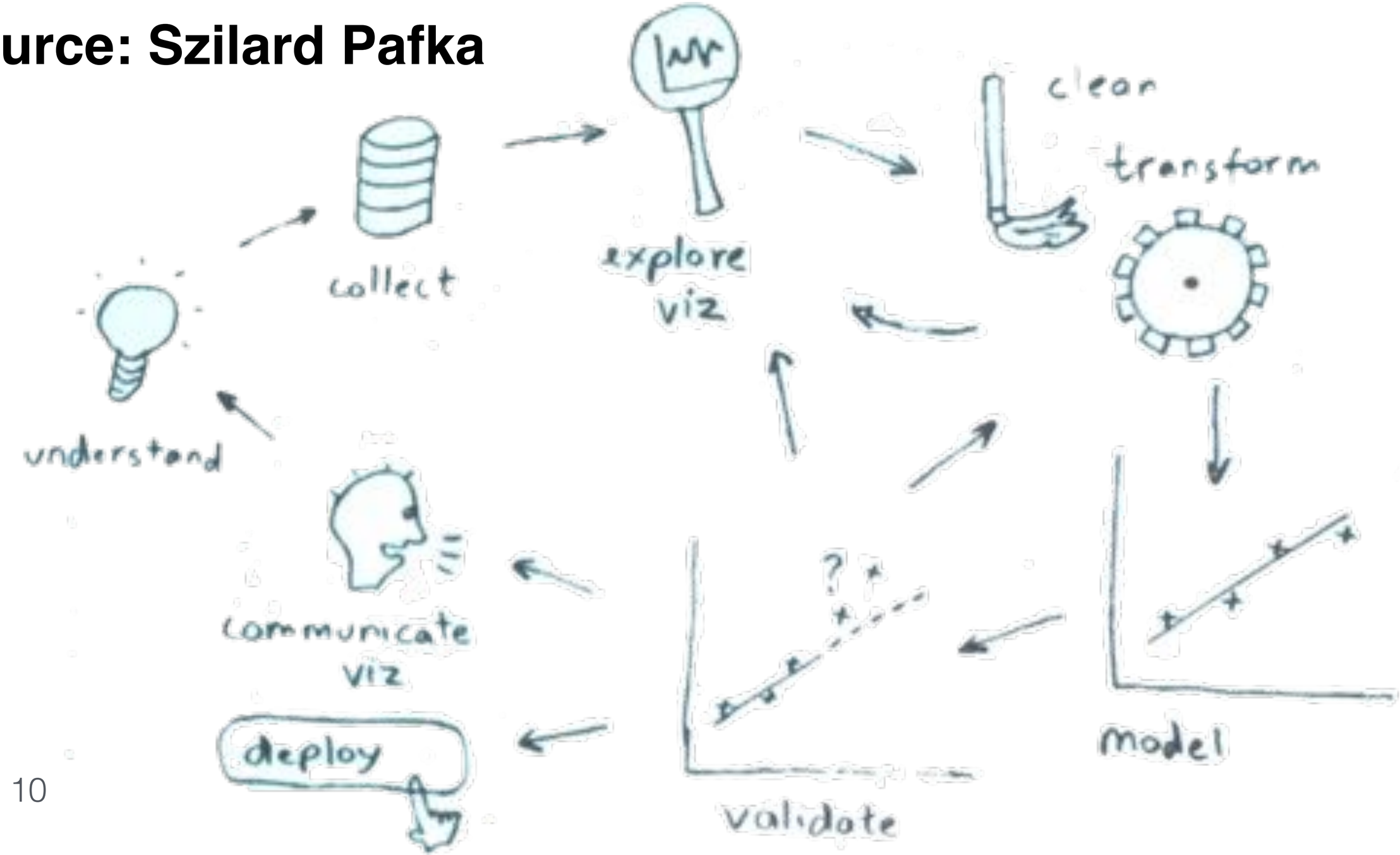
Some Things Don’t Change

Eduardo Ariño de la Rubia @TDWI Boston 2017



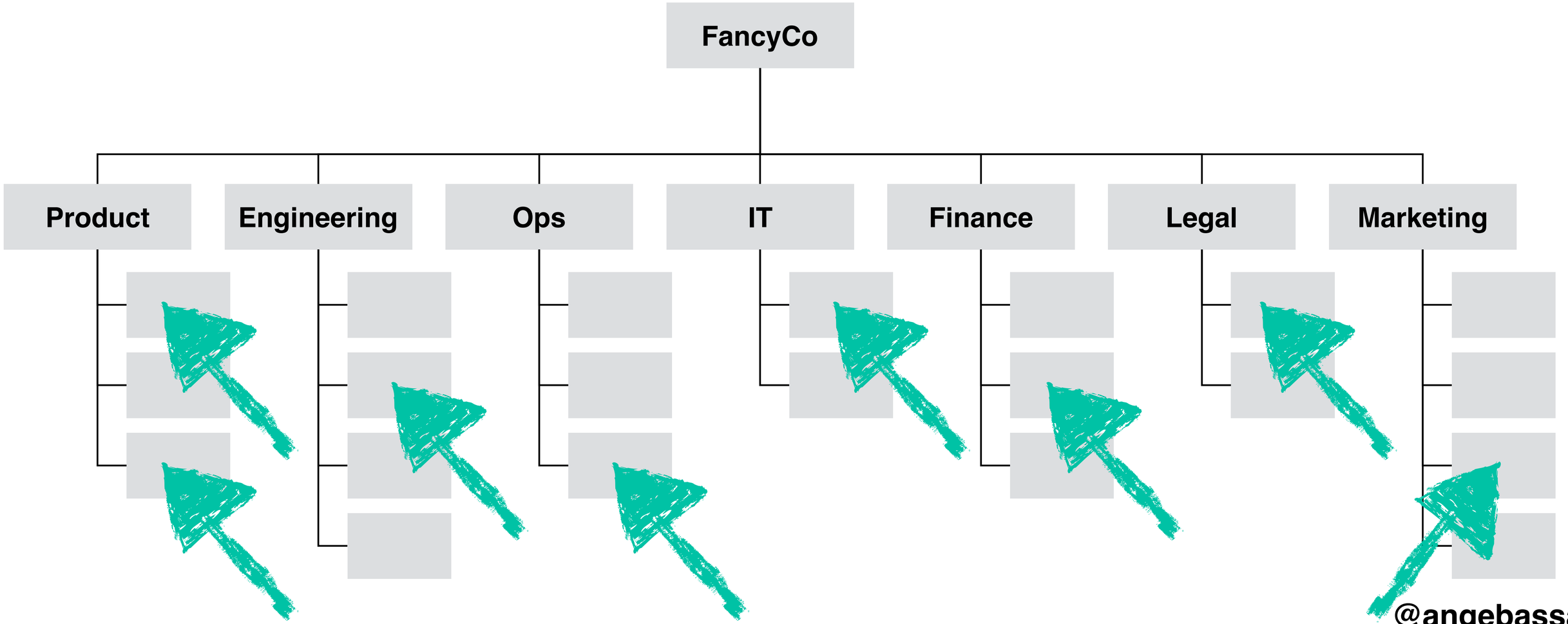
@angebassa

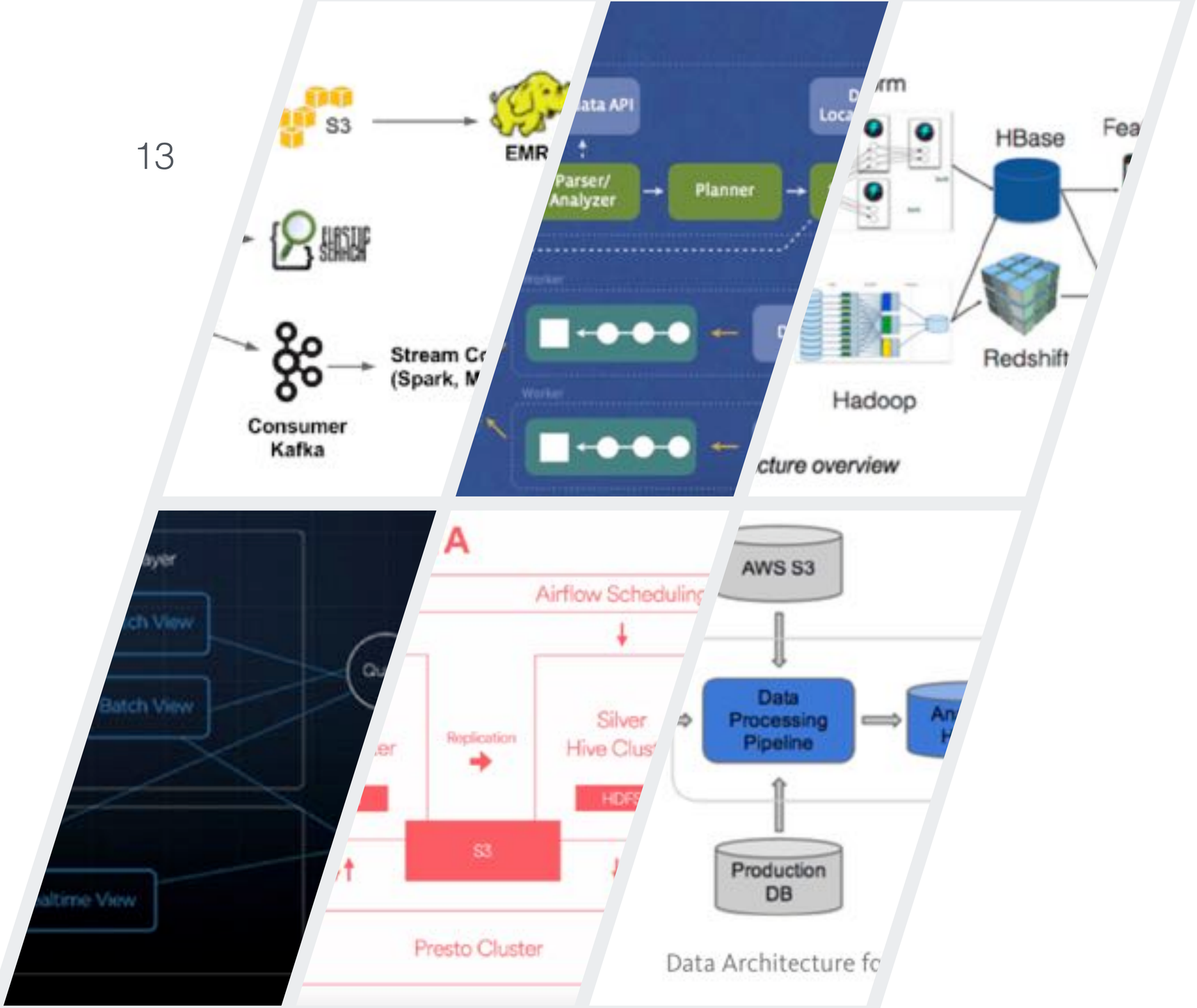
Source: Szilard Pafka



The Commercial Spectrum

Where Does Data Science Fit?





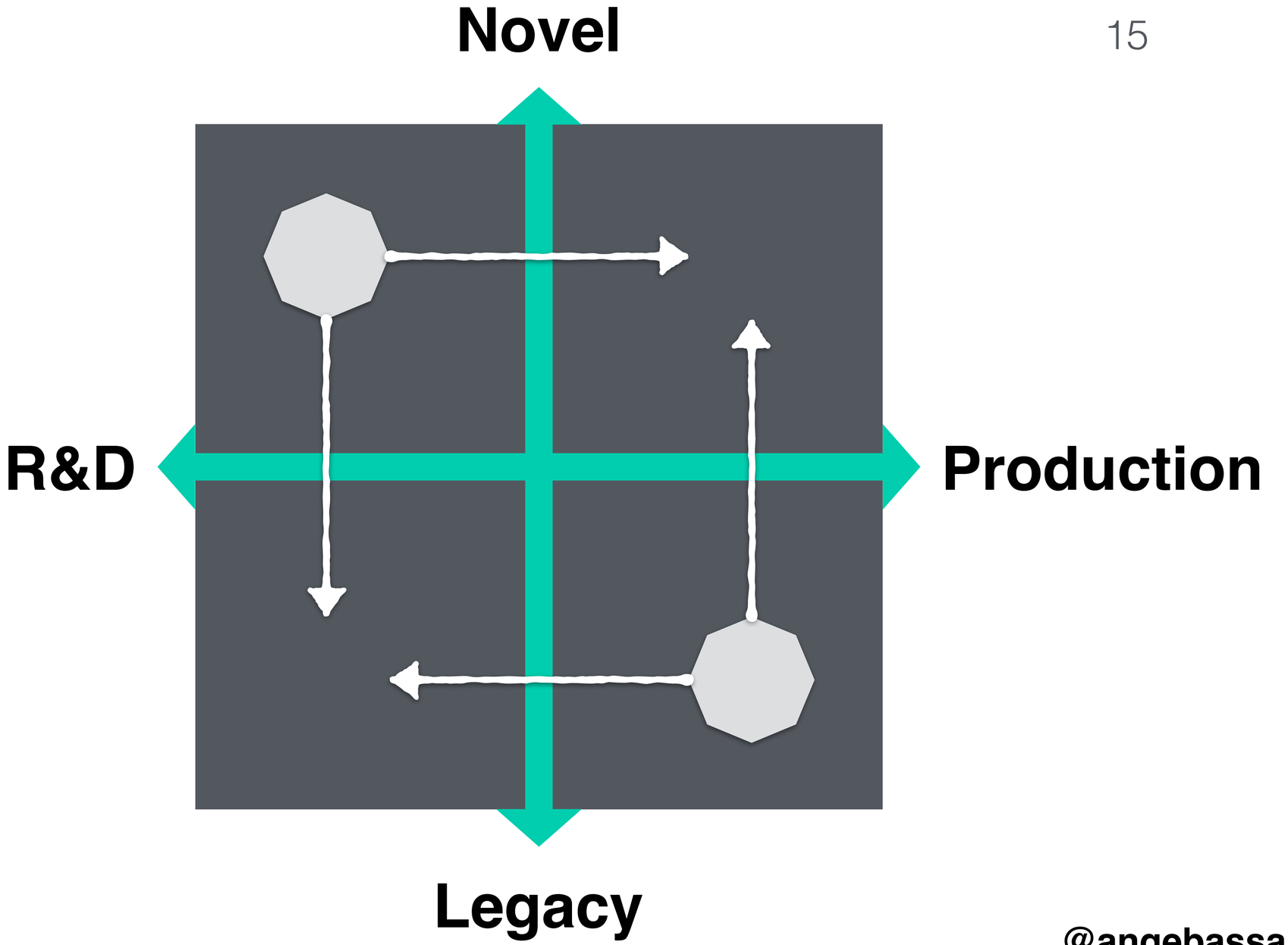
Some Things Change Quite A Bit

Pigeonholing Unicorns

Why is she here? We're not ready for analytics yet.

You design the algorithm, we will put it in production.

Blended Objectives



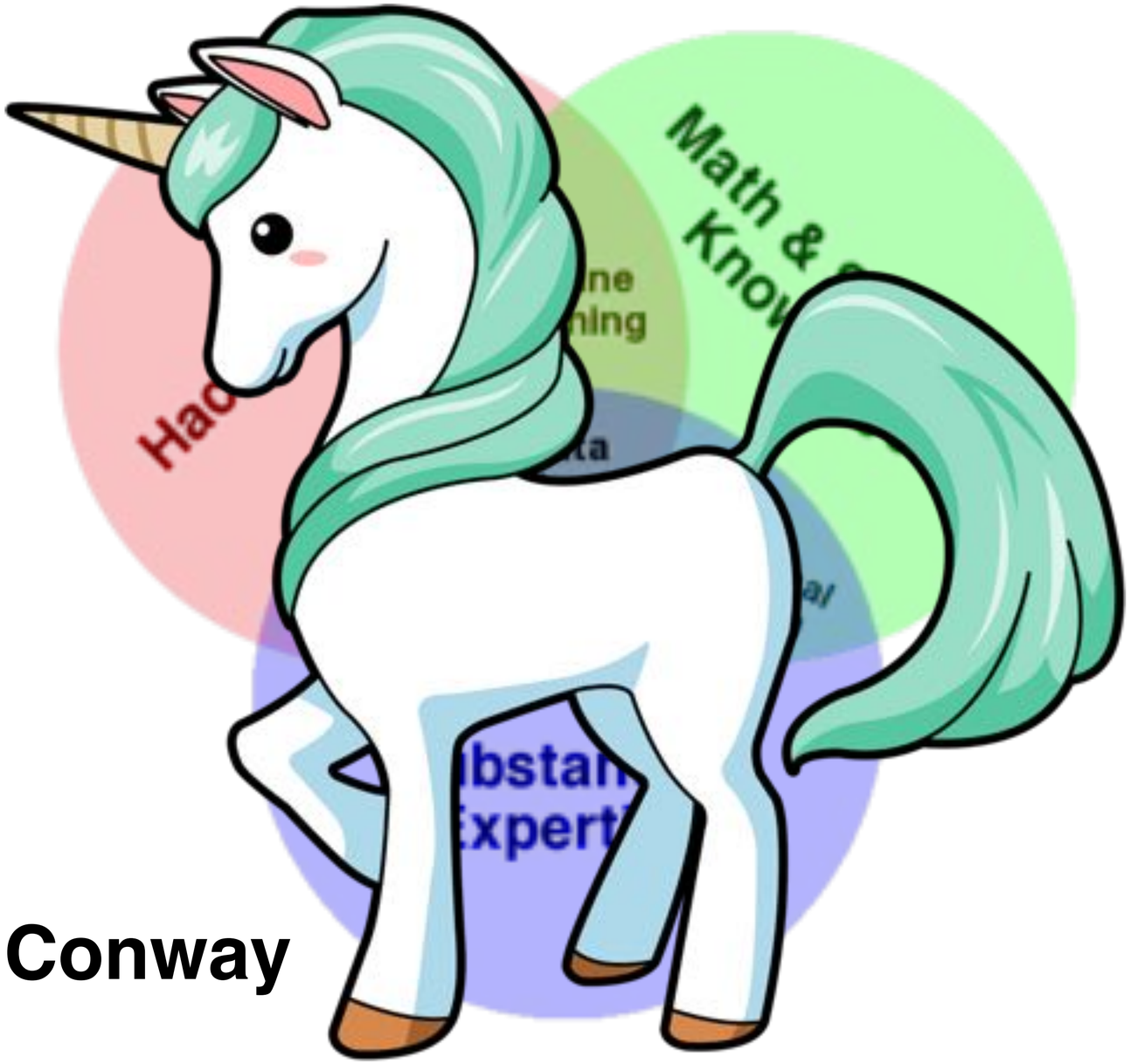
Agenda

Data Sciencing In Industry

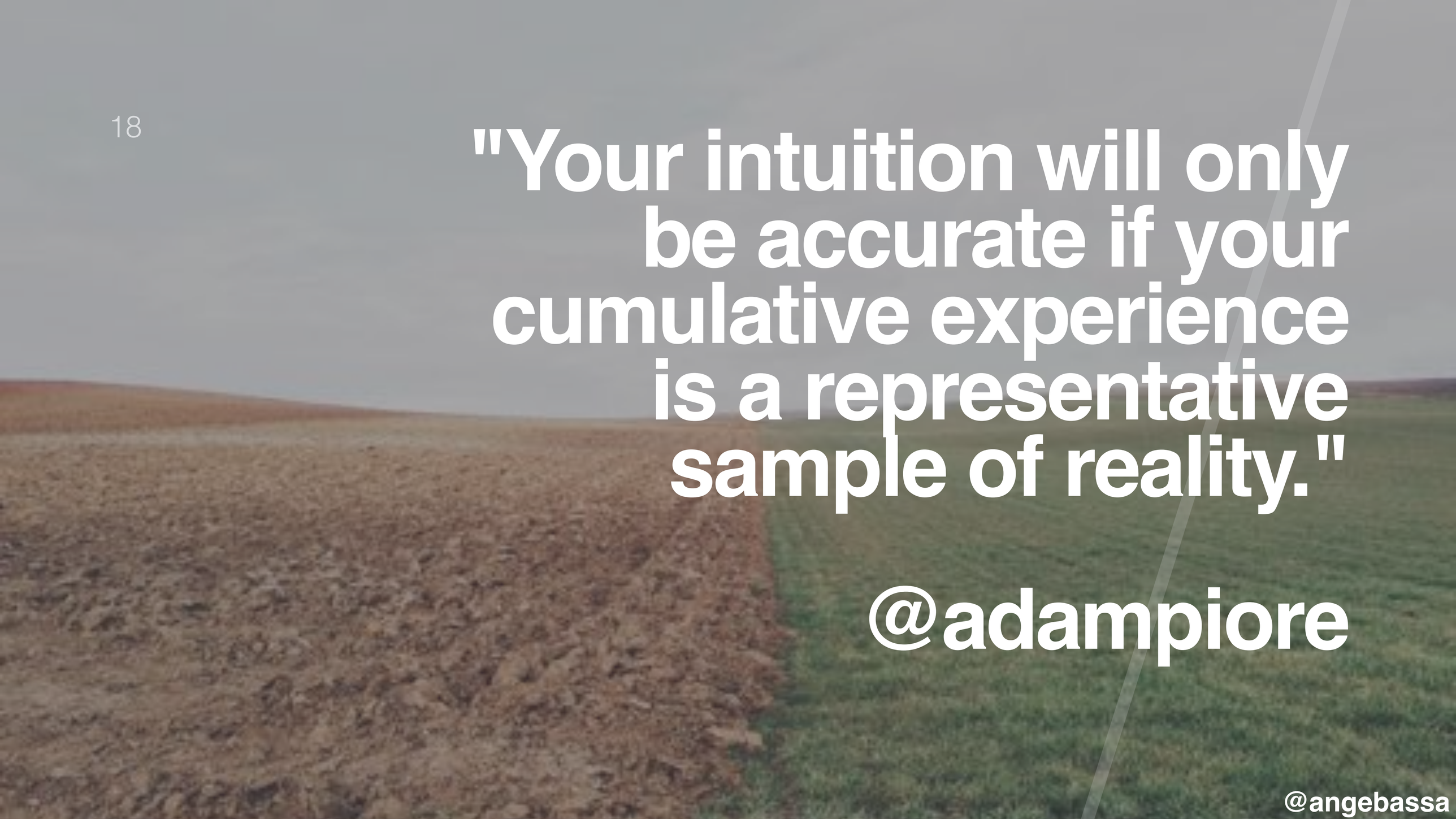
Building Effective Data Science Teams

Proper Care And Maintenance Of Data Teams

The Data Scientist Archetype



Source: Drew Conway

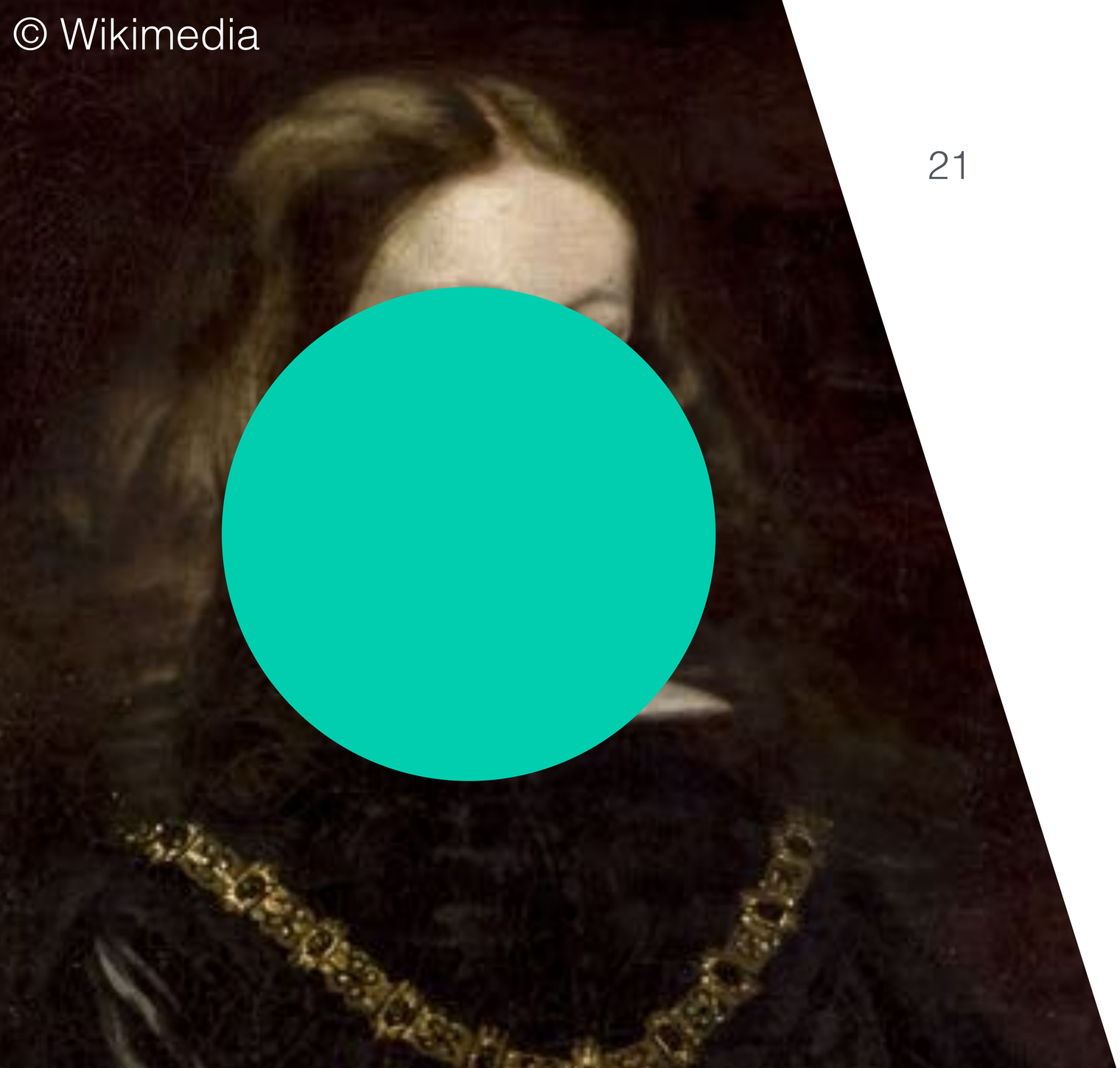


"Your intuition will only be accurate if your cumulative experience is a representative sample of reality."

@adampiore

Unicorns Don’t Really Exist

Have You Heard Of Super Chickens?

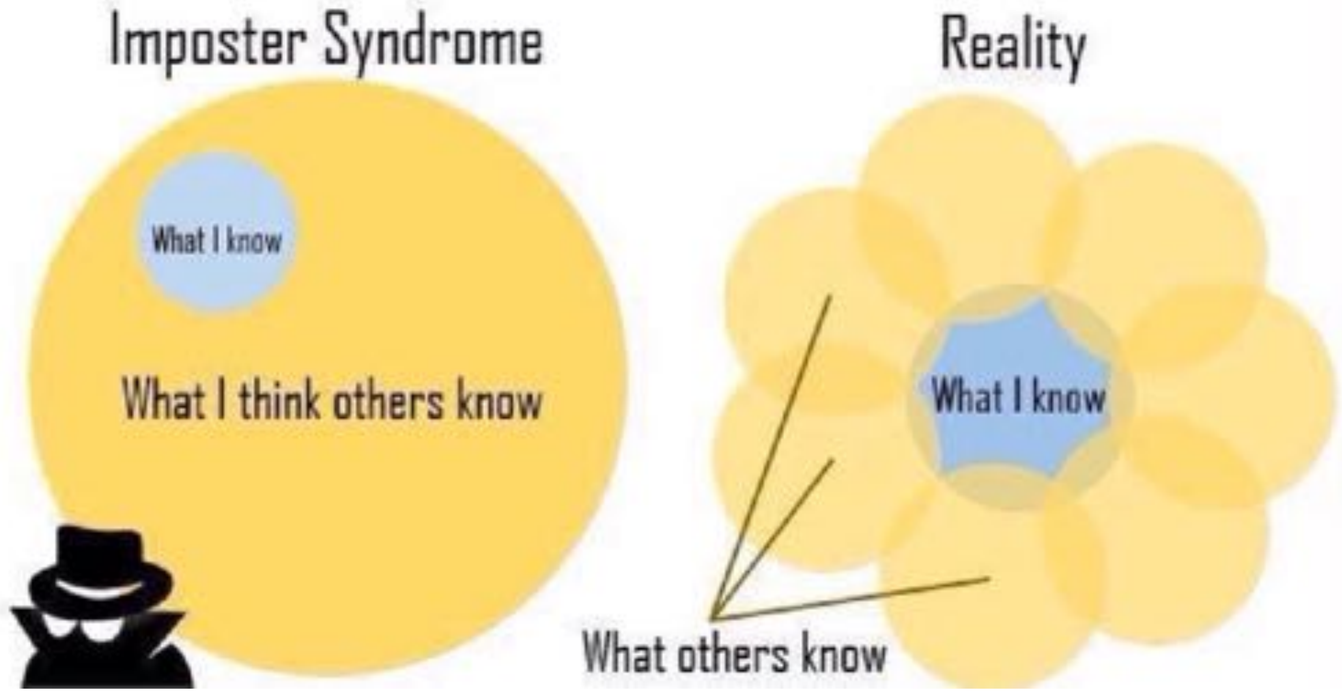


Dynasties And Intellectual Inbreeding

Mirror-tocracies[1] & Survivorship Bias

[1] Kara Swisher

Building A Large Reality



Source: David Whittaker

There’s More Than One Way To Skin A Cat

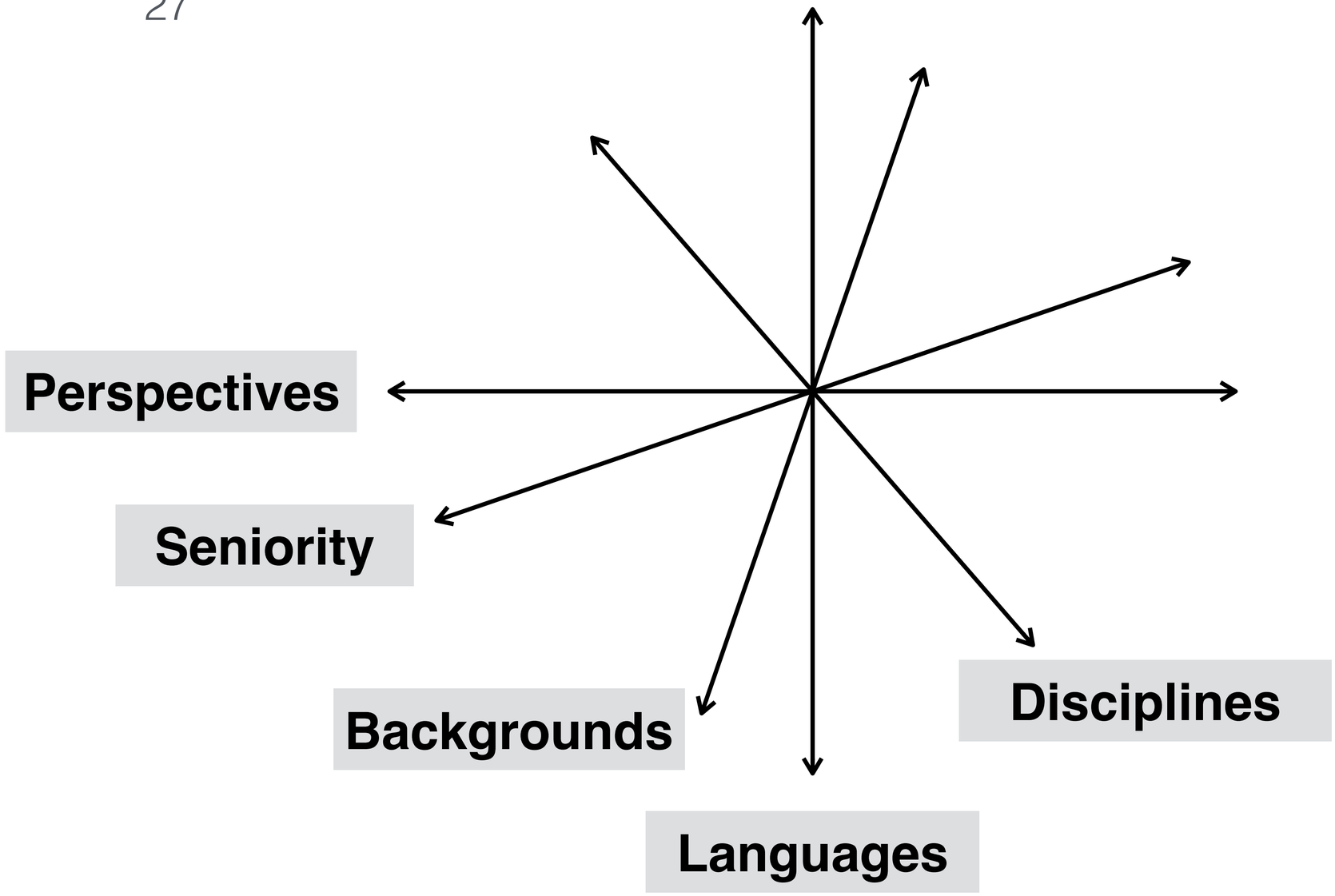
Mistakes Might Have Been Made

Common Blind Spots

Software Engineers
DLers
PhDs
Statisticians
Visualization Experts
Managers

Sampling
Rigorous Stats
Ship-It-ness
Unit Tests
Overfitting
P-hacking

Multiple-Objective Decision Analysis



How To Interview For The Best Intersections?

Tell me about your thesis/past project:

- **Can you teach me something about it?**
- **Whose ideas are you building on?**
- **Where does your idea/improvement start?**
- **Which ideas didn’t work out? Why?**
- **Who have you helped with this knowledge?**

Agenda

Data Sciencing In Industry

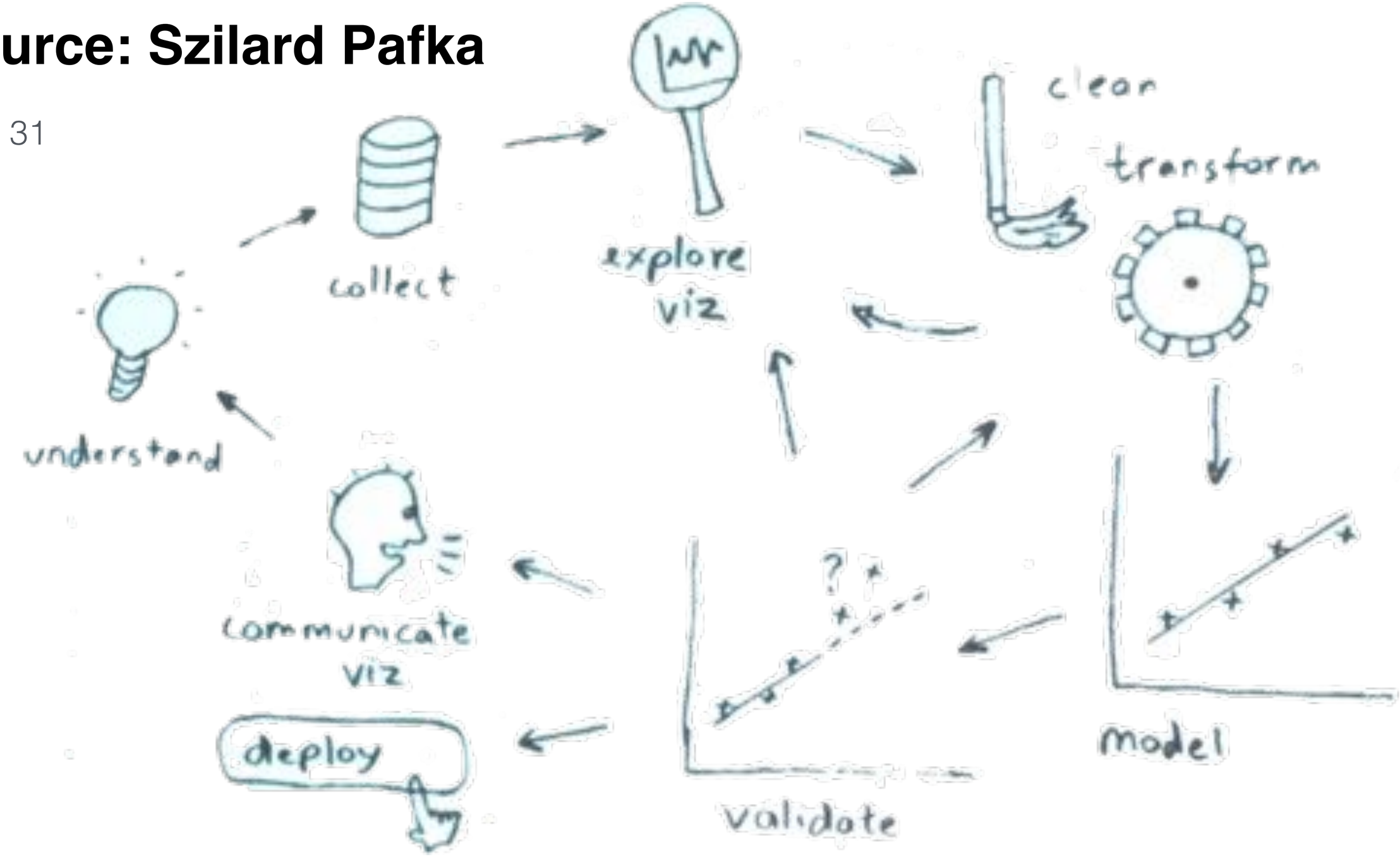
Building Effective Data Science Teams

Proper Care And Maintenance Of Data Teams

Great! You Have An Awesome Team In Place.

Now What?

Source: Szilard Pafka



Collection

Quality

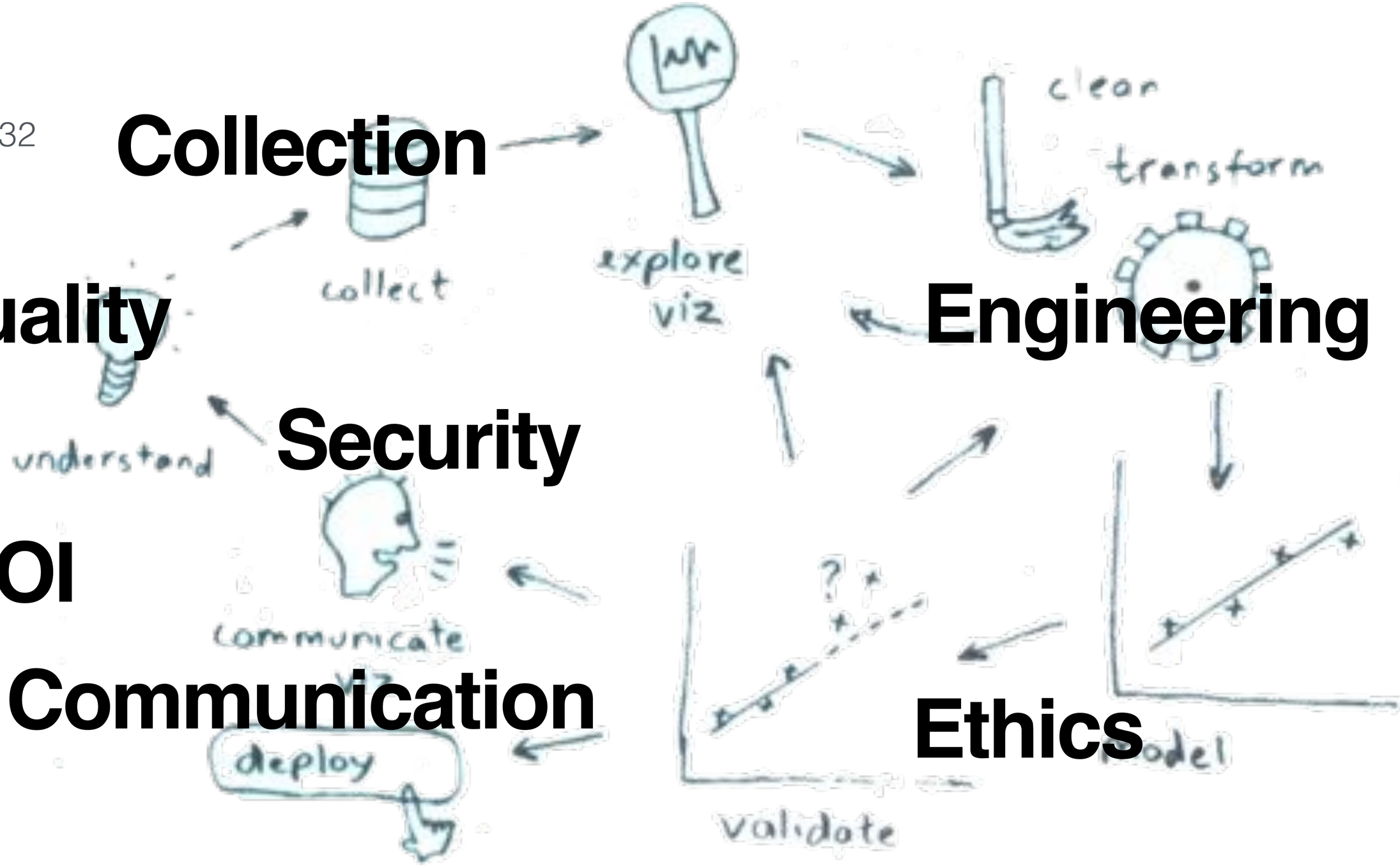
Security

ROI

Communication

Engineering

Ethics



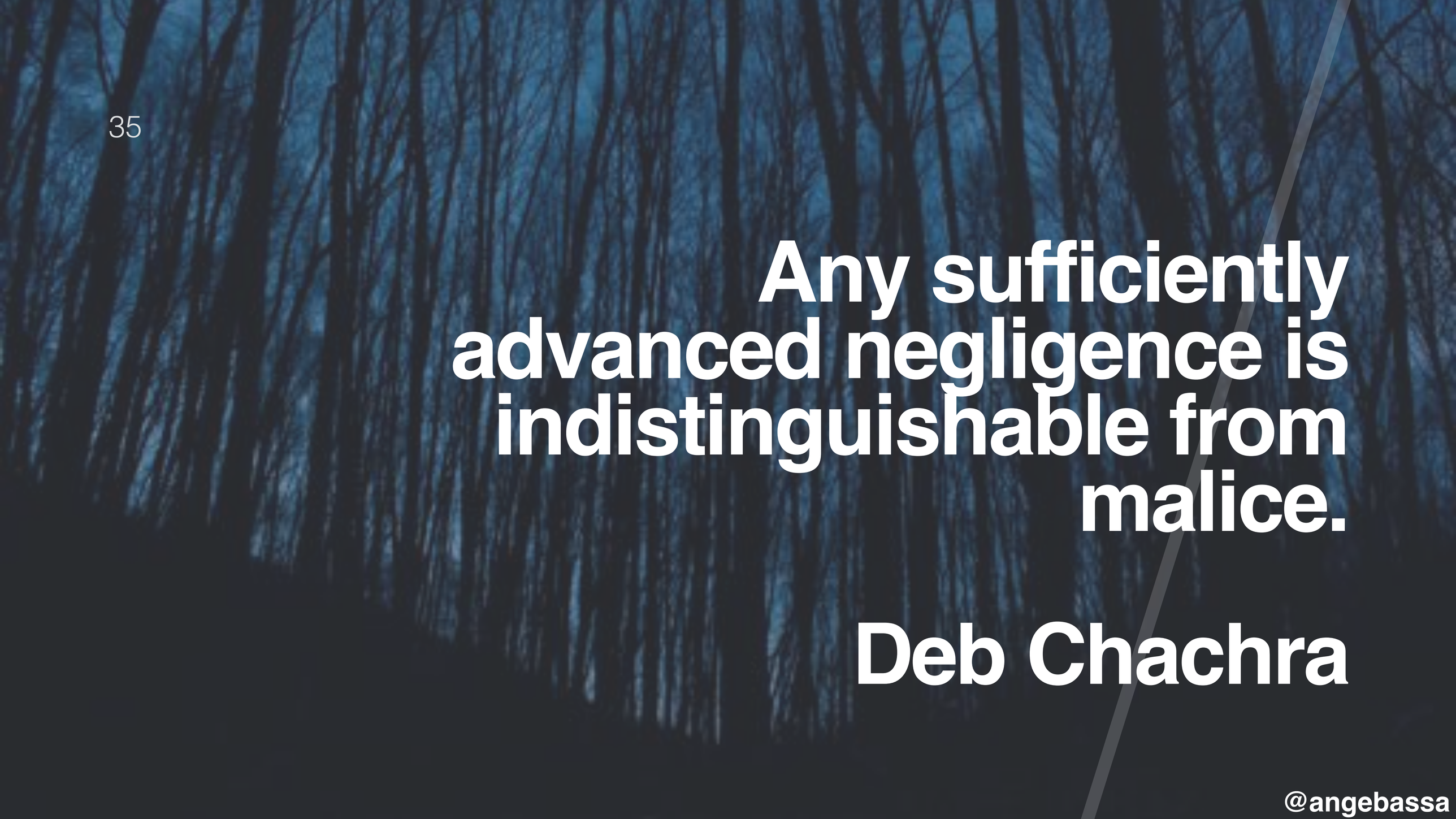
Data Team's Responsibilities

- Data Quality
- Data Collection
- Data Engineering
- Data Security
- Data Ethics
- Data Communication
- Data ROI

Why Do You Have To Care?

(Because No One Has Cared About It This Way Until Now)

@angebassa



Any sufficiently advanced negligence is indistinguishable from malice.

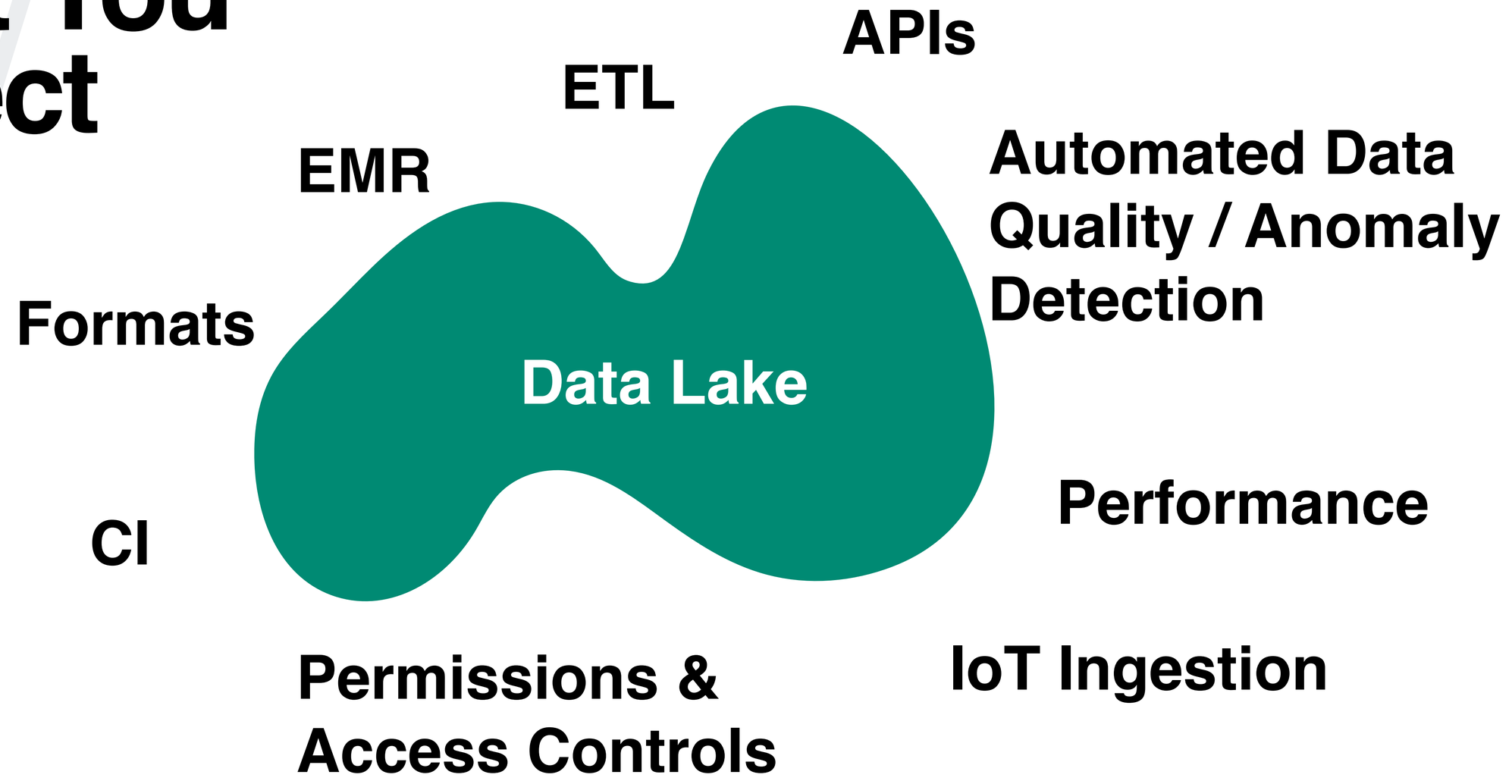
Deb Chachra

It’s Not That No One Cares

It’s That Non-Data Folks Unprepared To Care

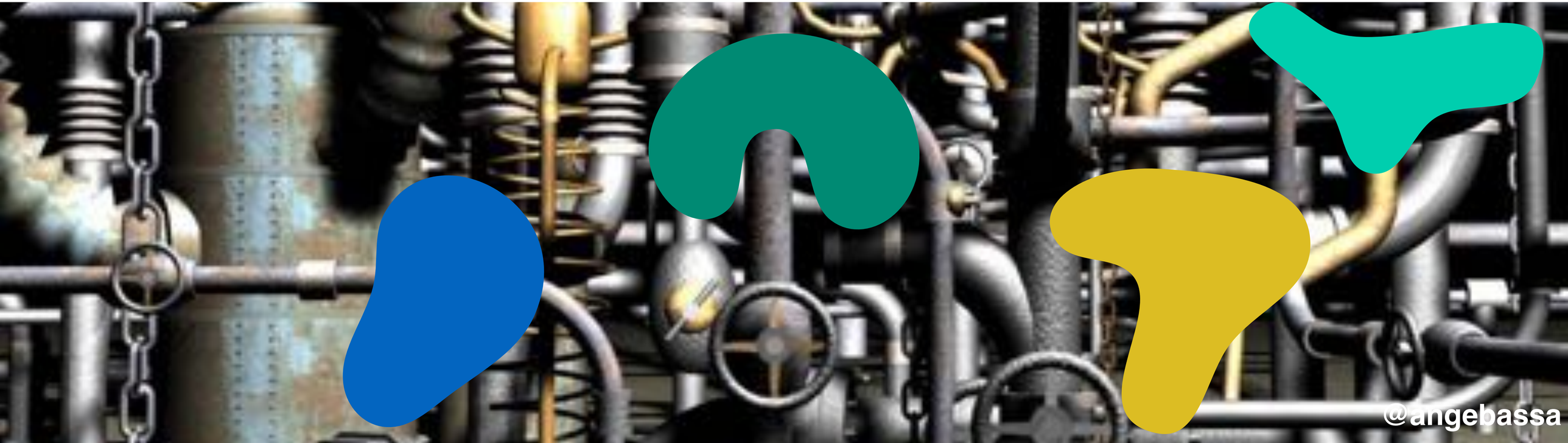
When Should You Start Caring?

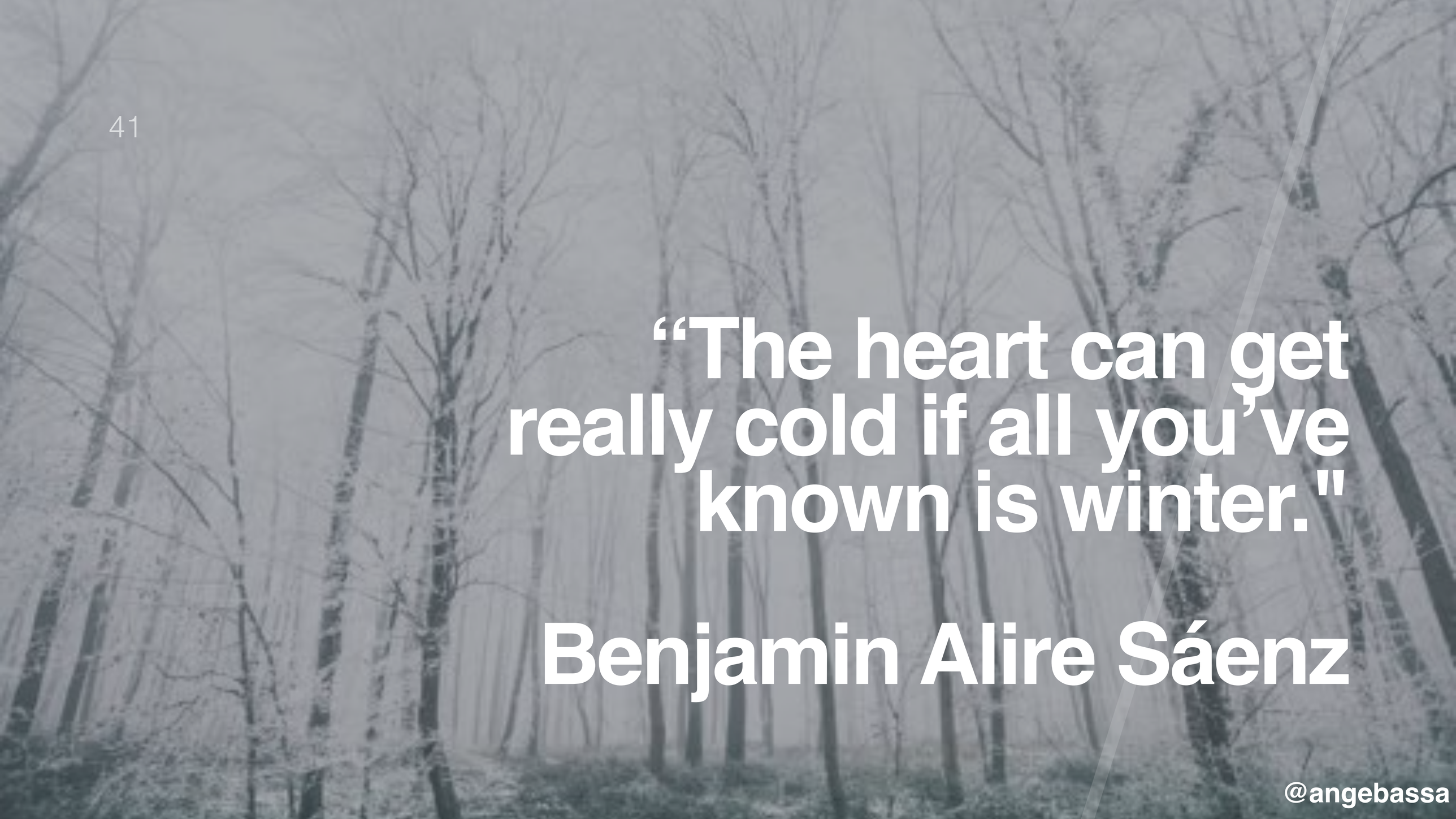
What You Expect



Mm-Hmm

What Actually Exists: A System Optimized For… Not-Analysis





“The heart can get really cold if all you’ve known is winter."

Benjamin Alire Sáenz

How To Care:

Leave Breadcrumbs

Imagine The Scene In Accounting

Documentation Is How You “Program” A Business

What Should You Document?

- Motivations for project/task
- Reasons to believe hypothesis
- Dead ends
- Paused projects
- Conjectures for next time (those gut feelz)
- *Meta-documentation:*
  - Taxonomies, Templates, and Heuristics
  - Rules of engagement (What gets stored where, who did what?)

What Will You Want To Know About The Data?

- In what format? — transformations, assumptions
- How fast? — volume, compression
- From whom? — chain of custody
- How trustworthy? — pre- and post-processing
- Where from? — performance, scale, access controls, speed

What Will You Want To Know About The Business?

- What does the business do? Really?
- How does it make money?
- How does it create value for customers?
- Who buys it? Who pays for it? Who are the competitors?
- What assumptions are baked in?
- Do you have enough data to validate these answers?

Having Junior Team Members Is A Great Forcing Function!

And We Come Full Circle: This Is Why You Build Your Teams This Way

Made with Unsplash, Wikimedia, and fair use of commercial images.
Thank You!
@angebassa