

Collapsing the IT stack

Clearing a path for AI adoption

Alan Morrison
Senior Research Fellow
Integrated Content | Emerging Tech



Outline for today's talk

Early adopters--winning the data war

Collapsing the IT stack

Diagnosing the problem

Progress on solutions

Conclusion



Early adopters--winning the data war

Largest change in market cap by company (2009 to 31 March 2018)

	Company name	Location	Industry	Change in market cap 2009-2018 (\$bn)	Market cap 2018 (\$bn)
1	Apple	United States	Technology	757	851
2	Amazon.Com	United States	Consumer Services	670	701
3	Alphabet	United States	Technology	609	719
4	Microsoft Corp	United States	Technology	540	703
5	Tencent Holdings	China	Technology	483	496
6	Facebook	United States	Technology	383(1)	464
7	Berkshire Hathaway	United States	Financial	358	492
8	Alibaba	China	Consumer Services	302(2)	470
9	JPMorgan Chase	United States	Financials	275	375
10	Bank of America	United States	Financials	263	307

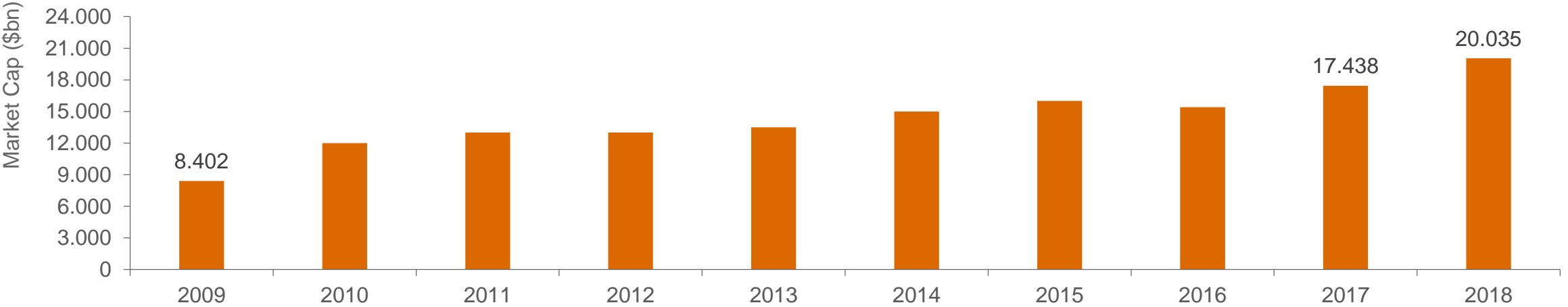
(1) Change in market cap from IPO date

(2) Market cap at IPO date

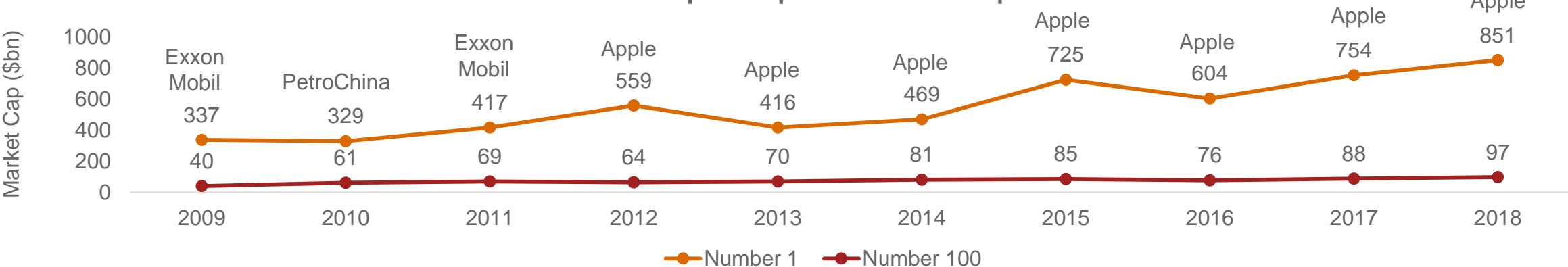
Source: Bloomberg and PwC analysis

Widening corporate inequality – Top versus bottom of Top 100

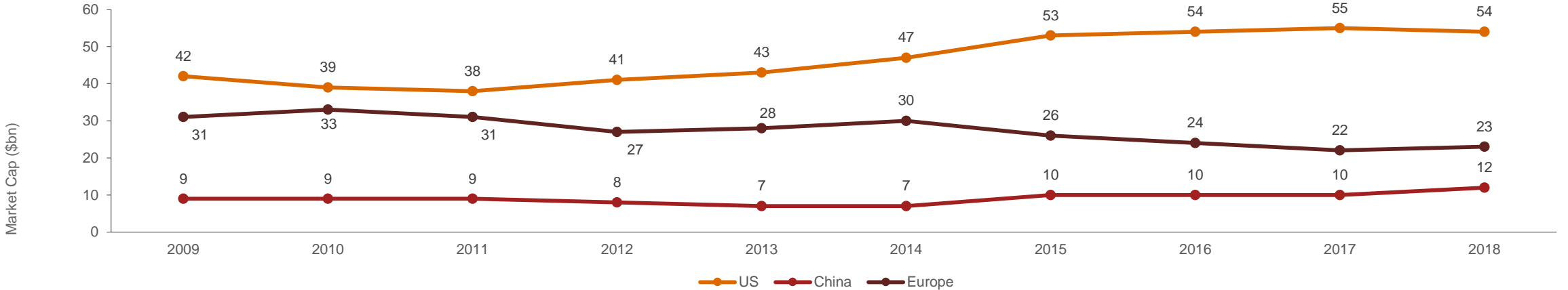
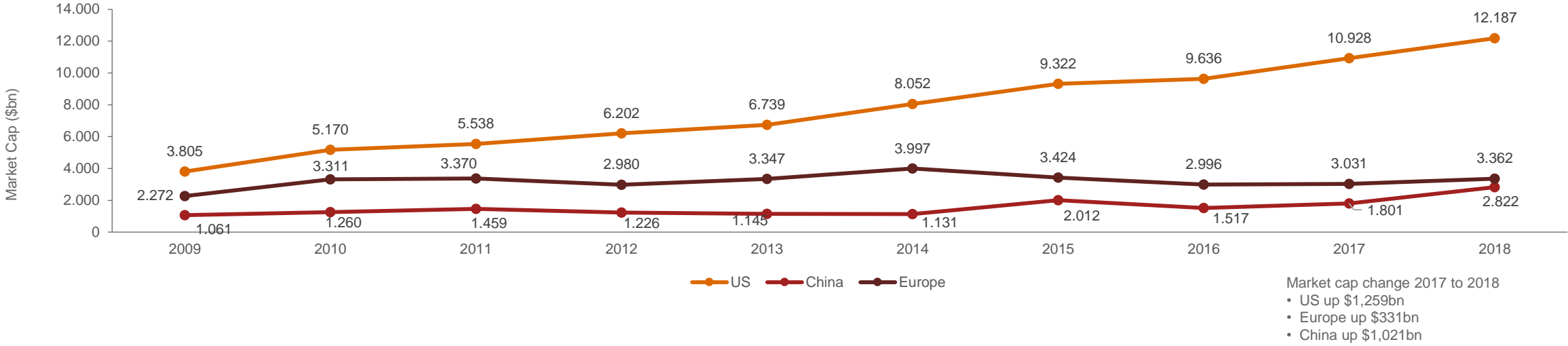
Total market cap of Top 100 companies as at 31 March



Market caps of top and bottom companies

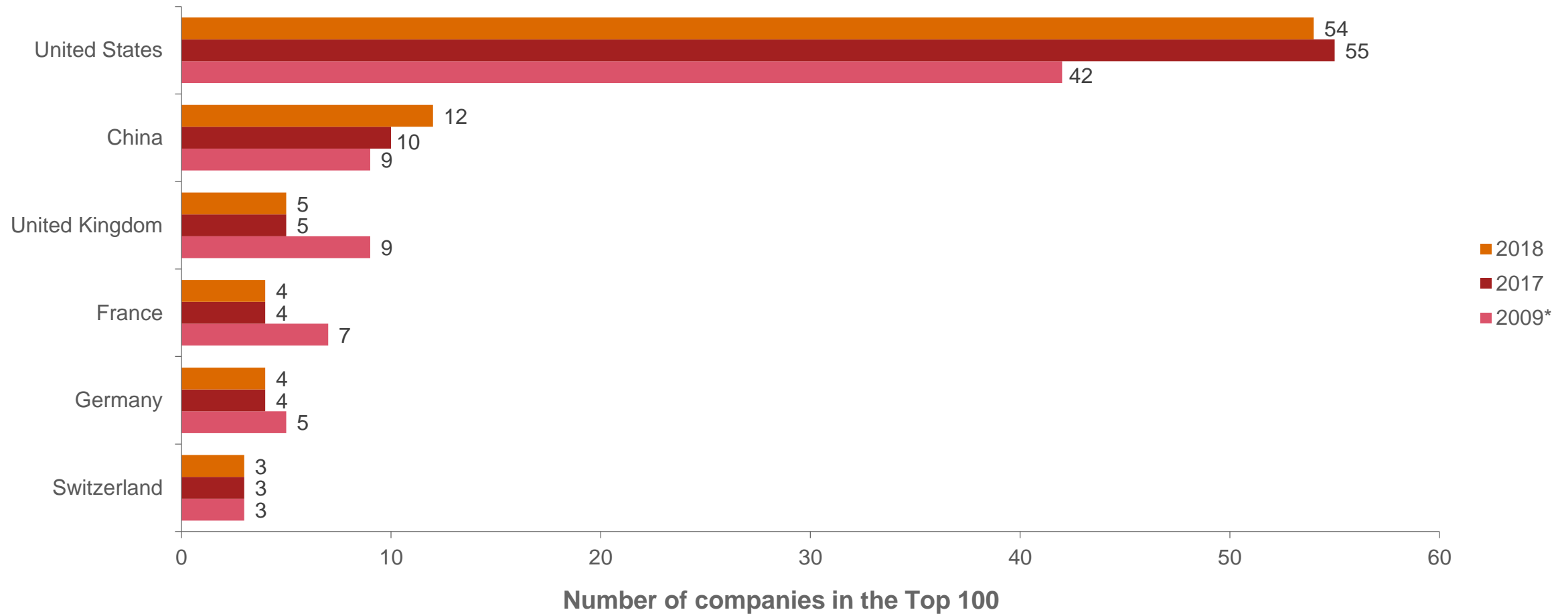


Widening corporate inequality in the Top 100—by country/region to 31 March



Source : Bloomberg and PwC analysis

Widening corporate inequality – by country to from 2009 to 31 March 2018



*2009 figures do not add to 100 due to seven companies in the 2009 Top 100 being in locations of domicile that are no longer in the Global Top 100

Source: Bloomberg and PwC analysis

San Francisco Bay Area now in Top 20 economies worldwide.... but for how long?

“The Bay Area has the **19th-largest economy in the world, ranking above Switzerland and Saudi Arabia....**

Startups, particularly those in the consumer-internet business, increasingly **struggle to attract capital in the shadow of Alphabet, Apple, Facebook et al.**”

--*The Economist*, “Why startups are leaving Silicon Valley,” 30 Aug 2018

Largest change in market cap by company (2009 to 31 March 2018)

	Company name	Location	Industry	Change in market cap 2009-2018 (\$bn)	Market cap 2018 (\$bn)
1	Apple	United States	Technology	757	851
2	Amazon.Com	United States	Consumer Services	670	701
3	Alphabet	United States	Technology	609	719
4	Microsoft Corp	United States	Technology	540	703
5	Tencent Holdings	China	Technology	483	496
6	Facebook	United States	Technology	383(1)	464
7	Berkshire Hathaway	United States	Financial	358	492
8	Alibaba	China	Consumer Services	302(1)	470
9	JPMorgan Chase	United States	Financials	275	375
10	Bank of America	United States	Financials	263	307

Known knowledge graph builders

Operator of Taobao and KG builder

Known KG builders

(1) Change in market cap from IPO date

(2) Market cap at IPO date

Source: Bloomberg and PwC analysis

Collapsing the IT stack



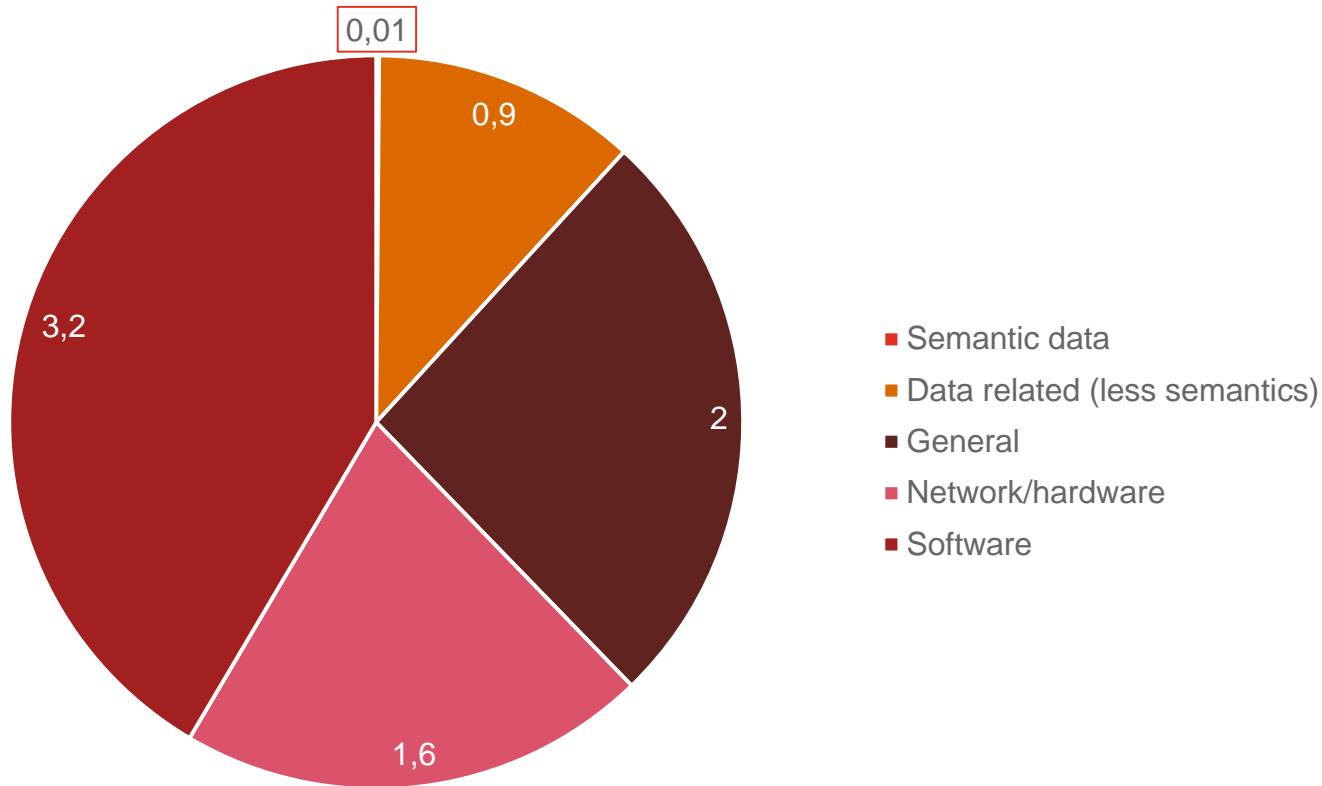
Wikimedia Commons, 2007

Most innovations are incremental, adding to the stack, with data as an afterthought (Type I)



Most of the IT workforce just adds to or keeps track of the sprawl

US IT workforce in 2016 (in millions)

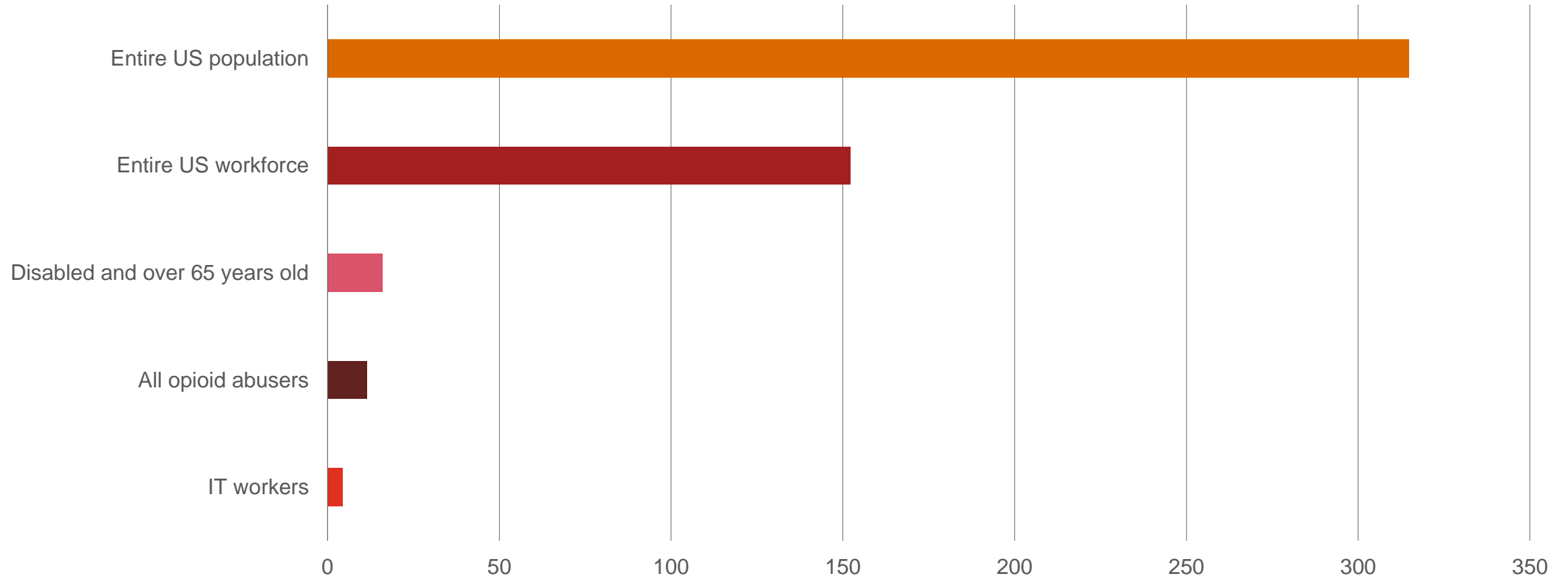


Total IT workforce = 7.7 million
(= 5 percent of the US overall workforce in 2016)

Sources: US Bureau of Labor Statistics and PwC estimates, 2018

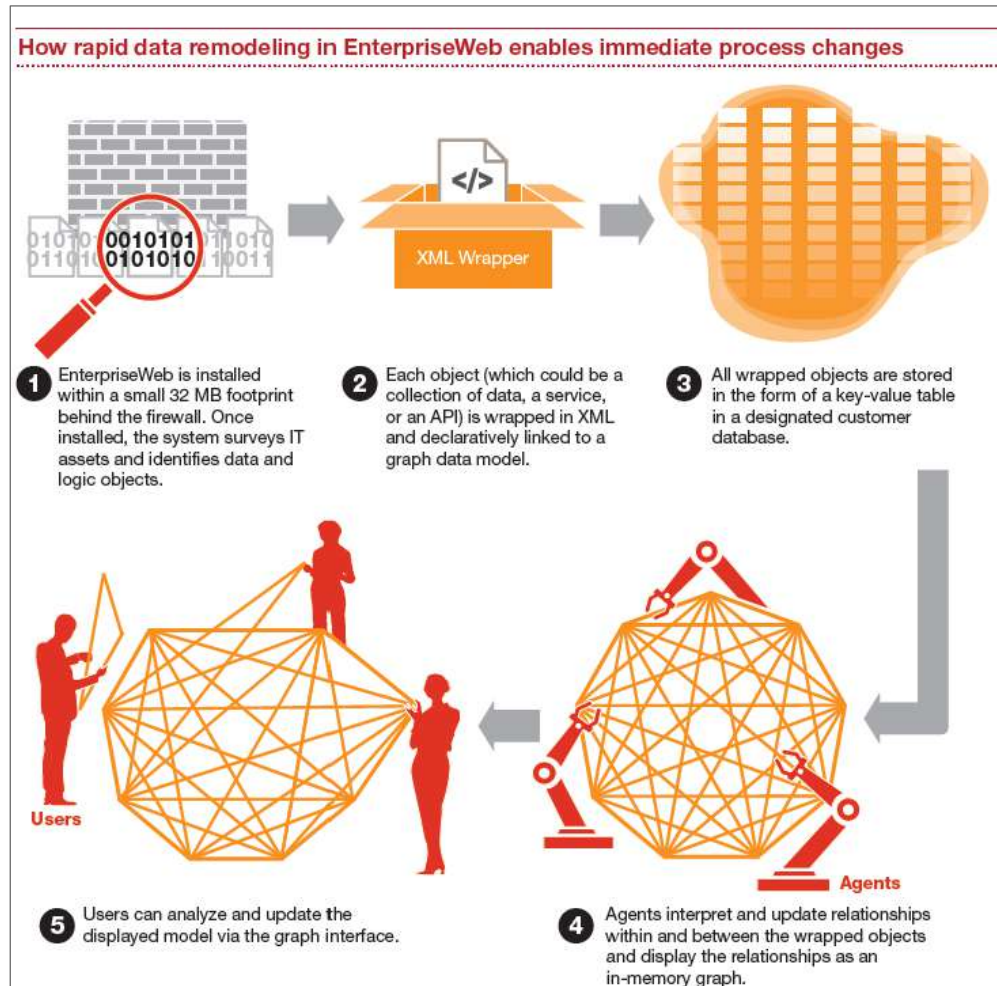
The US as a whole has more opioid abusers than it does IT workers

Number of IT workers in the US in 2016, in context (in millions)



US Census Bureau, Bureau of Labor Statistics, and Health and Department of Human Services, 2018

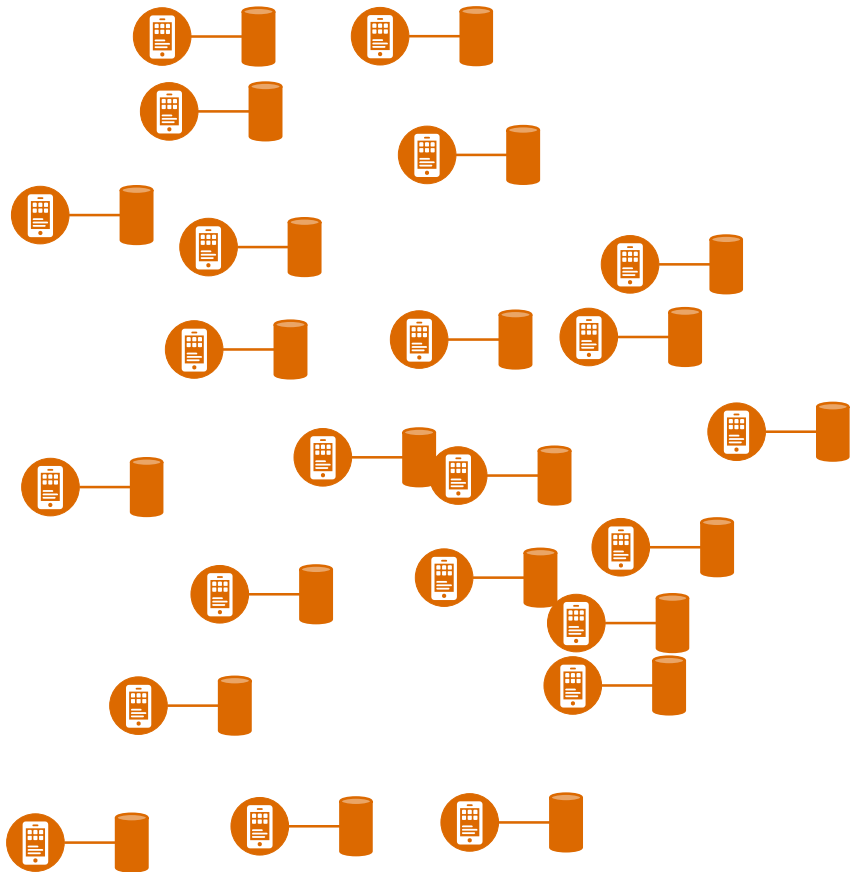
Object virtualization (Type II) manages complexity, just so IT can get its arms around the sprawl



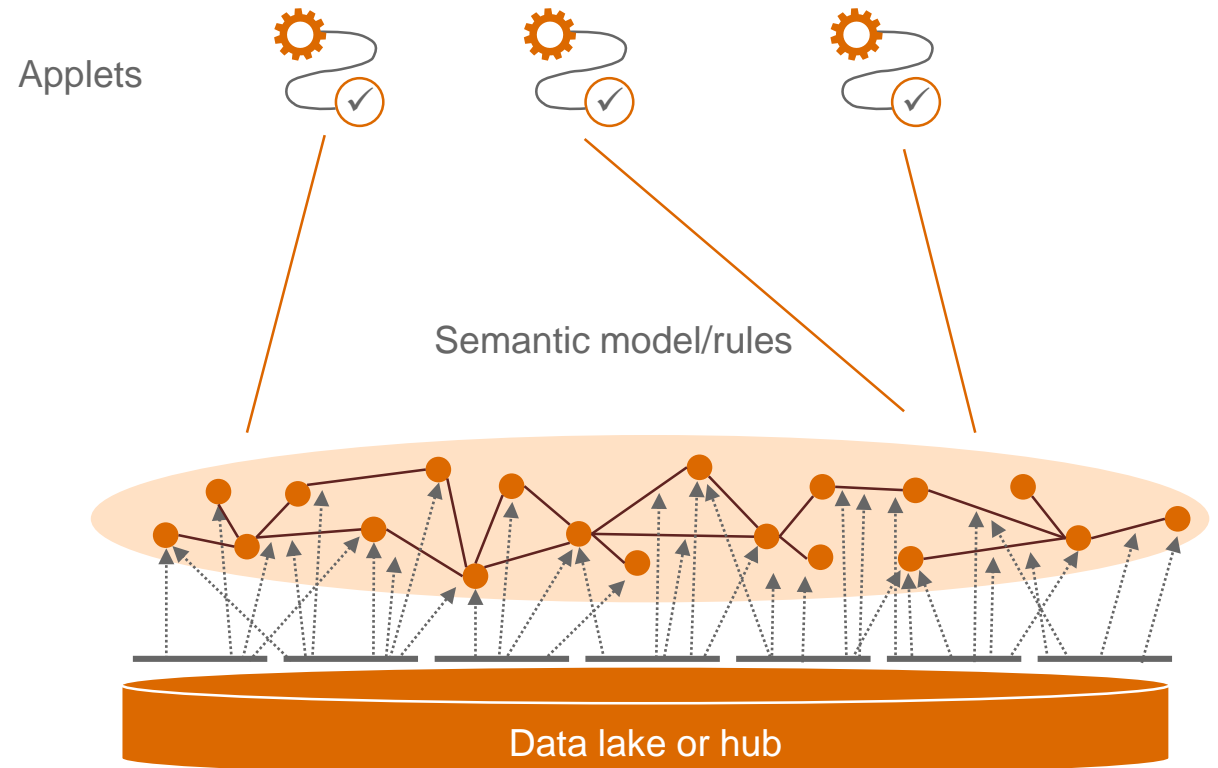
EnterpriseWeb and PwC, 2015

Type III: data-centric architecture reduces both application and database sprawl

App code trapped in Database orphans and models



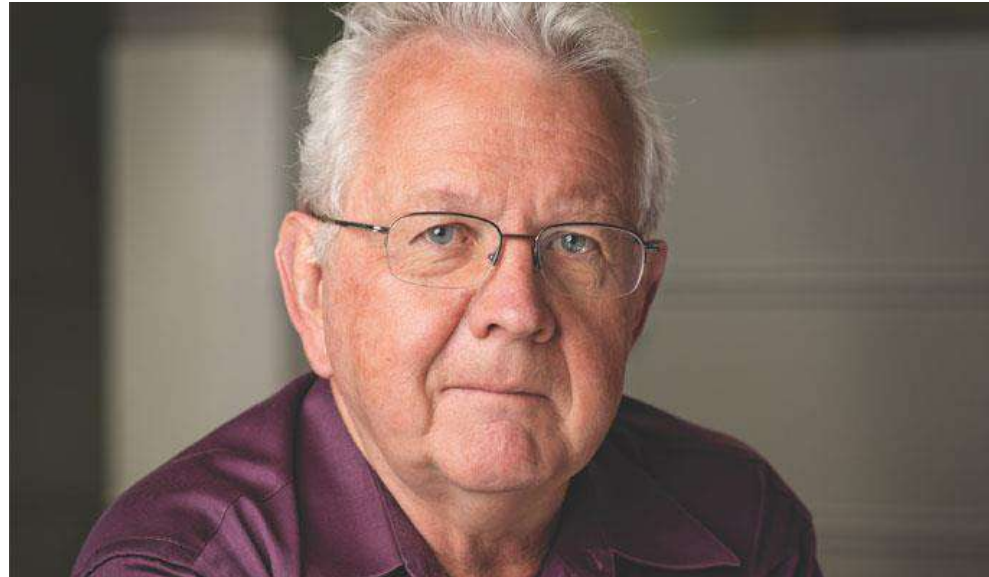
Applications for execution only, models exposed with the data



Identify and declare the few hundred business rules you need as a model

“In every company I’ve ever studied, **there are only a few hundred key concepts and relationships that the entire business runs on.** Once you understand that, you realize all of these millions of distinctions are just slight variations of those few hundred important things.”

--Dave McComb, author of *Software Wasteland*, quoted in *Strategy + Business*

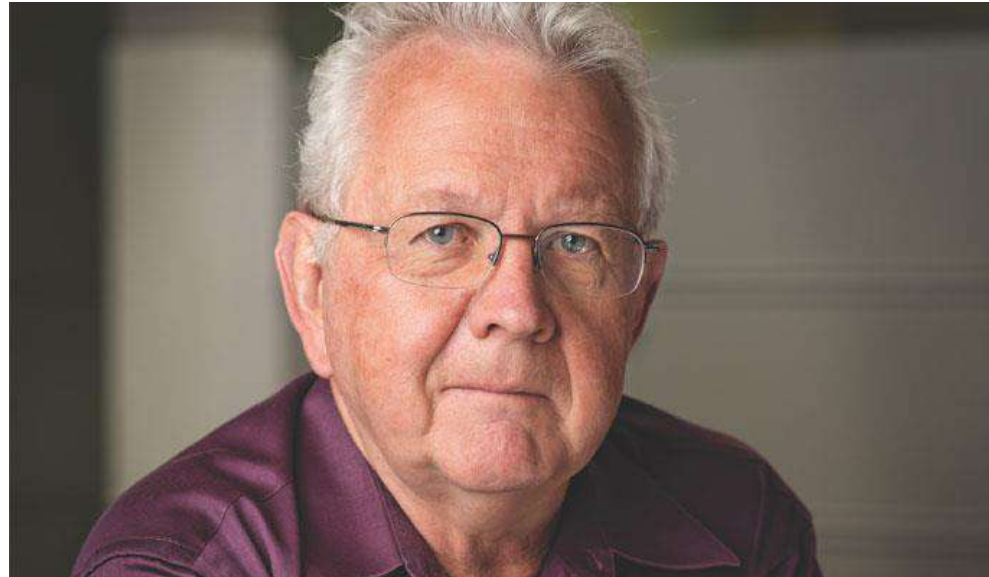


See “Are you Spending Way too Much on Software at <https://www.strategy-business.com/article/Are-You-Spending-Way-Too-Much-on-Software?>”

Call the model to reuse those rules whenever necessary

“You discover that **many of the slight variations aren’t variations at all**. They’re really the same things with different names, different structures, or different labels. So it’s desirable to describe those few hundred concepts and relationships in the form of **a declarative model that small amounts of code refer to again and again.**”

--Dave McComb (as previously cited)



See “Are you Spending Way too Much on Software at [https://www.strategy-business.com/article/Are-You-Spending-Way-Too-Much-on-Software?](https://www.strategy-business.com/article/Are-You-Spending-Way-Too-Much-on-Software?st=software)”

Diagnosing the bigger problem



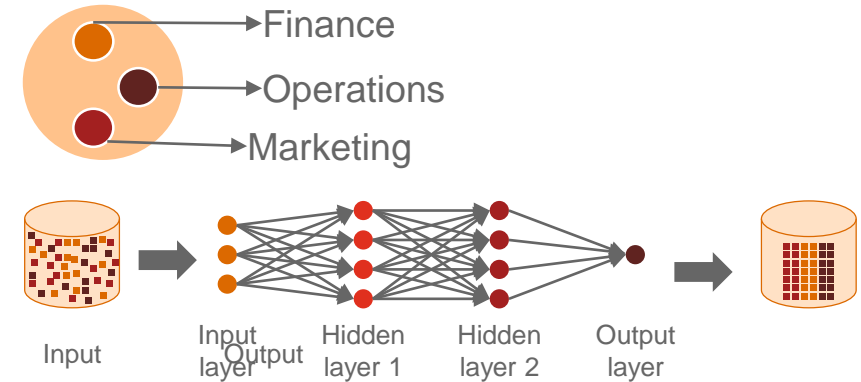
What AI needs versus what it has

What it needs: Contextualized, disambiguated, highly relevant and specific integrated data, flowing to the point of need

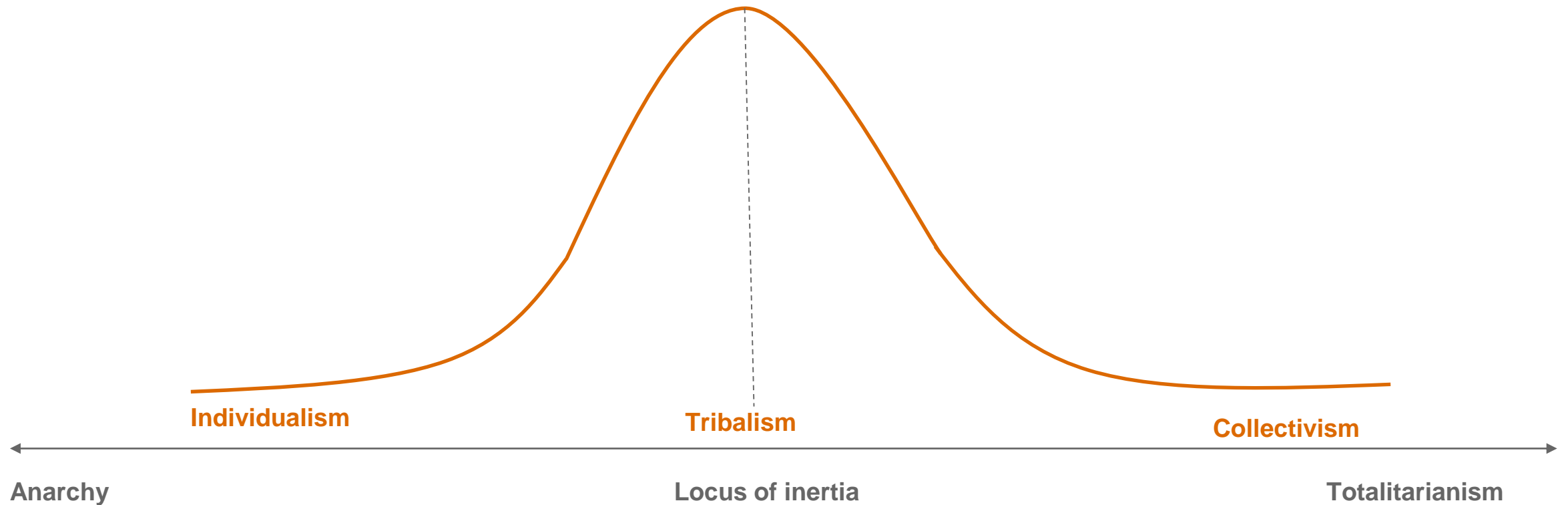
What it has: Single batch datasets cleaned up to be good enough by data scientists, who spend 80% of their time on cleanup

What it needs: Knowledge engineers, and many bold Data Visionaries in addition to big D Data Scientists, data-centric architects, pipeline engineers, specialists in many new data niches

What it has: A growing group of tool users versed only in probability theory, neural networks, python and R, including small D data scientists, engineers and architects, plus scads of entrenched application-centric developers



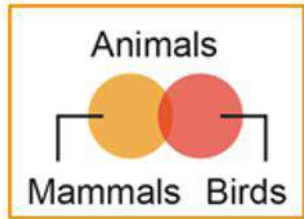
The real inhibitors to adoption aren't technological – they're rooted in tribal biases and resistance to change



Daniel Quinn, *Beyond Civilization* and Alice Linsley, "Daniel Quinn: A Return to Tribalism?", college-ethics.blogspot.com, 2018

Tribalism – Machine learning edition

Symbolists



Use symbols, rules, and logic to represent knowledge and draw logical inference

Favored algorithm
Rules and decision trees

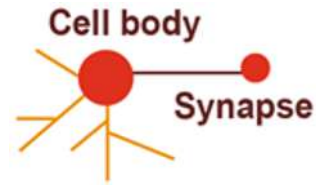
Bayesians



Assess the likelihood of occurrence for probabilistic inference

Favored algorithm
Naïve Bayes or Markov

Connectionists



Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons

Favored algorithm
Neural network

Evolutionaries



Generate variations and then assess the fitness of each for a given purpose

Favored algorithm
Genetic programs

Analogizers



Optimize a function in light of constraints (“going as high as you can while staying on the road”)

Favored algorithm
Support vectors

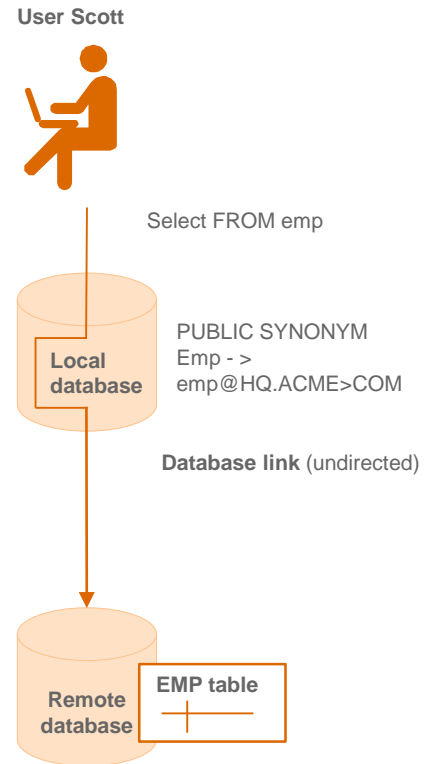
Source: Pedro Domingos, *The Master Algorithm*, 2015

More at “Machine learning evolution”: <http://usblogs.pwc.com/emerging-technology/machine-learning-evolution-infographic/>, PwC, 2017

Tribalism – Data integration edition

Trend toward more data centrality this way →

Relational database linkers

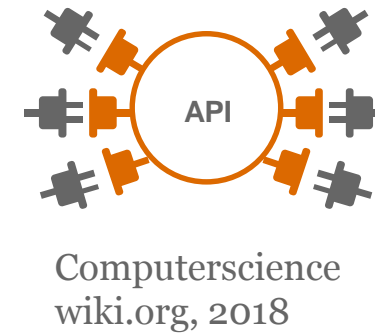


Oracle DBA's Guide, 2018

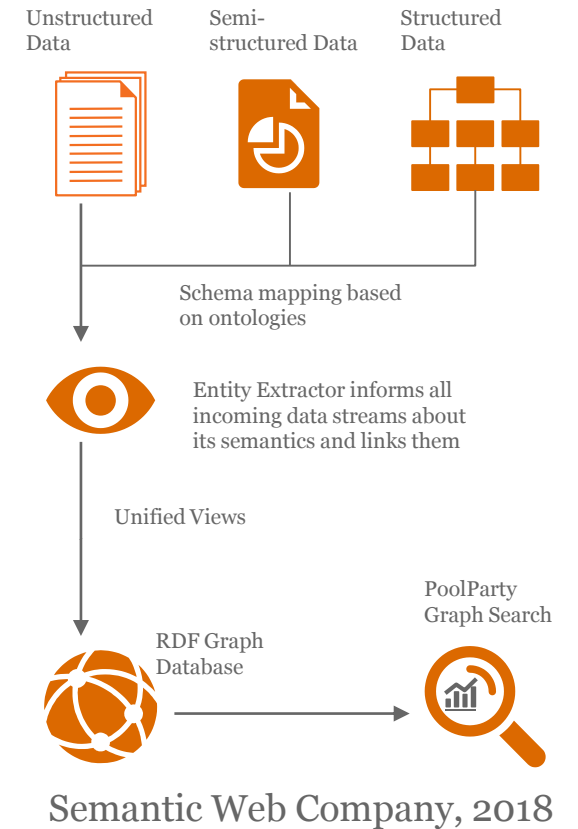
Application-centric ESB advocates



Application-centric RESTful developers



Data-centric knowledge graphers



Crossing the chasm between the tribes

Reducing the amount of unfamiliarity developers confront--familiar document means to achieve comparable ends to graph:

- **Semantic suites that use the JSON format and familiar hierarchies:** SWC's PoolParty is an example
- **GraphQL:** A popular *document* shape language that talks to APIs using SELECT-like statements and tree shapes; backend-agnostic; just uses a mental model for graph; addresses the API endpoint proliferation problem
- **Accessible web as database methods:** JSON-LD and Schema.org, etc. vocabularies
- **Document “schemas” via data objects:** JavaScript objects to developers = documents to NoSQL DB types; Object Data Modeling instead of database semantics
- **Mongoose or MongoDB JSON schema features + GraphQL:** MongoDB object modeling and querying that can be used for subdocument filtering within a GraphQL context
- **HyperGraphQL:** A GraphQL UI for Linked Data, restricted to certain tree-shaped queries
- **Universal Schema Language:** Mike Bowers' document/graph query and modeling language still in development
- **COMN:** Ted Hills' well-defined NoSQL + SQL data modeling notation

Progress on solutions



Types of logic most used in AI-enabled systems

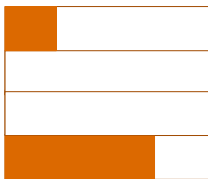
Previously dominant

Rule-based systems (includes KR)

“Handcrafted knowledge” is the term DARPA uses; rule-based programming + procedure replication in process automation, + some knowledge representation (KR)

- Strong on logical reasoning in specific concrete contexts
 - Procedural + declarative programming + set theory, etc.
 - Deterministic
- Can't learn or abstract
- Still exceptionally common and useful

Perceiving
Learning
Abstracting
Reasoning

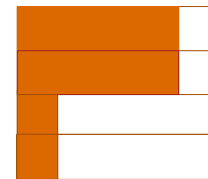


Example: Consumer tax software

On the rise and rapidly improving **Statistical machine learning**

- Probabilistic
- From Bayesian algorithms to neural nets (yes, deep learning also)
- Strong on perceiving and learning (classifying, predicting)
- Weak on abstracting and reasoning
- Quite powerful in the aggregate but individually (instance by instance) unreliable
- Can require lots of data

Perceiving
Learning
Abstracting
Reasoning



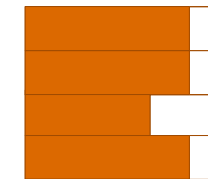
Example: Facial recognition using deep learning/neural nets

Nascent, just beginning

Contextualized, model-driven approach

- Contextualized modeling approach—allows efficiency, precision and certainty
- Combines power of deterministic, probabilistic and description logic
- Allows explanations to be added to decisions
- Accelerates the training process with the help of specific, contextual human input
- Takes less data

Perceiving
Learning
Abstracting
Reasoning



Example: Explains first how handwritten letters are formed so machines can decide based on these individual models—less data needed, more transparency.

John Launchbury of DARPA (<https://www.youtube.com/watch?v=N2L8AqkEDLs>), Estes Park Group and PwC research, 2017

Most automated knowledge graph – Diffbot?

“Diffbot’s crawler regularly refreshes the DKG with new information and its machine learning algorithms are smart enough to pass over sites with histories of producing ‘logically inconsistent’ facts.

“That’s one of the reasons why we fuse information together from different sources,’ Tung said. ‘Our scale is such that there’s minimal potential for errors. We’d bet the business on it.’

“Diffbot launched in 2008 and counts **28 employees** among its core staff of engineers and data scientists.”

--Mike Tung of Diffbot, quoted in *VentureBeat*

Diffbot claims an **automated** knowledge graph of **1 trillion + facts**, designed to grow without humans in the loop.

That compares with **1.6 billion crowdsourced facts** in Google’s knowledge graph, according to *VentureBeat*.

Kyle Wiggers, “Diffbot launches AI-powered knowledge graph of 1 trillion facts about people, places, and things,” *VentureBeat*, 30 August 2018

Versus more explicit, precise, contextualized meaning with a triadic, Peircean knowledge graph and less than **1M concepts**?

“There are many different approaches for distinguishing a logical basis for ontologies, but Peirce basically says to **base everything around 3s**, explains [Mike Bergman of Cognonto]. That is,

- 1. the object itself;**
- 2. what a particular agent perceives about the object;**
- 3. and the way that agent needs to try to communicate what that is.**

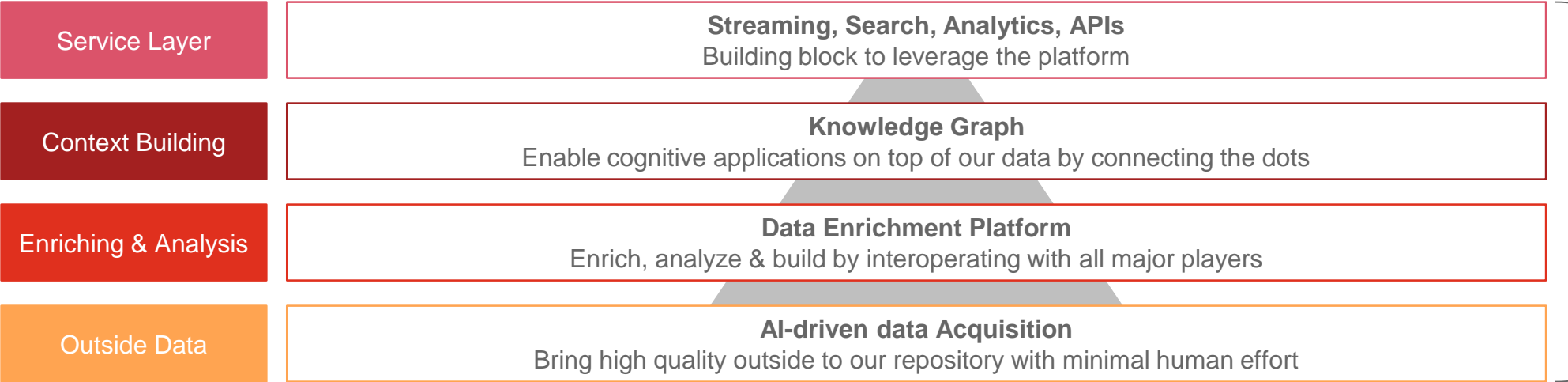
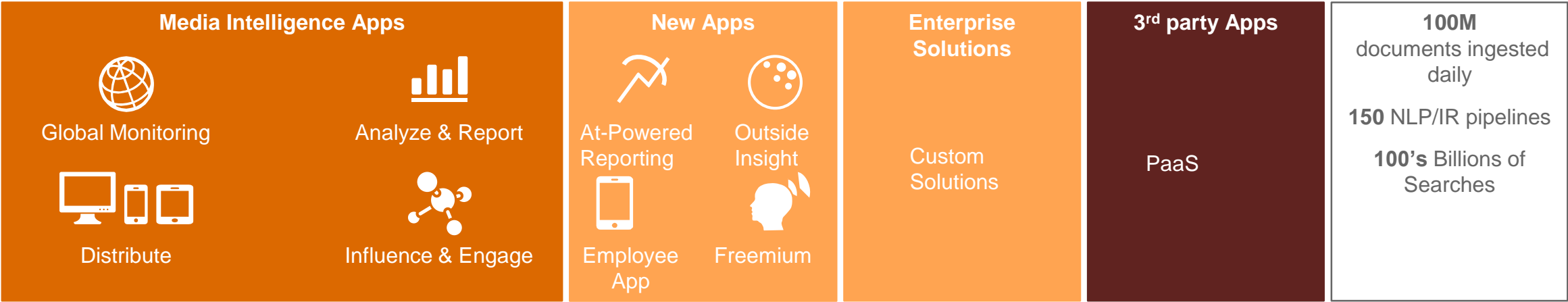
‘Without that triad it’s hard to ever get at differences of interpretation, context or meaning,’ he says, whether that be between something like events and activities or individuals and classes.

Once you adopt that mindset, a lot of things that seemingly were irreconcilable differences begin to fall away, and **the categorization of information becomes really very easy** and smooth....”

--Mike Bergman of Cognonto, quoted in *Dataversity*

Jennifer Zaino, “Cognonto Takes On Knowledge-Based Artificial Intelligence,” *Dataversity*, 23 November 2016

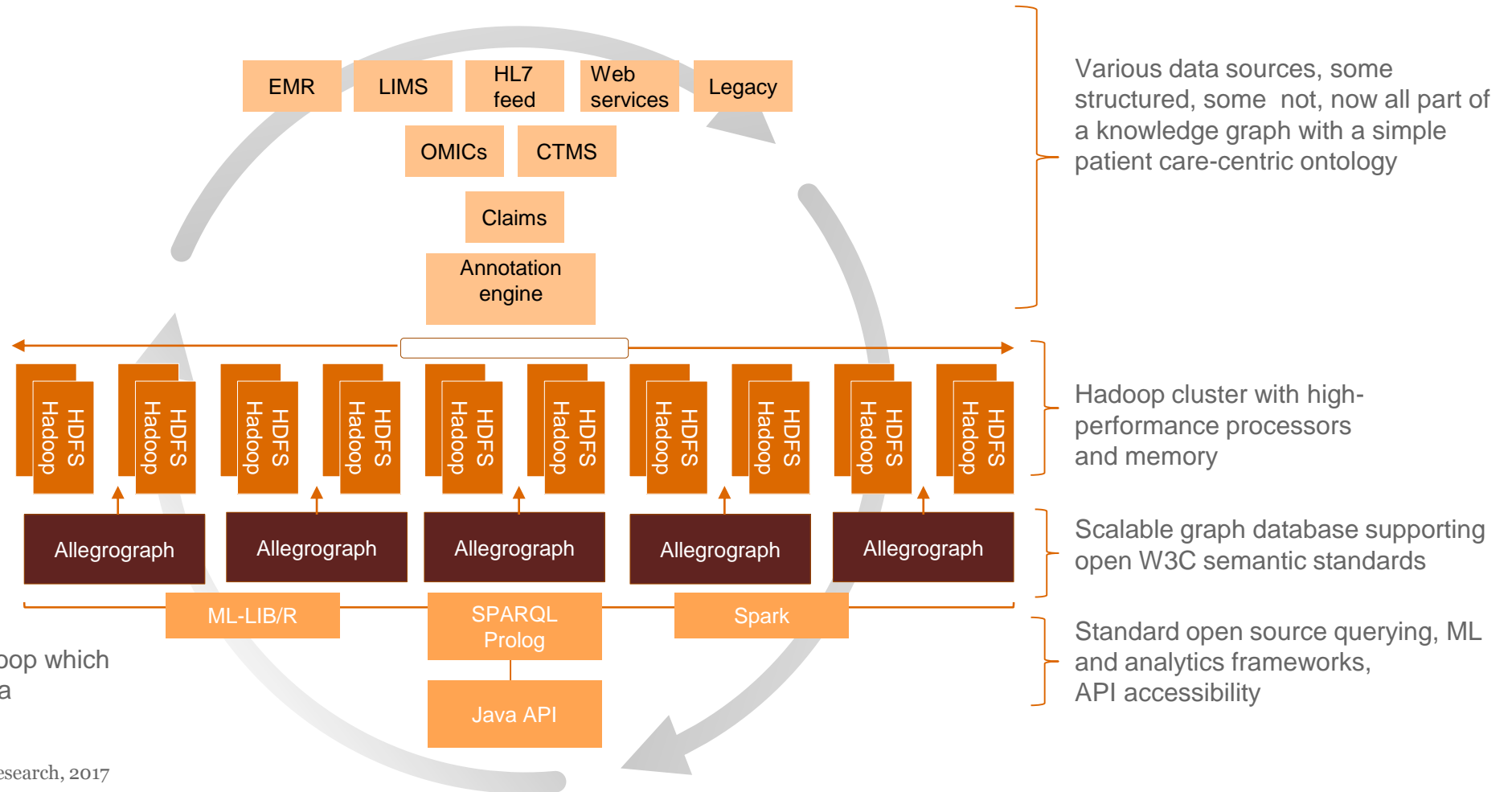
Contextual AI via a large knowledge graph at Fairhair.ai



Meltwater, 2018

Montefiore's semantic data lake

Doctors can query the graph or harness ML + analytics and receive answers from the system at the point of care via their handhelds.



The system also acts as a giant feedback-response or learning loop which learns from the data collected via user/system interactions.

Montefiore Health, Franz, Intel and PwC research, 2017

Siemens' industrial knowledge graph

Industrial Knowledge Graph

“Deep learning fails when it comes to context. Knowledge graphs can handle context and enable us to address things that deep learning cannot address on its own.”

--Michael May, Head of Company Core Technology, Data Analysis and AI, Siemens



1	09:00 – Analyze Turbine data hub
2	11:00 – Configure Configure turbine
3	12:00 – Maintain Master data Mgmt.
4	13:00 – Mitigate Financial Risk Analysis
5	15:00 – Contact Expert & Communities
6	18:00 – Guide Rules & Regulations

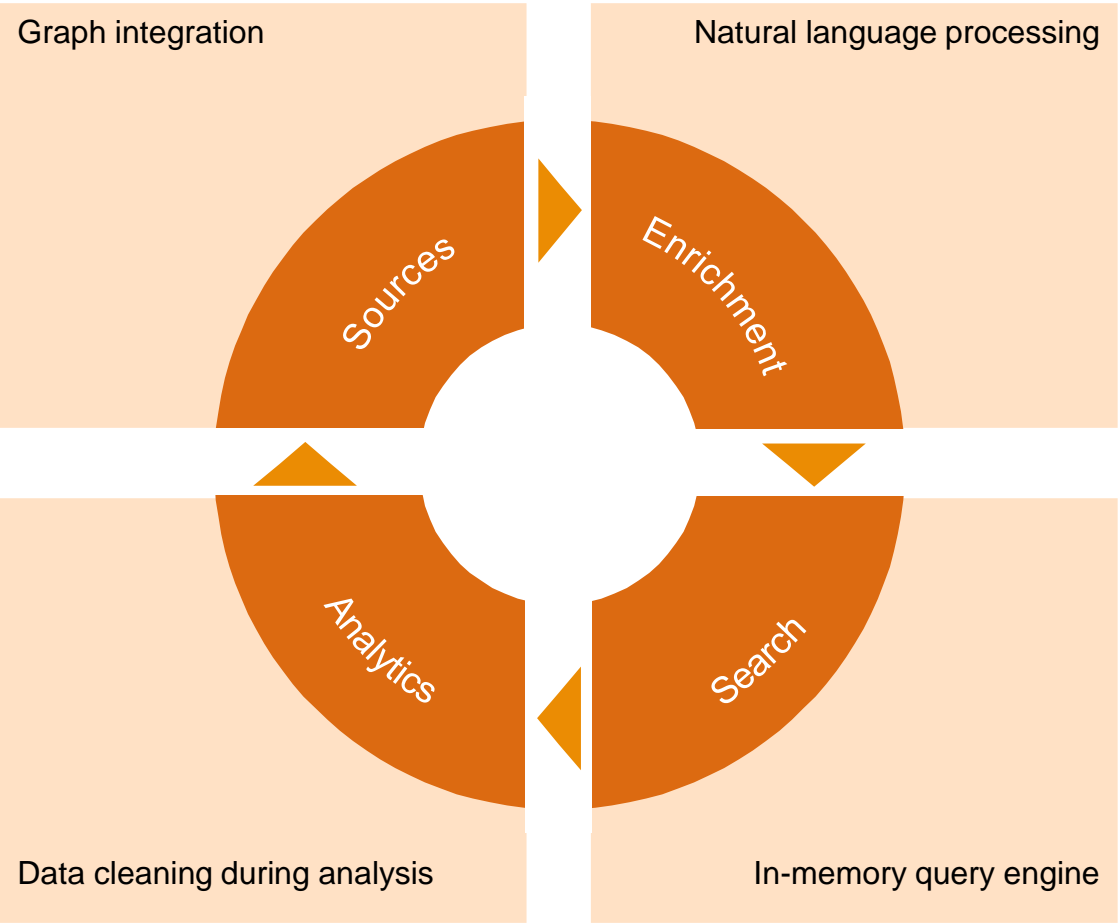


Pharma knowledge graphs for patient safety

Challenges



Solutions



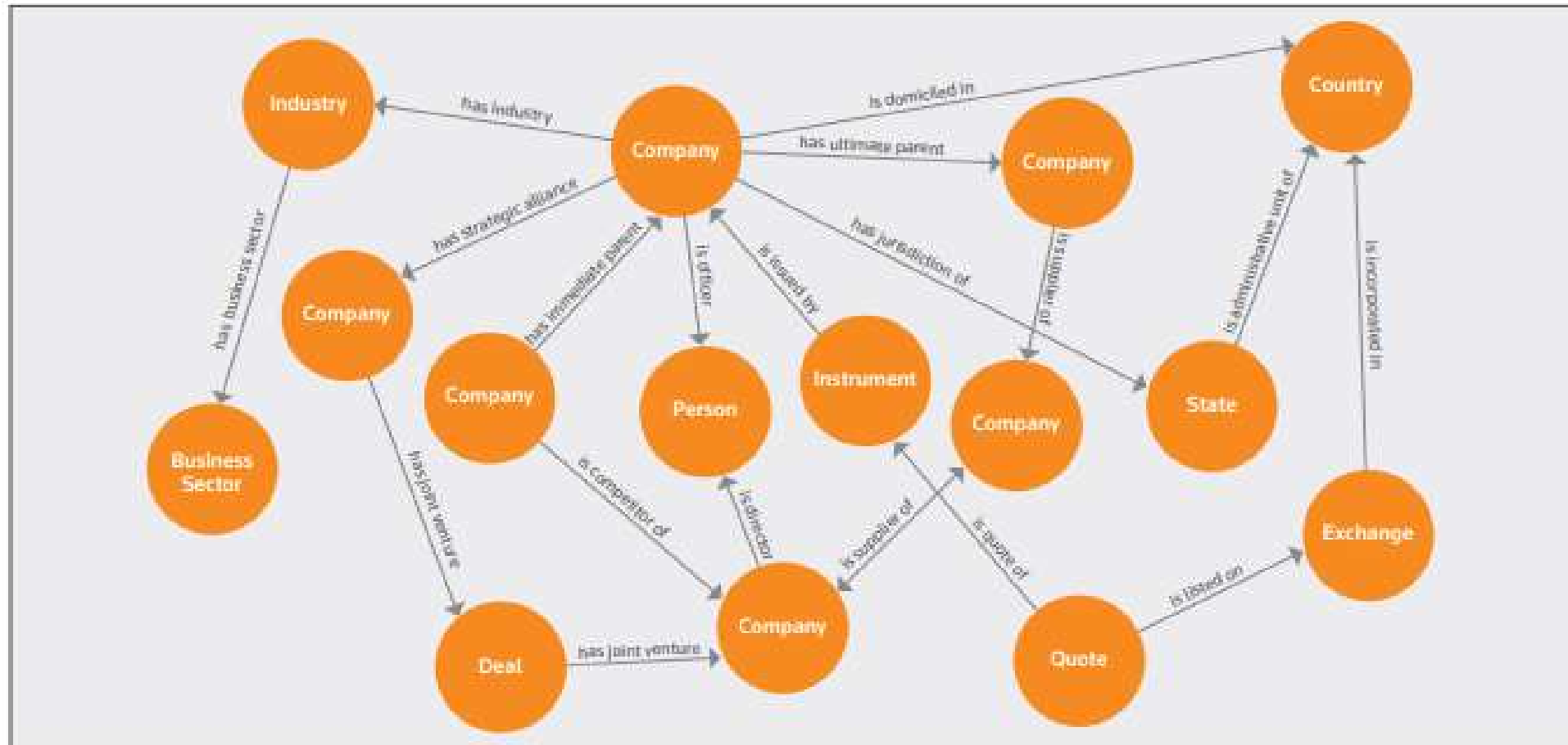
PwC and Cambridge Semantics, 2018

NuMedii's precision therapeutics knowledge graph



Ontotext and NuMedii, 2018

Thomson Reuters' financial knowledge graph as a service

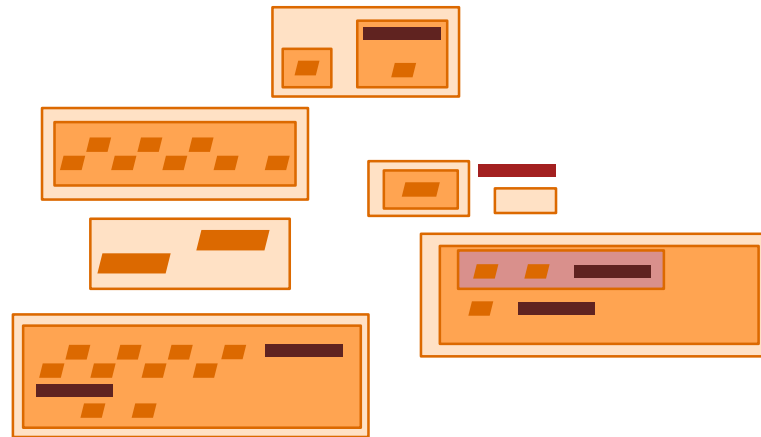


Thomson Reuters, 2018

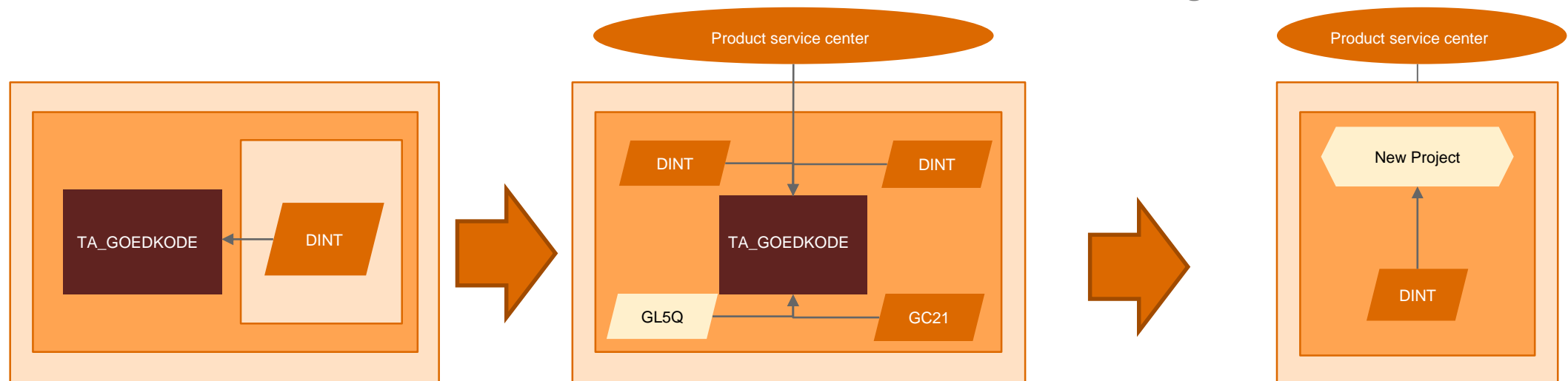
PwC | Collapsing the IT stack

Colryut Group's graph master data federation (Type II transformation)

In order to minimize dependencies between transformation projects, Belgian supermarket chain Colryut Group used a master data structuring, editing and visualization environment created by Tom Sawyer Software.



This graph visualization + data editing/filtering environment allows scalable and articulated governance at the data layer, as well as communications between groups in different parts of the organization, including IT and executive management.



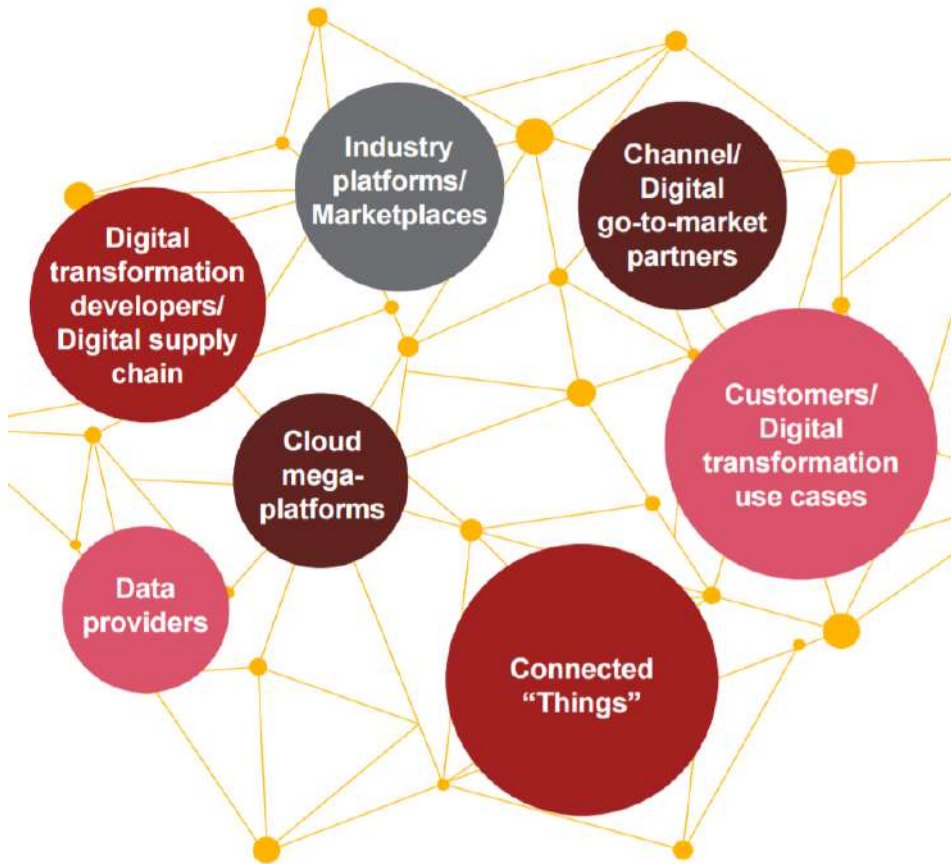
Conclusion and some suggestions



Tell your C-suite: The Third Wave of AI is missing half the data it needs

- **Relationship data** has long been overlooked, but *specifying relationships* is how you *build* context
- **Connected, relationship-rich data** will be seen as **the most important asset for companies**
- Can't have **governance** without **connected data**
- Can't have **connected, meaningful data** without a **semantic model**
- Can't **compete** in the **digital ecosystem** and cross boundaries without **meaningful data connectivity**
- When it comes to enabling the AI your company needs, **think semantic graph:**

The innovation graph must be semantic to scale

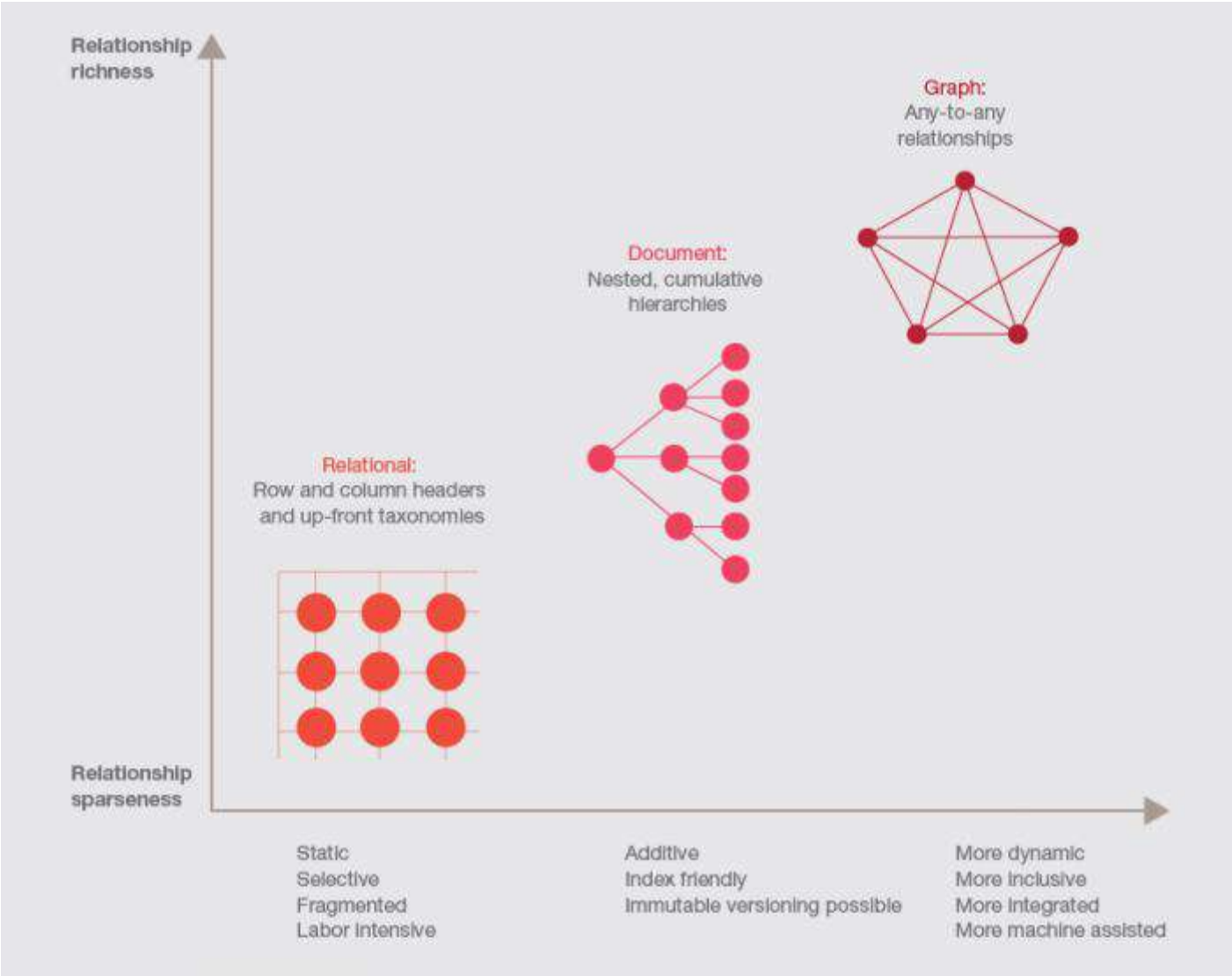


Source: Frank Gens, IDC Directions, 2017

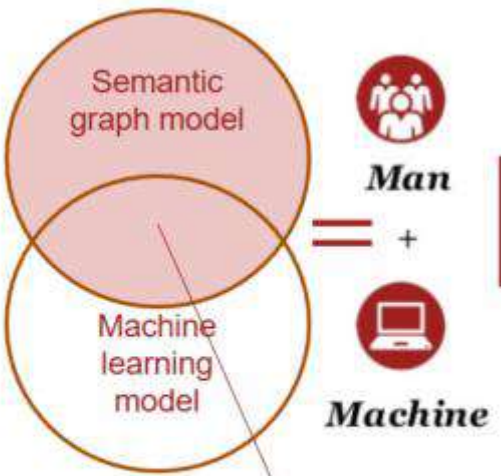
Once digitized (with the help of AI + blockchain, etc.), organizations play different roles than they've been accustomed to in the business ecosystem. Some because of their data collection heritage can become data providers.

Others take up roles in the data supply chain, or position themselves as industry platforms or marketplaces.

Document models can be a stepping stone to graphs



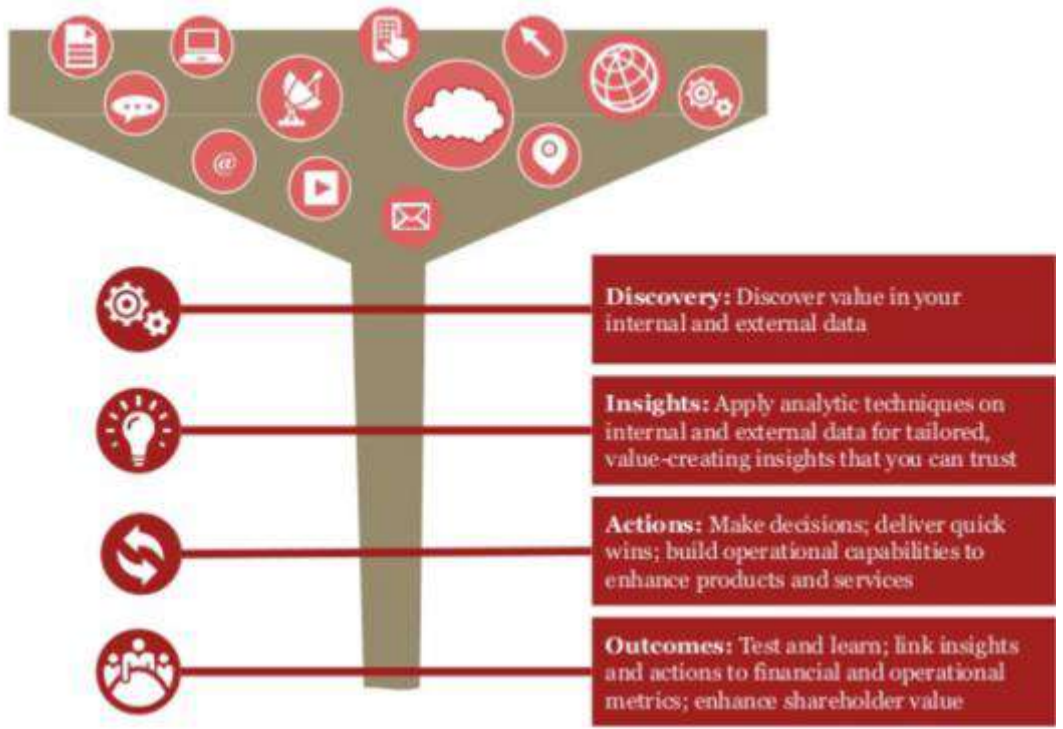
Contextual graphs + statistics methods = innovation at scale



Fully disambiguated combined model that can create actionable insights at scale



New Value



Questions or comments?

pwc.com

Alan Morrison
Sr. Research Fellow
PwC | Integrated Content | Emerging Tech
Mobile: +1(408) 205 5109
Email: alan.s.morrison@pwc.com
PricewaterhouseCoopers LLP
488 Almaden Blvd., Suite 1800
San Jose, CA 95110 USA
<http://usblogs.pwc.com/emerging-technology/author/amorrison009/>
<https://www.linkedin.com/in/alanmorrison>
[@AlanMorrison](#) [quora.com/profile/Alan-Morrison](https://www.quora.com/profile/Alan-Morrison)