"Datafication" — How Data and AI Are Changing Products and Organizations and Our Relationships With Them

Hugh Dubberly
Dubberly Design Office

New technology, its applications, and their effects may suggest new ideas and thus new language.

Historical examples

steam engine*

iron horse

horseless carriage

personal computer

computer science

smart phone

search engine

New examples

digital transformation
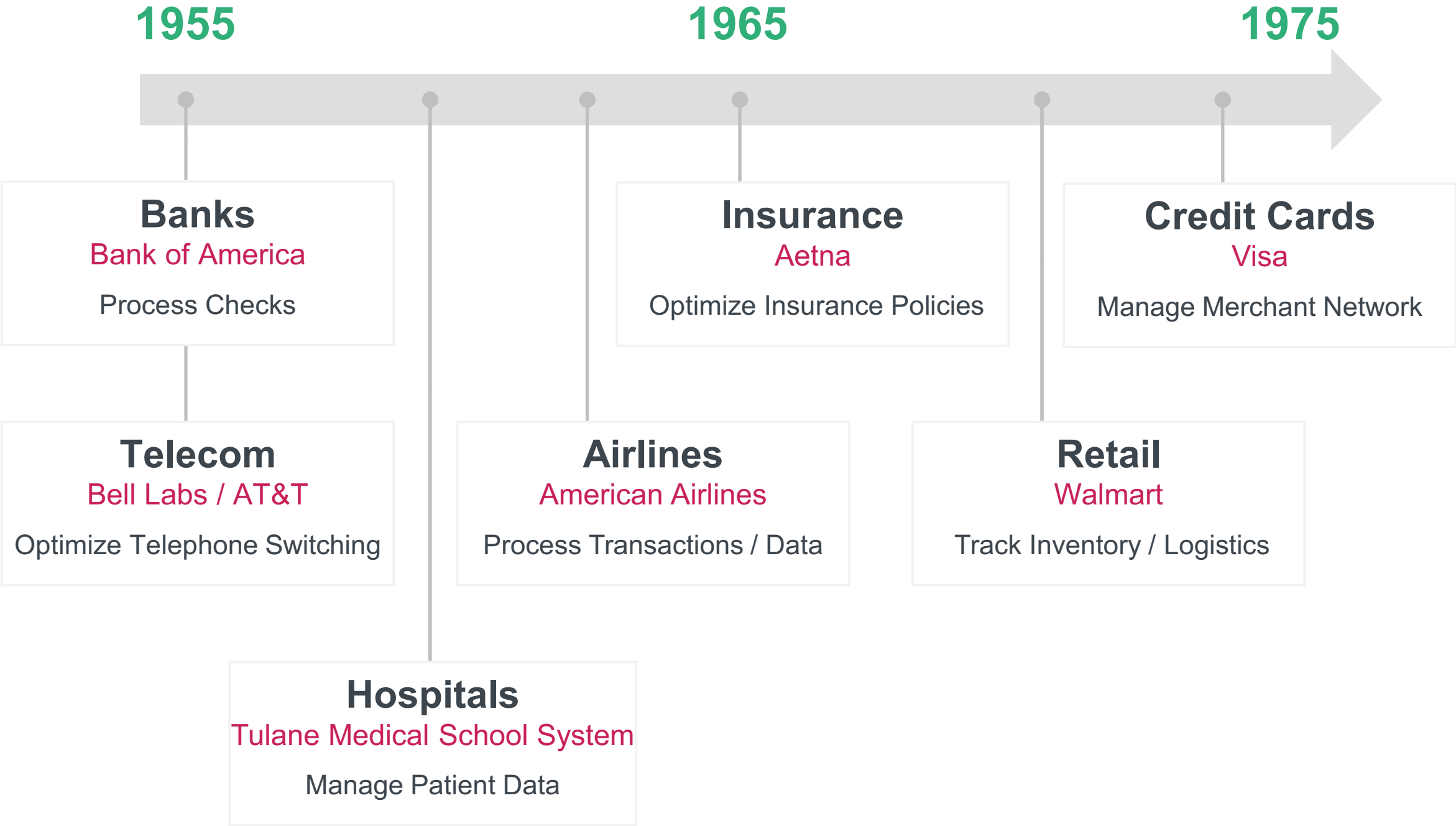
datafication

digital twin

data refinery

data science

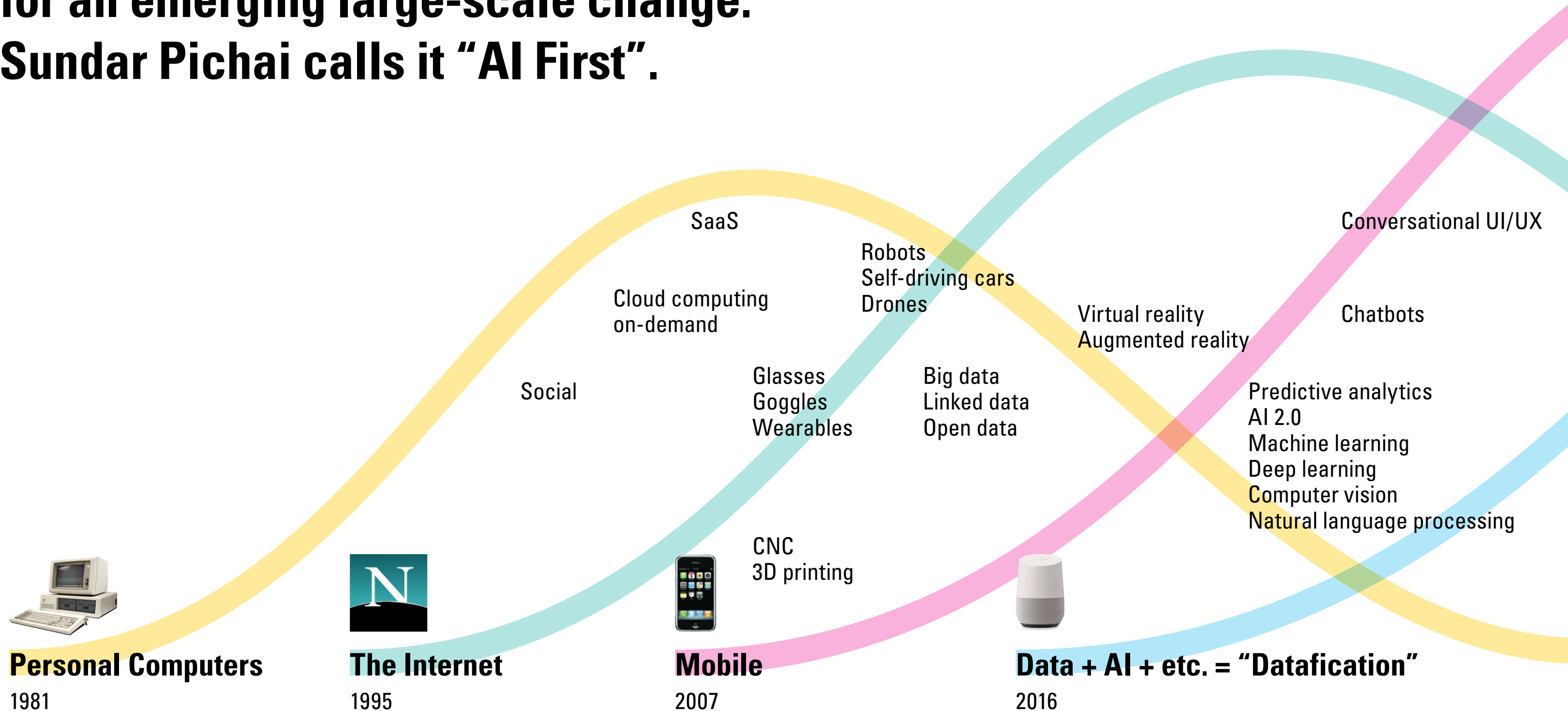data-animated organization

self-driving organization

*See ingenuity, i.e, in-born, thus engineer

Consultants use "digital transformation" to describe the process by which organizations adopt information technology.

1955   **1965**   **1975**

Banks
Bank of America
Process Checks

Insurance
Aetna
Optimize Insurance Policies

Credit Cards
Visa
Manage Merchant Network

Telecom
Bell Labs / AT&T
Optimize Telephone Switching

Airlines
American Airlines
Process Transactions / Data

Retail
Walmart
Track Inventory / Logistics

Hospitals
Tulane Medical School System
Manage Patient Data

Source: https://www.recode.net/2018/5/30/17385116/mary-meeker-slides-internet-trends-code-conference-2018

"Datafication" is a clumsy new word, for an emerging large-scale change. Sundar Pichai calls it "AI First".

SaaS

Conversational UI/UX

Robots
Self-driving cars
Drones

Cloud computing
on-demand

Virtual reality
Augmented reality

Chatbots

Social

Glasses
Goggles
Wearables

Big data
Linked data
Open data

Predictive analytics
AI 2.0
Machine learning
Deep learning
Computer vision
Natural language processing

CNC
3D printing

Personal Computers

The Internet

Mobile

Data + AI + etc. = "Datafication"

1981

1995

2007

2016

What it means:

"Datafication" refers to a new type of organizational change, an extension of "digital transformation",
providing
greater customer engagement
through improved personalization
and
increased foresight about operations + environment
through the development of "digital twins".

E.g., Netflix tracks what you watch; finds patterns; recommends similar shows. They collect data to build models to make predictions to improve personalization.

historical data → algorithms → patterns (models) → predictions → personalization

new data → more algorithms
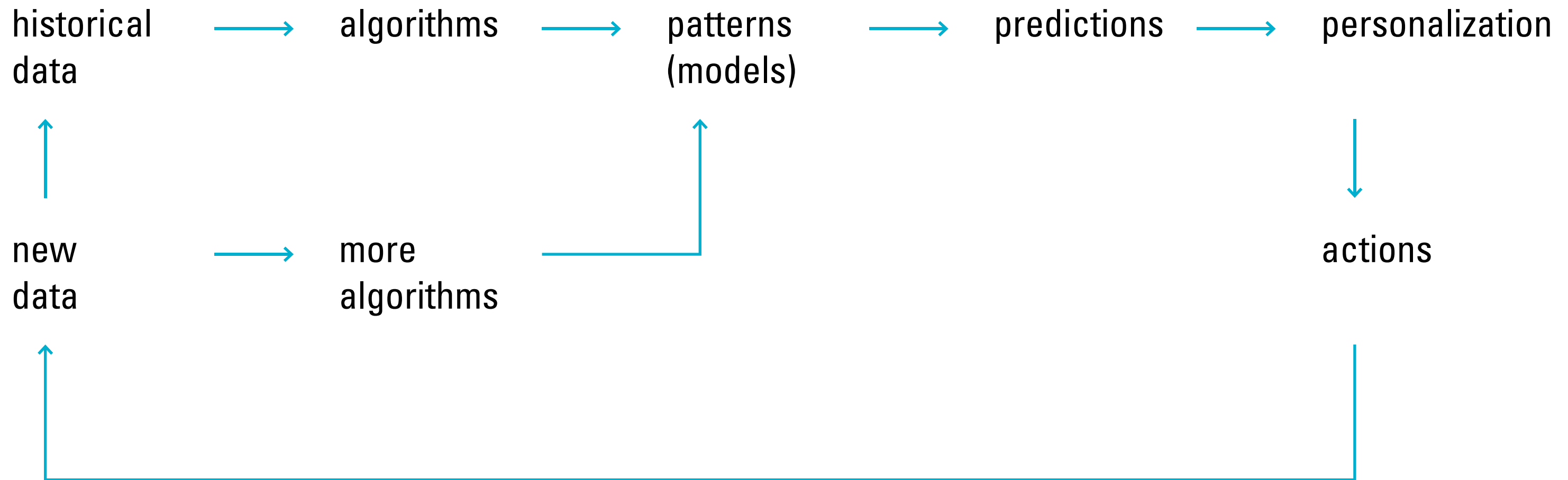
actions

This pattern (data-models-predictions) is key to digital native companies; their data and models (huge "digital twins") are key business assets.

– **Google:** Search + AdWords + User Profiles + CRM
Web Graph + Knowledge Graph + Social Graph + Buyer-journey Graph

– **Facebook:** User Profiles + "Friends" + Newsfeed + Ads + CRM
Social Graph + User-journey Graph

– **Amazon:** Product Profiles + User Profiles + CRM
Product Graph ("People Who Bought This Also Bought That")

As "physical-native" companies undergo "digital transformation", "digital twins" of physical plant + environment will be key business assets.
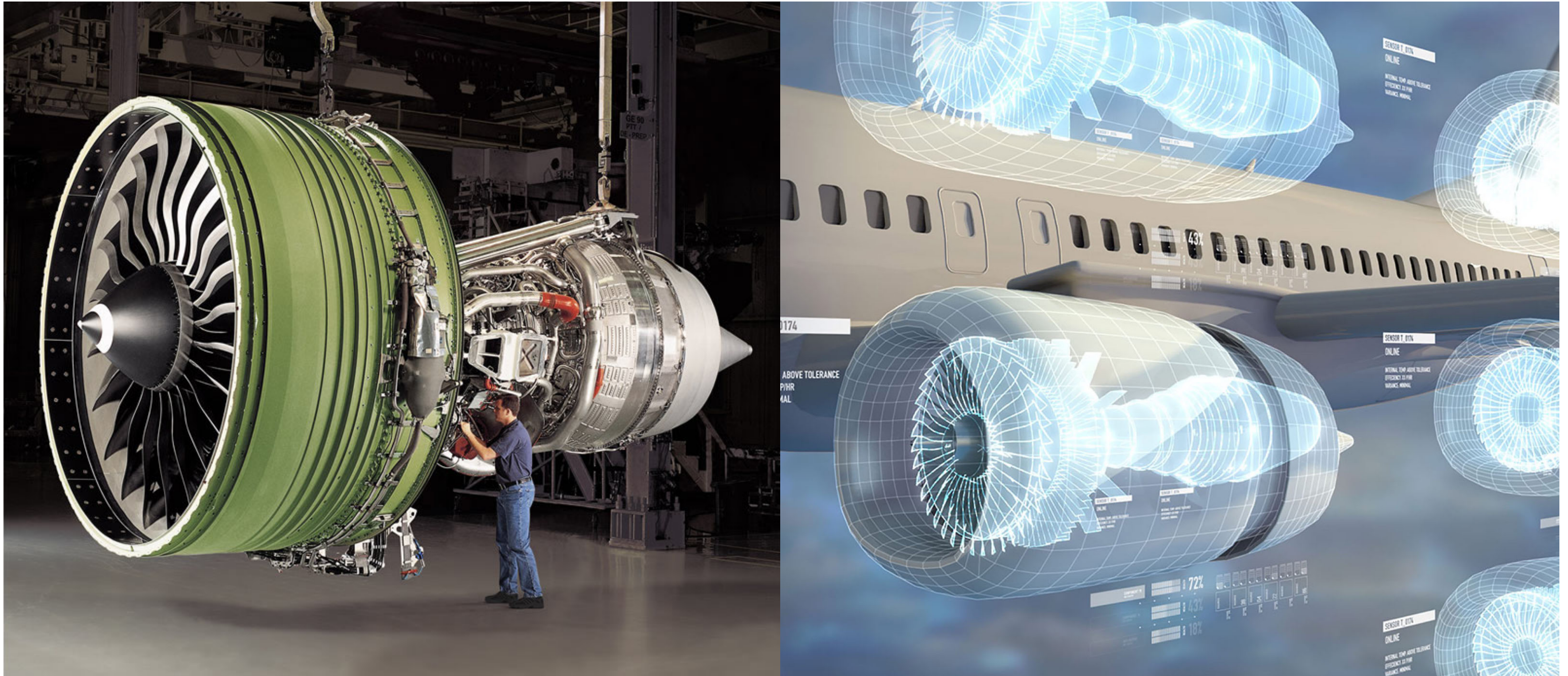
This new "digital transformation" — "datafication" — builds on trends, each affecting the others, leading to "combinatorial innovation".
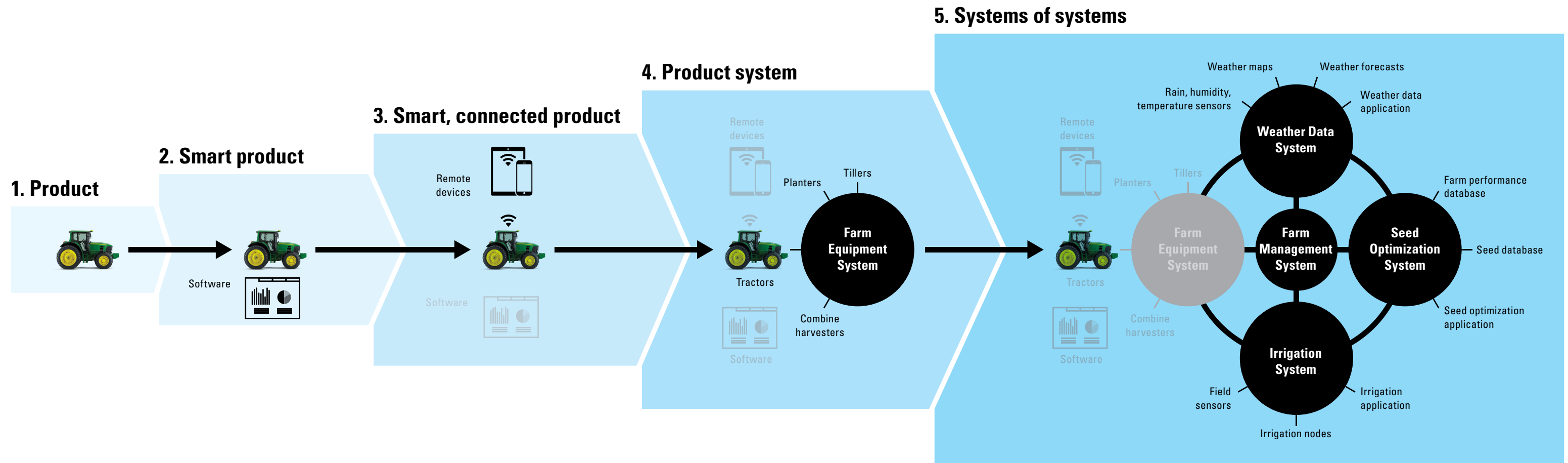
- **Sensor Revolution** — printing sensors on chips; installing measurement capability all around us.

- **Smart Things** — adding "intelligence" to everything, by building in microprocessors.

- **Internet of Things (IoT)** — connecting sensors and smart things to the cloud.

- **Big Data** — recording everything that happens in the physical world and online.

- **Cloud Computing** — putting massive resources online, so that the marginal cost of computation falls to zero.

- **AI (ML, DL, NLP, CV)** — algorithms (often run in the cloud), making sense of the measurements we record.

"Datafication"

Let's look at a few signals of the emerging, large-scale change.

# "...smart, connected products are transforming competition" and "redefining industry boundaries"

1. Product

2. Smart product
Software

3. Smart, connected product
Remote devices
Software

4. Product system
Remote devices
Planters
Tillers
Tractors
Combine harvesters
Software
Farm Equipment System

5. Systems of systems
Remote devices
Planters
Tillers
Tractors
Combine harvesters
Software
Farm Equipment System
Weather maps
Weather forecasts
Rain, humidity, temperature sensors
Weather data application
Weather Data System
Farm performance database
Farm Management System
Seed Optimization System
Seed database
Seed optimization application
Irrigation System
Field sensors
Irrigation application
Irrigation nodes

— **Michael Porter**, HBR, 2014

Farms are becoming automated factories.
Plants are attached to sensors, connected to networks, generating data.
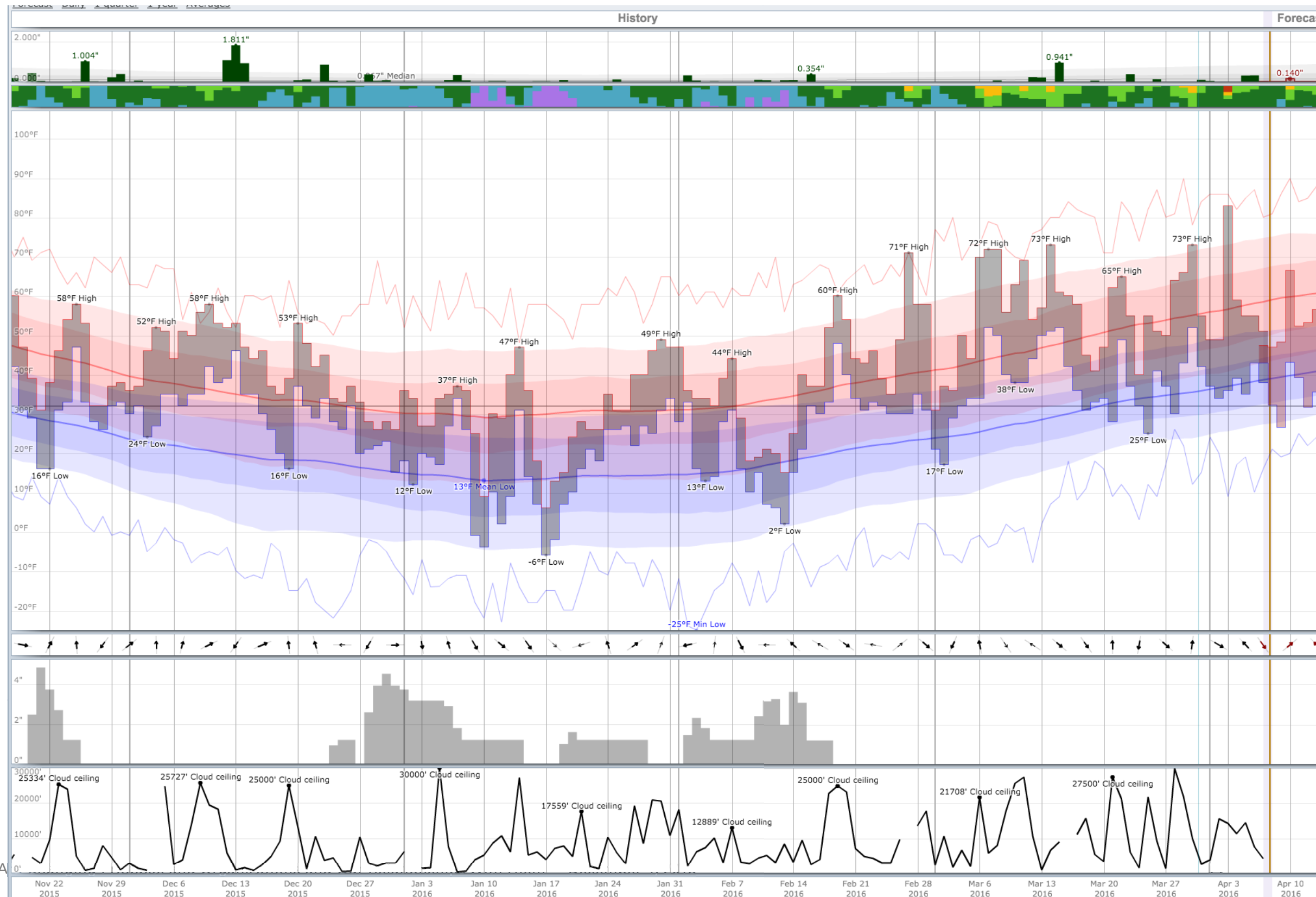
Macro view: processed satellite images of crop growth over time,
e.g., central Iowa, March 29 to October 23, in 8 day increments.

Algorithms automatically align images,
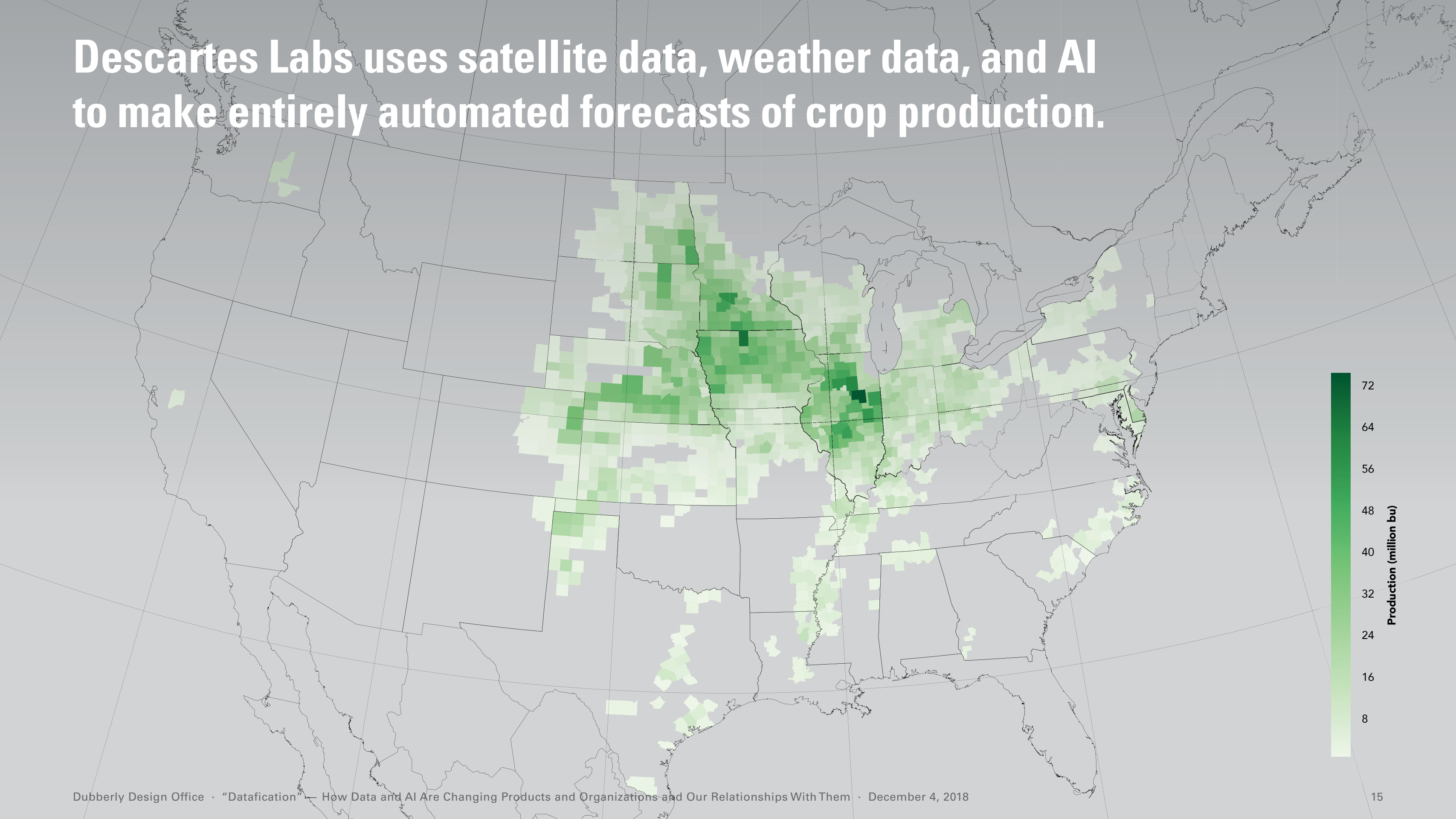remove clouds,
and detect vegetation.
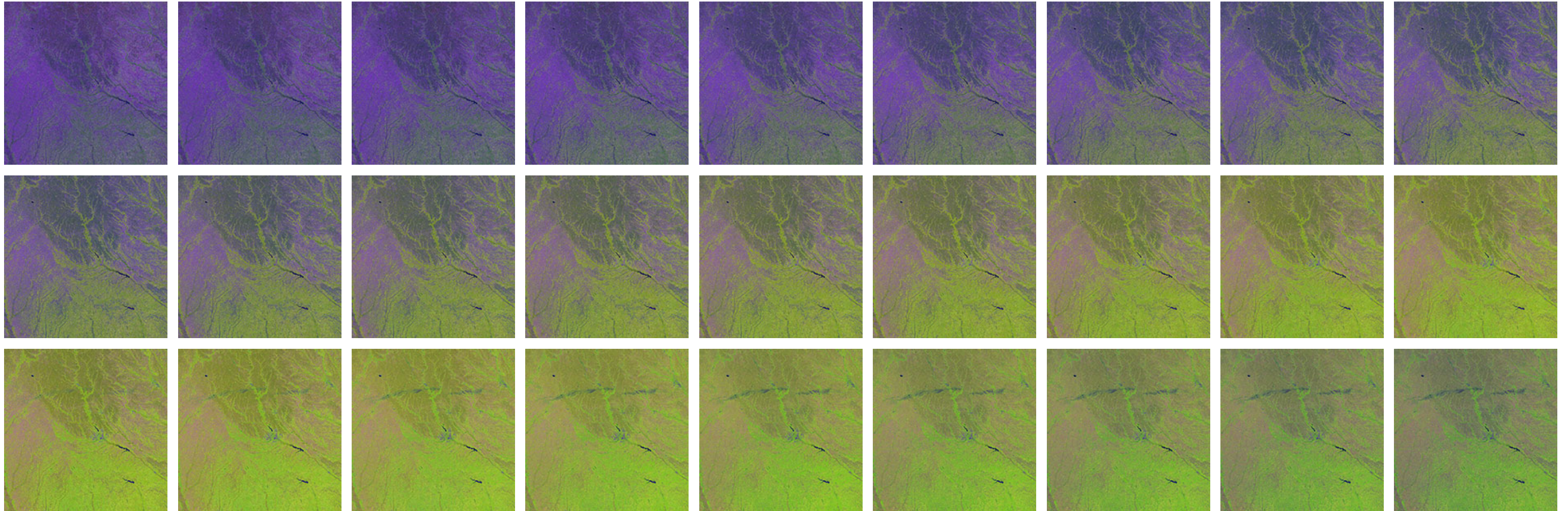
Daily weather data can augment AI.

Precipitation,
temperature,
wind direction and speed,
snow cover,
and cloud cover
can aid forecasting.

Descartes Labs uses satellite data, weather data, and AI to make entirely automated forecasts of crop production.

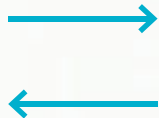Production (million bu)

72

64

56

48

40

32

24

16

8

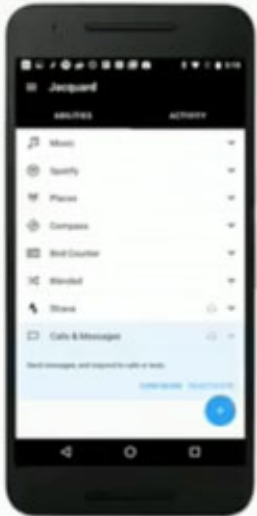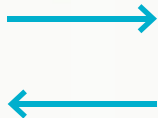# Sensors will be everywhere — all around you, on you, even in you. E.g., Google + Levi's connected denim smart jacket
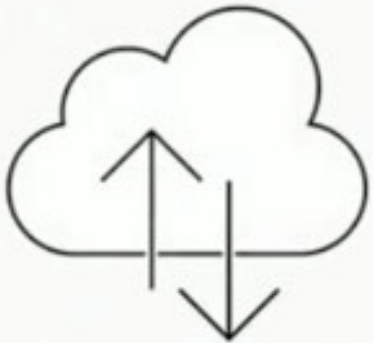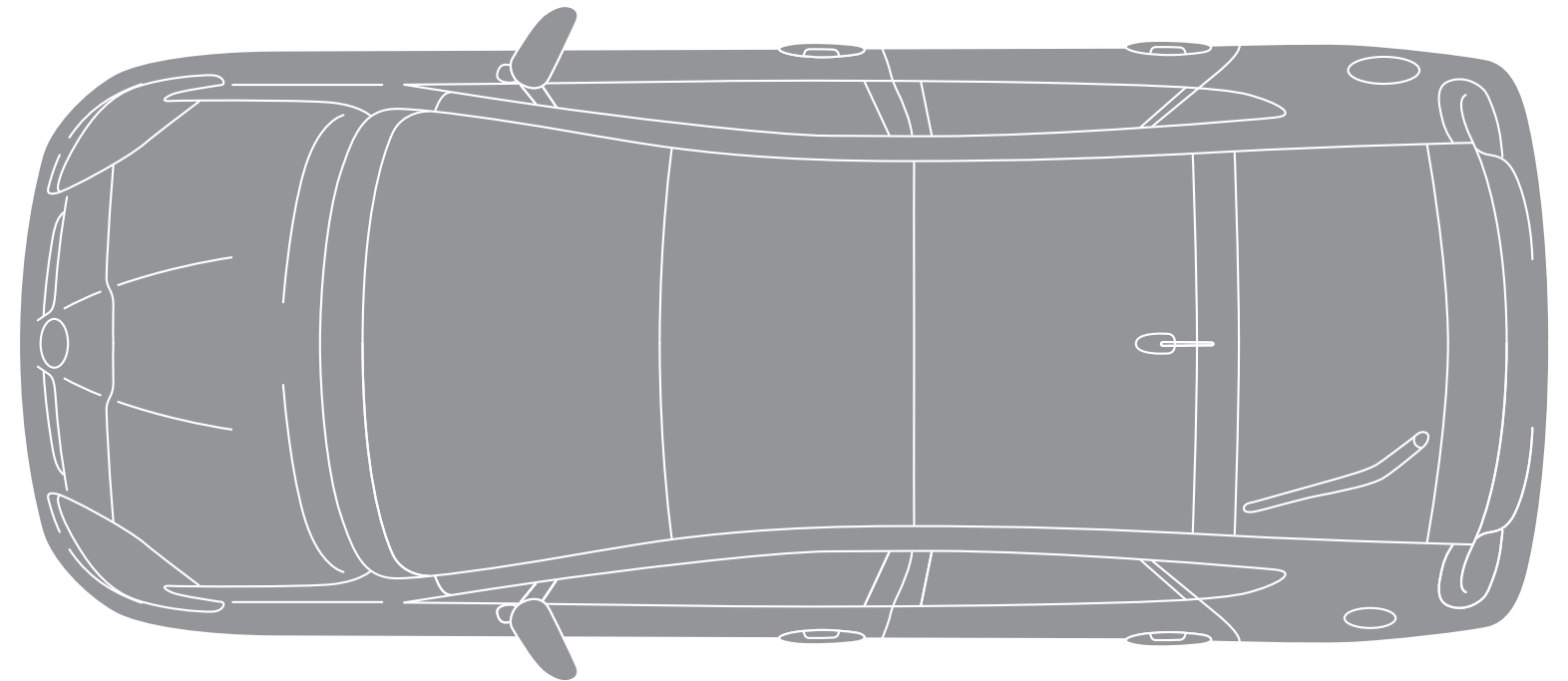
Jacquard Woven
Gesture Sensor

Jacquard Tag

Jacquard App

Jacquard Services

Today's average car has:

-1 engine

~7 small motors
(windows, wipers, fans)

~30 micro processors
(up to 100 for luxury cars) [1]

~60-100 sensors
(growing to 200 by 2020) [2]

~100 million lines of code
(up from 2 million lines in a generation) [3]

And it produces
"terabytes of data per car per day" [4]

Sources:
[1] http://www.nytimes.com/2010/02/05/technology/05electronics.html
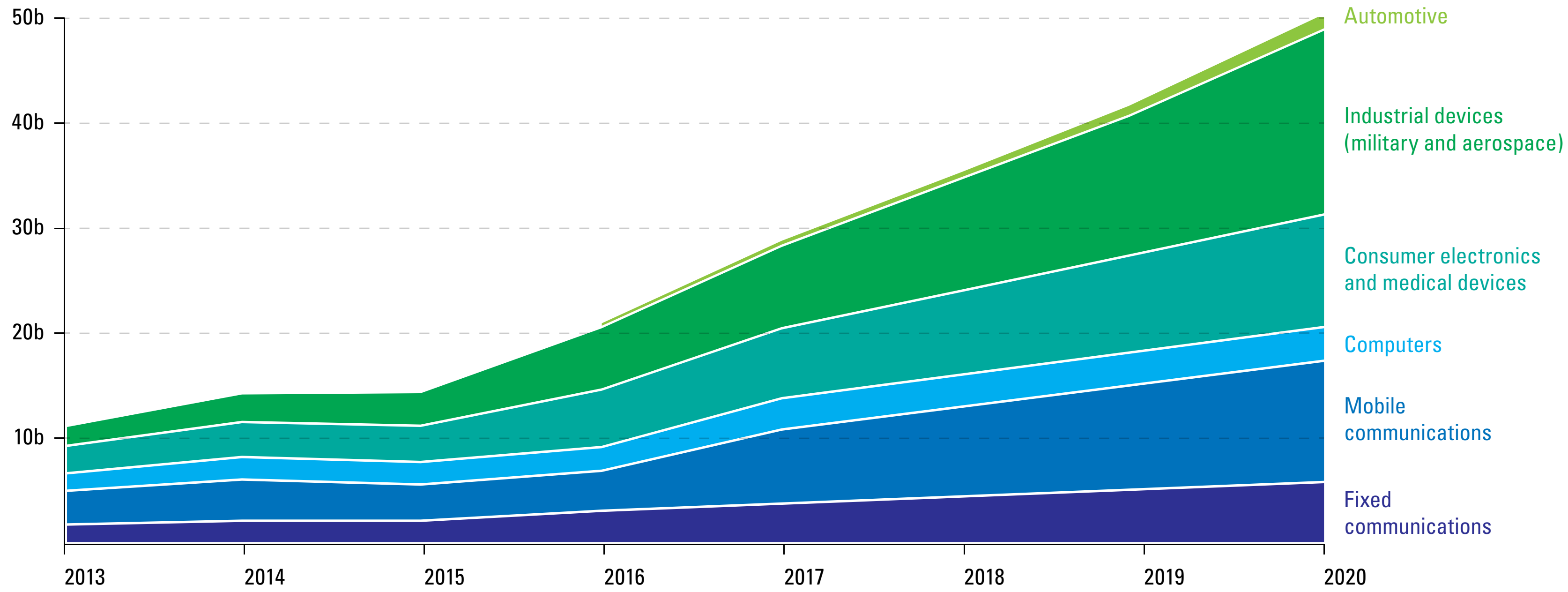[2] http://www.automotivesensors2015.com/
[3] https://leithporsche.com/news/What+Makes+the+2017+Porsche+Panamera+Different3F+Computer+Code/7659/
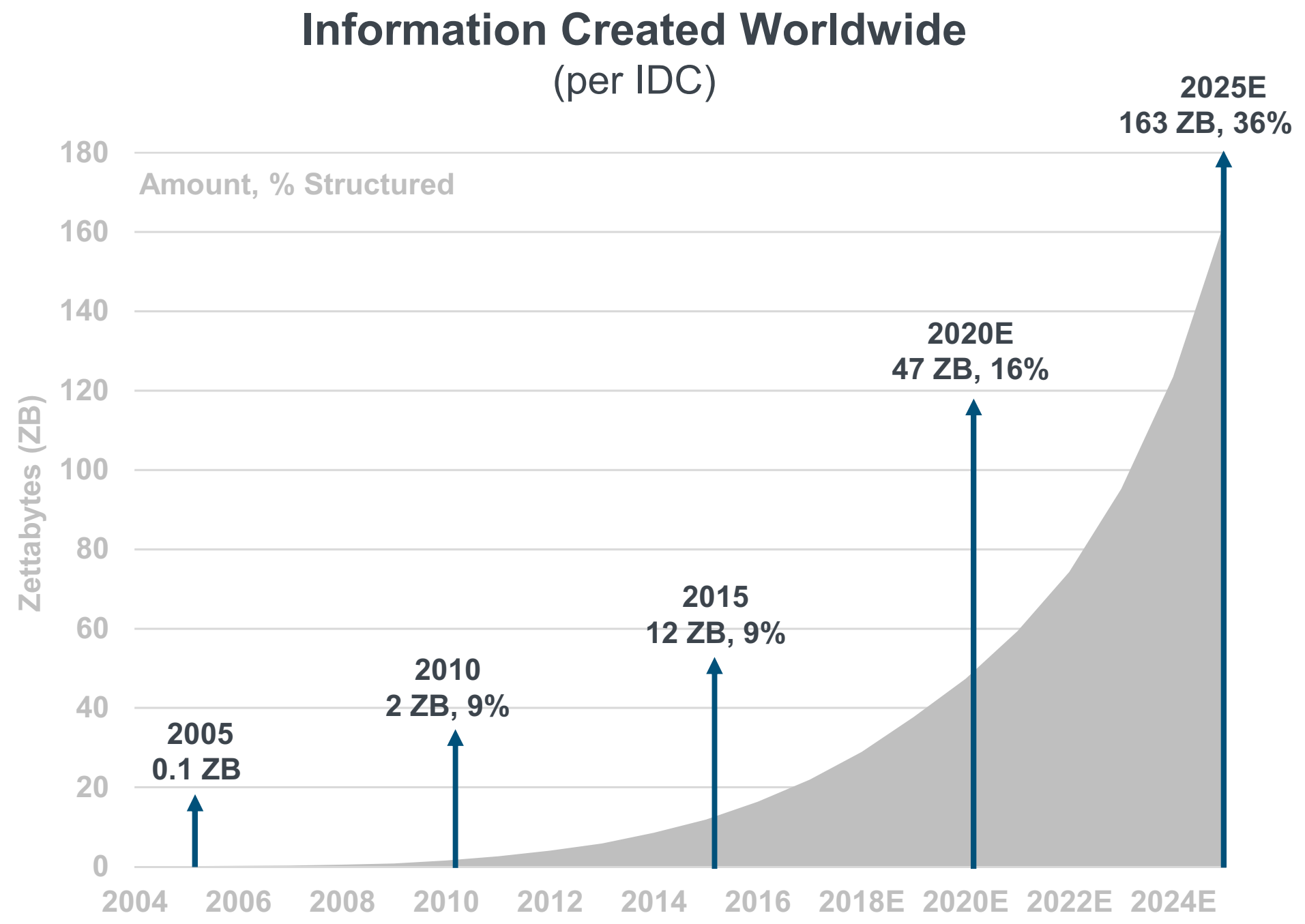[4] Parrish Hanna, Global Director of HMI at Ford (personal communications)

By 2020, ~50 billion devices will be connected to the Internet.

Automotive

Industrial devices
(military and aerospace)

Consumer electronics
and medical devices

Computers

Mobile
communications

Fixed
communications

50b
40b
30b
20b
10b

2013   2014   2015   2016   2017   2018   2019   2020

Sources: The Economist and Cisco

Growth of information created worldwide (per IDC)

Information Created Worldwide
(per IDC)

2025E
163 ZB, 36%

2020E
47 ZB, 16%

2015
12 ZB, 9%

2010
2 ZB, 9%

2005
0.1 ZB

Amount, % Structured

Zettabytes (ZB)

180
160
140
120
100
80
60
40
20
0

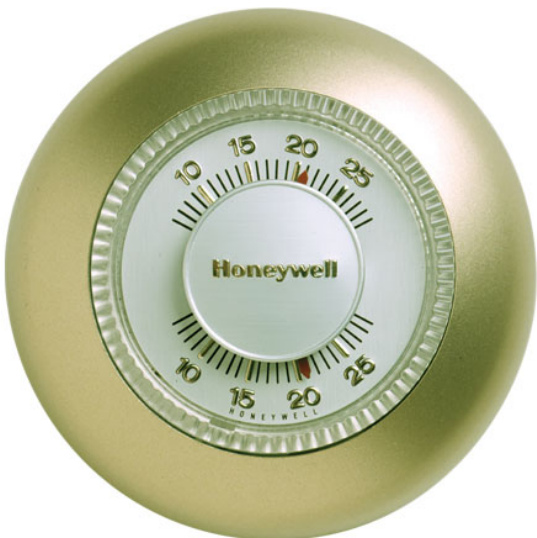2004   2006   2008   2010   2012   2014   2016   2018E   2020E   2022E   2024E

Source: https://www.recode.net/2018/5/30/17385116/mary-meeker-slides-internet-trends-code-conference-2018

"Datafication" is changing the nature of products.

From
**Physical artifacts
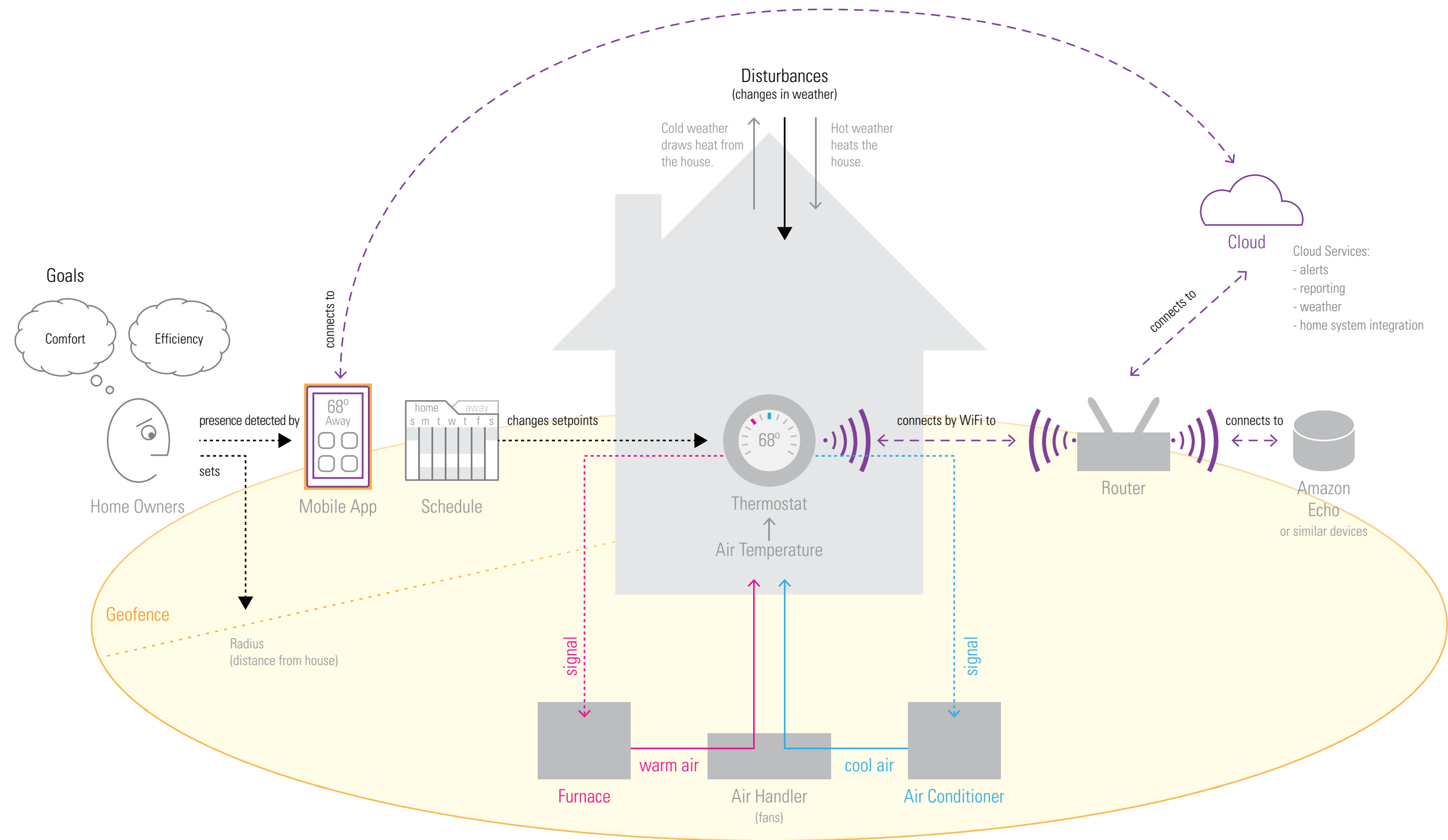— objects**

To
**Adaptive systems
— ecologies**

Product Design   Human Factors   Interaction Design   Service Design
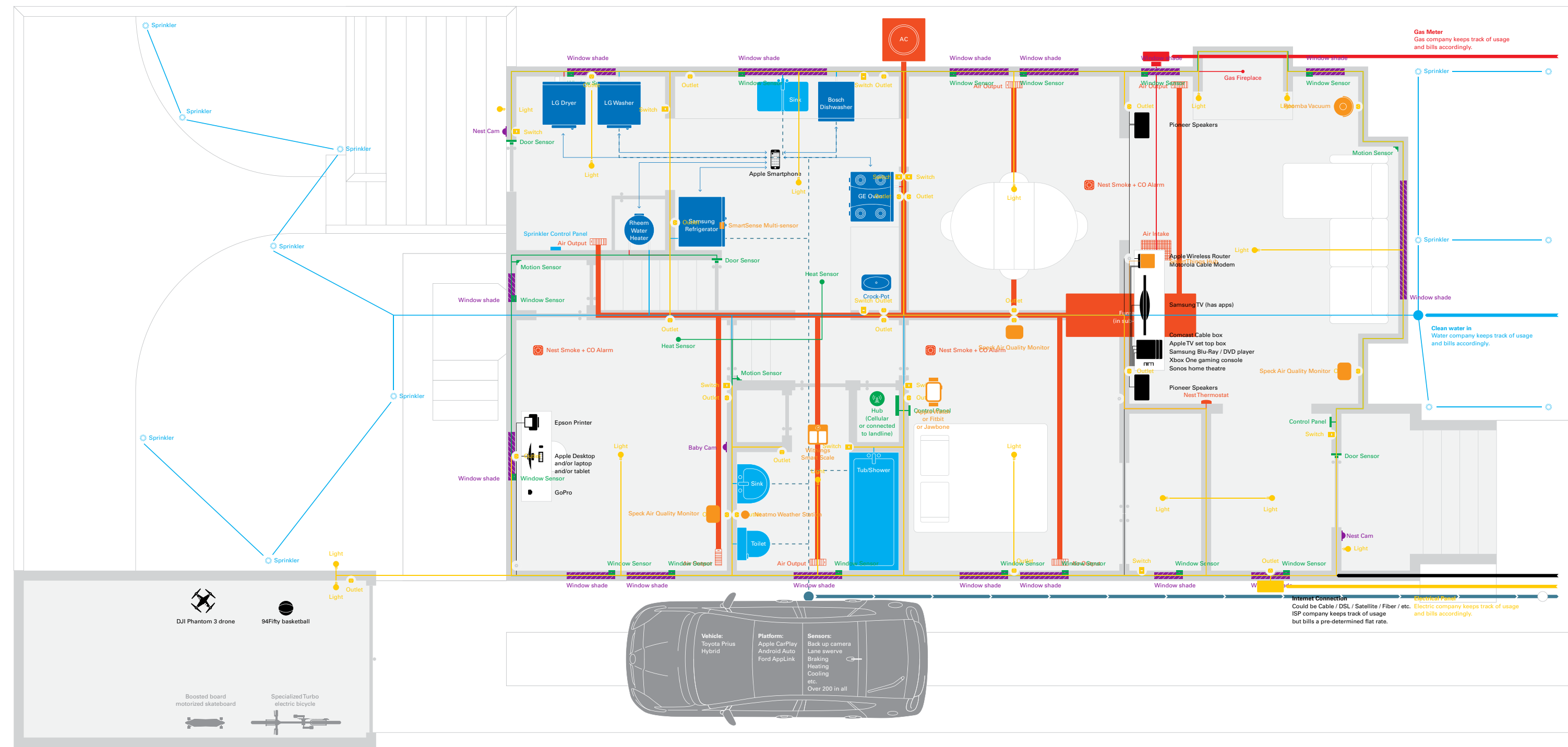
Focus Groups   Usability Studies   Data-driven Design   Model-driven Design

A smart thermostat gathers together a larger network of products, services, people, and their interactions.

Disturbances
(changes in weather)

Cold weather draws heat from the house.

Hot weather heats the house.

Cloud

Cloud Services:
- alerts
- reporting
- weather
- home system integration

Goals

Comfort

Efficiency

connects to

connects to

68°
Away

home   away
s m t w t f s

changes setpoints

68°

connects by WiFi to

connects to

presence detected by

sets

Home Owners

Mobile App

Schedule

Thermostat

Air Temperature

Router

Amazon Echo
or similar devices

Geofence

Radius
(distance from house)

signal

signal

warm air

cool air

Furnace

Air Handler
(fans)

Air Conditioner

The HVAC network is part of an even larger home ecology.

"It seems to me that the notion of machine that was current in the course of the Industrial Revolution — and which we might have inherited — is a notion, essentially, of a machine without goal, it had no goal 'of', it had a goal 'for'.

And this gradually developed into the notion of machines with goals 'of', like thermostats, which I might begin to object to because they might compete with me.

Now we've got the notion of a machine with an underspecified goal, the system that evolves. This is a new notion, nothing like the notion of machines that was current in the Industrial Revolution, absolutely nothing like it. It is, if you like, a much more biological notion, maybe I'm wrong to call such a thing a machine; I gave that label to it because I like to realise things as artifacts, but you might not call the system a machine, you might call it something else."
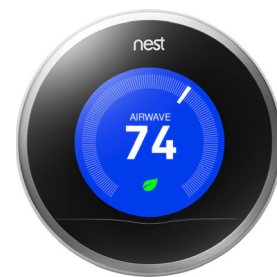
— **Gordon Pask**, 1972 (See Usman Haque, 2007)

We might call it a "product-service ecology".

> "...networks of products, services, technology, people, and collective and collaborative interaction are generating value for the populations they serve."
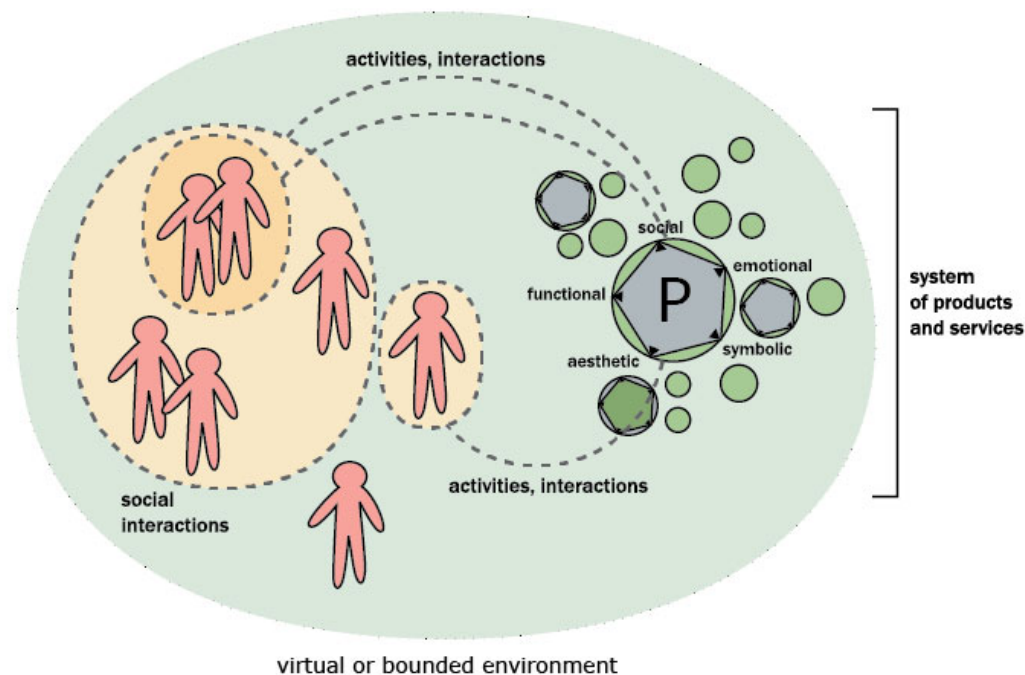
— **Jodi Forlizzi**, HCII, CMU, 2008

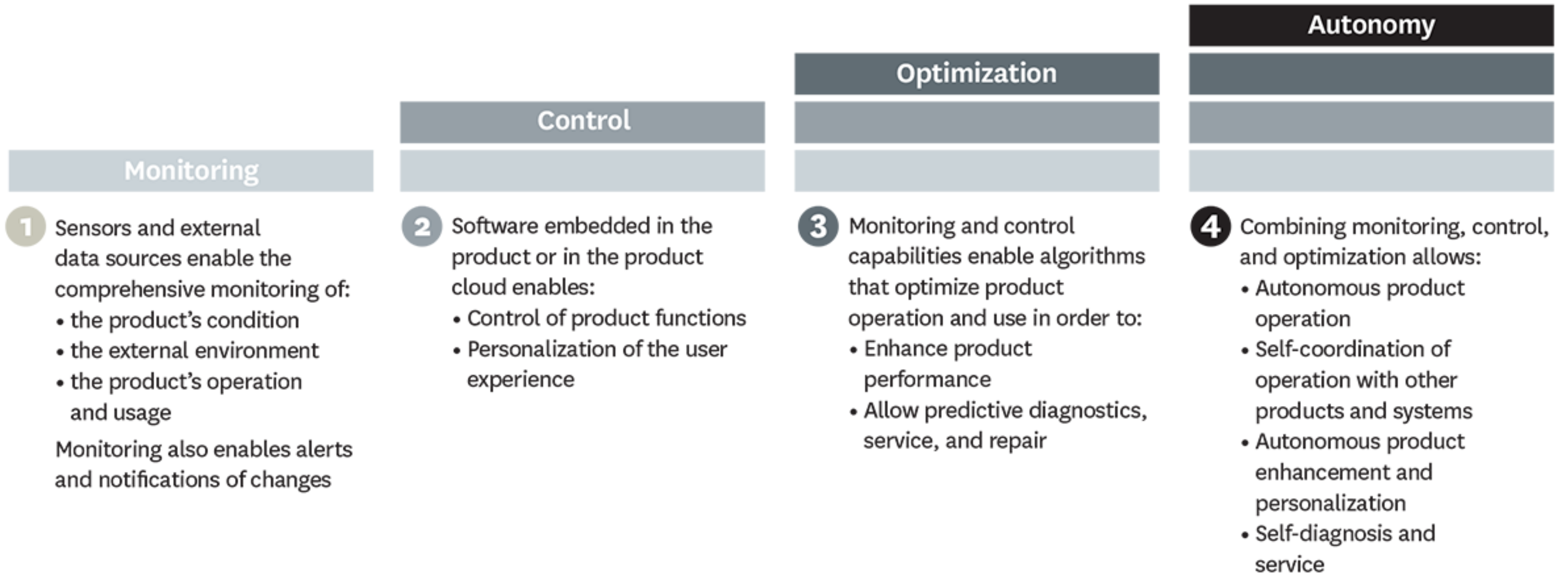# Yet only recently, have we begun to talk about data as a material — as a component of products.
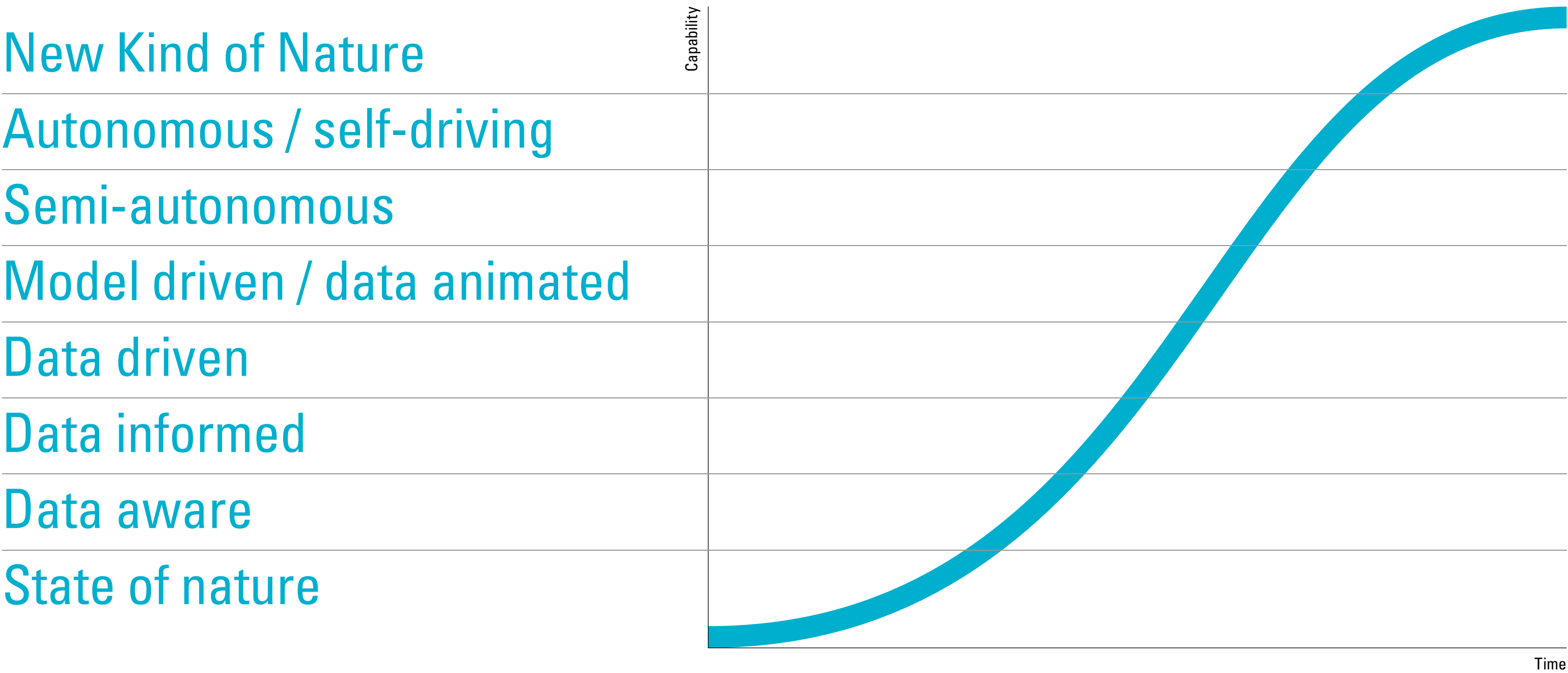
A Model of Data-driven Design

DATA AWARE

DATA INFORMED

DATA DRIVEN
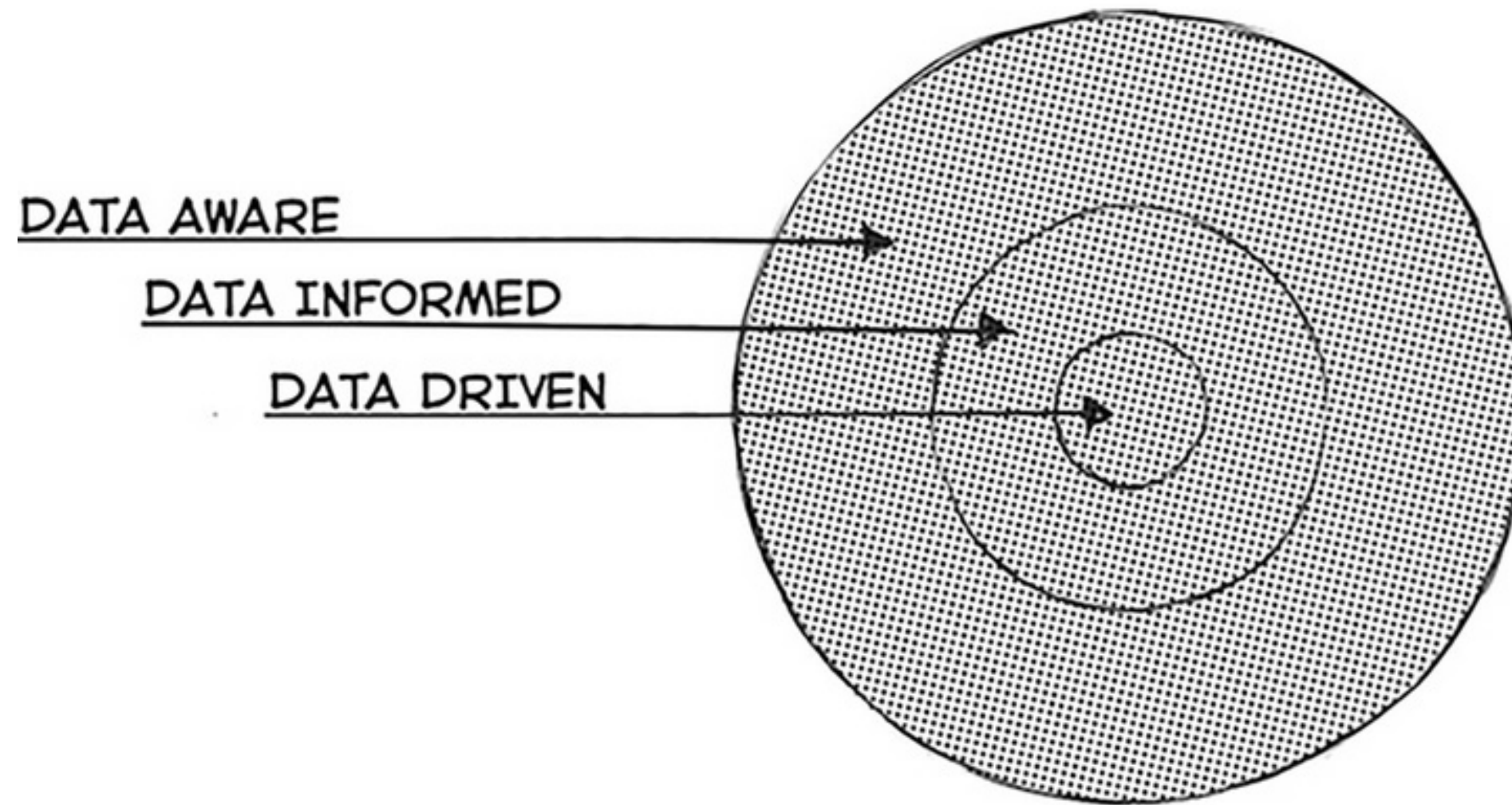
— **Elizabeth Churchill et al.,** *Designing with Data: Improving the User Experience with A/B Testing,* 2017

"Capabilities of Smart, Connected Products"

		Autonomy
	Optimization	
Control		
Monitoring		

1 Sensors and external data sources enable the comprehensive monitoring of:
- the product's condition
- the external environment
- the product's operation and usage

Monitoring also enables alerts and notifications of changes

2 Software embedded in the product or in the product cloud enables:
- Control of product functions
- Personalization of the user experience

3 Monitoring and control capabilities enable algorithms that optimize product operation and use in order to:
- Enhance product performance
- Allow predictive diagnostics, service, and repair

4 Combining monitoring, control, and optimization allows:
- Autonomous product operation
- Self-coordination of operation with other products and systems
- Autonomous product enhancement and personalization
- Self-diagnosis and service

— **Michael Porter**, HBR, 2014

A model of the "datafication" learning curve.

New Kind of Nature

Autonomous / self-driving

Semi-autonomous

Model driven / data animated

Data driven

Data informed

Data aware

State of nature

Capability

Time

Example:
An individual managing diabetes

State of nature

Metabolism happens on its own.
The individual is unaware.
(But the body *is* aware.)

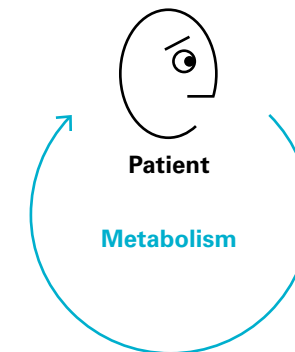New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware

State of nature

Patient

Metabolism

The individual learns diet + exercise affect weight, blood glucose, and general health.

Health

Patient

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
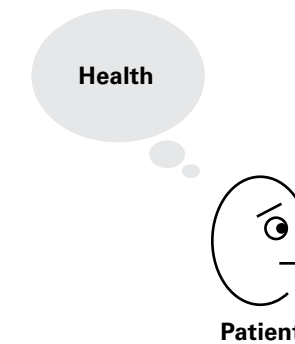Model driven / data animated
Data driven
Data informed
Data aware — reflecting, possible sensing

State of nature

The individual counts carbs and steps; choices start to affect behavior.

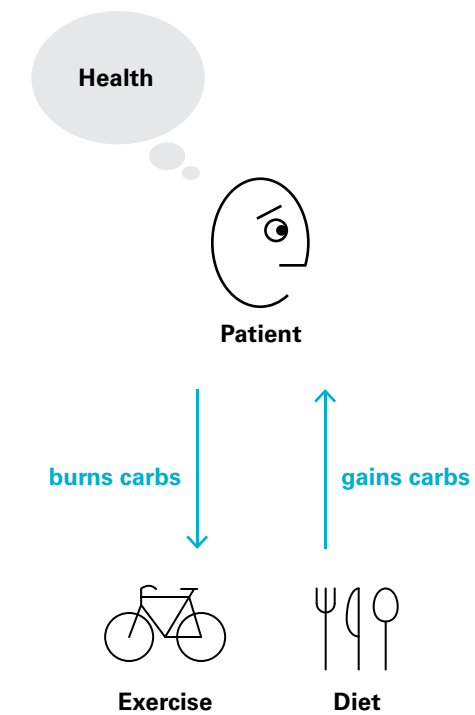New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
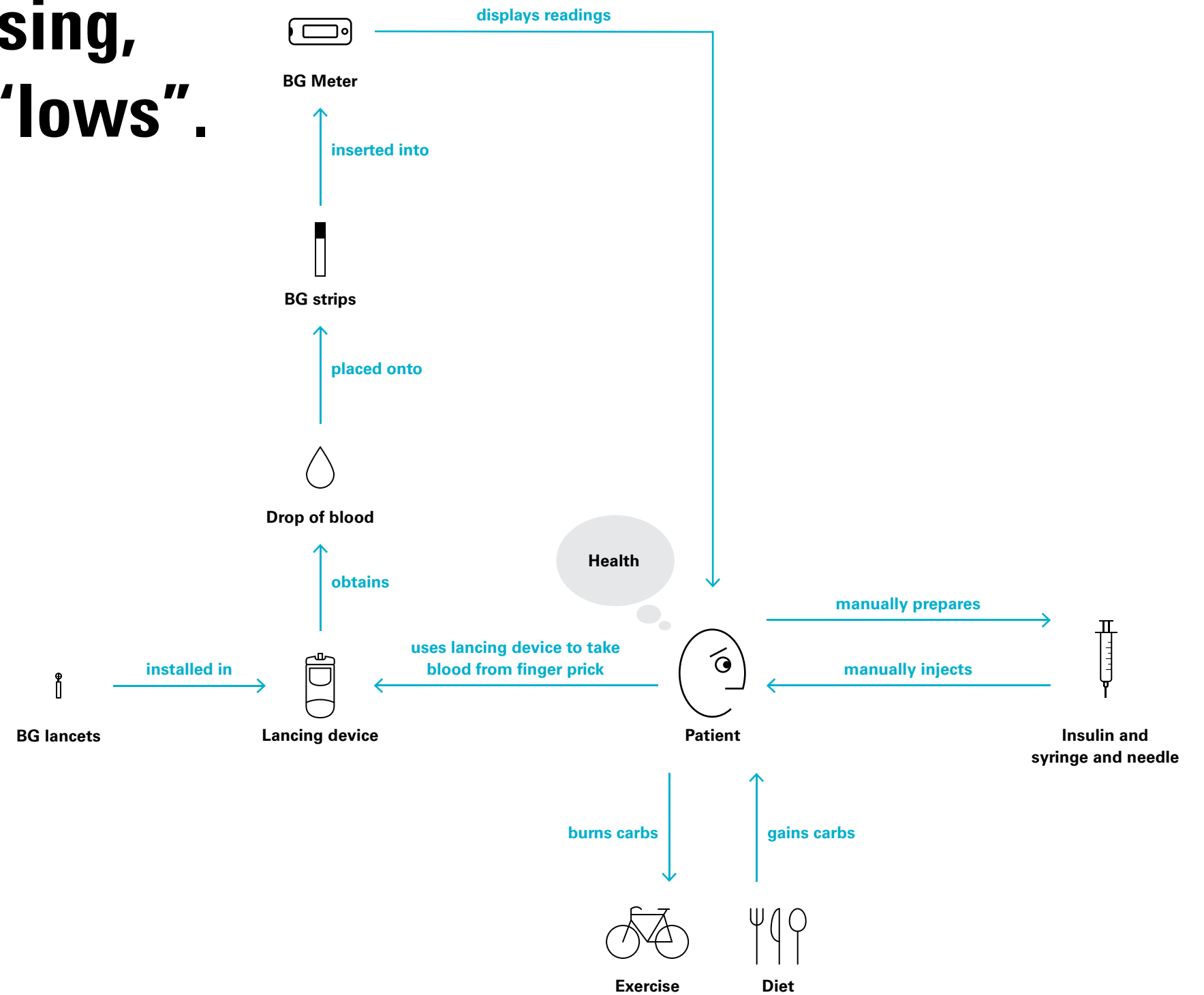Data informed — sensing, reflecting, possible acting, "open loop"
Data aware

State of nature

Health

Patient

burns carbs   gains carbs

Exercise   Diet

Measuring BG drives insulin dosing, reducing "highs" and avoiding "lows".

BG Meter

displays readings

inserted into

BG strips

placed onto

Drop of blood

obtains

Health

manually prepares

uses lancing device to take
blood from finger prick

manually injects

installed in

BG lancets   **Lancing device**   **Patient**   **Insulin and
syringe and needle**

burns carbs   gains carbs

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
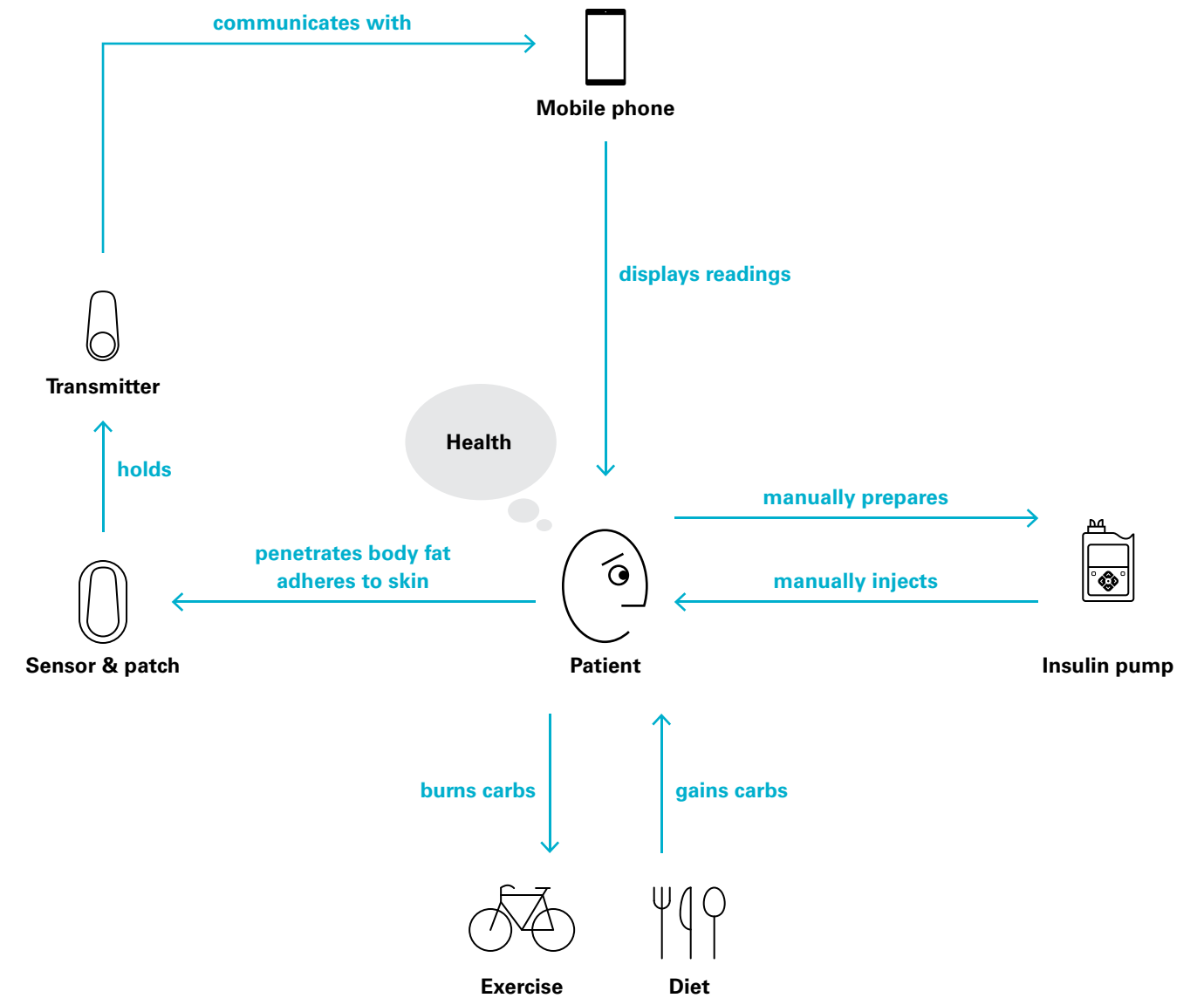Data driven — sensing, reflecting, acting on feedback (human closes the loop)
Data informed
Data aware

State of nature

Exercise   **Diet**

Model driven / data animated
A CGM recognizes "trends" and provides prompts and warnings.

communicates with

Mobile phone

displays readings

Transmitter

Health

holds

manually prepares

penetrates body fat
adheres to skin

manually injects

Sensor & patch

Patient

Insulin pump

New Kind of Nature

burns carbs

gains carbs

Autonomous / self-driving
Semi-autonomous
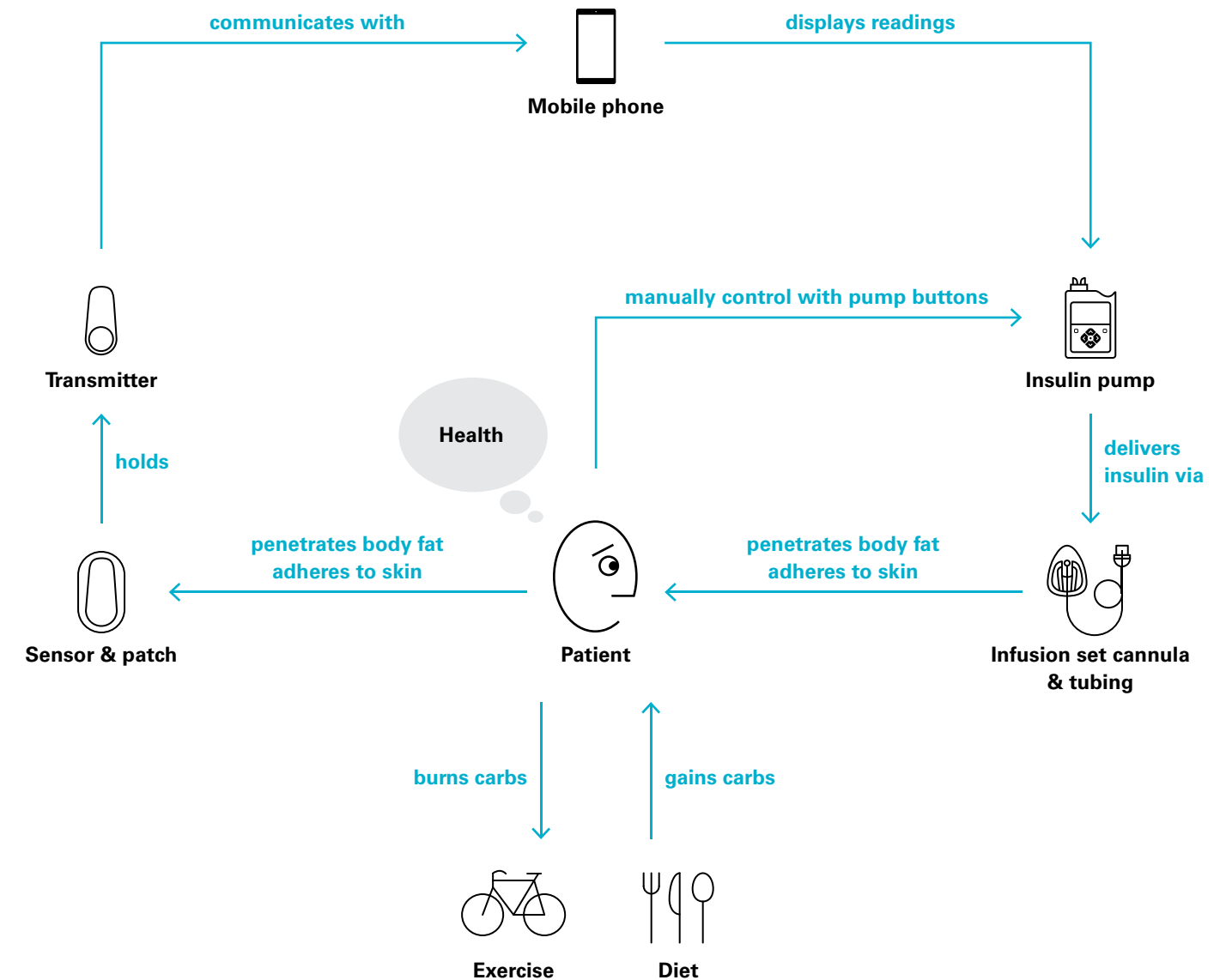Model driven / data animated — collecting data history, recognizing patterns, predicting outcomes
Data driven
Data informed
Data aware

Exercise

Diet

State of nature

Semi-autonomous
CGM connects to a pump; person still inputs carbs and exercise.

communicates with → Mobile phone → displays readings

Transmitter

manually control with pump buttons → Insulin pump

Health

holds

delivers insulin via

penetrates body fat adheres to skin

Sensor & patch

penetrates body fat adheres to skin

Patient

Infusion set cannula & tubing

New Kind of Nature

burns carbs

gains carbs

Autonomous / self-driving
Semi-autonomous — closed loop with minimal human inputs
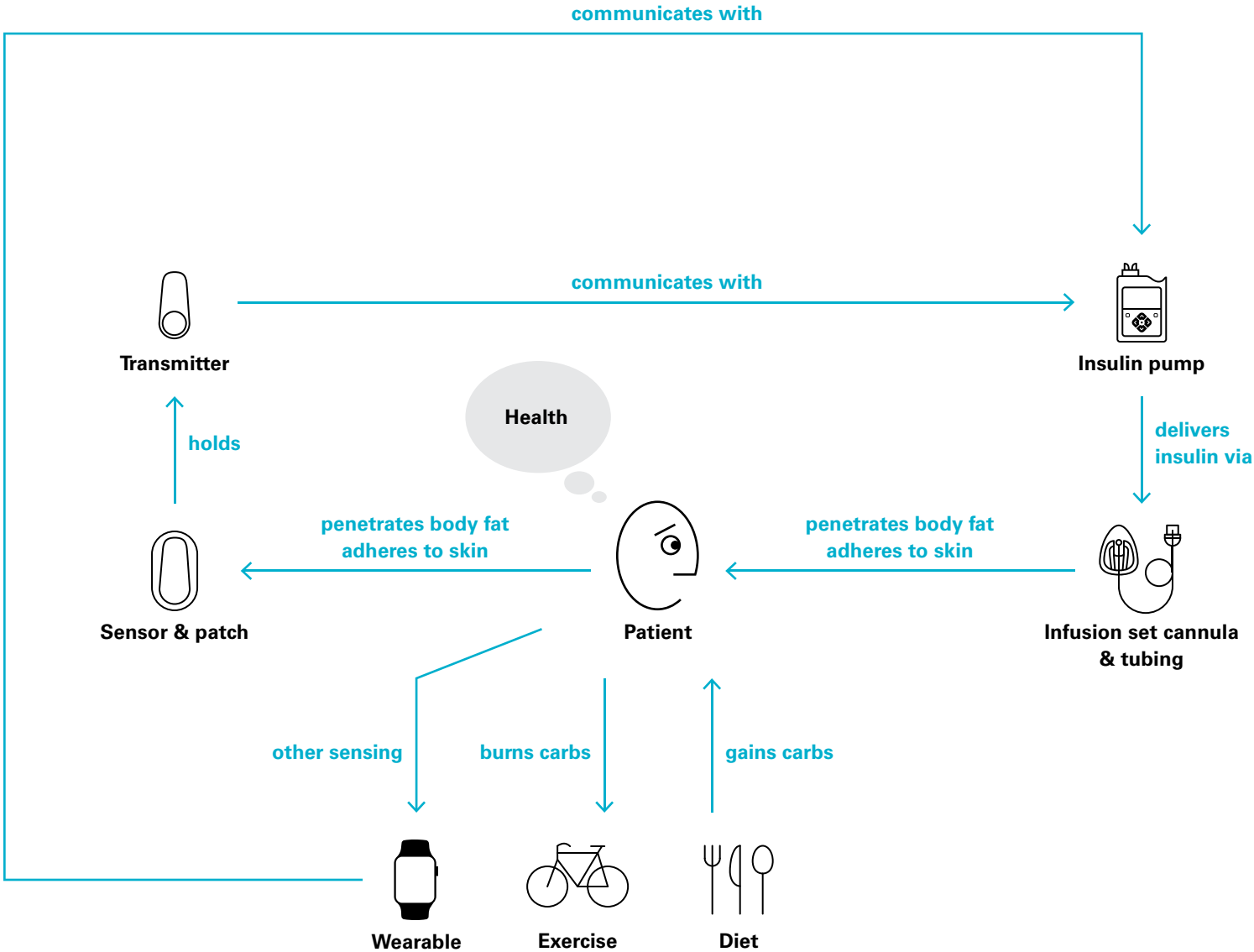Model driven / data animated
Data driven
Data informed
Data aware

Exercise

Diet

State of nature

Autonomous / self-driving

A true "artificial pancreas";
the system measures carbs + exercise.
(We're not here, yet.)

communicates with

communicates with

Transmitter

Health

holds

Insulin pump

delivers
insulin via

penetrates body fat
adheres to skin

penetrates body fat
adheres to skin

Sensor & patch

Patient

**Infusion set cannula
& tubing**

New Kind of Nature

Autonomous / self-driving — closed loop, no human input
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware

other sensing

burns carbs

gains carbs

Wearable

Exercise

Diet

State of nature

New Kind of Nature
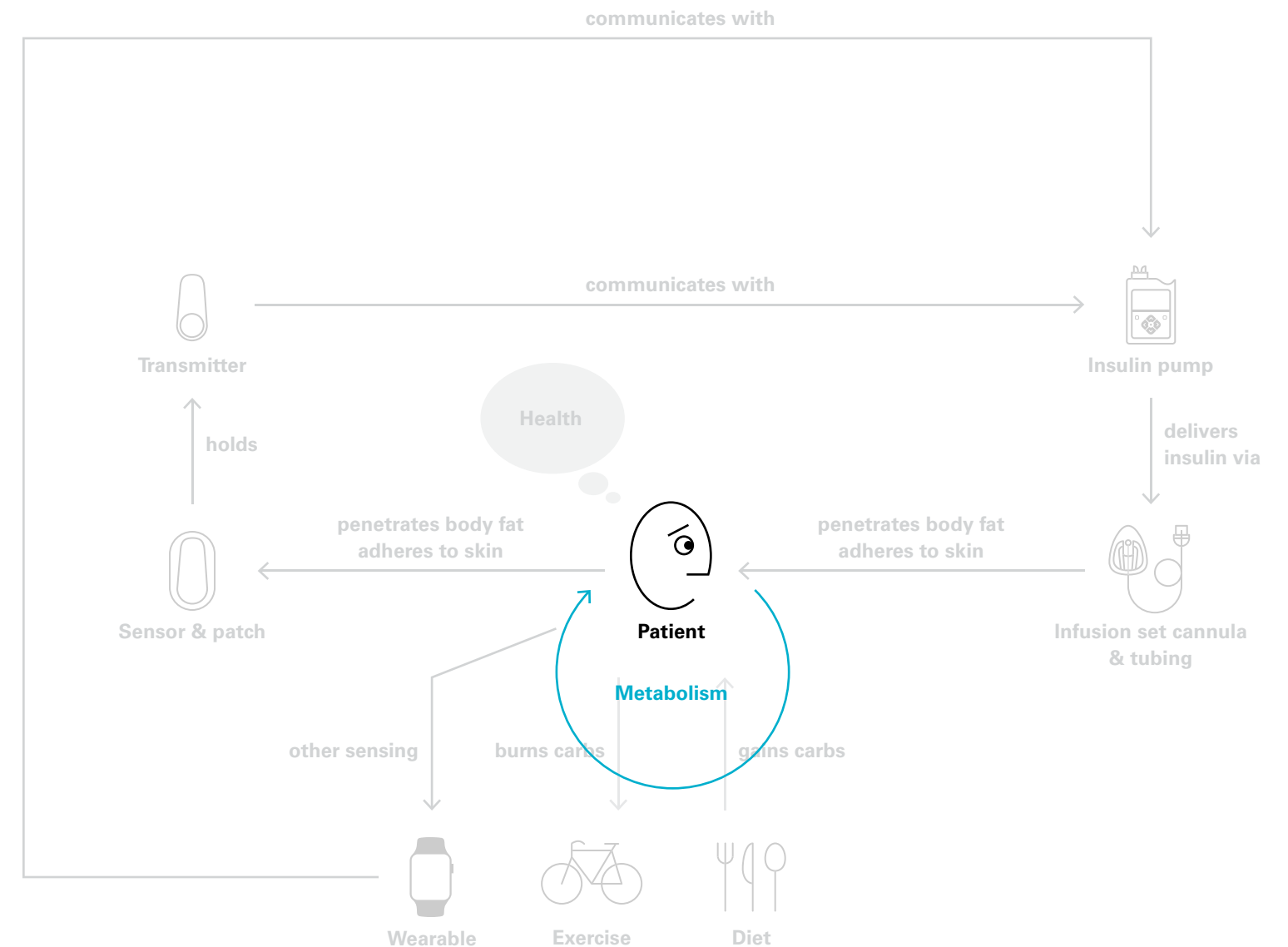
Metabolism happens on its own.
## The individual is unaware.
(But the system *is* "aware".)

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware

State of nature

communicates with

communicates with

Transmitter

Insulin pump

Health

holds

delivers
insulin via

penetrates body fat
adheres to skin

penetrates body fat
adheres to skin

Sensor & patch

Patient

Infusion set cannula
& tubing

Metabolism

other sensing

burns carbs

gains carbs

Wearable

Exercise

Diet

Medical products don't stand alone; they join product-service ecologies, as pharma and device organizations undergo "datafication".
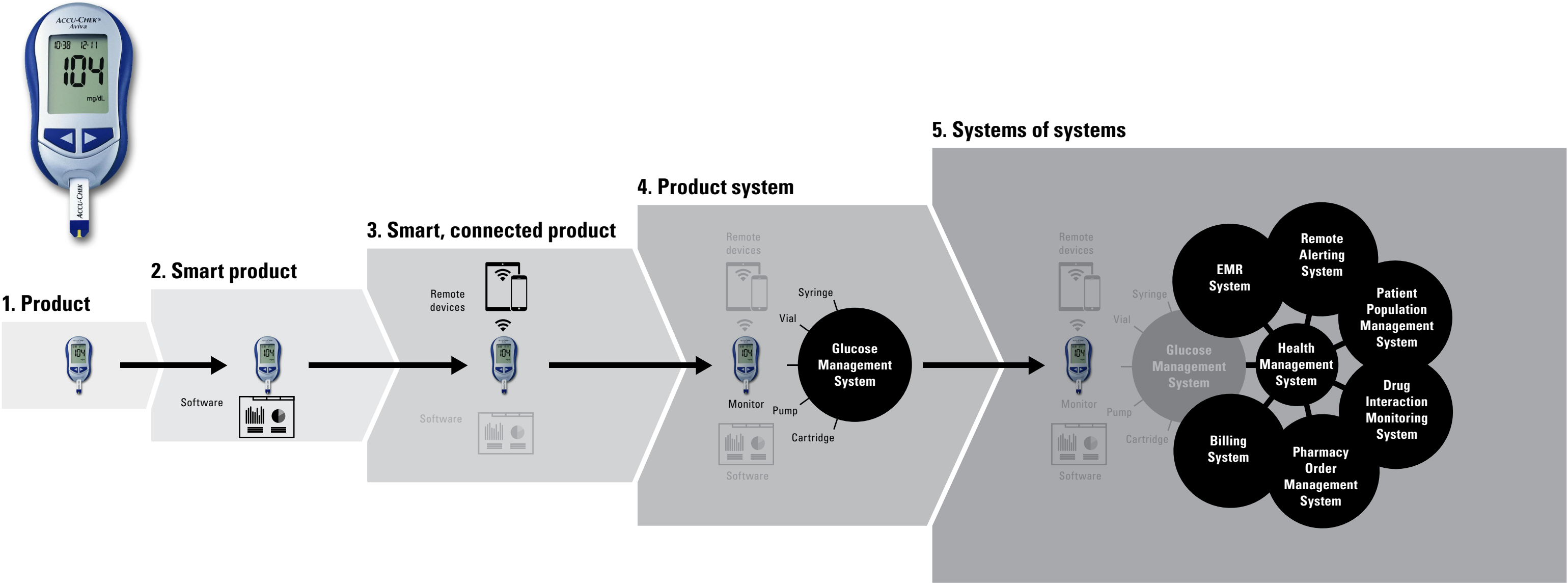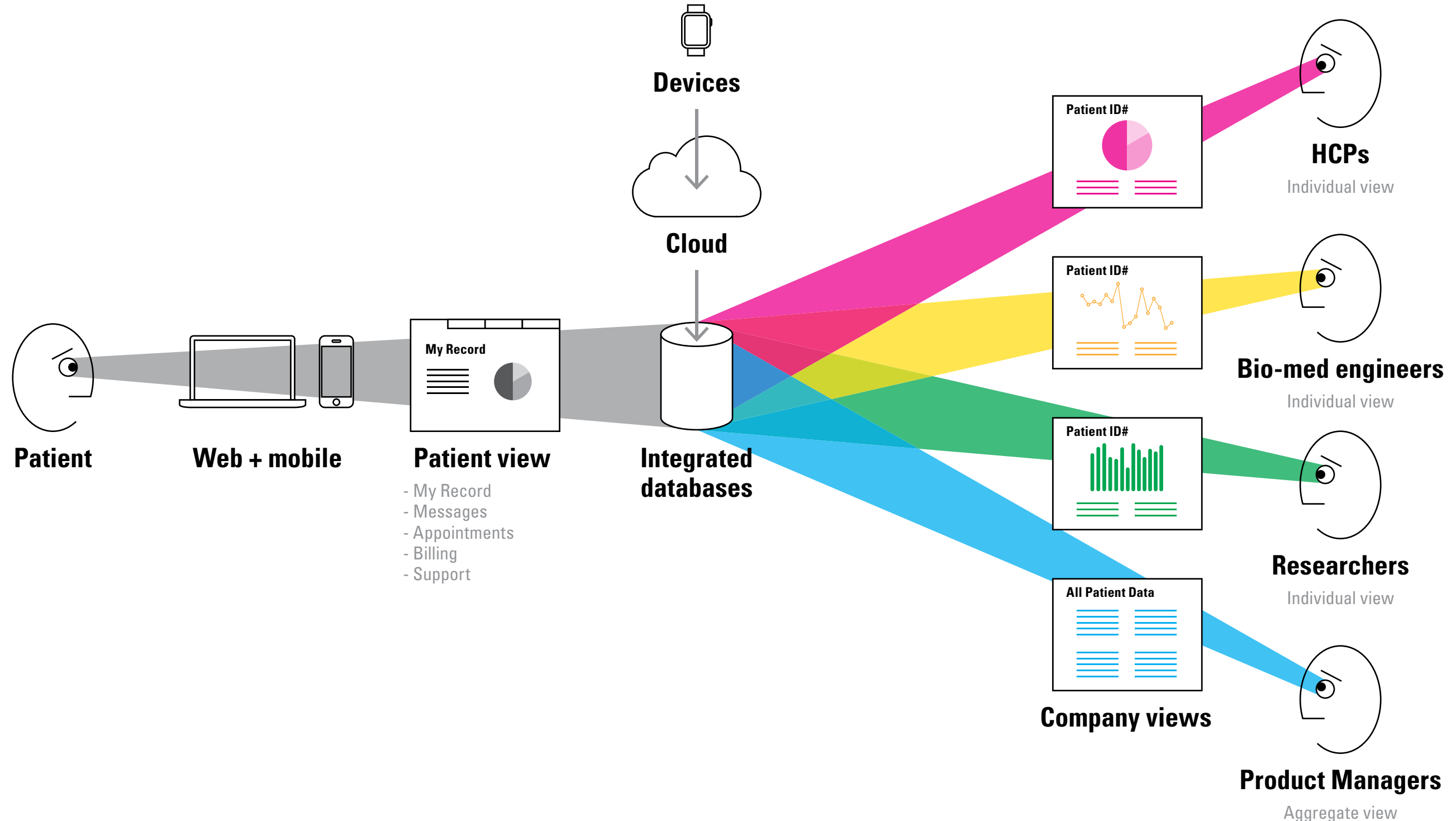

1. Product
Glucometer + processor = computer that can run apps; e.g., bolus calculator, calorie estimator and tracker.

2. Smart product
Software

3. Smart, connected product
Remote devices

Software

When a glucometer connects to a smart-phone, cost can come down, because the meter can build on the phone's processor and display. Plus data can be shared with family and HCPs.

4. Product system
Remote devices

Syringe
Vial
Glucose Management System
Monitor
Pump
Cartridge

Software

A near continuous glucose monitor can be coupled with an insulin pump, forming a glucose management system.

5. Systems of systems
Remote devices

Syringe
Vial
Glucose Management System
Monitor
Pump
Cartridge

Software

EMR System
Remote Alerting System
Patient Population Management System
Health Management System
Drug Interaction Monitoring System
Billing System
Pharmacy Order Management System

The glucose management system can connect with many other systems, such as EMRs, remote alerting, patient population management, drug interaction monitoring, pharmacy order management, and billing.

"Datafication" of medical services will reveal patterns in populations; that will change healthcare and organizations that deliver it.

Devices

Cloud

Patient

Web + mobile

Patient view
- My Record
- Messages
- Appointments
- Billing
- Support

My Record

Integrated databases

Patient ID#

Patient ID#

Patient ID#

All Patient Data

Company views

HCPs
Individual view

Bio-med engineers
Individual view

Researchers
Individual view

Product Managers
Aggregate view

Example:
An organization managing a population of individuals with diabetes

Individuals are treated in isolation; no concept of "patient population".

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware

State of nature

Patient

HCP
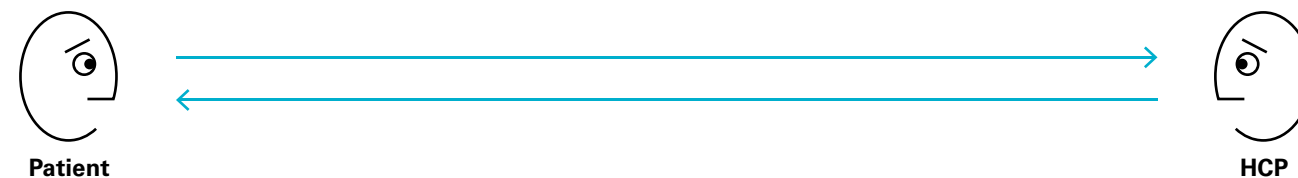
HCPs establish standards of care; fees paid for services. Managing risks and costs become a concern.

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware — reflecting, possible sensing

State of nature

Data informed
Systematic measurement of outcomes begins; payers compare organizational performance.
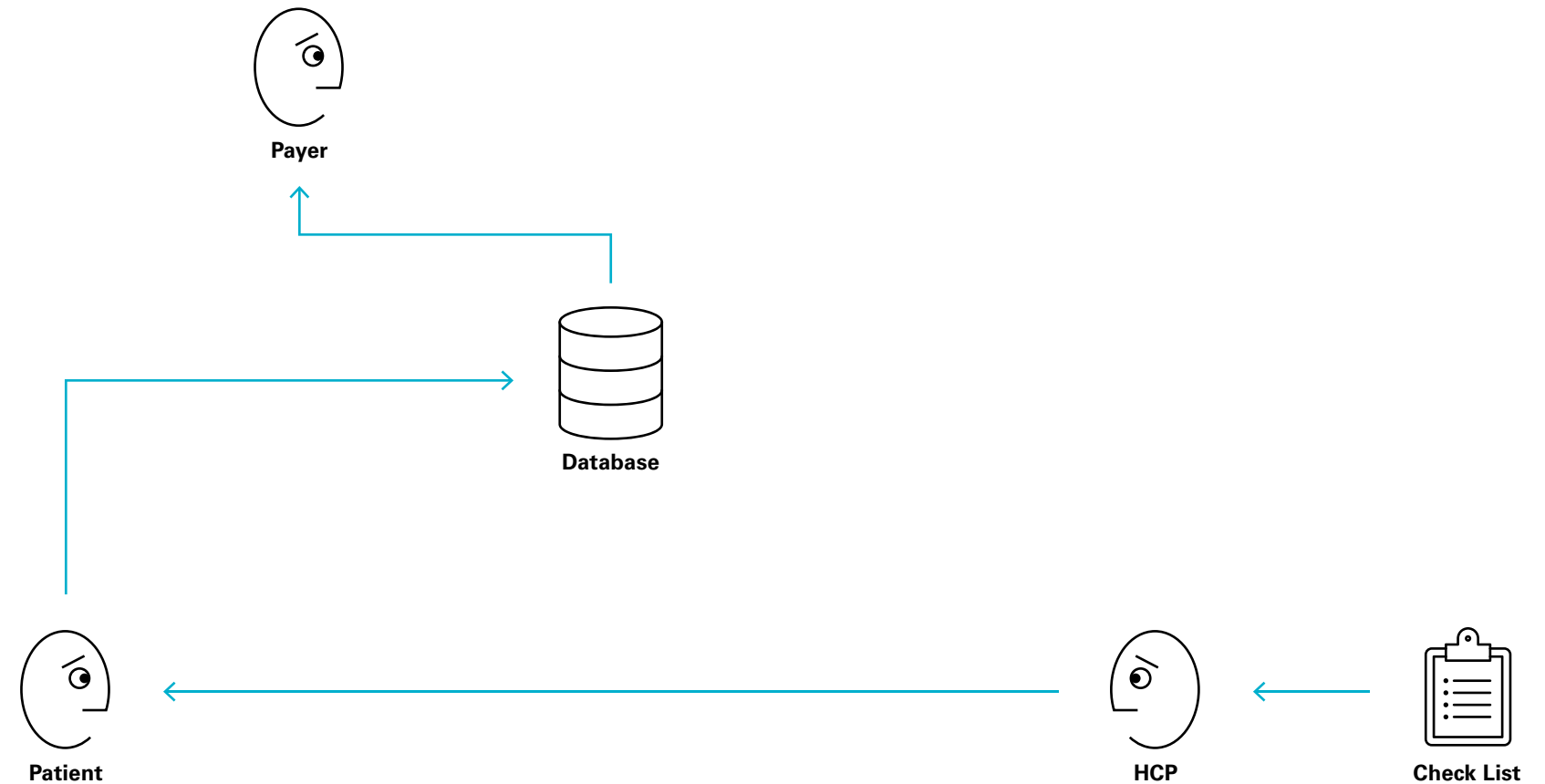
Payer

Database

Patient

HCP

Check List

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed — sensing, reflecting, possible acting, "open loop"
Data aware

State of nature

"Pay-for-performance" begins (feedback); "outcomes" begin to drive HCP actions.
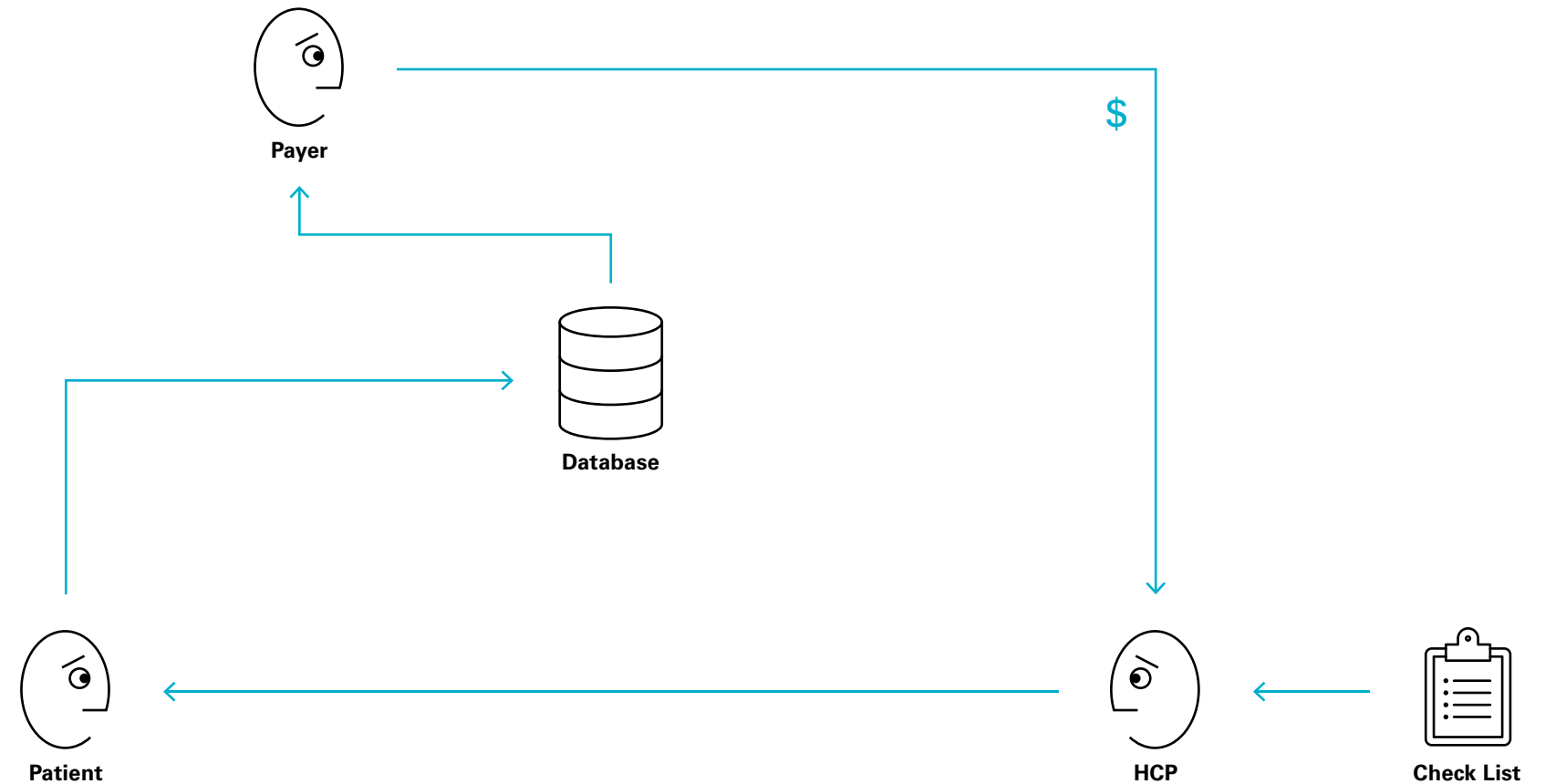
Payer

$

Database

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven — sensing, reflecting, acting on feedback (human closes the loop)
Data informed
Data aware

State of nature

Patient

HCP

Check List

Model driven / data animated

Historical data creates risk models; system refers "high risk" folks to HCPs.
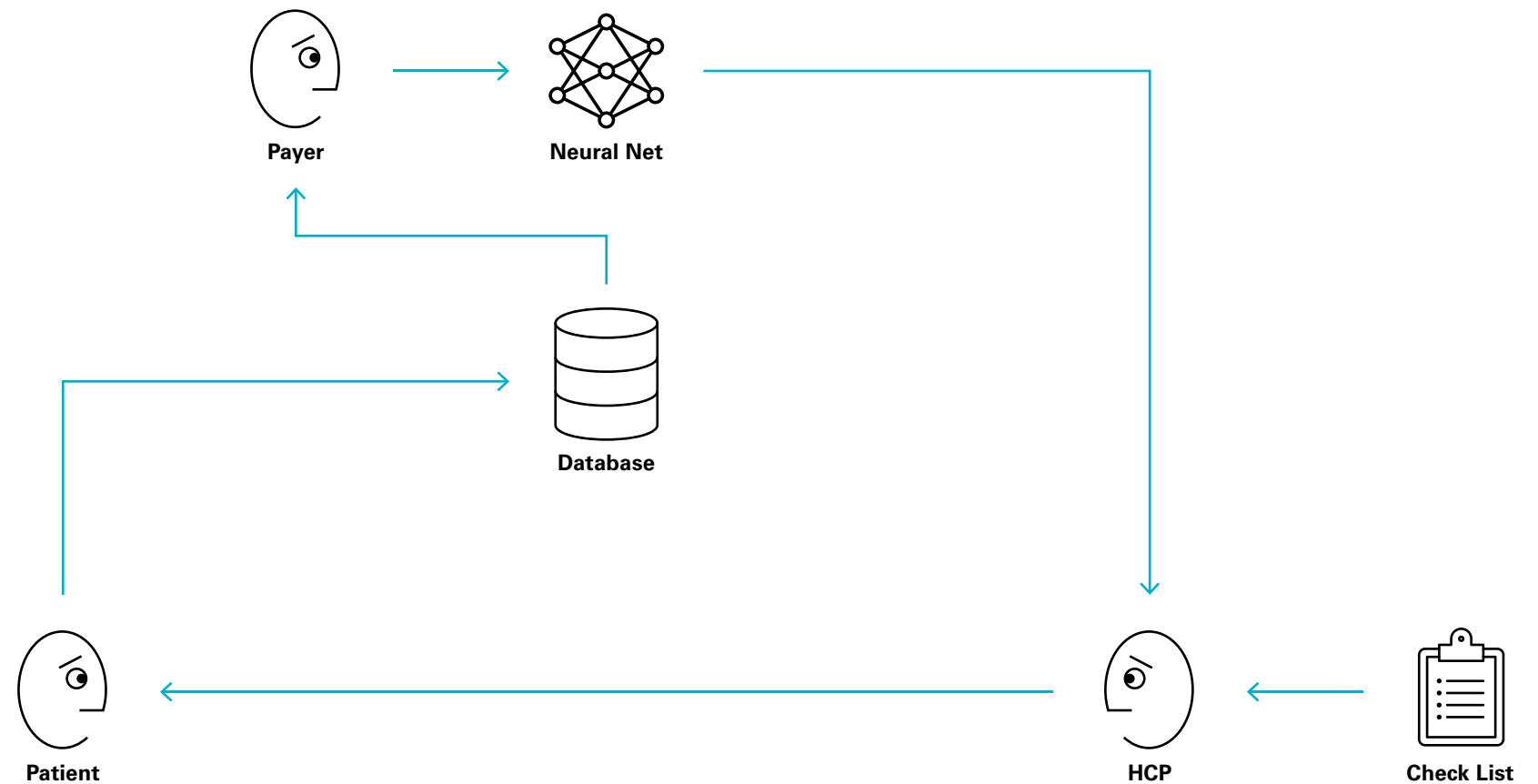
Payer

Neural Net

Database

New Kind of Nature

Patient

HCP

Check List

Autonomous / self-driving
Semi-autonomous
Model driven / data animated — collecting data history, recognizing patterns, predicting outcomes
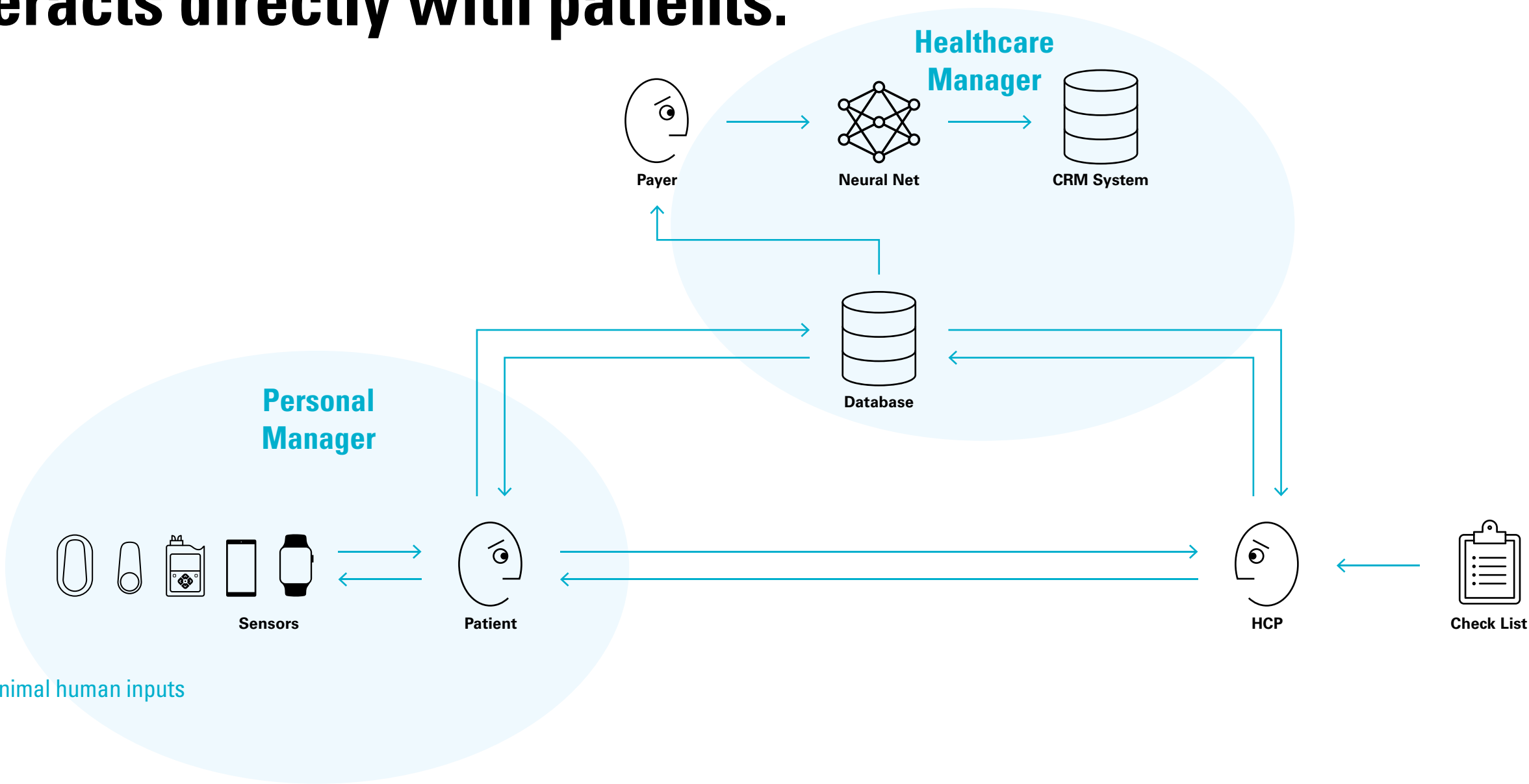Data driven
Data informed
Data aware

State of nature

Semi-autonomous

Healthcare embraces sensors and CRM; the system interacts directly with patients.

Healthcare Manager

Payer

Neural Net

CRM System

Database

Personal Manager

Sensors

Patient

HCP

Check List

New Kind of Nature

Autonomous / self-driving
Semi-autonomous — closed loop with minimal human inputs
Model driven / data animated
Data driven
Data informed
Data aware

State of nature

Autonomous / self-driving
The state (or other payers) measure behavior and offer incentives for following guidelines.

Healthcare Manager

Payer · Neural Net · CRM System

Personal Manager

Database

Sensors · Patient · State · Check List

New Kind of Nature

Autonomous / self-driving — closed loop, no human input
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware
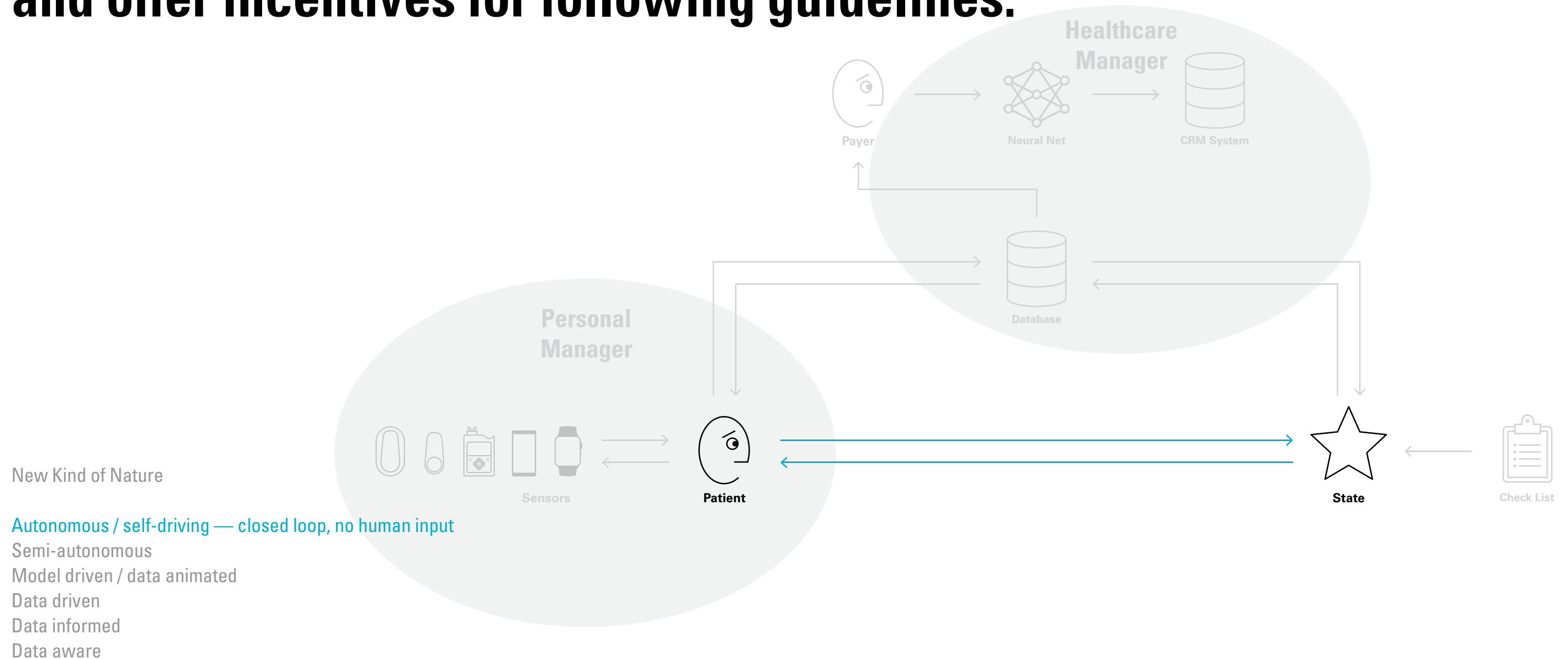
State of nature

New Kind of Nature

Management happens largely on its own.
But how will the managers be managed?
Who will set the goals and guidelines?

New Kind of Nature

Autonomous / self-driving
Semi-autonomous
Model driven / data animated
Data driven
Data informed
Data aware

State of nature
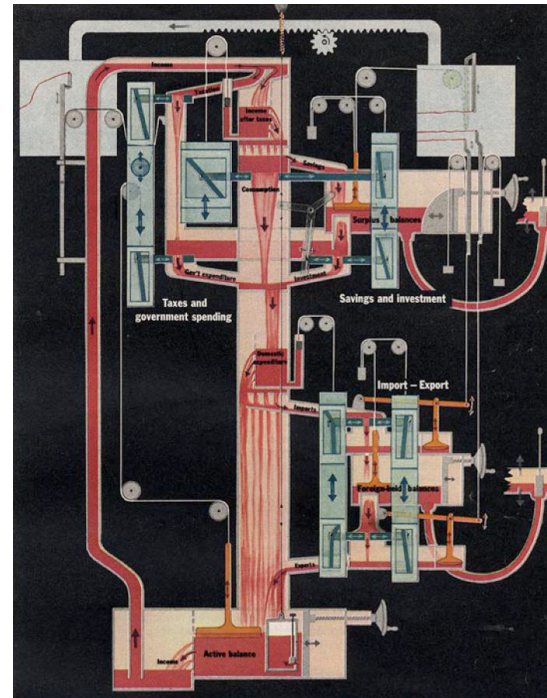
**Patient Population
Health Management System**

Patient Population

"Datafication" leads to "digital twins".

Building dynamic simulations is not a new idea,
but now supercomputers in the cloud enable models to be digital.

Mississippi River Basin Model

Built between 1943–1966
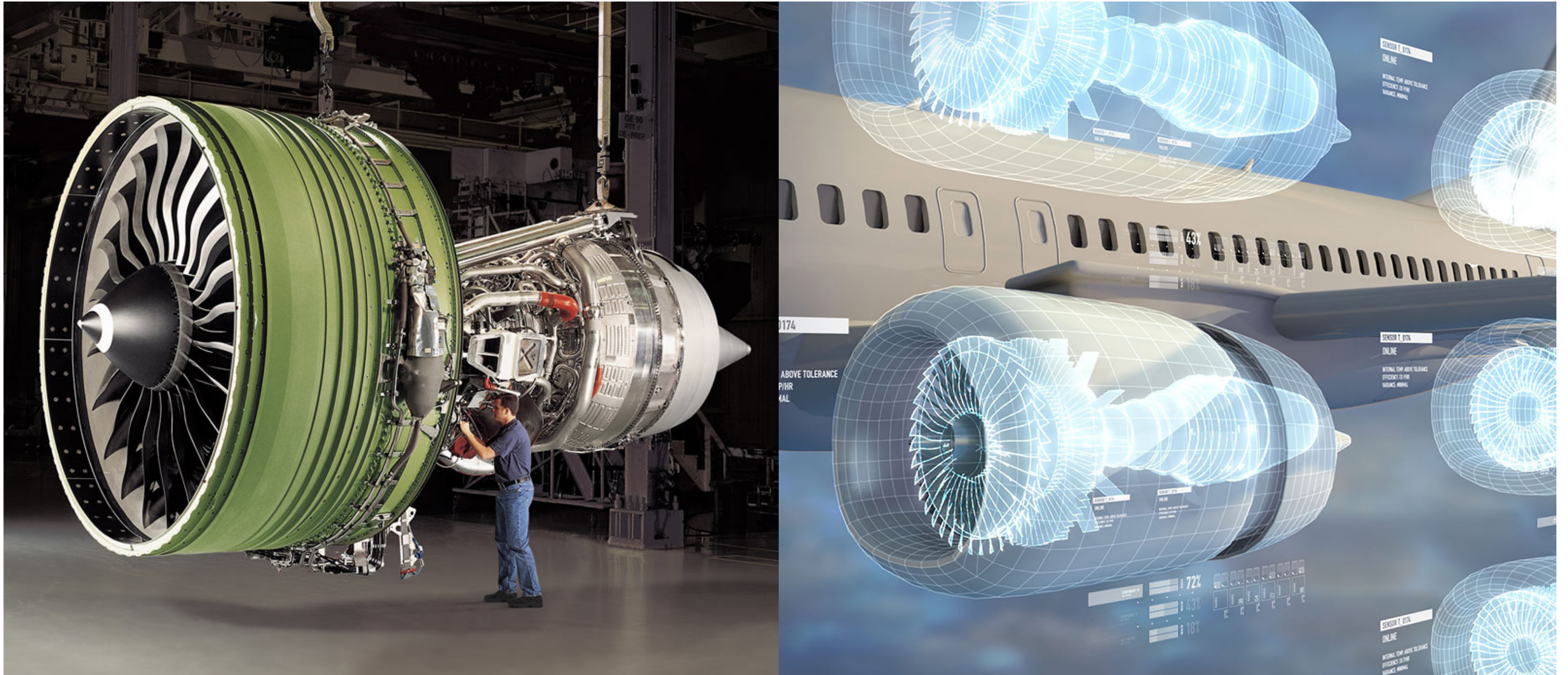

MONIAC
(Monetary National
Income Analogue
Computer)

Built 1949

U.S. Army Corps of Engineers Bay Model

Built in 1957

A "digital twin" is a system of connected models,
plus realtime data, simulating physical operations.

"Digital native" companies already record most user actions. Here is Max Schrems with the 1,200-page profile Facebook collected on him.

Aggregated profiles are already forming; e.g., Facebook and credit info. The Chinese Social Credit Score is another aggregate profile.

Income + Expenses
Liquid Assets
Possessions
Resource Use
Insurance

Schools + Courses
Employees + Jobs
Skills + Expertise

Hobbies
Media
Performances
Sports

×

Create/Participate
Curate/Coach
Comment
Consume

Work

Finance

Interests

Grooming + Prevention
Body Systems
Emotions + Affect
Exercises + Diet

Civic

Wellbeing

Social

Voting + Registration
Juries + Offices
Licenses, Visas, Passports
Taxes, Fees, Finances

Family + Friends
Roommates + Neighbors
Colleagues +
Acquaintances
Vendors + Organizations

×

Meetings
Conversations
Agreements
Transactions

Drilling into a sub-categories shows the potential for detail;

today lab tests can measure over 150 analytes;
more are in development.

Work

Interests

Social

Wellbeing

ce

c

Grooming + Prevention

Body Systems

Emotions + Affect

Exercises + Diet

Cardiovascular Blood

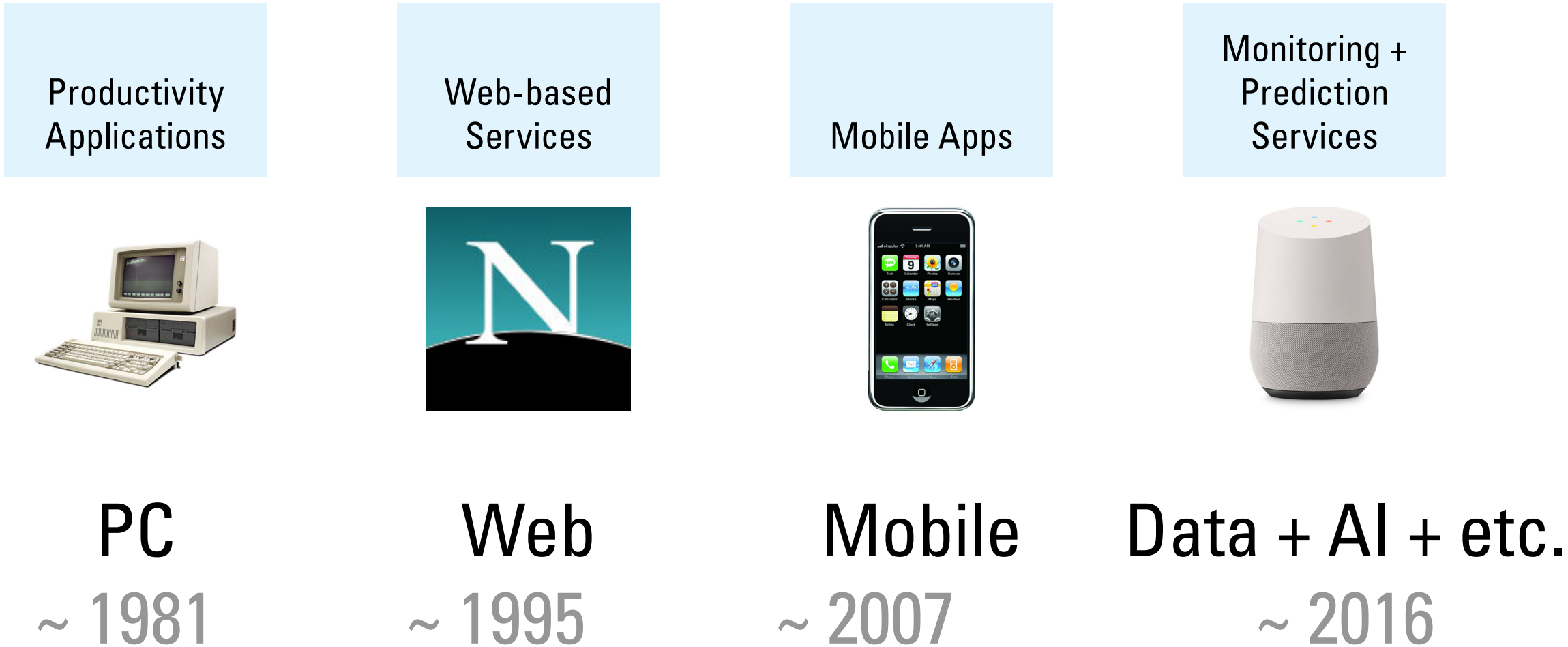|			
	(hydrocortisone)		Normal, PM: 3–17 µg/dL
	17 Hydroxyprogesterone * See also: ovaries		Man, normal: .06–3 mg/L Woman (follicular phase), normal: .2–1 mg/L
	Angiotensin-converting enzyme (ACE)		Normal: 23–57 U/L
Pituitary gland	Growth hormone		At peak: 5–45 ng/mL Between peaks: < 5 ng/mL
	Follicle-stimulating hormone (FSH)		Prepubertal: < 1 – 3 IU/L Adult male: 1–8 IU/L Adult female (follicular & luteal phase): 1–11 IU/L Adult female (ovulation): 6–26 IU/L Post-menopausal female: 30–118 IU/L
	Adrenocorticotropic hormone (ACTH)		Normal: 20–80 pg/mL
	Prolactin		Female, normal: < 20 ng/mL Male, normal: < 15 ng/mL
	Blood Glucose		Hypoglycemia: < 3 mmol/l Normal: 3.6–5.8 mmol/l Normal, post-meal: <10 mmol/l Hyperglycemia: > 7 mmol/l (chronicly)
	Luteinizing hormone (LH)		Female (peak): 20–75 IU/L emale (post-menopausal): 15–60 IU/L
	Insulin absorption		-
	Plasma osmolality		Normal: 275–295 mOsm/kg
Blood: lipids	Total cholesterol		Desirable: < 200

Measuring changes across populations produces models; combined with realtime data, they become "digital twins" of you.
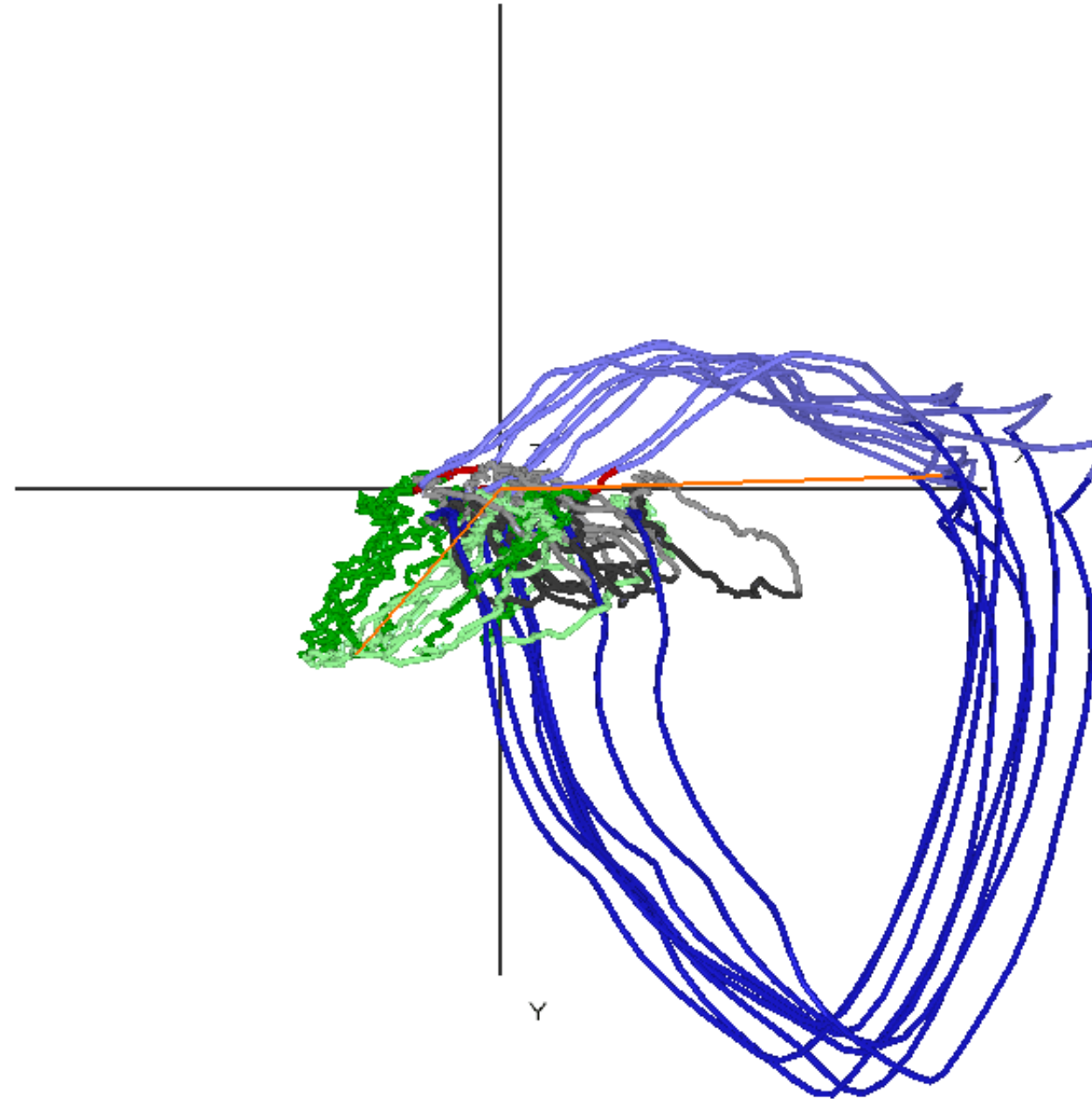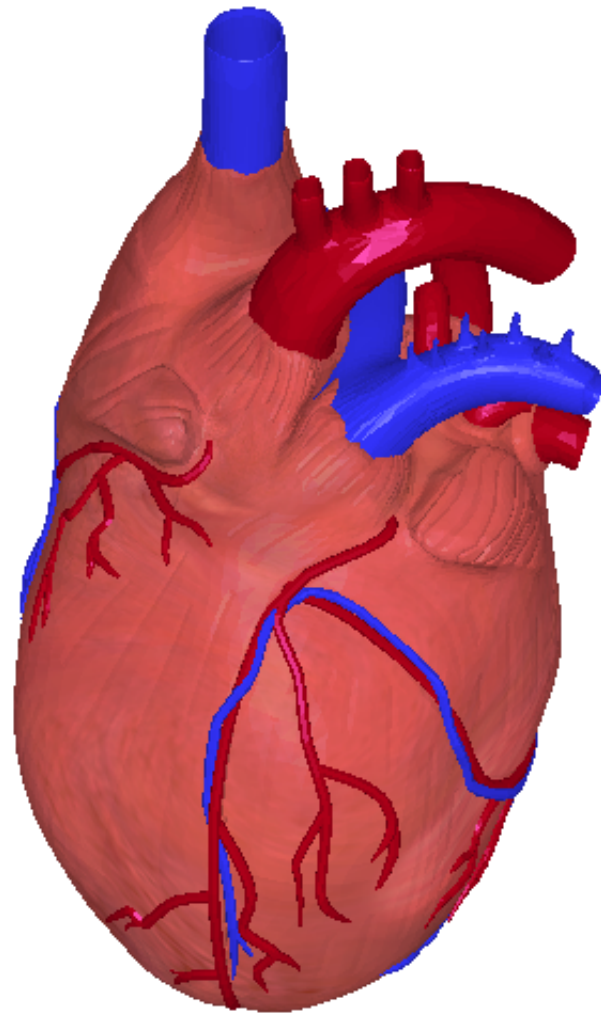
"Digital twins" require a platform — a "data refinery".

In each era, the dominant technology is a "platform"— a system on which others can build apps and stacks.

Productivity Applications	Web-based Services	Mobile Apps	Monitoring + Prediction Services

PC	Web	Mobile	Data + AI + etc.
~ 1981	~ 1995	~ 2007	~ 2016

"Datafication" also suggests a stack — a series of technology layers — each adding value and creating opportunity.

Prediction algorithms — recognizing "patterns of daily living,"
reasoning about sequences of events and what is likely to happen.

Change-detection algorithms — recognizing events (changes in objects)
and sending alerts when a threshold is reached.

Pattern-recognition algorithms — recognizing objects,
teasing "meaning" out of masses of data.

Programmable APIs — making archives accessible
for online machine-based queries.

Multi-modal archives — connecting data from multiple sources,
so that it can be correlated.

Data pipelines — collecting data in a central repository,
cleaning and wrangling it, so that it can be retrieved and used.

Sensor arrays — measuring the environment,
by deploying and connecting foundational technology.

More formally, the "datafication" stack requires a cloud-based supercomputer supporting a "data refinery".
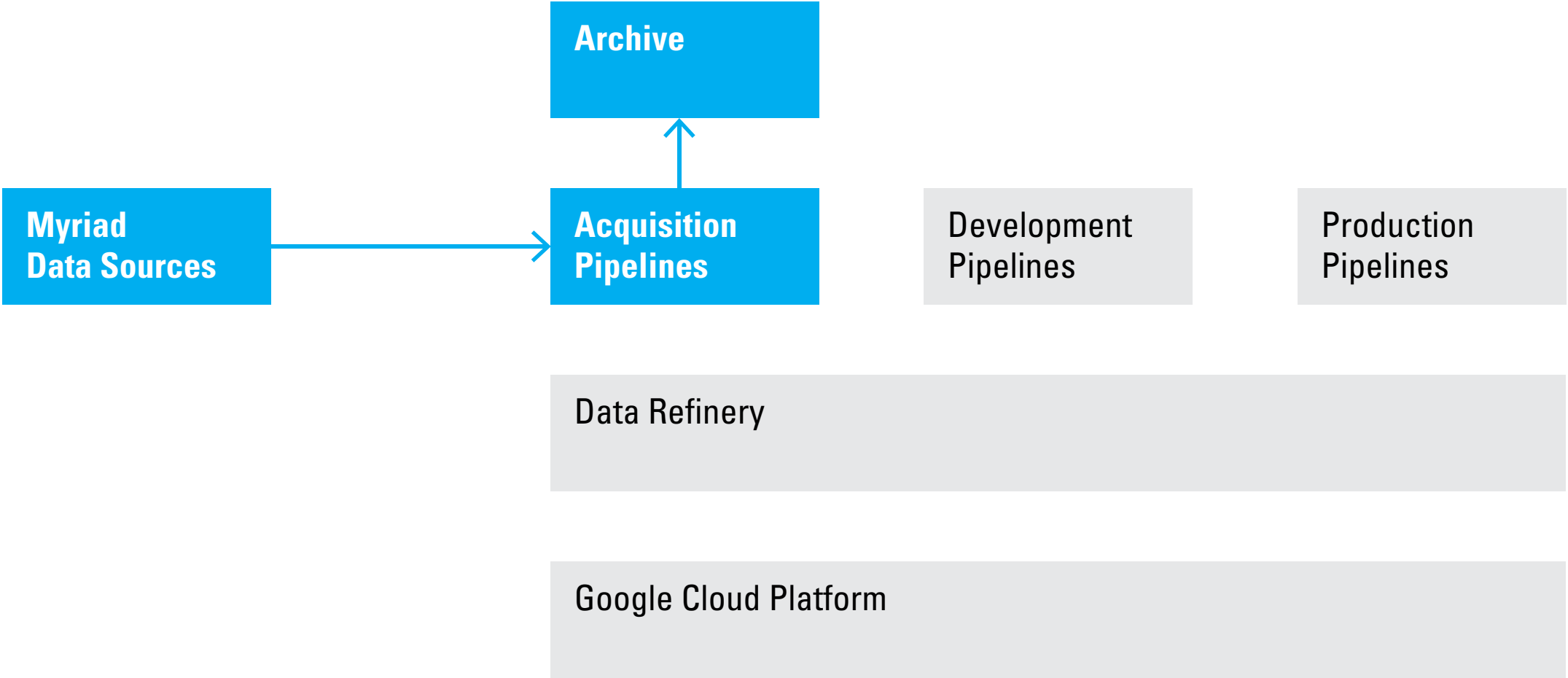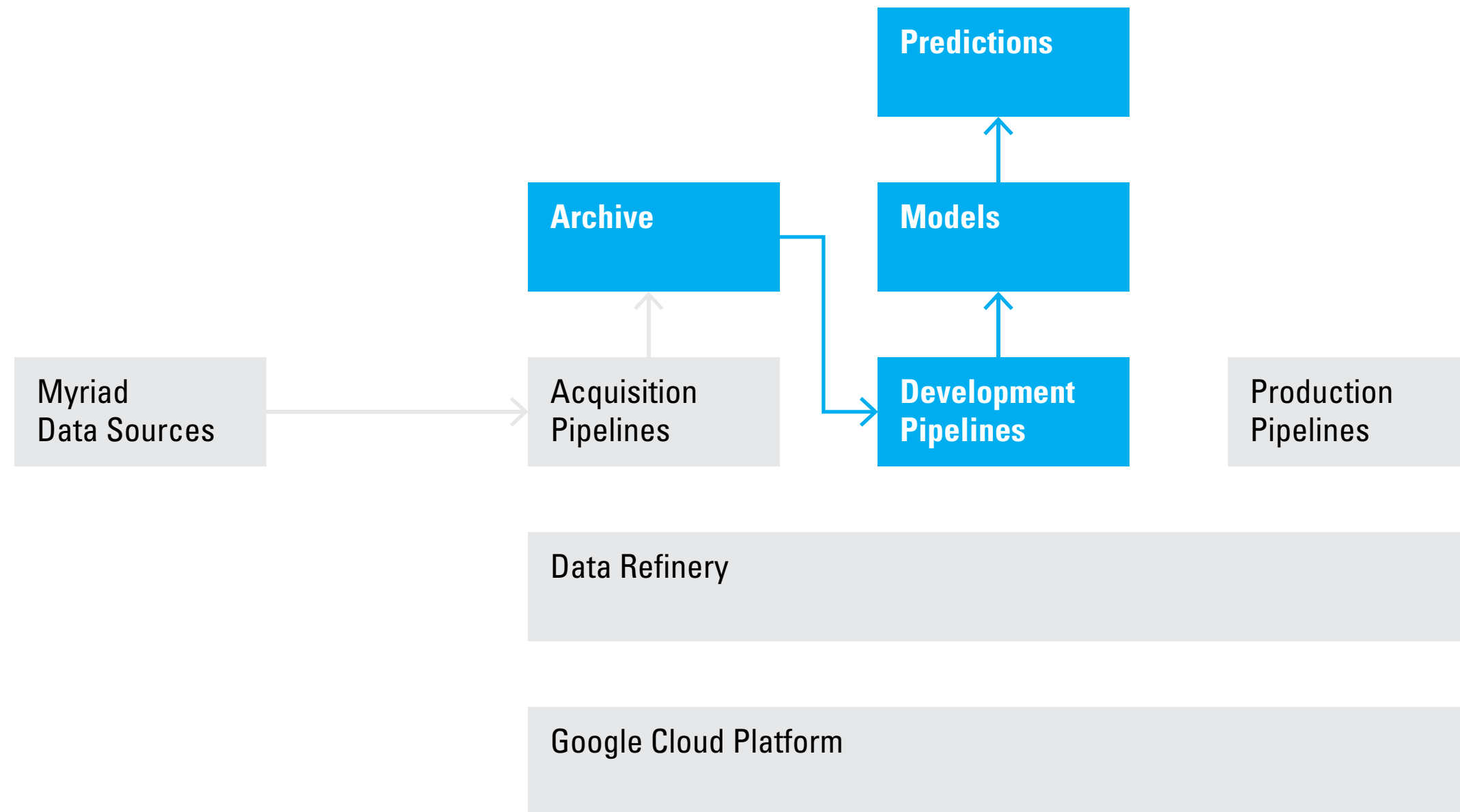
Data Refinery

Google Cloud Platform

The "data refinery" supports a series of pipelines
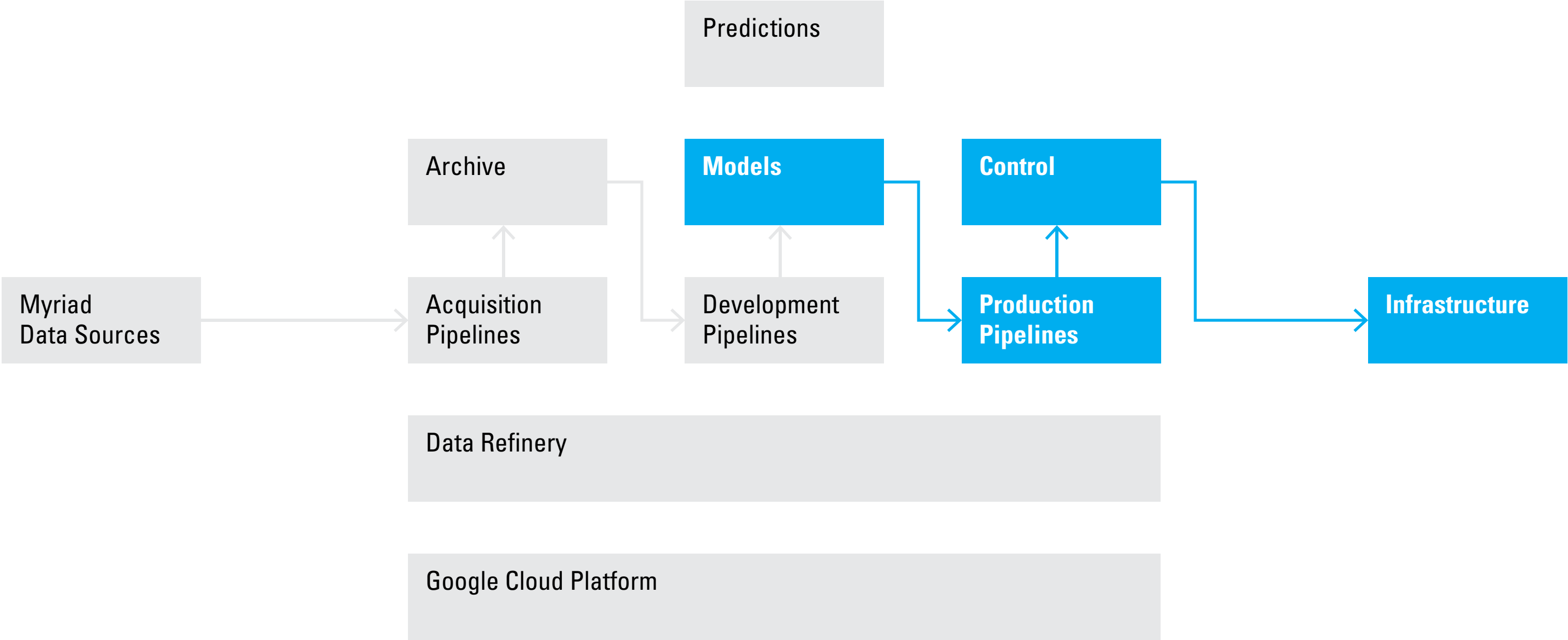# for chaining events (automated transforms) at massive scale.

Acquisition
Pipelines

Development
Pipelines

Production
Pipelines

Data Refinery

Google Cloud Platform
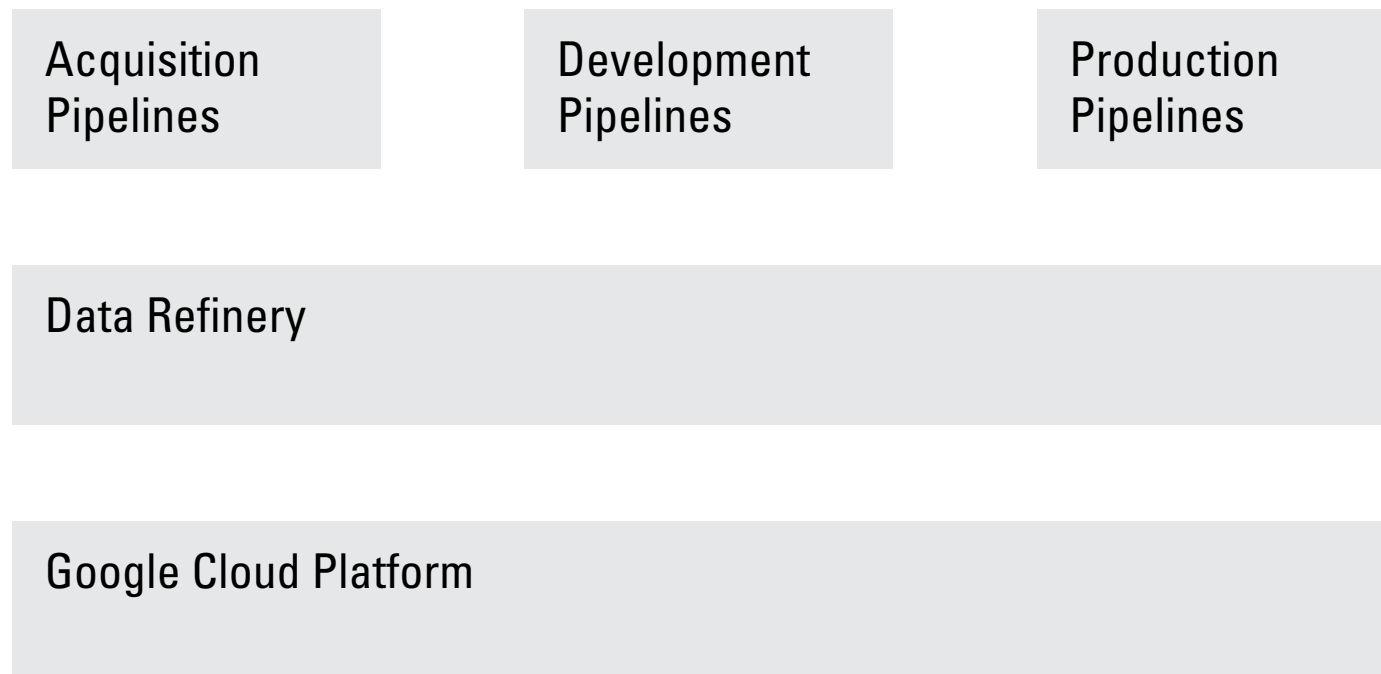
A "data refinery" builds an archive of data on users, operations, and environment.

```
                        ┌─────────────┐
                        │   Archive   │
                        └─────────────┘
                               ▲
                               │
┌─────────────┐        ┌─────────────┐   ┌─────────────┐   ┌─────────────┐
│   Myriad    │───────▶│ Acquisition │   │ Development │   │ Production  │
│ Data Sources│        │  Pipelines  │   │  Pipelines  │   │  Pipelines  │
└─────────────┘        └─────────────┘   └─────────────┘   └─────────────┘

        ┌──────────────────────────────────────────────────────┐
        │  Data Refinery                                        │
        └──────────────────────────────────────────────────────┘

        ┌──────────────────────────────────────────────────────┐
        │  Google Cloud Platform                                │
        └──────────────────────────────────────────────────────┘
```

Large data sets enable development of models and predictions; the more data, the more accurate the results.

Models can be put into production, creating control systems that drive operations.

Predictions

Archive

Models

Control

Myriad
Data Sources

Acquisition
Pipelines

Development
Pipelines

**Production
Pipelines**

Infrastructure

Data Refinery

Google Cloud Platform

As production systems accumulate data and results, they improve their models — effectively "learning".
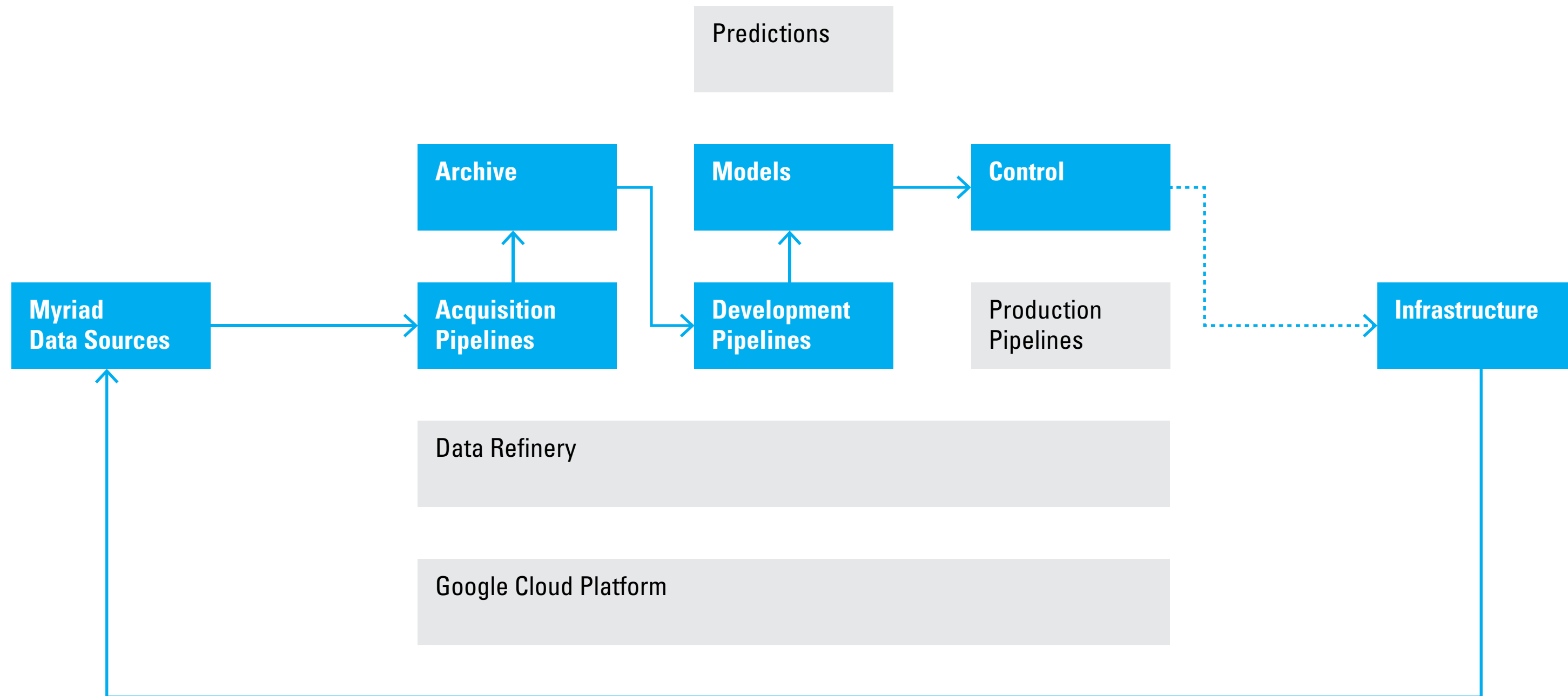
How do organizations get there?

- Start with small, low cost trials
- Expand early successes
- Consider a couple of moonshots
- In short, develop a portfolio
- Develop a data governance structure
- Develop a data infrastructure (centralize? federate?)
- Choose a platform!

"Datafication" may lead to a new type of organization.

Self-driving cars are very nearly real.
Are self-driving organizations next?
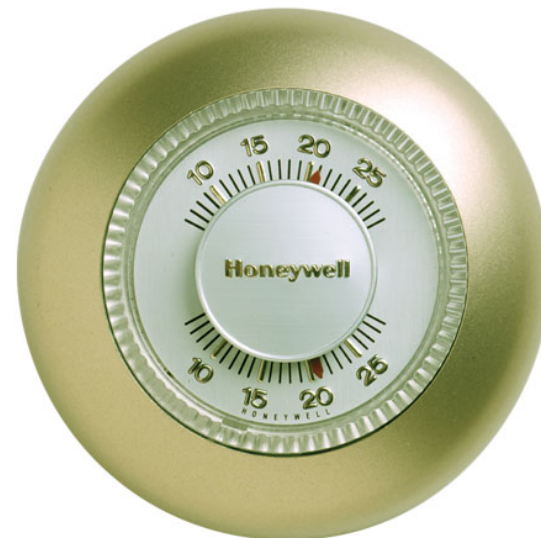
The first semi-autonomous organizations are already here.

amazon

Google

NETFLIX


facebook
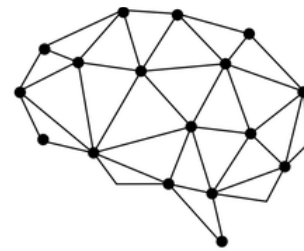
Cambridge Analytica
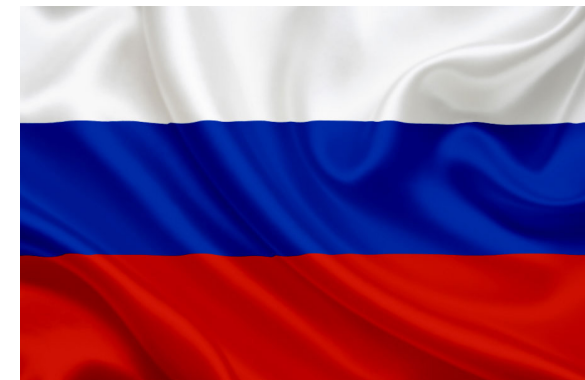
Internet Research Agency

Al Gore says the sustainability revolution will have a greater effect on the economy than the industrial revolution.

It requires the "datafication" of industry and government.

"*We're in the early stages of a sustainability revolution.... empowered by new digital tools, including the internet of things and machine learning,...*"
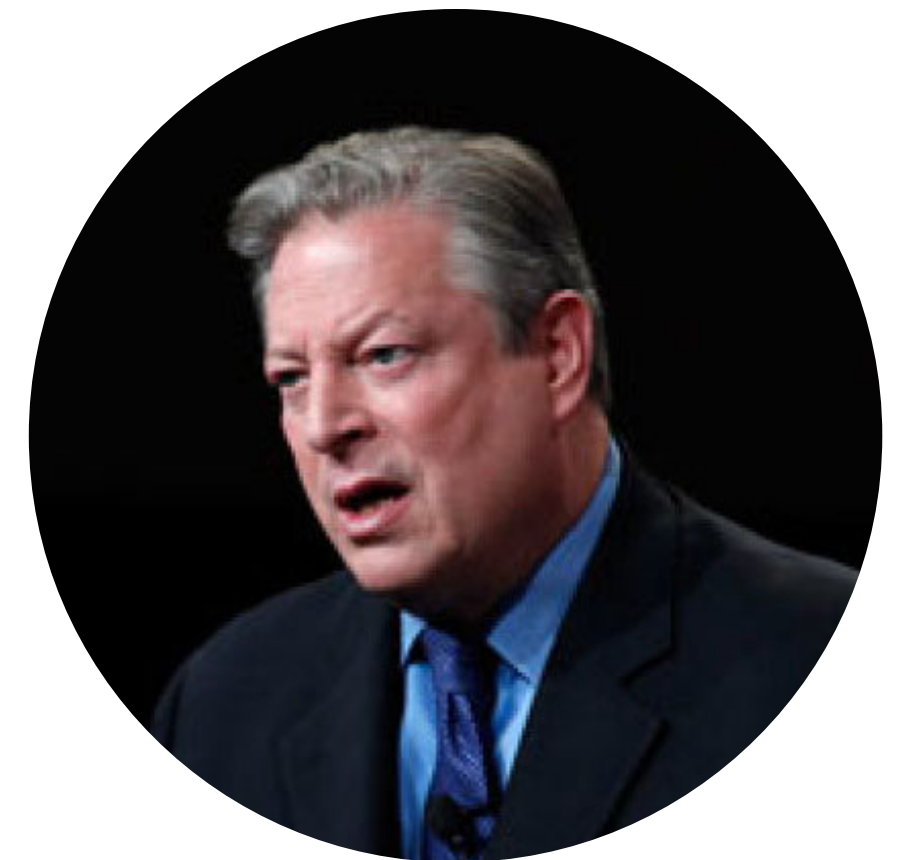
— **Al Gore,** WiReD, April 25, 2018

Michael Lewis frames government as a system for managing risk.

Again, reforming government requires "datafication";
it may turn government into a "data-animated" organization.
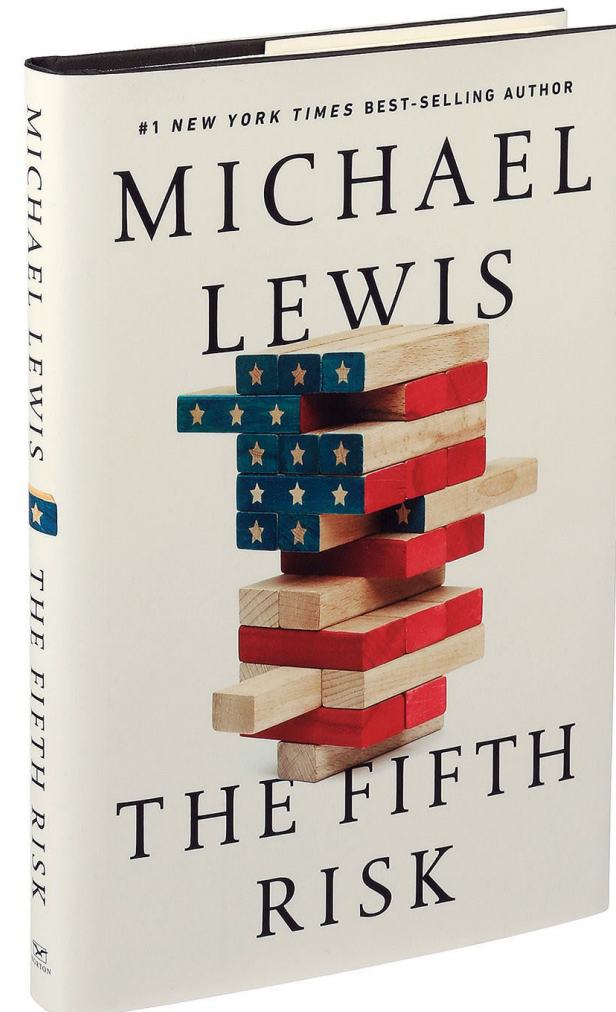
"Datafication" poses opportunities and threats; we need to consider not only rights but also our responsibilities to each other — both as individuals and in our roles within organizations.

Organizations

Individuals

Responsibilities   Rights
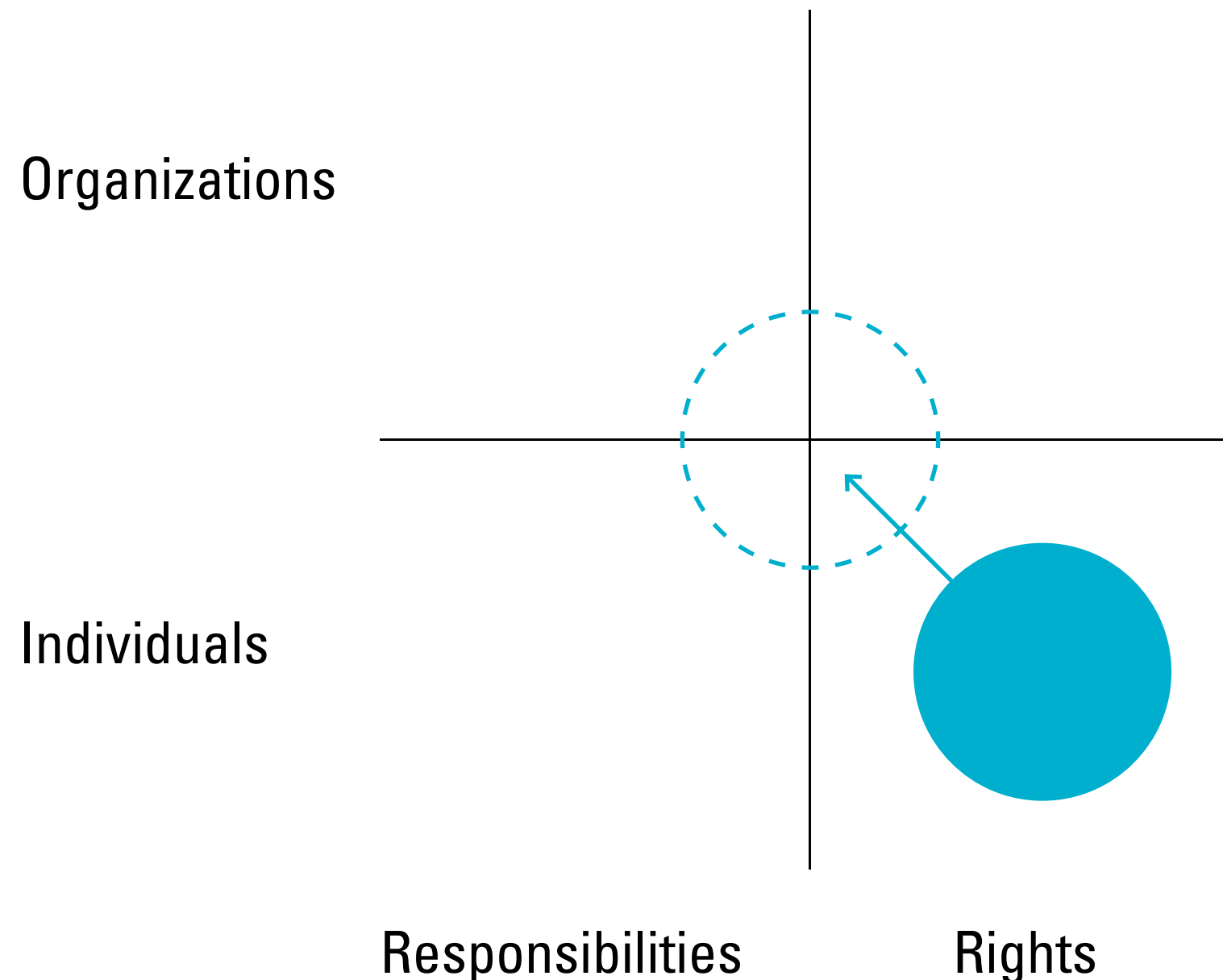
Special thanks to
Todd Elliott
Judy Ohara
Eric Knudtson
Ryan Reposar