CONNECT TO THE FUTURE

Interreg
North Sea Region
Like!
European Regional Development Fund

EUROPEAN UNION

Topkring
Dienstverlening
Gemeenten

VNG

VDP | Vereniging Directeuren
Publieksdiensten

follow us: www.northsearegion.eu/like

Data-Driven Organizations

Dr. John Dong

About Me

- Associate professor at University of Groningen
- Research fellow of Groningen Digital Business Center (GDBC)
- Director of master programme in change management
- Director of digital business focus area
- PhD in information systems
- Senior editor of *Information Technology & People*
- Associate editor of *Information & Management*
- Chair of big data analytics track at *Wuhan International Conference on E-Business (WHICEB)*

...

McKinsey Global Institute

The age of analytics: Competing in a data-driven world

December 2016 | Report

Private Sectors

AIRBNB
The largest accommodation provider owns no real estate.

FACEBOOK
The most popular media provides no content.

NETFLIX
The largest growing television network lays no cables.

@BUSINESSMINDSET101

ALIBABA
The most valuable retailer has no inventory

INSTAGRAM
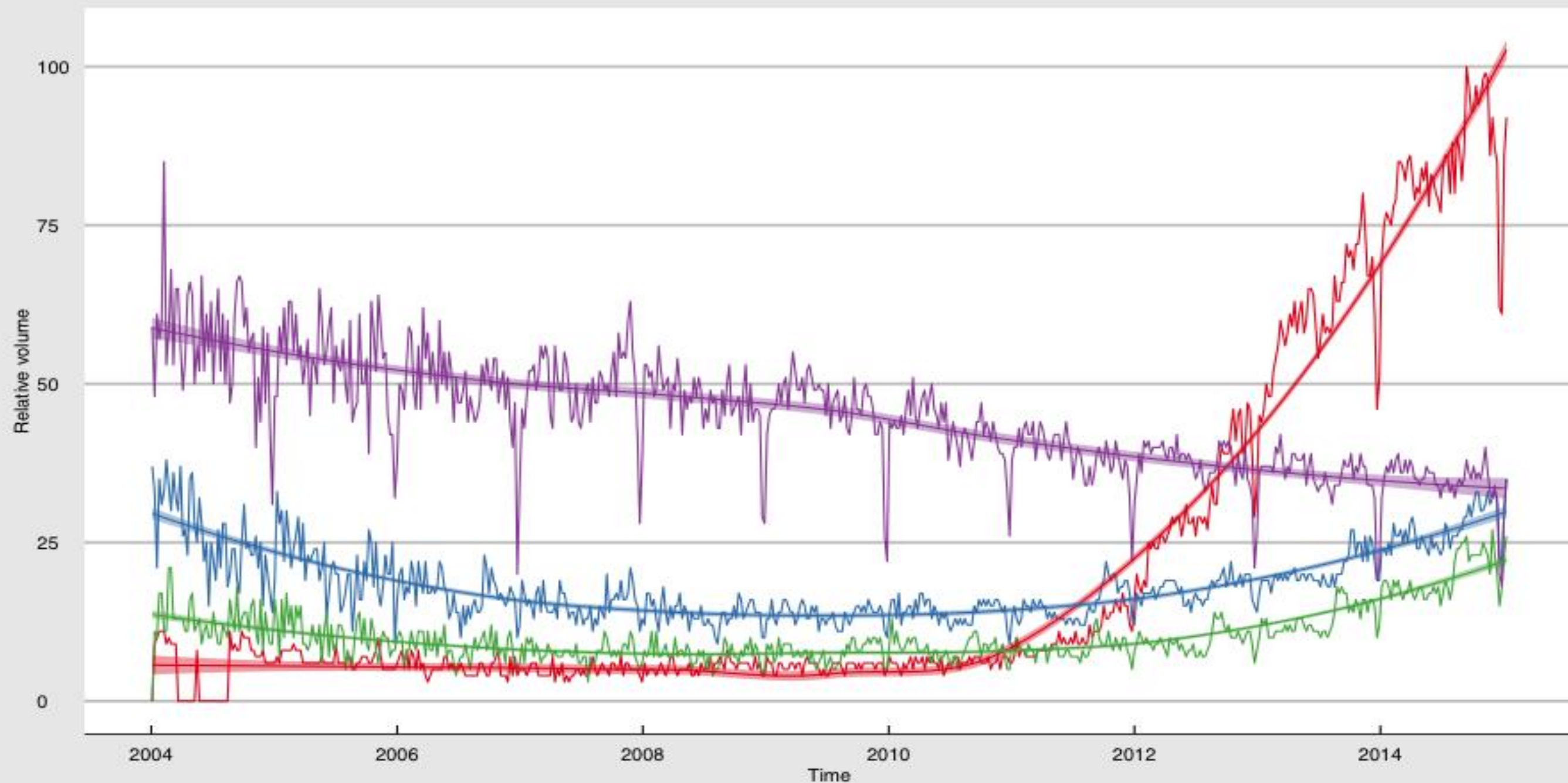The most valuable photo company sells no cameras

UBER
The largest taxi company owns no vehicles.

The most valuable asset that any company can have is DATA!

Big Data

Google Search volume for data related phrases over time

Search phrase — Big Data — machine learning — data science — business intelligence

What Is It

The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume, variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.

Analytics

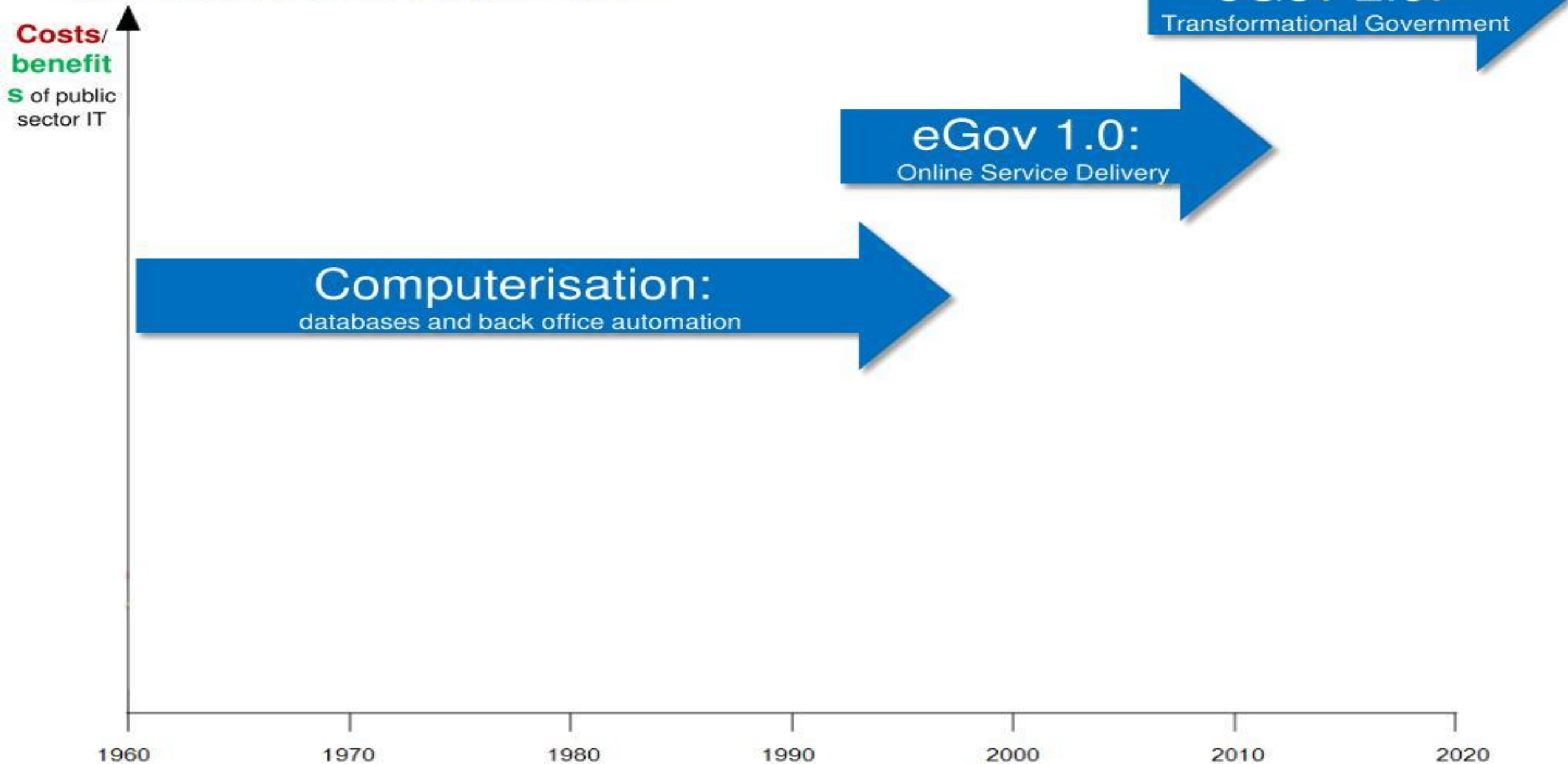| Table 1. BI&A Evolution: Key Characteristics and Capabilities			
	Key Characteristics	**Gartner BI Platforms Core Capabilities**	**Gartner Hype Cycle**
BI&A 1.0	DBMS-based, structured content • RDBMS & data warehousing • ETL & OLAP • Dashboards & scorecards • Data mining & statistical analysis	• *Ad hoc* query & search-based BI • Reporting, dashboards & scorecards • OLAP • Interactive visualization • Predictive modeling & data mining	• Column-based DBMS • In-memory DBMS • Real-time decision • Data mining workbenches
BI&A 2.0	Web-based, unstructured content • Information retrieval and extraction • Opinion mining • Question answering • Web analytics and web intelligence • Social media analytics • Social network analysis • Spatial-temporal analysis		• Information semantic services • Natural language question answering • Content & text analytics
BI&A 3.0	Mobile and sensor-based content • Location-aware analysis • Person-centered analysis • Context-relevant analysis • Mobile visualization & HCI		• Mobile BI

E-Government

A brief history of e-Government
From automation to transformation

Costs/benefits of public sector IT

eGov 2.0: Transformational Government

eGov 1.0: Online Service Delivery

Computerisation: databases and back office automation

1960   1970   1980   1990   2000   2010   2020

E-Government 2.0

- From digitization/digitalization to digital transformation

- The more quality and accurate data is available, the better the public decisions

- The faster data is available, the faster public decisions

Big Data Analytics

Table 2. BI&A Applications: From Big Data to Big Impact

	E-Commerce and Market Intelligence	E-Government and Politics 2.0	Science & Technology	Smart Health and Wellbeing	Security and Public Safety
Applications	• Recommender systems • Social media monitoring and analysis • Crowd-sourcing systems • Social and virtual games	• Ubiquitous government services • Equal access and public services • Citizen engagement and participation • Political campaign and e-polling	• S&T innovation • Hypothesis testing • Knowledge discovery	• Human and plant genomics • Healthcare decision support • Patient community analysis	• Crime analysis • Computational criminology • Terrorism informatics • Open-source intelligence • Cyber security
Data	• Search and user logs • Customer transaction records • Customer-generated content	• Government information and services • Rules and regulations • Citizen feedback and comments	• S&T instruments and system-generated data • Sensor and network content	• Genomics and sequence data • Electronic health records (EHR) • Health and patient social media	• Criminal records • Crime maps • Criminal networks • News and web contents • Terrorism incident databases • Viruses, cyber attacks, and botnets
	Characteristics: Structured web-based, user-generated content, rich network information, unstructured informal customer opinions	Characteristics: Fragmented information sources and legacy systems, rich textual content, unstructured informal citizen conversations	Characteristics: High-throughput instrument-based data collection, fine-grained multiple-modality and large-scale records, S&T specific data formats	Characteristics: Disparate but highly linked content, person-specific content, HIPAA, IRB and ethics issues	Characteristics: Personal identity information, incomplete and deceptive content, rich group and network information, multilingual content
Analytics	• Association rule mining • Database segmentation and clustering • Anomaly detection • Graph mining • Social network analysis • Text and web analytics • Sentiment and affect analysis	• Information integration • Content and text analytics • Government information semantic services and ontologies • Social media monitoring and analysis • Social network analysis • Sentiment and affect analysis	• S&T based domain-specific mathematical and analytical models	• Genomics and sequence analysis and visualization • EHR association mining and clustering • Health social media monitoring and analysis • Health text analytics • Health ontologies • Patient network analysis • Adverse drug side-effect analysis • Privacy-preserving data mining	• Criminal association rule mining and clustering • Criminal network analysis • Spatial-temporal analysis and visualization • Multilingual text analytics • Sentiment and affect analysis • Cyber attacks analysis and attribution
Impacts	Long-tail marketing, targeted and personalized recommendation, increased sale and customer satisfaction	Transforming governments, empowering citizens, improving transparency, participation, and equality	S&T advances, scientific impact	Improved healthcare quality, improved long-term care, patient empowerment	Improved public safety and security

Unbalanced Speed

- Technology evolves and is adopted very rapidly, but policymaking practices change slowly
- Primary focus on creating and gathering, rather than using
- Policymaking concerns the long term and cannot adapt to the speed of data gathering
- The speed of policymaking is constrained by public administration and political dynamics

Shortage of Resources

- Government is often lack of sufficient resources and investments in big data analytics
- Strong conservative culture against change
- Shortage of financial investment
- Without top management support
- Lack of qualified skills, internally and externally
- Zero tolerance for experimentation and failure

People, Not Technology

- The most challenging part is transforming people, rather than adopting new technology
- Difficult in integrating big data analytics with current work practices
- Fear of being replaced, though it is useful
- Passive, active and aggressive resistance behavior
- Knowledge gap between administrative and technical staffs (translation is needed)

Chase of Size

- How bigger doesn't always mean better in policymaking
- Substantial representativeness problems
- More "noise" than "meaning"
- Better is not objective, but context-specific

Privacy

- The privacy dilemma: If we speak of data it needs to mirror reality as closely as possible; in order to avoid privacy issues, we need to distort data
- Privacy essentially challenges the assumption on accuracy of big data
- Privacy requires more sophisticated technical and analytical approaches than what we have now
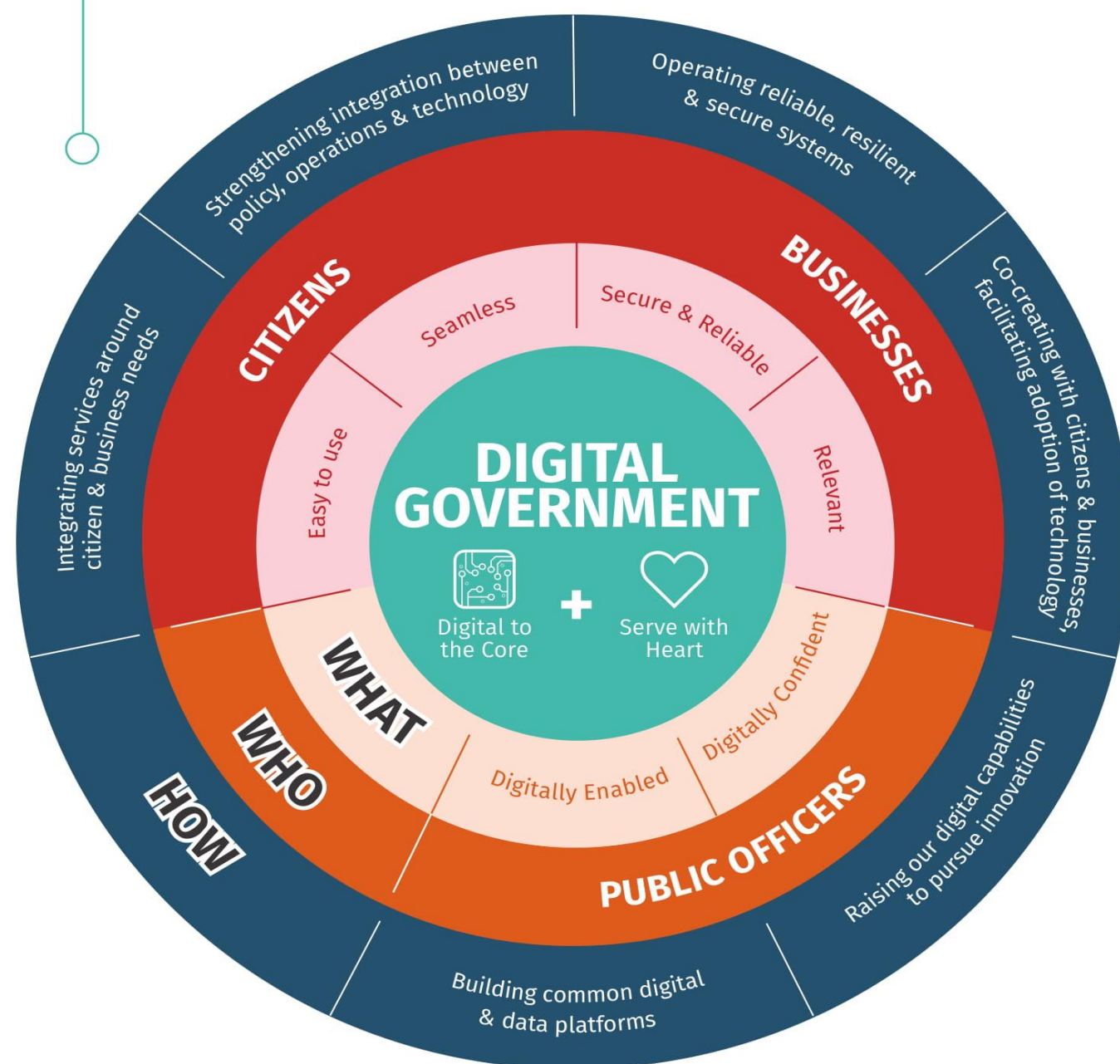
To Do List

1.

2.

3.

A Framework

DIGITAL GOVERNMENT BLUEPRINT

Strengthening integration between policy, operations & technology

Operating reliable, resilient & secure systems

Co-creating with citizens & businesses, facilitating adoption of technology

Raising our digital capabilities to pursue innovation

Building common digital & data platforms

Integrating services around citizen & business needs

CITIZENS

BUSINESSES

PUBLIC OFFICERS

Seamless

Secure & Reliable

Easy to use

Relevant

Digitally Enabled

Digitally Confident

WHAT

WHO

HOW

DIGITAL GOVERNMENT

Digital to the Core + Serve with Heart

Solutions

- Investing sufficient resources
- Gaining support from top management
- Developing systematic data-centric strategies
- Offering training and transition plans
- Accumulating a diverse, qualified workforce
- Embedding data in work practices
- Fostering a changing, innovative culture
- Holding reasonable risk tolerance

Case

311 on-demand services

- The city of Houston is ranked the 4th in the list of top 10 largest metro areas in the U.S. with a population of 7 million people, and the 30th largest economy in the world with annual gross metropolitan product (GMP) of 500 billion USD
- In February 1997, The U.S. Federal Communication Commission (PCC) created the 311 number for non-emergency police and government service with the goal of relieving congestion on the emergency number 911
- Many cities, including Houston, since have adopted 311 as a way of centralizing public service issues for city government to streamline processes and more effectively respond to their citizens' service and information needs

Case

311 on-demand services

- Sufficient resource investment: The city has significantly invested in enterprise systems, cloud computing, big data analytics and IT workforce, supporting transition from a call center to 360 self-service channels allowing citizen engagement
- Data-driven strategic vision: Mayor Annise Parker outlined strategic vision of "data-driven government" in 2010 and created the Performance Improvement Division and Portal in 2011 for continuous improvement of public services
- Integration of data and work: Houston 311 localizes big data analytics use at the department level, making leaders and departments more responsive to the need for citizens' needs and some department to achieve new levels of operational flexibility and efficiency

Case

311 on-demand services

- Big data analytics empowers the government for self-organization: monitoring emerging trends across 2 million calls and to gain new insights into the potential cost-saving efficiency for 22 departments
- Big data analytics enables effective service process: 1300 → 450 FTEs for 1.4 → 2.2 million citizens; quick decisions on effective solutions on the on-demand services delivery
- Big data analytics also moves citizens from physical channels to digital channels: 311 hotline is more costly than website or mobile app in digitization and accuracy of data collection → annual cost savings of 500,000 USD