Use of Big Data sources in monitoring of socio-economic processes:

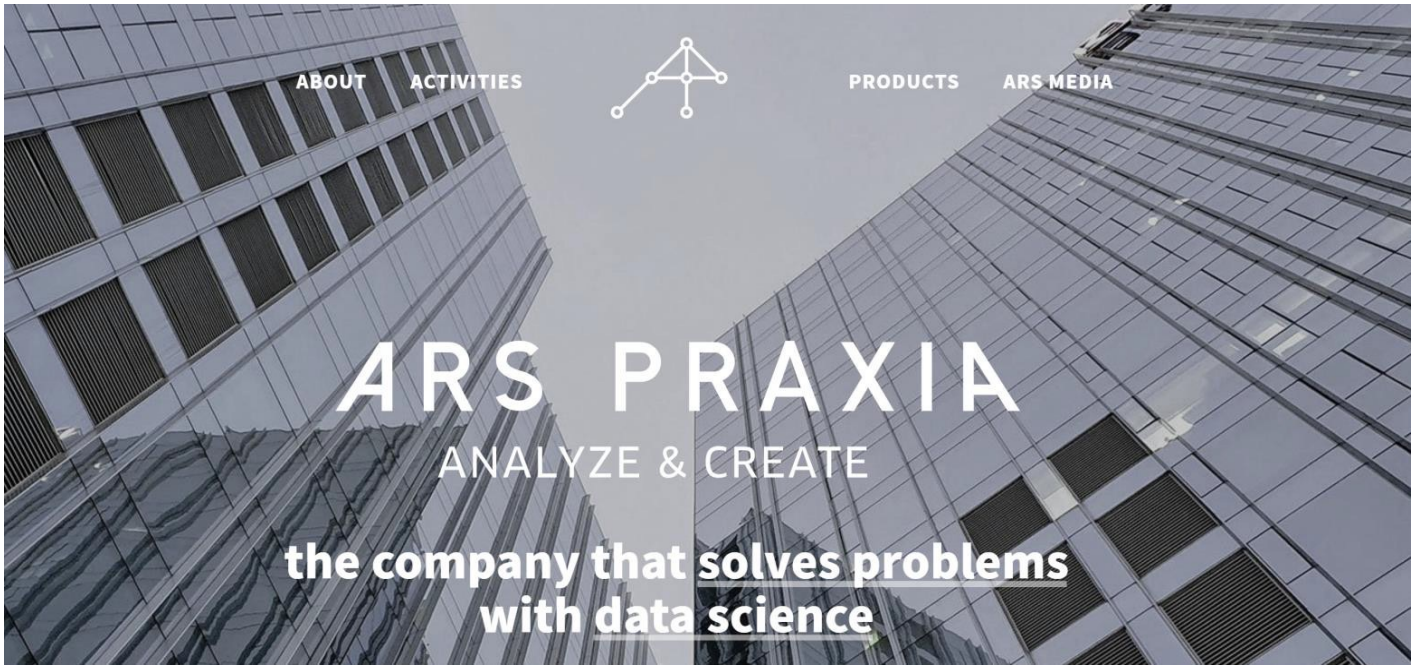
Case from Korea

Leo Dhohoon Kim

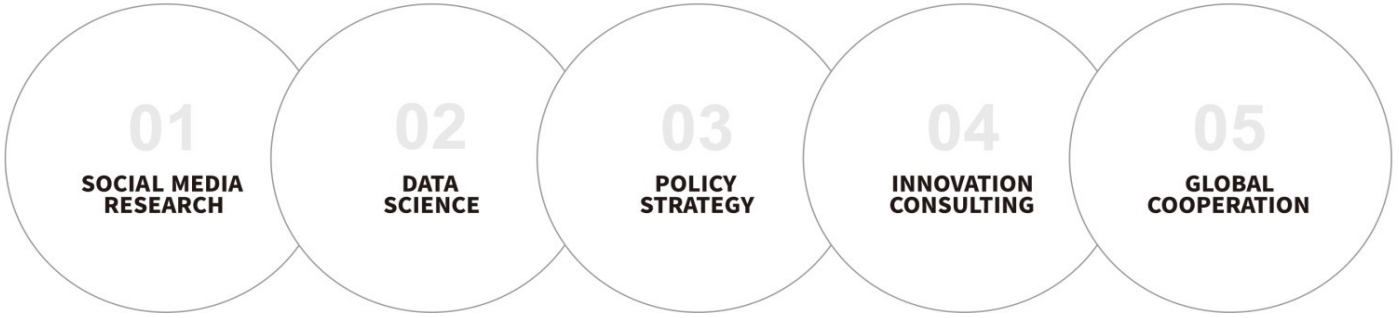
CEO of Ars Praxia

INTRO

Introduction



Ars Praxia (formerly Treum) is a company that specializes in solving problems by using data science. This type of work used to be called “consulting”. However, connecting data analysis and social context requires a new level of professional ability. Ars Praxia has accumulated unique expertise by assuming tasks that are difficult to solve with existing strategic or design consulting approaches. Our vision is to “practice” (praxia) “intelligence” (Ars) by critically integrating knowledge of humanities and engineering (mainly computer science), and to utilize new methods of innovation consulting.



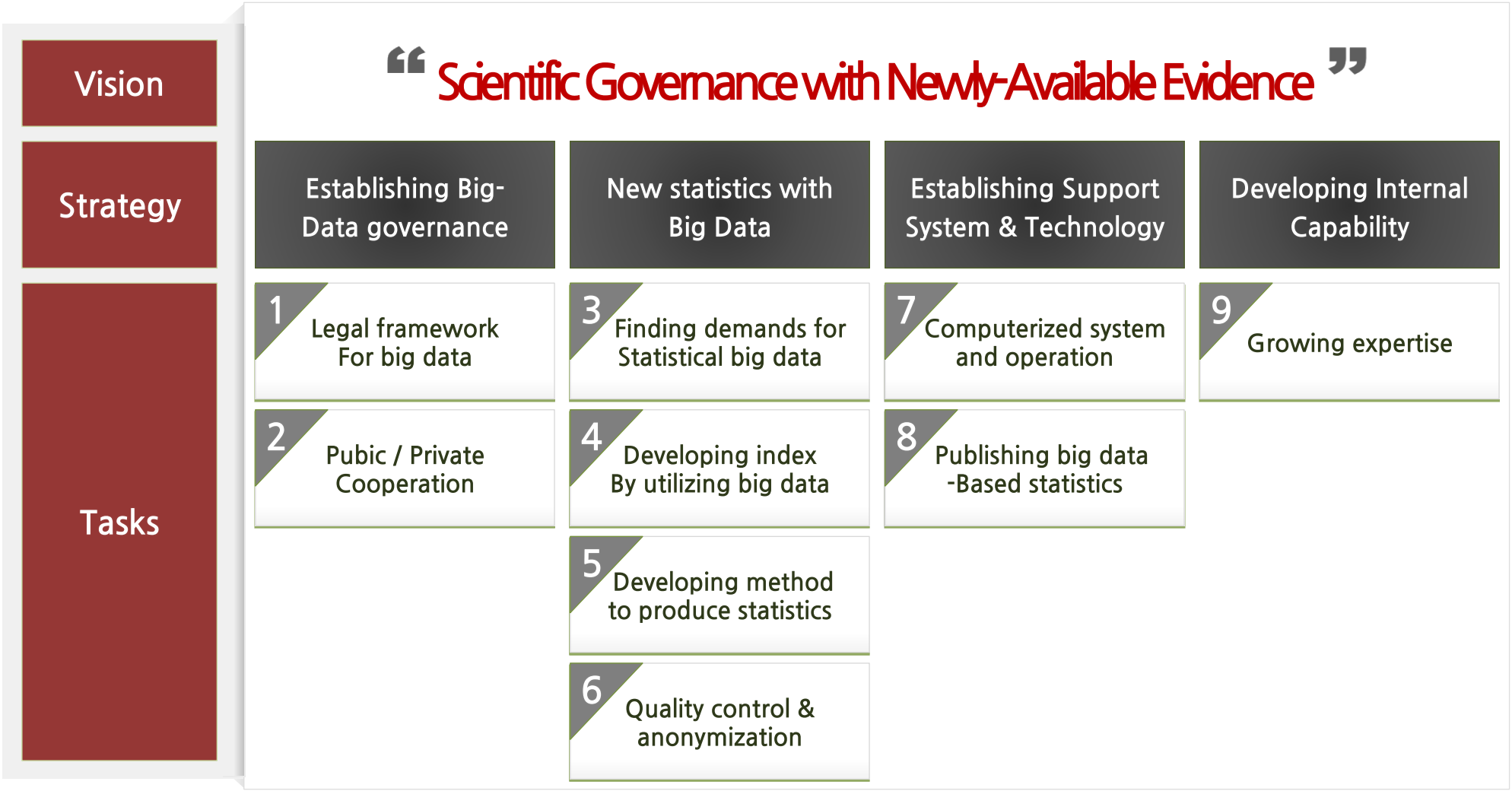
Introduction

- **In 2014, Ars Praxia proposed 29 big data candidates for official statistics**
- **They include price, financial status, sentiment of actors and mobility of population**

Category	Sector	No	Proposed index	Content
Price	On-line	3	On-line daily price index	Measurement of on-line price of key category goods
	POS	4	POS price index	Measurement of transaction price in retail stores
	Necessity	5	T-price index	Measurement of food price in retail stores
Finance	Loan	6	Household loan trend	Measurement of household security and loan
	Overdue	7	Household overdue payment	Measurement of overdue security and loan
		8	Credit card overdue payment	Measurement of overdue credit card debt
		9	Health insurance overdue	Measurement of overdue health insurance payment
		10	Household electricity overdue	Measurement of overdue household electricity payment
Sentiment	Tourism	18	Tourism leading indicator	Measurement of tourism sentiment by internet search
	Employment	19	Employment composite indicator	Measurement of employment trend by internet search
Population	Mobility	20	Mobility of population	Tracking mobility of target population thru telecom. data

Introduction

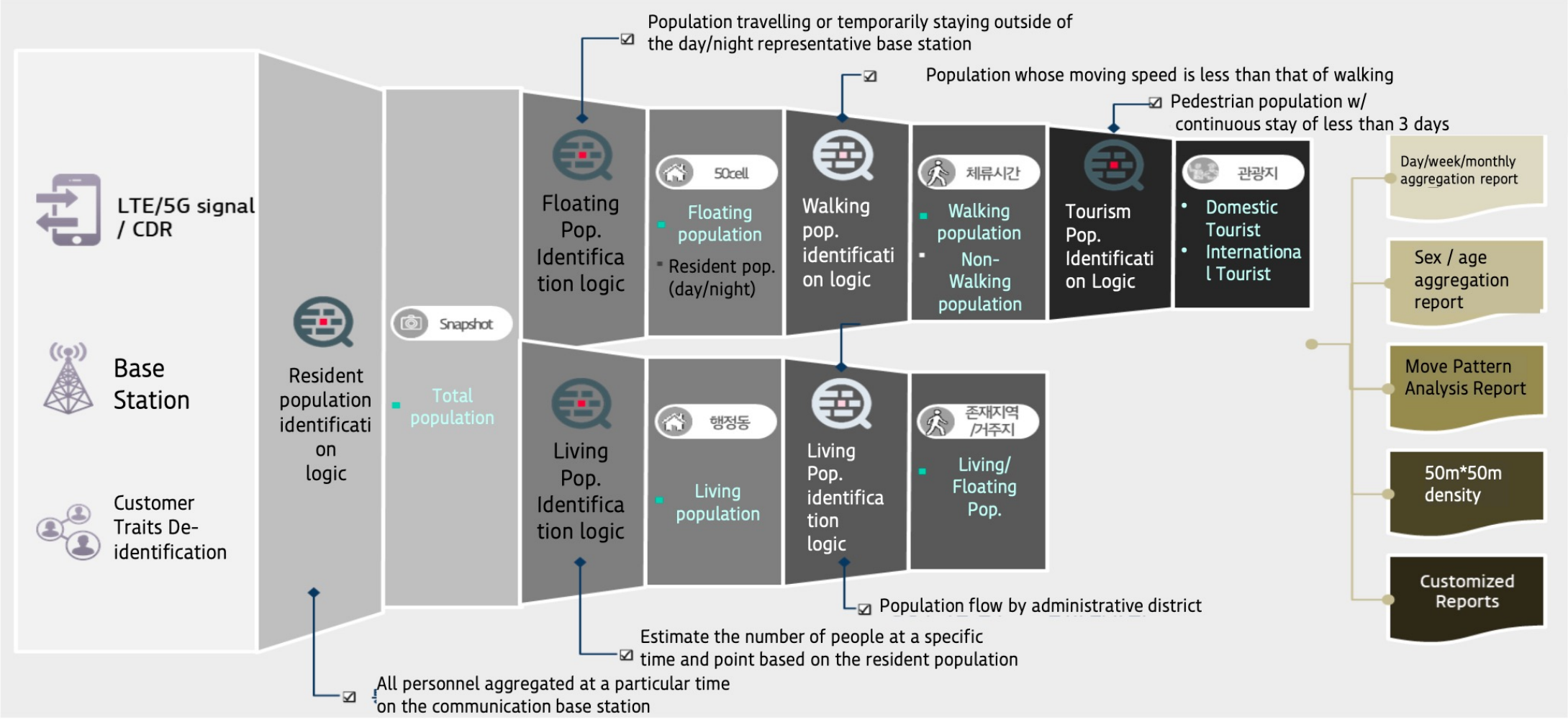
- **Since 2014, Korean government has been dedicated to evidence-based policy**
- **Big data played an integral part of decision-making as “experimental statistics”**



Developed practices

Developed practices

Based on de-identification, step-by-step logic is applied to process data to prepare for sales



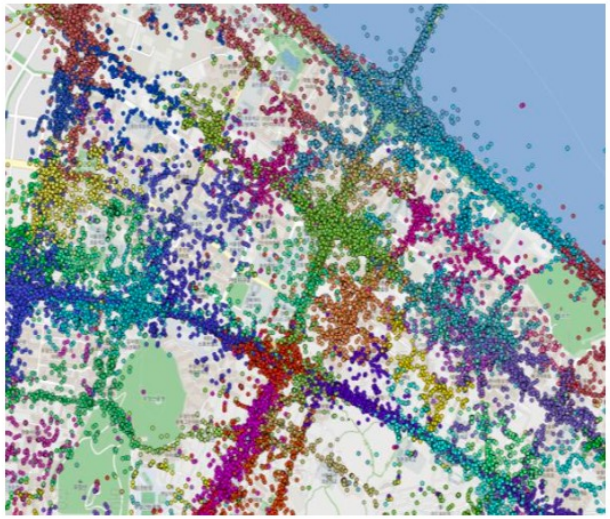
Source: Korea Telecom (2023)

Developed practices

Step1) Data collection/preprocessing

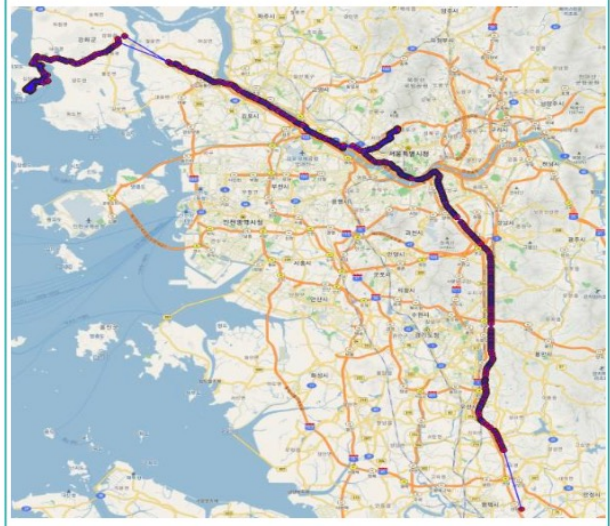
1. Base station location-based signal collection
2. Location measurement based on signal strength, adjacent base station, quality
3. Adding user attributes, primary location measurement data is generated

Step2) Usage of strategy App. and GPS data



1. Adjusting location estimation with the aid of GPS
2. Final adjustment of primary location measurement data using GPS data collected and preprocessed from KT strategy APP(WhoWho) and wireless quality analysis system(WiNG)

Step3) Usage of MDT GPS

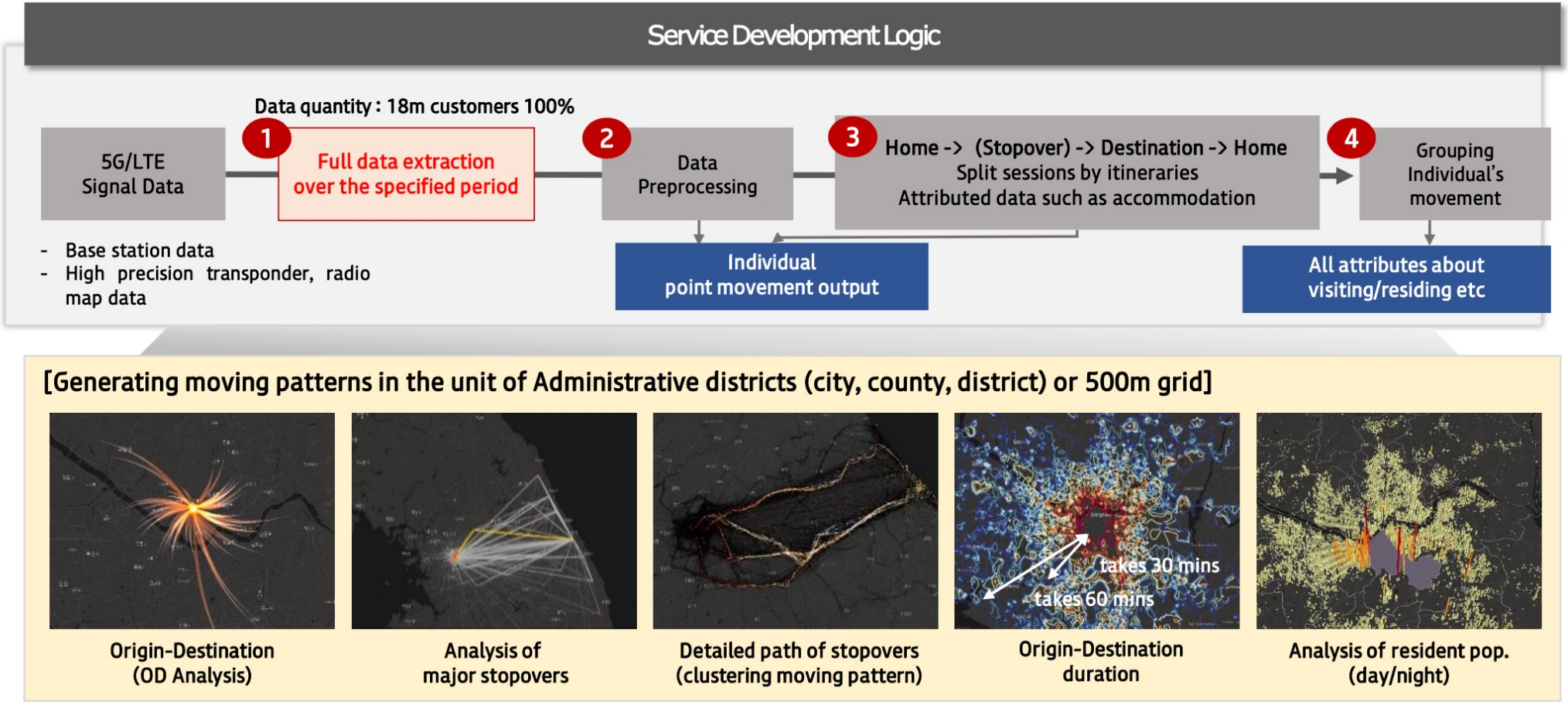


1. Positioning actual location using GPS and base station as primary means
2. Combining MDT GPS location estimation similar to the actual movement
   - MDT GPS : high quality GNSS data of the mobile terminal

Source: Korea Telecom (2023)

Developed practices

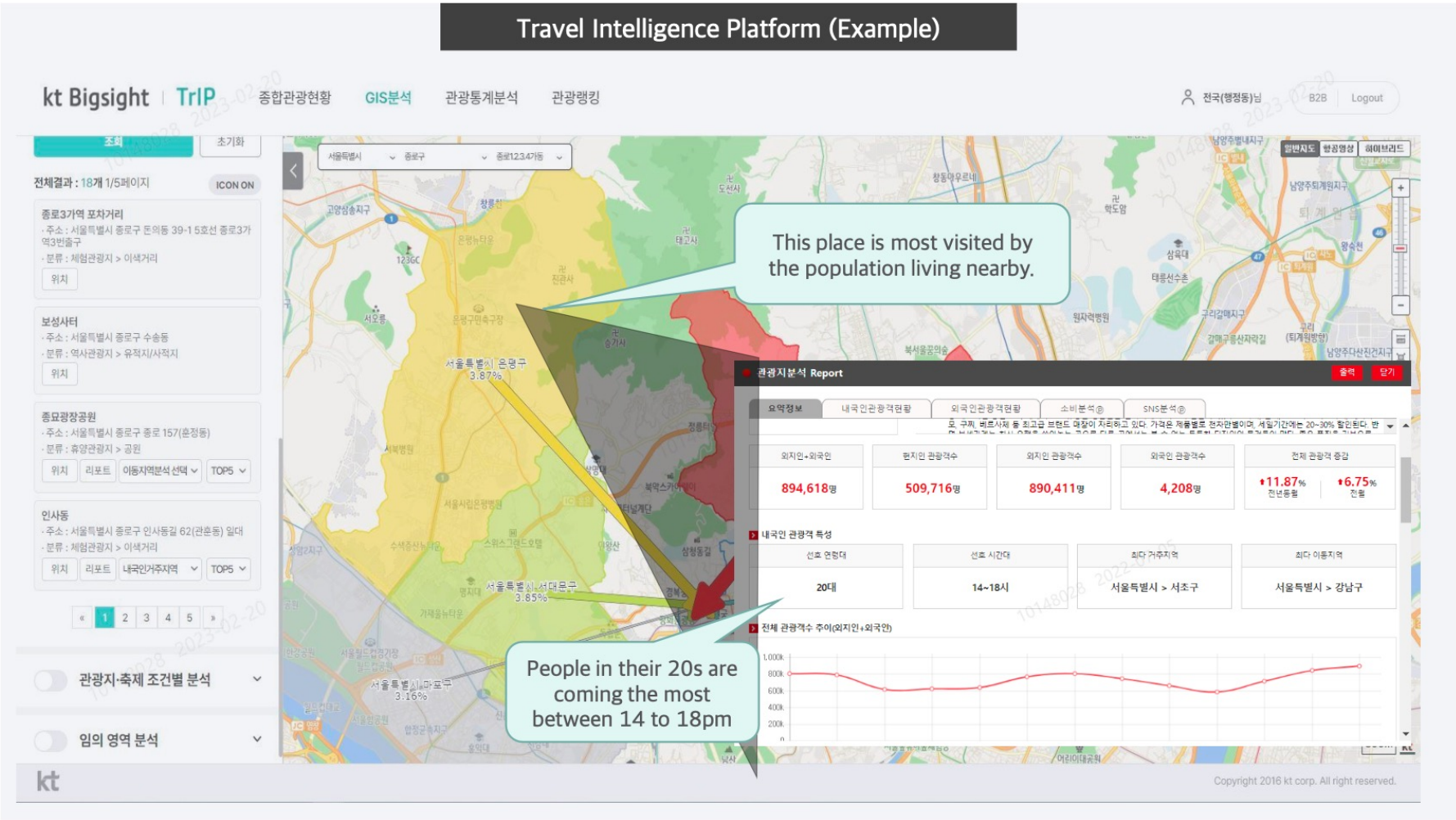
We visualize the overall and detailed movement patterns of the population and individual clusters(15mins dashboard) utilizing the massive point data of 18m customers, thus leading users to meaningful insights



Source: Korea Telecom (2023)

Developed practices

To establish effective policies for tourists visiting the sights and festivals and accurately predict the demand for tourism, we provide tourist analysis tools and reports based on our big data.



Source: Korea Telecom (2023)

Developed practices

Historical Data-Based Analytics Service(2017~) (One-Time Analysis)

Big Data

「Movement Pattern Analysis」
「Sales Structure Analysis」
「Life Stage Analysis」
「Social Media Analysis」
etc.

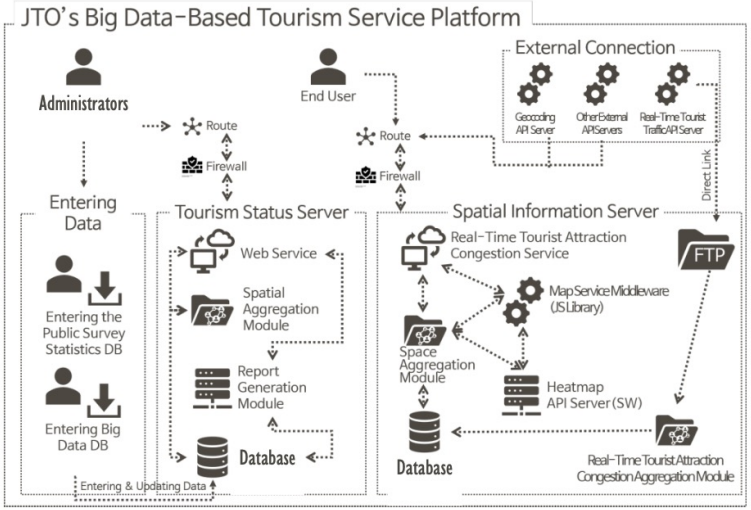
Survey Statistics

「Survey on Jeju Visitors」
「Status of Public Tourist Attractions」
「Casino Permit Status」
etc.

Personalized Services Based on Real-Time Data(2020~) (Continuous Platform Services)

Jeju Tourism Big Data Service Platform

"Real-Time" & "Personalized"



Gain Data Leadership(2021~)

Culture, Sports, and Tourism JEJU Tourism Bigdata Center

"Opening" & "Distribution"

Developed practices

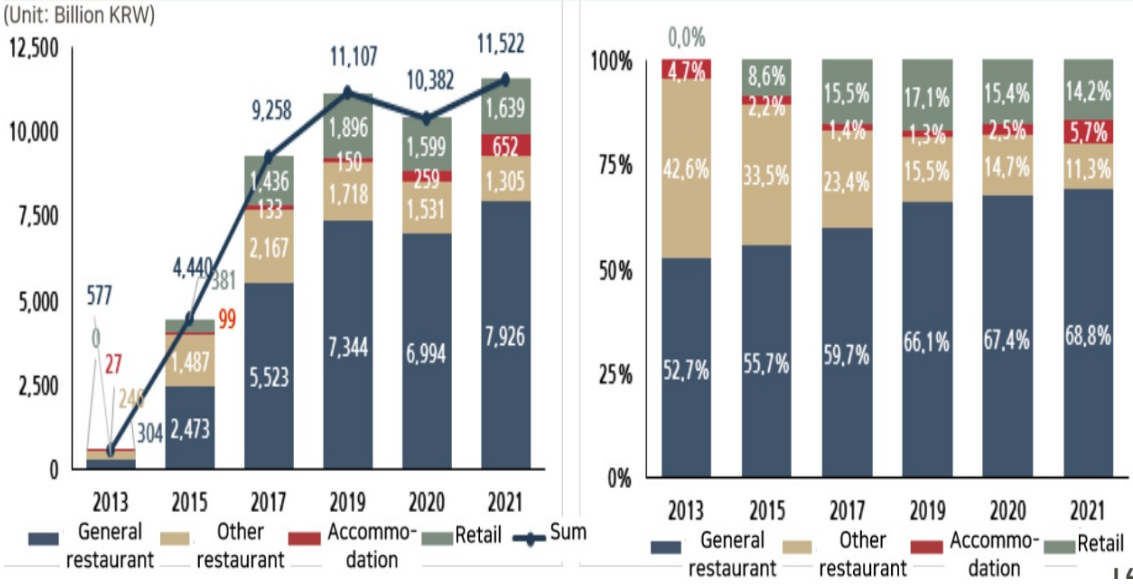
Big Data Center Construction Project

Promote data production and construction by stages before, during, and after high-demand travel

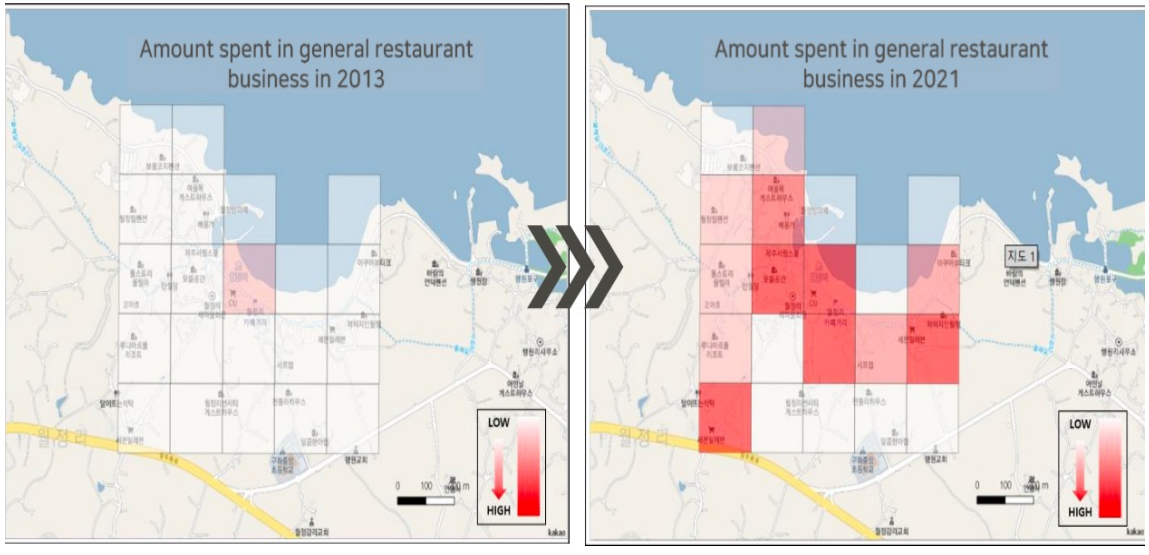
Production and construction of 54 types of data with high demand in the Jeju tourism industry

Before Travel	During Travel: Mov.	During Travel: Consumption	After Travel
◎ VisitJeju Tourism content search ◎ SNS Keyword Data ◎ Rent-a-Car Reservation Data	◎ Kickboard ◎ Rent-a-Car ◎ Trekking	◎ Duty-free Shop ◎ F&B Ranking ◎ Visitors to tourist destinations	◎ Call center data ◎ Review of portal site ratings
	Fusion Data	◎ JTO Data + Accommodation Data	

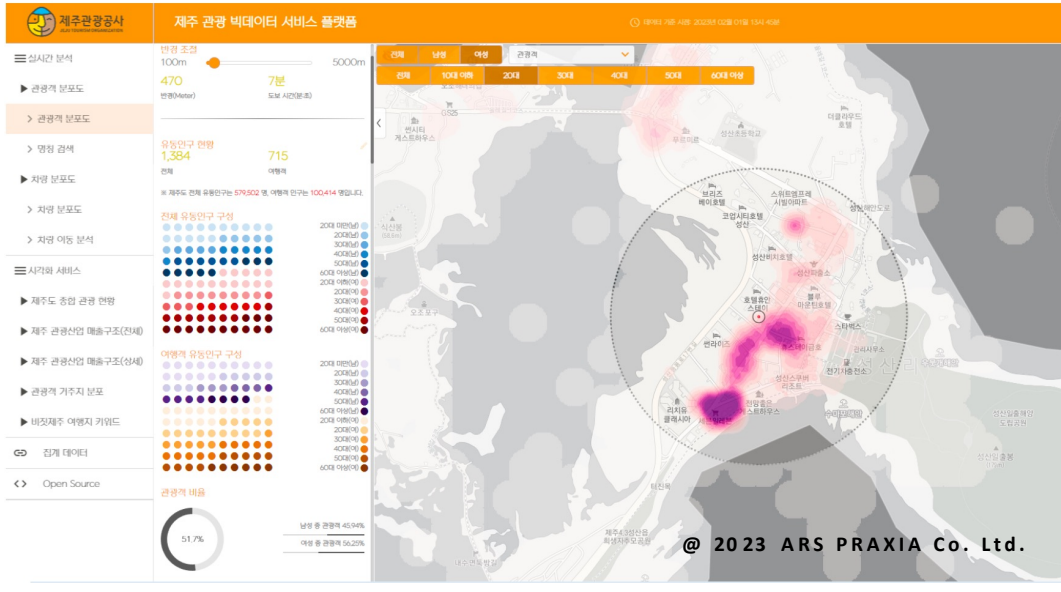
Ex3: Analysis of card consumption scale and ratio by industry



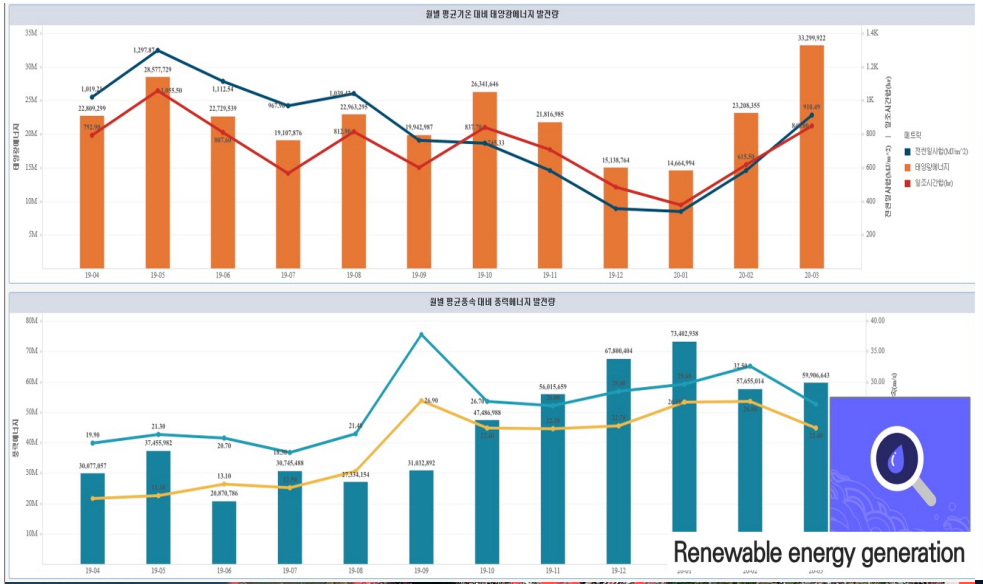
Ex1: Analysis of Card Consumption by Industry by Year

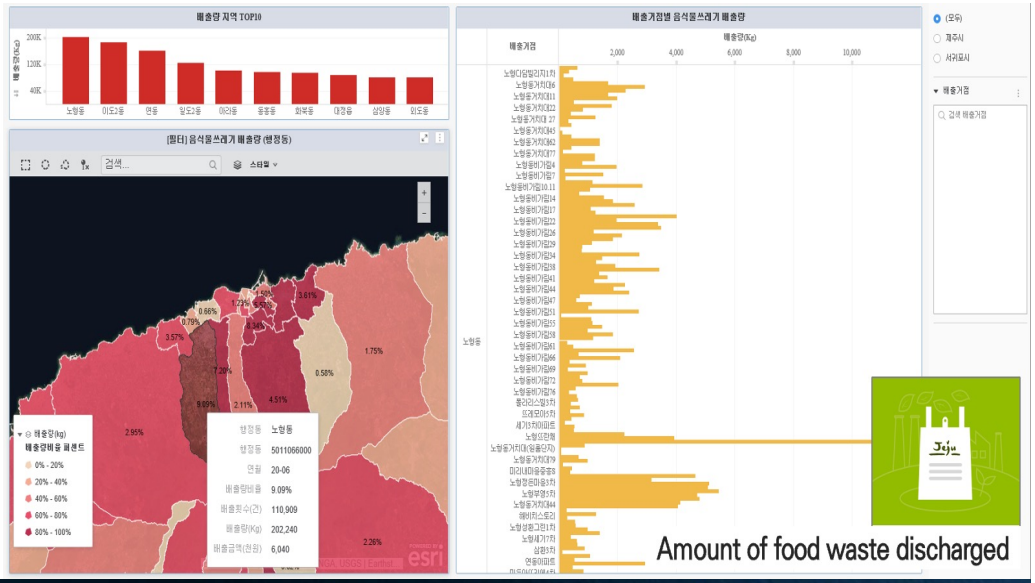


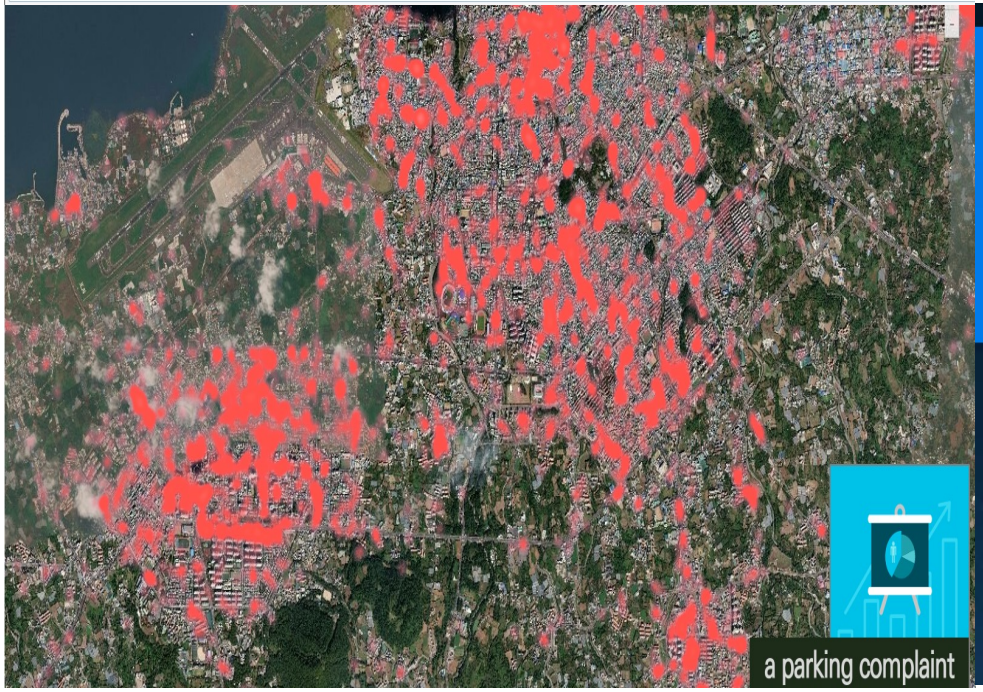
Big Data Service Platform > Real-time analysis > Tourist Distribution Map



Developed practices







Developed practices

Jeju towards Smart Island:
Drone-based Administrative Services Innovation

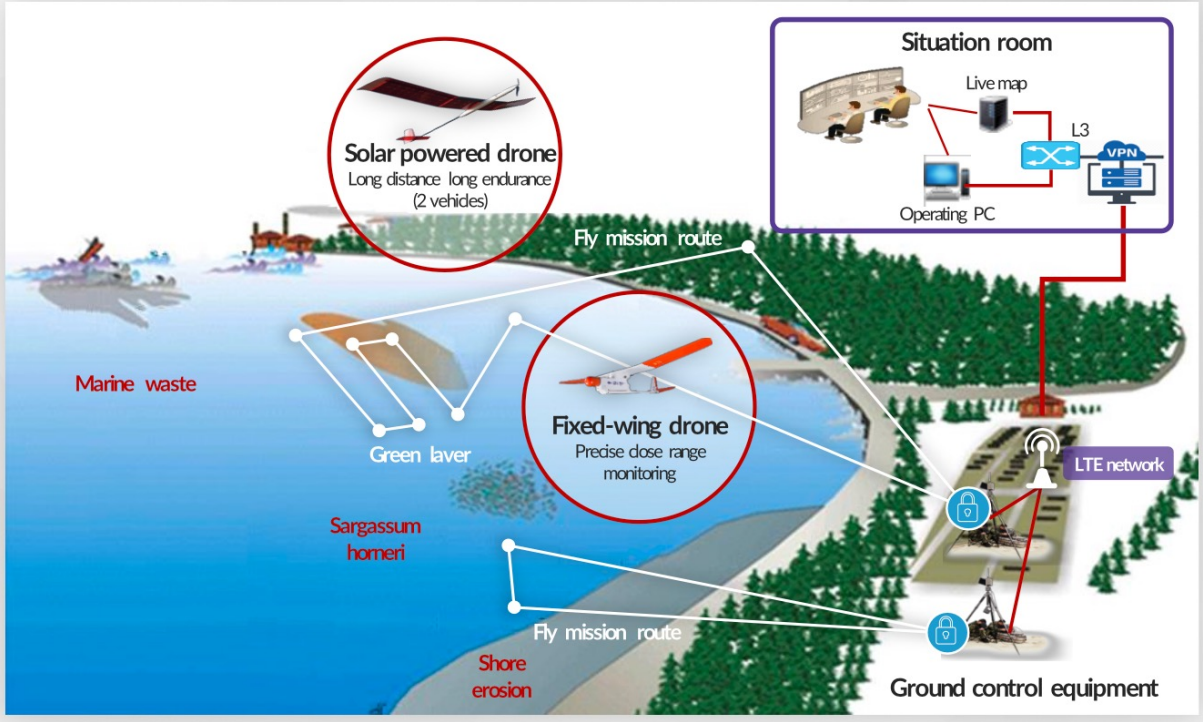
First Subtask
: Key demonstrations in marine environmental monitoring

Alteration of solar-powered and fixed-wing drones
- Secure space and load mission equipment
- Optimize the propulsion system
- Optimize power supply

Realization of remote control and data transmission
- Discussing communication protocols between modules
- Building control and monitoring software
- Implementing real-time video transmission

Automatic object detection using AI
- Selecting AI object detection algorithms
- Constructing learning data
- Optimizing and deploying services

Real time drone mapping and visualization
- Applying real-time mapping modules
- Applying real-time visualization modules
- Implementing integrated administrative services



Extended analysis

Extended analysis

- By linking people's behavior and speeches, we can understand their underlying desire & motivation
- In doing so, we can forecast people's reaction and movement

Analysis of people's behavior

Act & Speak

"What do they talk about?"

- Main subject of concern
- Key concepts

Think

"What do they have in mind?"

- Connotation and context
- Frame & sentiment

Forecasting

"How people will move?"

- Orientation of movement
- Emerging social phenomenon

Extended analysis

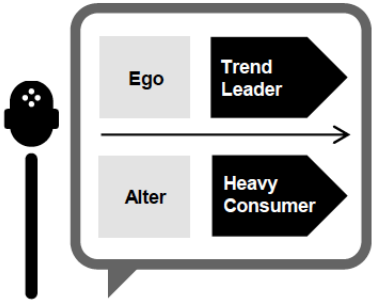
- On-line monitoring can quickly capture weak signal that might proliferate in the near future
- Integration of big data & statistics can develop powerful tool to understand "social physics" of actors

Research approach

On-line Monitoring

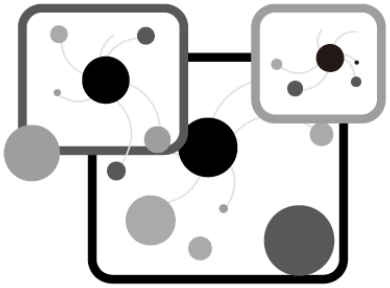
- Social media analytic tools
- Dashboard of aggregate data

Statistical research



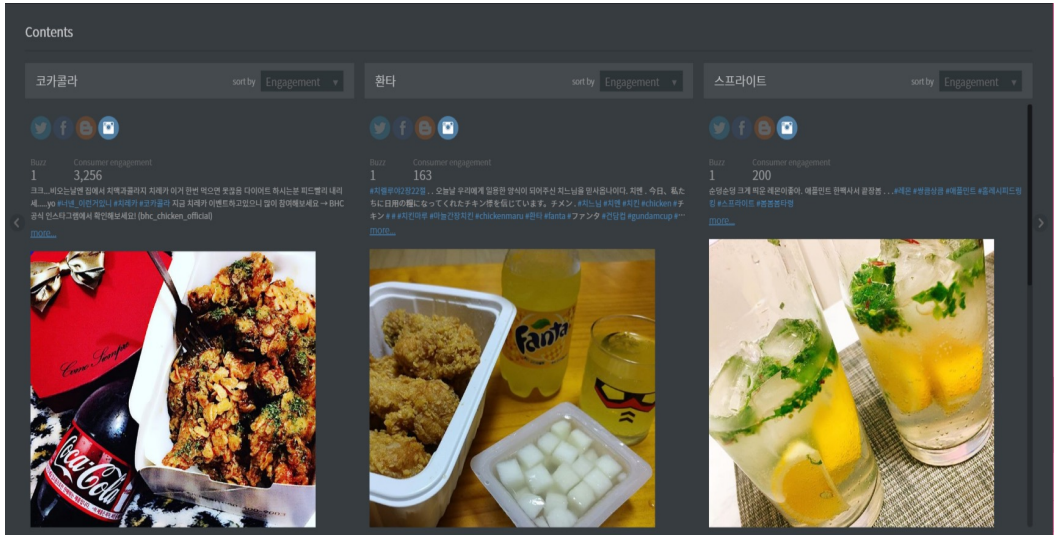
- Demographic characteristics
- Social relations

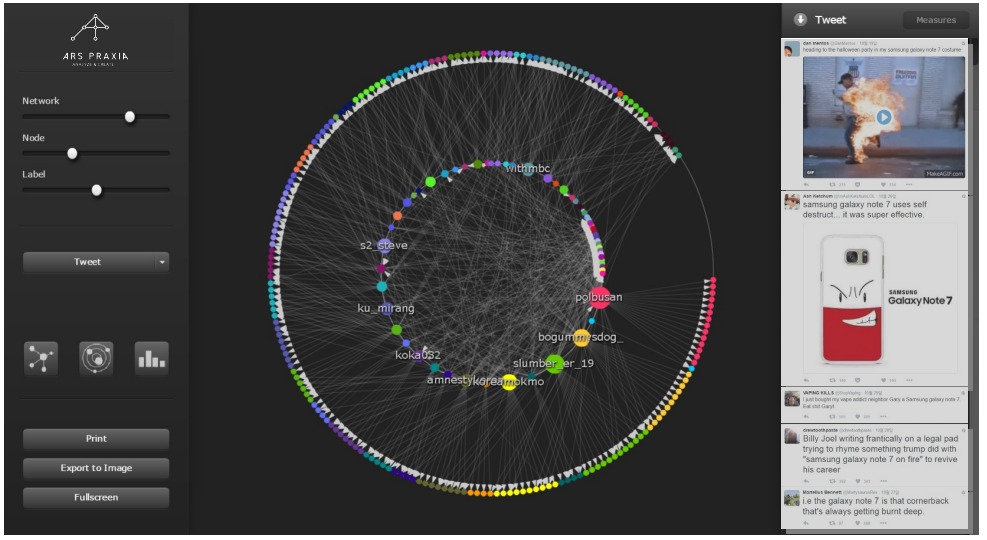
Social physics & context

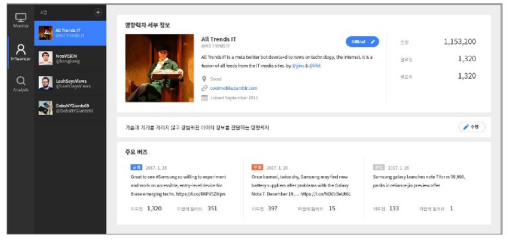


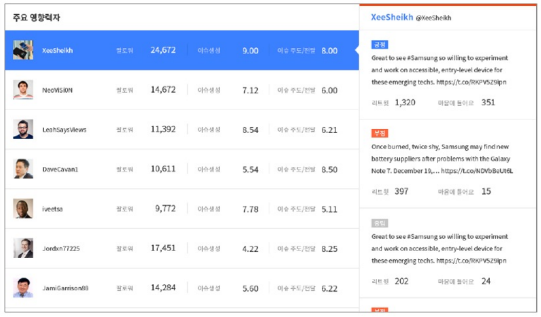
- Evaluation of social movement
- Comprehending underlying context

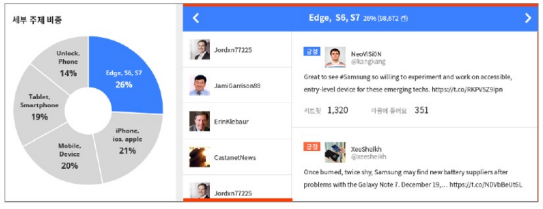
Extended analysis

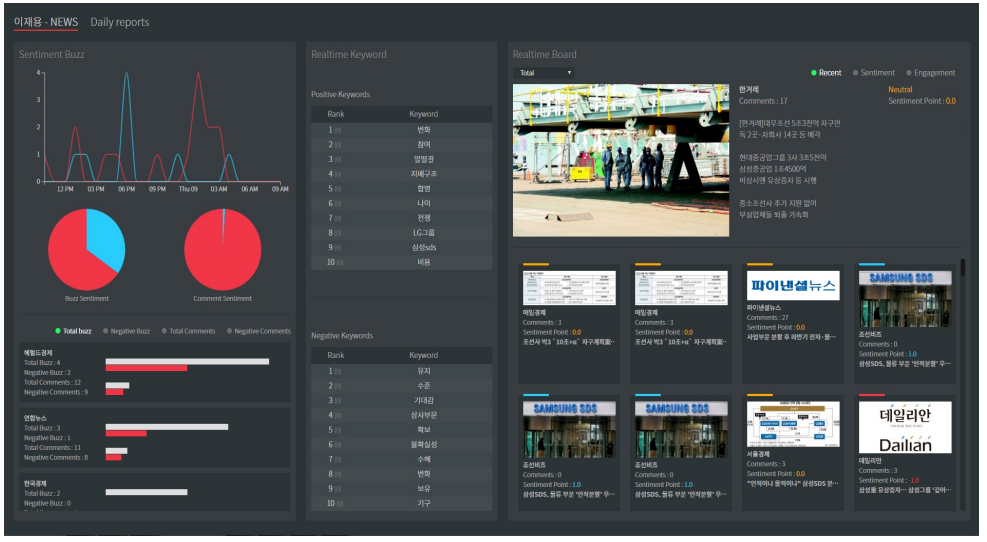






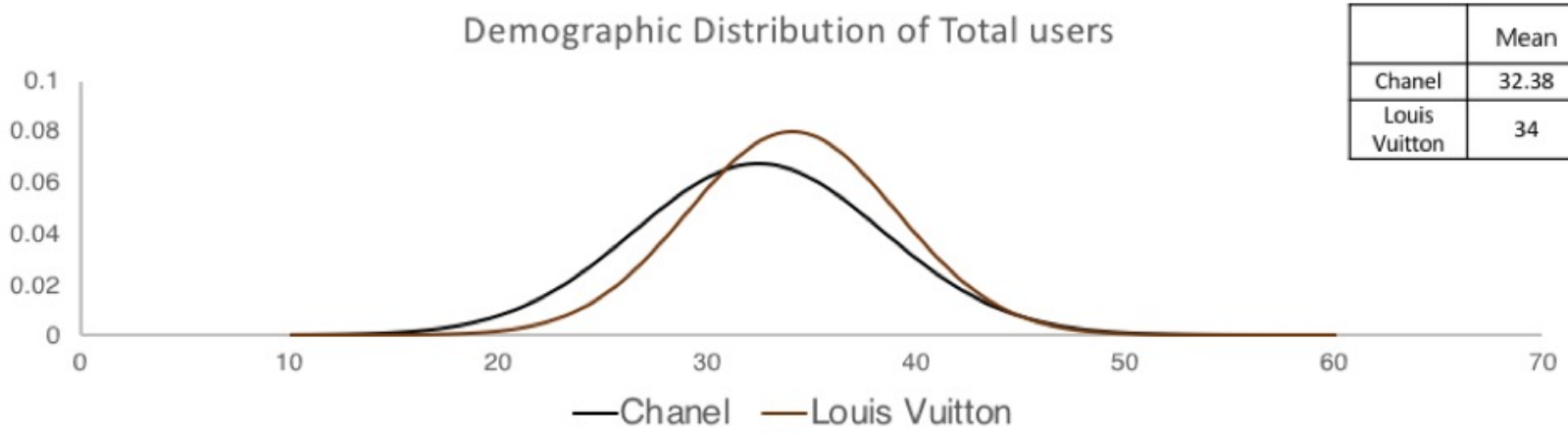




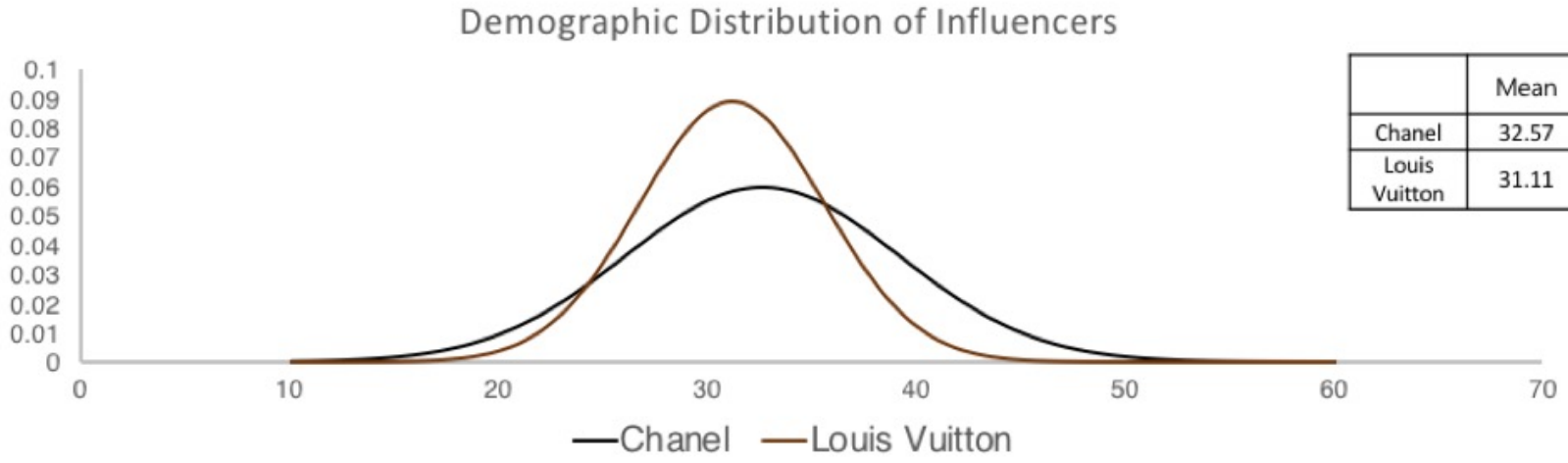


Extended analysis

- **With trained data sets, we can infer the demography of on-line users**
- **In doing so, we can evaluate the consumption pattern & trend of brands**



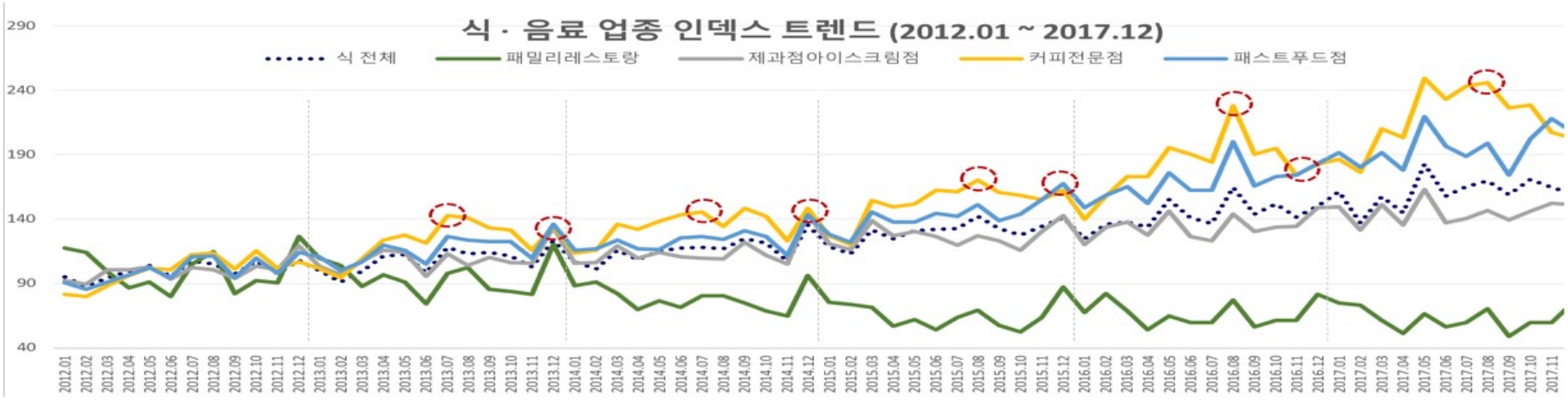
	Mean	Standard deviation
Chanel	32.38	5.95
Louis Vuitton	34	5

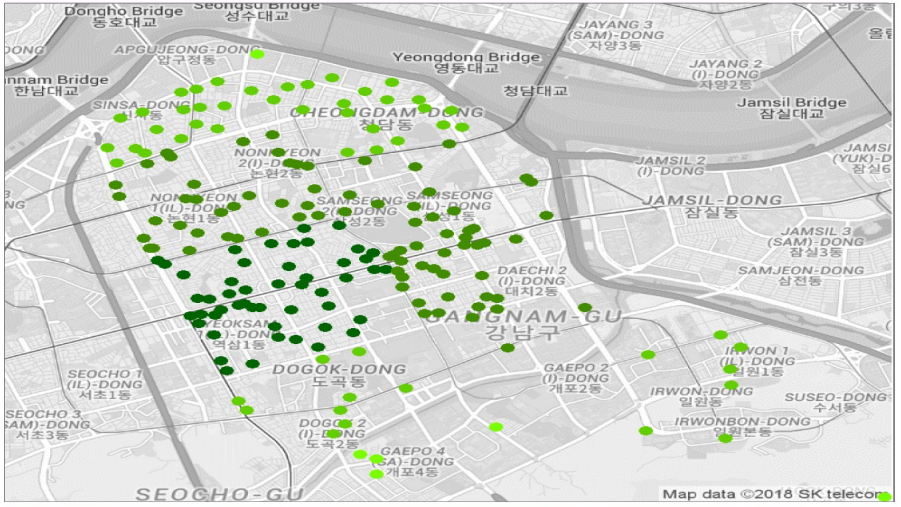


	Mean	Standard deviation
Chanel	32.57	6.57
Louis Vuitton	31.11	4.44

Extended analysis

- **Integration of on-line data, AI (machine learning) and transaction data brings analytic innovation**
- **We are opening an era of ”micro-behavior analysis”**



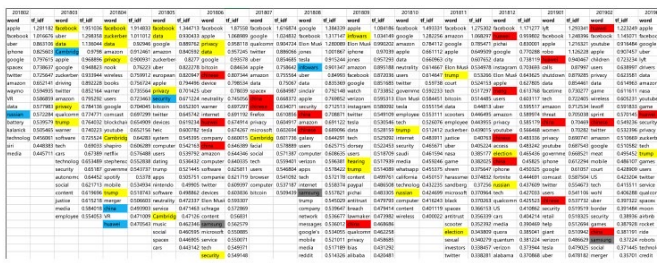


Extended analysis

Could we anticipate US-China trade dispute?

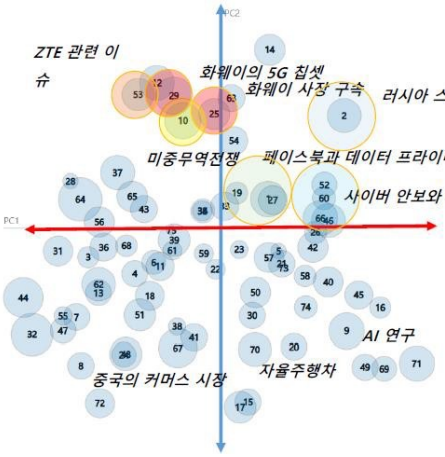
- 미국 매체 분석 결과는 다음과 같음

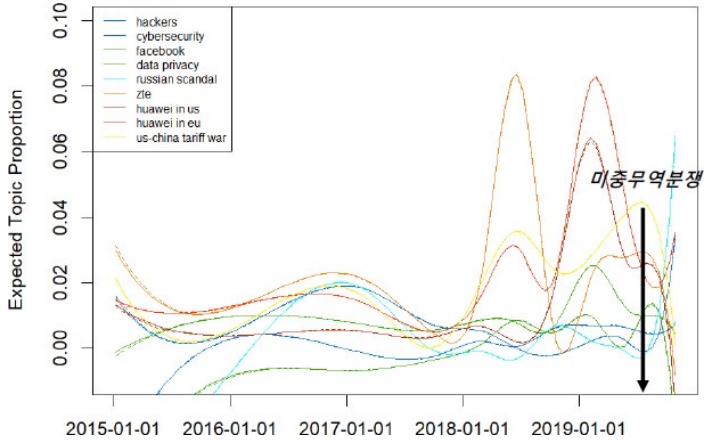
1. 2016.1-2019.9 기간 미국 5개 언론 매체(NYT, WP, WSJ, MIT Tech Review, Wired)의 텍스트데이터에대해 TF-IDF 및 키워드 필터링을 적용
2. 미국 정부의 화웨이 부사장 Meng 체포 시점(2018.12)으로부터 8개월 전(2018.4)부터 China 및 Huawei에 대한 부정적 논조 및 언급 키워드가 상위 리스트로 부상함
3. 2018년 2월에 **Russian** 정보 스캔들이 상위 리스트에 언급되고, 이어 3월에 **Cambridge Analytica**의 페이스북 개인정보 유출에 대한 논란이 이슈화 됨. 5월에 **security**이슈가 처음으로 상위키워드로 부상하고, 6월에 **China** 및 **Huawei**를 국가 안보 이슈와 연관지어 부정적으로 언급하기 시작함. 8월에는 중국과 고조되는 갈등을 **infowars**에서 국가안보 관련 음모이론으로 유통함

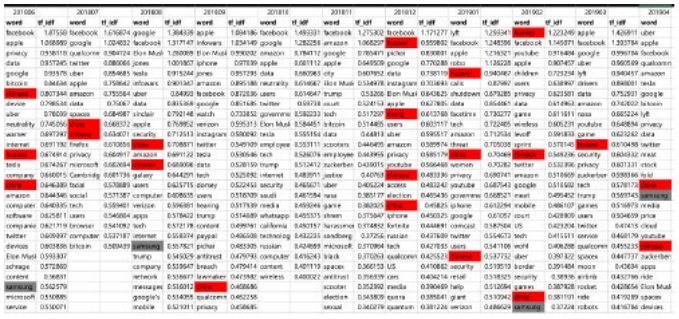


파란색: 최초 언급 키워드 / 노란색: 이슈 연관 키워드 / 빨간색: 중국(화웨이) 관련 키워드 / 회색: 한국 관련 키워드

* 자세한 내용은 별도 자료(엑셀)로 첨부



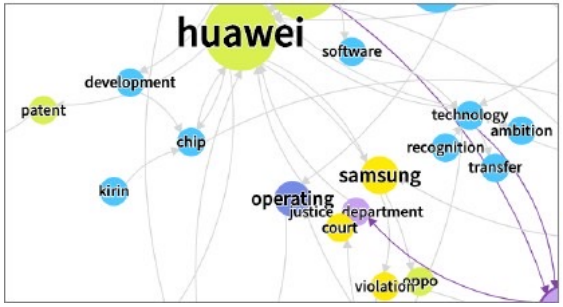




미국 뉴스 매체

중국 및 한국 기업(삼성) 언급 추세

적색: 화웨이
회색: 삼성



중국 뉴스 매체

화웨이 및 삼성의 연결 관계:

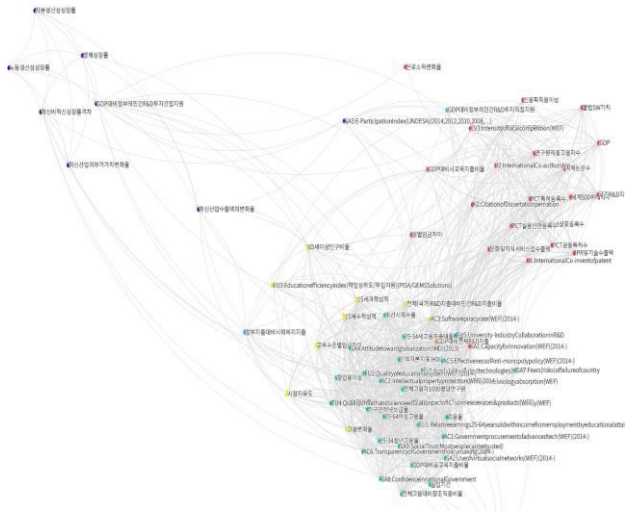
1) 기술적 경쟁자
2) 사법적 견제 대상

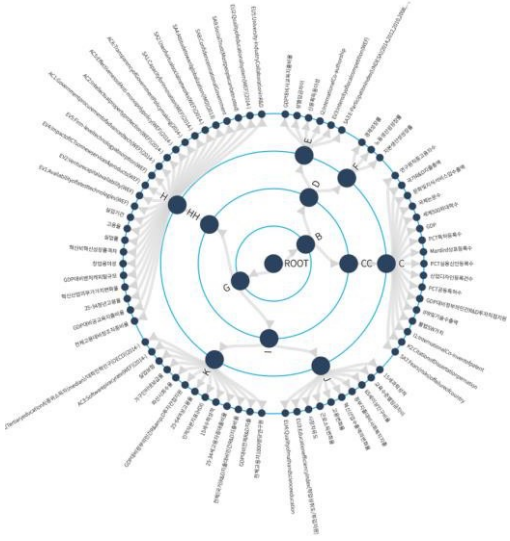
Extended analysis

National innovation index & diagnosis

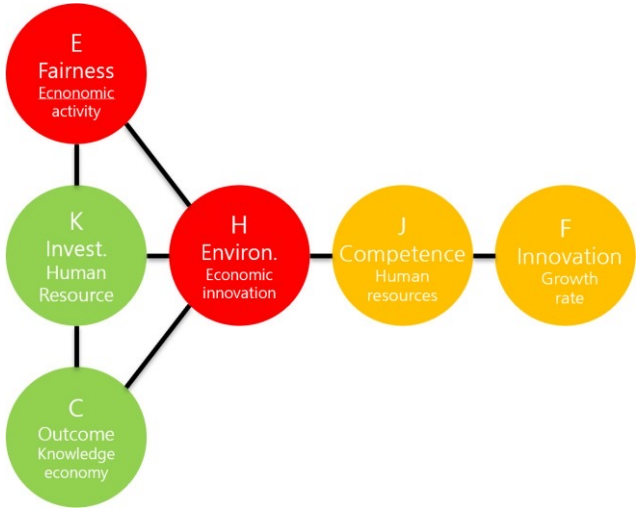
I1 : International Co-invent of patent
I2 : International Co-authorship
K2 : Citation of Dissertation per nation
EV1 : Availability of latest technologies (WEF)
EV2 : Venture capital availability (WEF)
EV3 : Intensity of local competition (WEF)
EV4 : Impact of ICTs on new services & products
EV5 : Firm-level technology absorption (WEF)
AC1 : Government procurement of advanced tech
AC2 : Intellectual property protection (WEF) (2014
AC3 : Software piracy rate (WEF) (2014-)
AC4 : ICT use & gov't efficiency (WEF) (2014-)
AC5 : Effectiveness of Anti-monopoly policy (WEF
AC6 : Transparency of Government Policymaking
SA1 : Capacity for innovation (WEF) (2014-)
SA2 : Use of virtual social networks (WEF) (2014-)

I1 : International Co-invent of patent
I2 : International Co-authorship
K2 : Citation of Dissertation per nation
EV1 : Availability of latest technologies (WEF)
EV2 : Venture capital availability (WEF)
EV3 : Intensity of local competition (WEF)
EV4 : Impact of ICTs on new services & products
EV5 : Firm-level technology absorption (WEF)
AC1 : Government procurement of advanced tech
AC2 : Intellectual property protection (WEF) (2014
AC3 : Software piracy rate (WEF) (2014-)
AC4 : ICT use & gov't efficiency (WEF) (2014-)
AC5 : Effectiveness of Anti-monopoly policy (WEF
AC6 : Transparency of Government Policymaking
SA1 : Capacity for innovation (WEF) (2014-)
SA2 : Use of virtual social networks (WEF) (2014-)
SA3 : E-Participation Index (UNDESA) (2014,2012
SA4 : Attitude toward globalization (IMD) (2013)



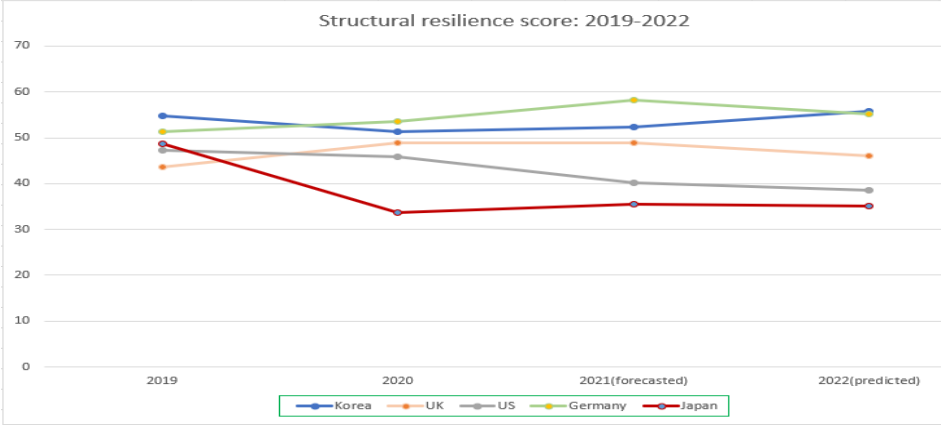


1	United States	1.070
2	Germany	0.868
3	Sweden	0.851
4	United Kingdom	0.849
5	Norway	0.828
6	Japan	0.815
7	Switzerland	0.815
8	Netherlands	0.784
9	Denmark	0.773
10	Luxembourg	0.769
11	Finland	0.763
12	Iceland	0.751
13	Belgium	0.724
14	France	0.709
15	Austria	0.702
16	Canada	0.691
17	Israel	0.688
18	Korea	0.666
19	Australia	0.665



Measuring systemic resilience & developing future strategy

	Korea	UK	US	Germany	Japan
2019	54.76	43.58	47.22	51.35	48.64
2020	51.27	48.83	45.77	53.47	33.61
2021(forecasted)	52.38	48.83	40.12	58.26	35.46
2022(predicted)	55.87	46.05	38.52	55.25	35.09



	주요 변인	세부 데이터	Korea	UK	US	Germany	Japan
정치	정치적 통제력(지지율) 7.5	Approval rating	11.51	11.18	7.70	10.50	1.81
	전자정부 역량(ICT governance) 7.5	EGDI (E-government devlopment index)					
외교	글로벌 리더십 7.5	Governance	4.70	8.18	2.15	13.13	5.29
	글로벌 소프트파워 7.5	Soft power index					
경제	무역대응(무역의존도 X 무역감율) - 3	Trade 1.5	1.09	0.75	0.17	1.19	0.55
		Trade dependency 1.5					
	실업대응 - 3 (Benefits in unemployment/실업률)	Benefits in unemployment 1.5	1.96	0.67	1.71	1.05	1.46
		Unemployment rate 1.5					
	폐업대응(시장) - 2	BCI	0.10	0.13	0.06	0.18	0.13
	금융 취약성(역산) - 2	Financial vulnerability	1.53	1.47	2.00	1.23	0.90
	경제 신뢰도 - 2	Economic confidence (FDI)	0.86	0.44	0.20	0.52	0.55
	소비의욕 - 3	Domestic demand 1.5	0.89	0.47	0.88	0.66	0.30
		CCI 1.5					
	경제 공정성 - 3 (Poverty preventionx equiable education x labor market access	poverty prevention 1	1.66	1.29	0.98	1.53	1.59
		Equitable education 1					
		Labor market access 1					
산업	산업투자 (의료 R&D) - 15	Health expenditure 5	11.56	2.75	8.06	6.53	8.44
		investment 5					
		R&D Spending 5					
사회	(의료)사회안전 인프라 - 8	Physicians per 1000 inhabitants 2	7.74	3.78	2.22	4.53	3.27
		Beds per 1000 inhabitants 1					
		Health policy 2					
		household out of pocket expenses 2					
		spending on preventive health programs 1					
	사회적 공정성 - 8	Social inclusion 2	3.02	4.42	2.92	4.70	1.60
		non-discrimination 1					
		Gender wage gap 2					
		Employment rates by gender 2					
		Employment by education level 1					
	교육 혁신 - 4	Government expenditure on education	0.70	1.44	1.50	1.71	1.06
		HDI (Human development index)					
문화	사회적 신뢰(정부+의료+시민+언론) - 10	trust in government 2.5	4.39	5.37	5.39	5.42	4.75
		Community 2.5					
		Health Care index 2.5					
		Media trust 2.5					
	사회참여도(준수+시민참여) - 5	Compliance (마스크착용률x외출자제) 2.5	4.17	3.70	2.57	2.36	3.38
		citizens, participatory competence 2.5					

2022				
Korea	UK	US	Germany	Japan
55.87	46.05	38.52	55.25	35.09

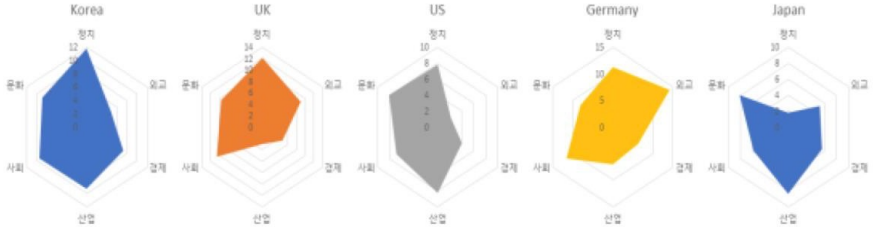
Supeior
Good
Moderate
Underperform
Poor

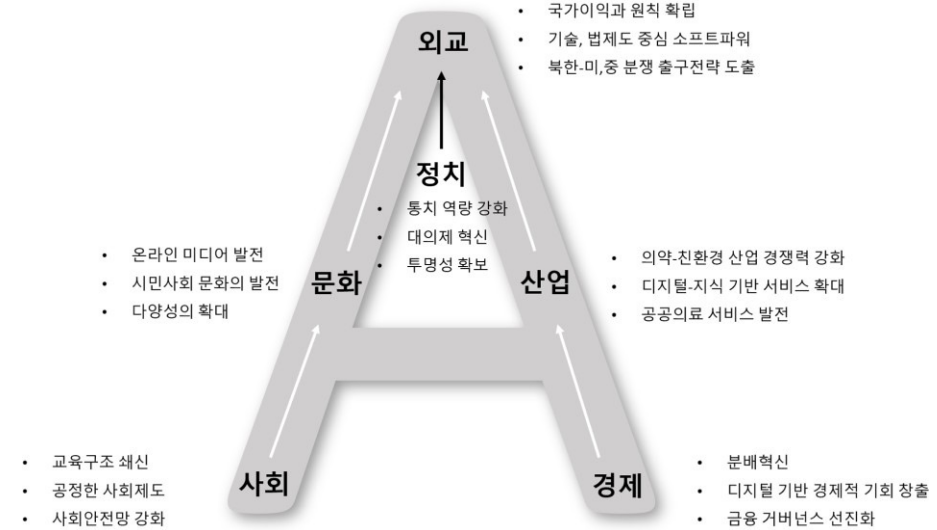
[그림 11] Transformation: 코로나 이후의 시스템 회복탄력성 지표 예측(2022)

종합점수

KOR	GBR	USA	DEU	JPN
52.06686	49.87486	38.50812	59.70563	35.2817

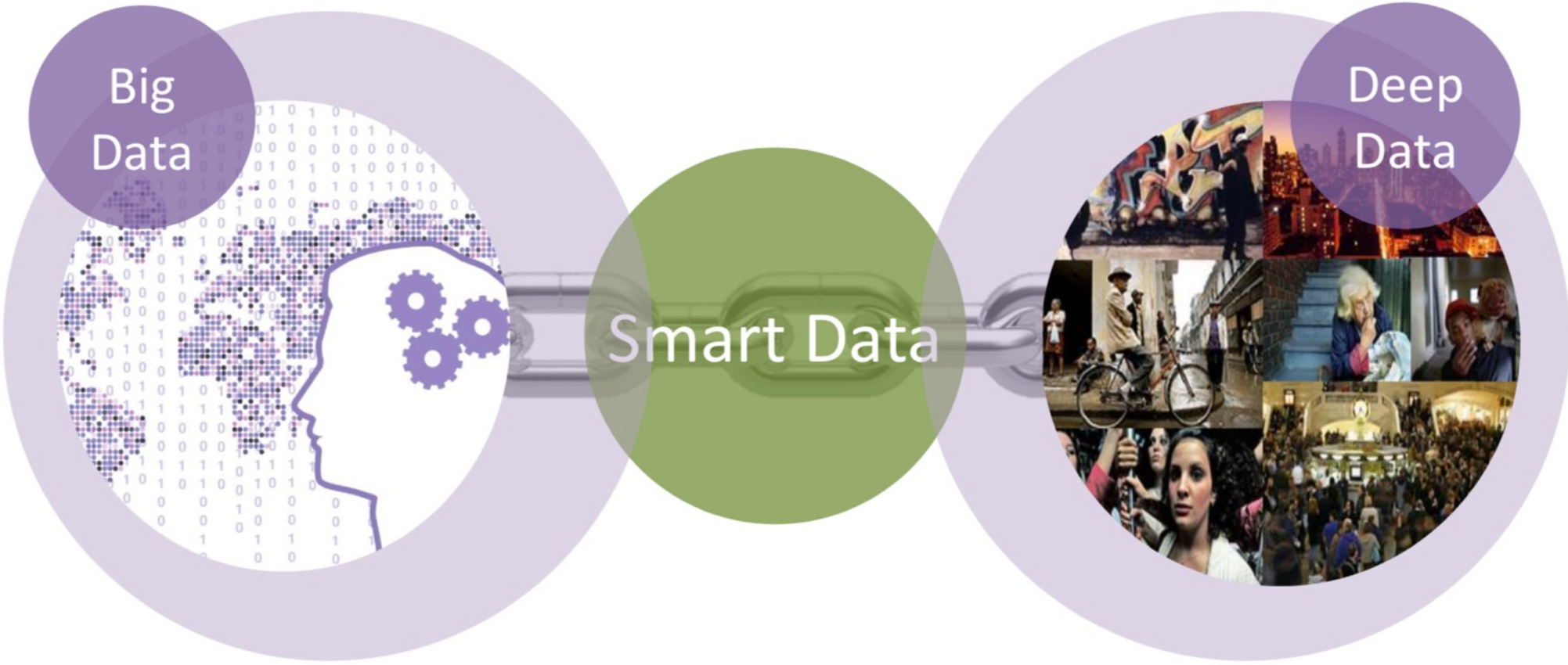
영역별 점수





Intellectual concerns

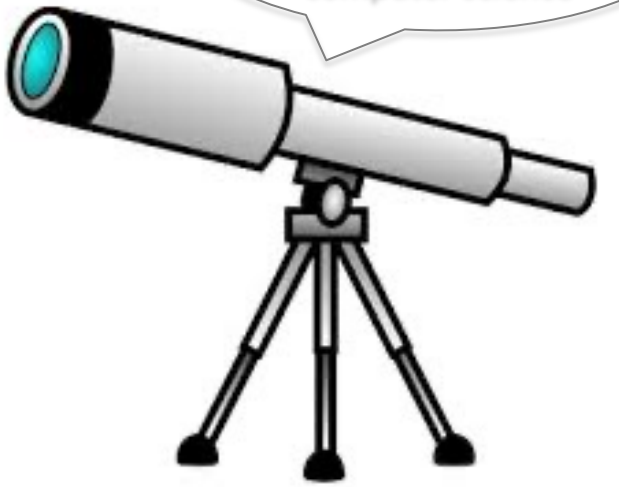
Intellectual concerns



Intellectual concerns

Big Data

Macroscopic analysis of data through statistics and computer science

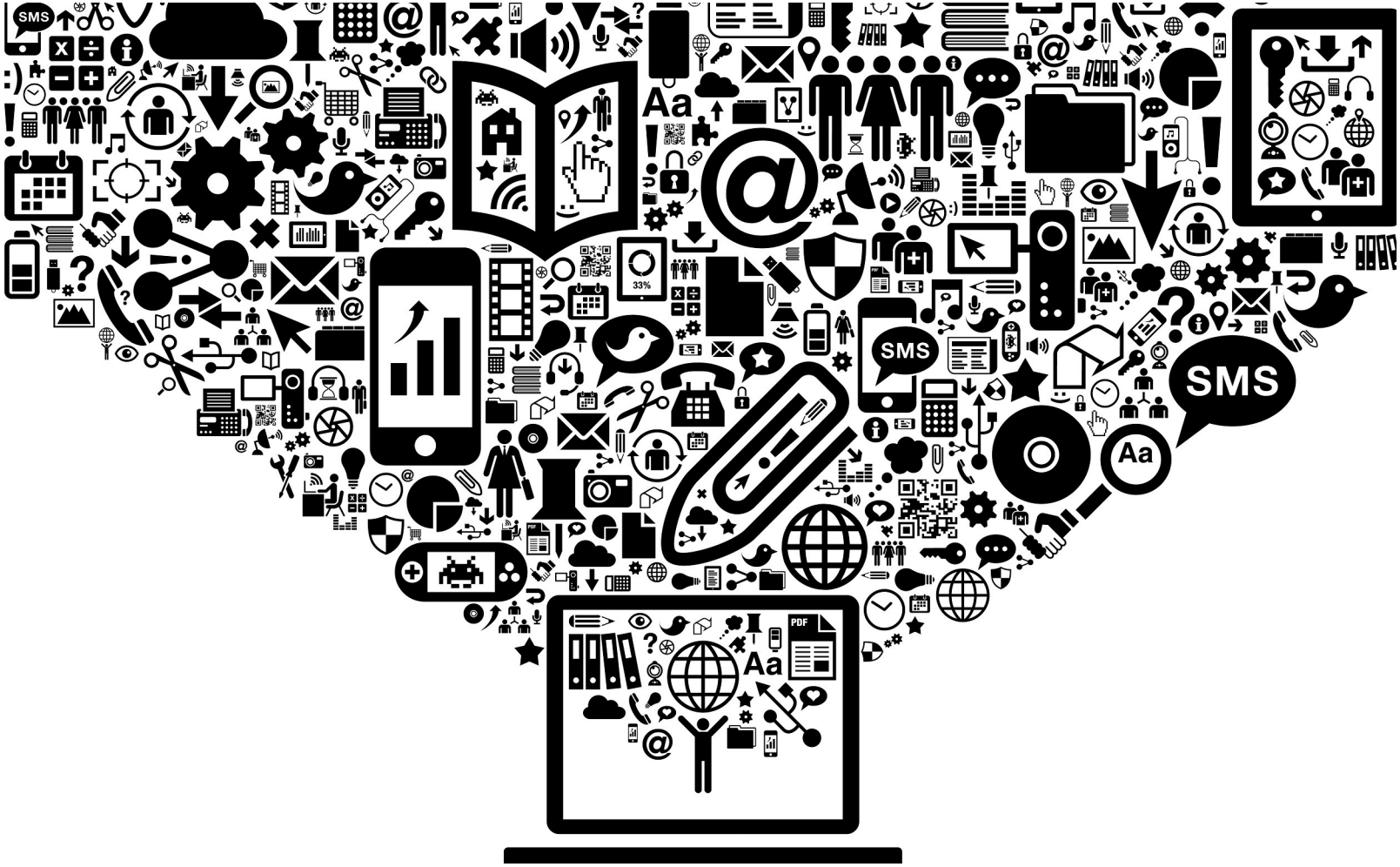


Smart Data Platform

Deep Data

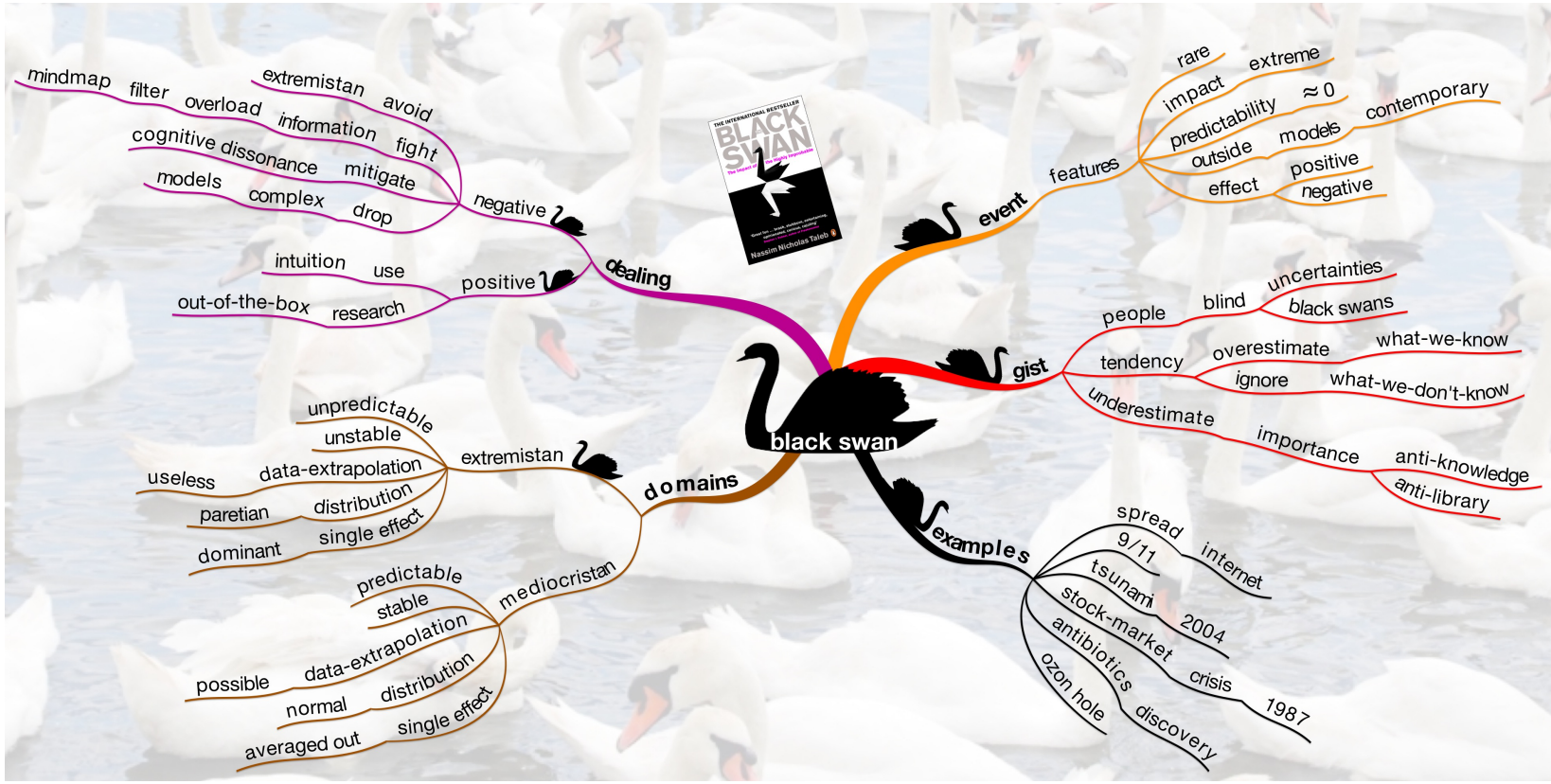
Ethnographic practice to comprehend social context

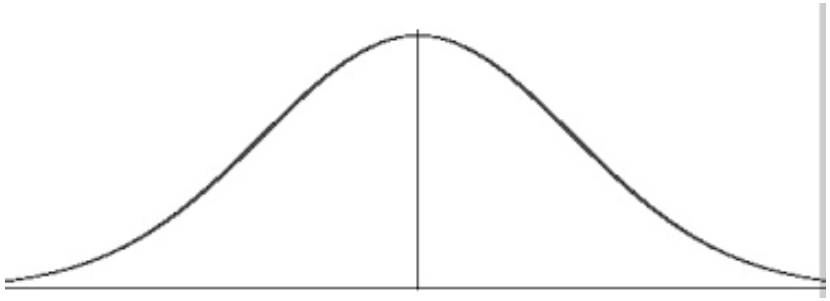
What matters: Theory > methodology > model > method > technology

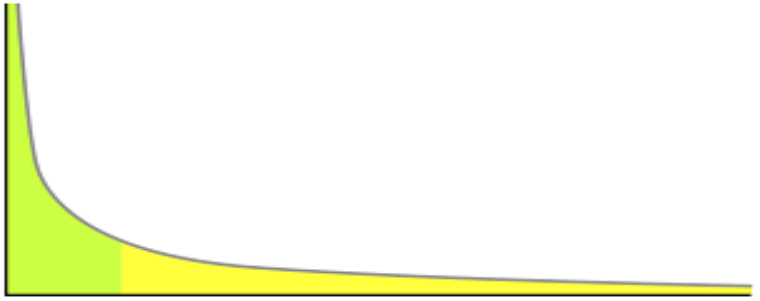


Theory → Methodology → Modeling → (Pre)Processing → Analysis → Prediction → Alternative

Normal distribution vs. Power law. And “black swan”








“Wave is coming”

Thank you