WORLD BANK GROUP
Digital Development

DIGITAL DEVELOPMENT PARTNERSHIP

Harnessing better data for development

Morning seminar Series, Tokyo,
29 January 2019

OVERVIEW

2018
Information and Communications for Development
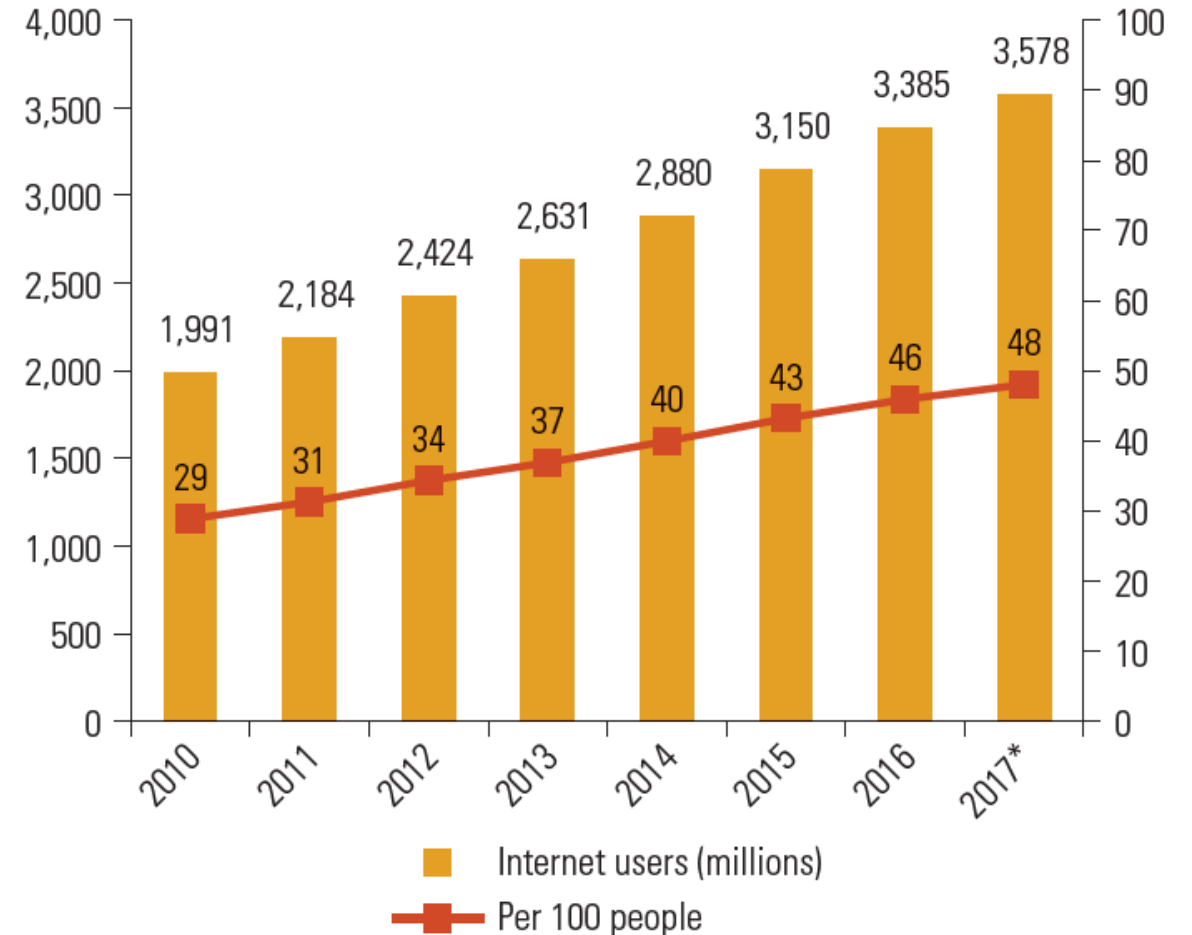
Data-Driven Development

WORLD BANK GROUP

IC4D

Why Data-Driven Development?

Data Deluge

- Over 90% of data that exists today was created in the last two years
- Cross-border data flows exceed US$3 trillion each year
- More than 2.7 million emails are sent, and 60'000 GB of data transmitted, each *second*
- A single autonomous vehicle will create 4'000 GB of data for each hour of driving

The challenge is to extract value from this data and to put it to work— for governments, firms and individuals.

Growth in internet users, globally

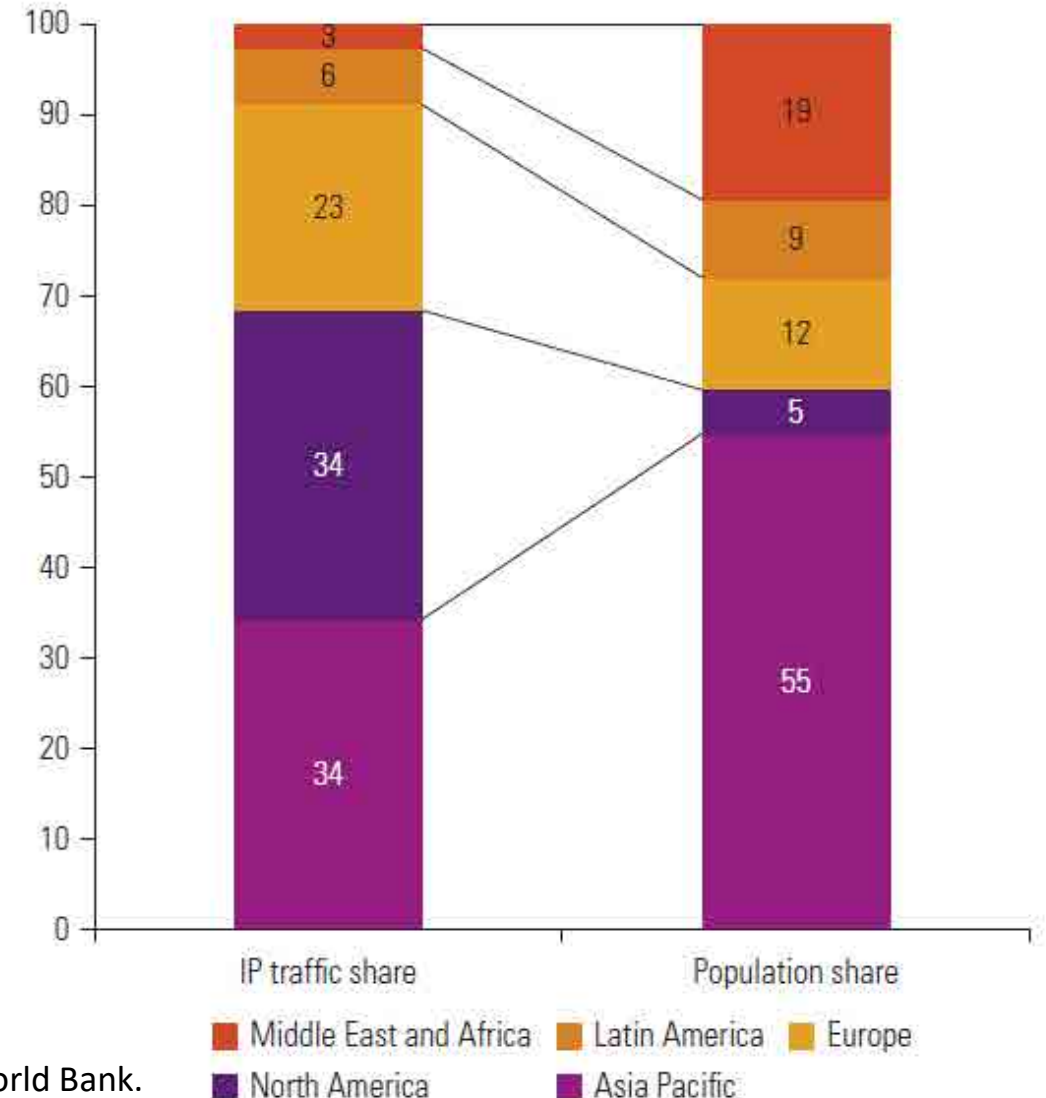Internet users (millions)
— Per 100 people

Source: ITU World Telecommunication/ICT Indicators Database

Report Structure

Executive Summary

1. Overview: Data, the fuel of the future
2. Supply: Data connectivity and capacity
3. Better data for doing good
4. People and data
5. Firms and data
6. Policies for the data economy

Data for development indicators

Bibliography

a. Distribution of global internet protocol traffic and population, 2015 (percent)

Middle East and Africa | Latin America | Europe
North America | Asia Pacific

IP traffic share:
- Asia Pacific: 34
- North America: 34
- Europe: 23
- Latin America: 6
- Middle East and Africa: 3

Population share:
- Asia Pacific: 55
- North America: 5
- Europe: 12
- Latin America: 9
- Middle East and Africa: 19

Source: CISCO, World Bank.

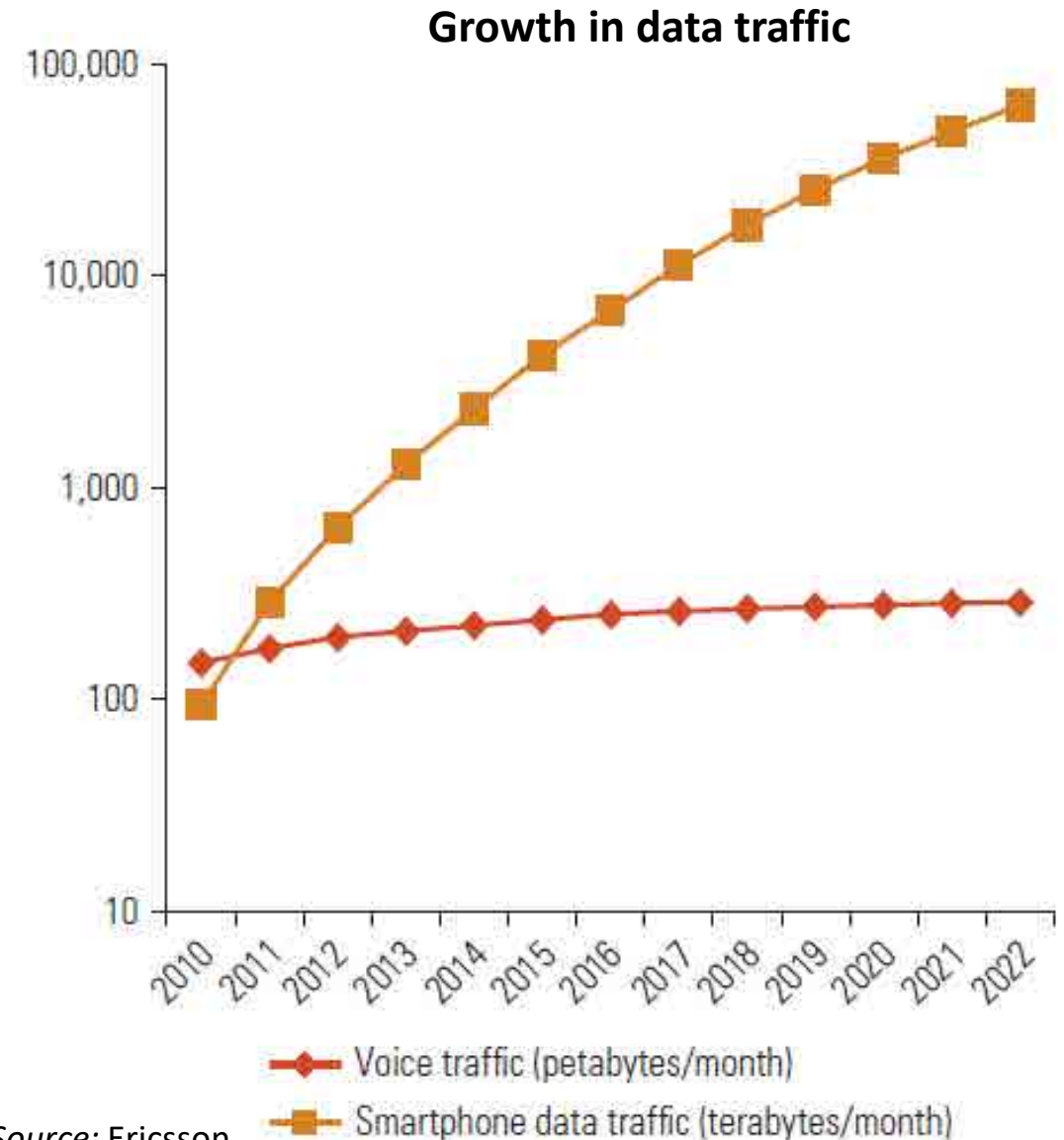# Where is all the data coming from?

Historically

- **Fixed networks,** optimized for voice, provided the majority of traffic
- **Public Telecom Operators** were national champions of their respective markets

Currently

- **Mobile data,** especially from smartphones, provides a growing share of global traffic
- Traffic is shifting away from network operators to **Over-the-top content creators**
- **Video streaming** (eg Youtube, Netflix) provides largest share of traffic in most markets

In future

- **"Machines" will be the main source of traffic --** Internet of Things, automated vehicles, sensor networks, data centers etc
- Traffic patterns will be closer to **population shares**

Growth in data traffic

Legend:
- Voice traffic (petabytes/month)
- Smartphone data traffic (terabytes/month)

Source: Ericsson.

Data-driven firms, by market capitalization

Rank	Company	Country	Market capitalization (US$ billions)	2016 revenue (US$ billions)
1	Apple	United States	801	218
2	Google / Alphabet	United States	680	90
3	Microsoft	United States	540	86
4	Amazon	United States	476	136
5	Facebook	United States	441	28
6	Berkshire Hathaway	United States	409	215
7	Exxon Mobil	United States	346	198
8	Johnson & Johnson	United States	342	76
9	Tencent	China	335	22
10	Alibaba	China	314	21
	Top 10 total		4,684	1,090
	Data-driven companies as percent of top 10		76.6	55.1

Source: Adapted from Meeker 2017, with market capitalization data from CapIQ and valid for May 26, 2017.
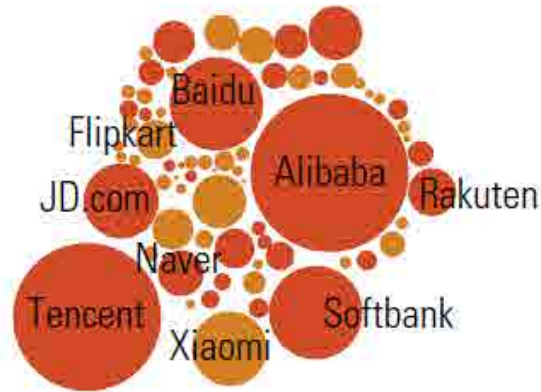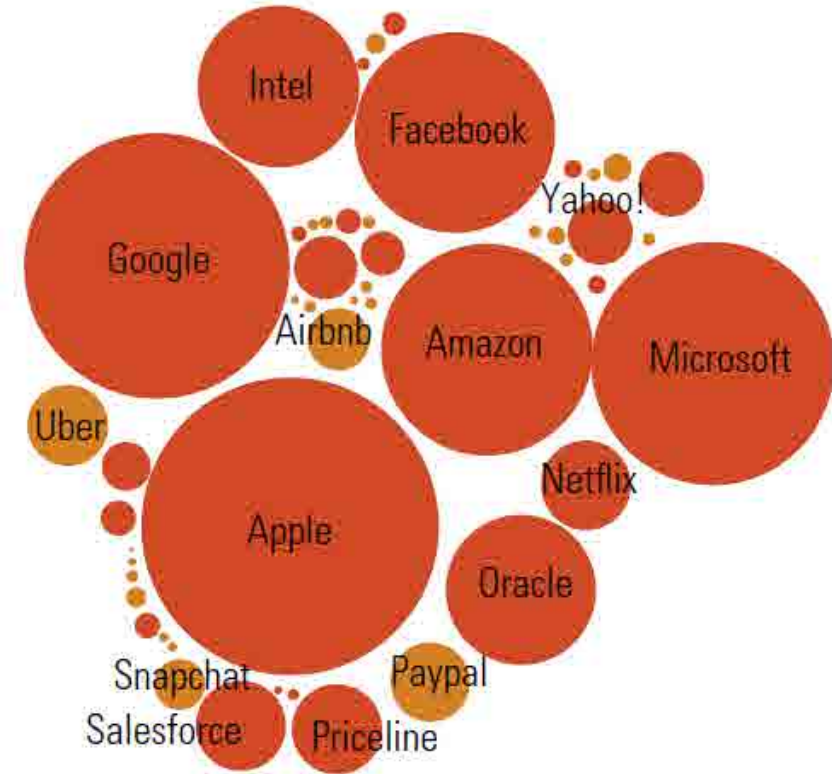Note: Data-driven companies are shown in red.

Europe, Africa & LAC in danger of missing the data revolution

North America

Intel
Facebook
Yahoo!
Google
Airbnb
Amazon
Microsoft
Uber
Netflix
Apple
Oracle
Snapchat
Paypal
Salesforce
Priceline

Asia

Baidu
Flipkart
JD.com
Alibaba
Rakuten
Naver
Tencent
Softbank
Xiaomi

Europe

SAP
Spotify

Africa and Latin America

Naspers

Digital enterprises (US$1bn plus revenue), by region

63: $2.8 trillion

42: $670 billion

27: $161 billion

3: $61 billion

● Publicly listed   ● Privately owned

What types of data are likely to be useful for development?

The data market

- **Personal data is created through an individual's actions**
  - For instance, making an online payment
- **Through digitizing analogue records**:
  - Such as medical records
- **Through devices**:
  - Such as call data records and "data exhaust"
- **As phones, sensors, cameras and computers multiply,** our digital trail grows ever larger
- Our **digital profiles** represent our online personas
  - As our digital history grows ever richer, marketeers are better able to respond to and anticipate our needs
- Companies use **Artificial Intelligence** to better predict our desires and action
  - More data equals stronger AI

Health
- Medical history
- Prescriptions and vaccinations
- Fitness tracking

Government
- Identification number and identity
- Address
- Civil information (birth, marriage, and so on)
- Legal records

Web
- Email
- Browsing and search history
- Content (social profiles, posts, photos, and so on)
- Contacts, followers, friends

Mobile phone
- Number and preferred network
- Call data records
- Location data (GPS)
- Social media contacts
- Purchasing history

Financial
- Accounts
- Transactions
- Debts
- Investments
- Insurance

Other
- Home information
- Travel
- Vehicle information
- Inferred data, created using other data points

Responsible Use of Big Data & AI

- **SDG 1: No Poverty**
  - Using nightlights data, combined with high-res satellite imagery, for rapid estimation of poverty levels in Africa
- **SDG 2: Zero hunger**
  - Using Big Data on climate patterns to predict drought in Colombia
- **SDG 6: Clean water and sanitation**
  - Sensor networks to monitor water quality in coastal waters off Singapore
- **SDG 8: Decent work and economic growth**
  - Using LinkedIn data to understand labor market trends and enhance vocational education
- **SDG 10 Reduced inequalities**
  - Sentiment analysis of local language radio broadcasts in Uganda
- **SDG 13: Climate action**
  - Using mobile signal strength (attenuation) to enhance road network resilience to flash floods in Senegal
- **SDG 16: Peace, justice and strong institutions**
  - Using AI to improve eGovernment services in Estonia

1. LAWFUL, LEGITIMATE AND FAIR USE
Data should be obtained, collected, analysed or otherwise used through lawful, legitimate and fair means, taking into account the interests of those individuals whose data is being used.

2. PURPOSE SPECIFICATION, USE LIMITATION AND PURPOSE COMPATIBILITY
Any data use must be compatible or otherwise relevant, and not excessive in relation to the purposes for which it was obtained.

3. RISK MITIGATION AND RISKS, HARMS AND BENEFITS ASSESSMENT
A risks, harms and benefits assessment that accounts for data protection and data privacy as well as ethics of data use should be conducted before a new or substantially changed use of data (including its purpose) is undertaken.

4. SENSITIVE DATA AND SENSITIVE CONTEXTS
Stricter standards of data protection should be employed while obtaining, accessing, collecting, analysing or otherwise using data on vulnerable populations and persons at risk, children and young people or any other data used in sensitive contexts.

5. DATA SECURITY
Robust technical and organizational safeguards and procedures should be implemented to ensure data management throughout the data lifecycle and prevent any unauthorized use, disclosure or breach of personal data.

6. DATA RETENTION AND DATA MINIMIZATION
Data access, analysis or other use should be kept to the minimum amount necessary to fulfill the purpose of data use.

7. DATA QUALITY
All data-related activities should be designed, carried out, reported and documented with an adequate level of quality and transparency.

8. OPEN DATA, TRANSPARENCY AND ACCOUNTABILITY
Appropriate governance and accountability mechanisms should be established to monitor compliance with relevant law, including privacy laws and the highest standards of confidentiality, moral and ethical conduct with regard to data use.

9. DUE DILIGENCE FOR THIRD PARTY COLLABORATORS
Third party collaborators engaging in data use should act in compliance with relevant laws, including privacy laws as well as the highest standards of confidentiality and moral and ethical conduct.
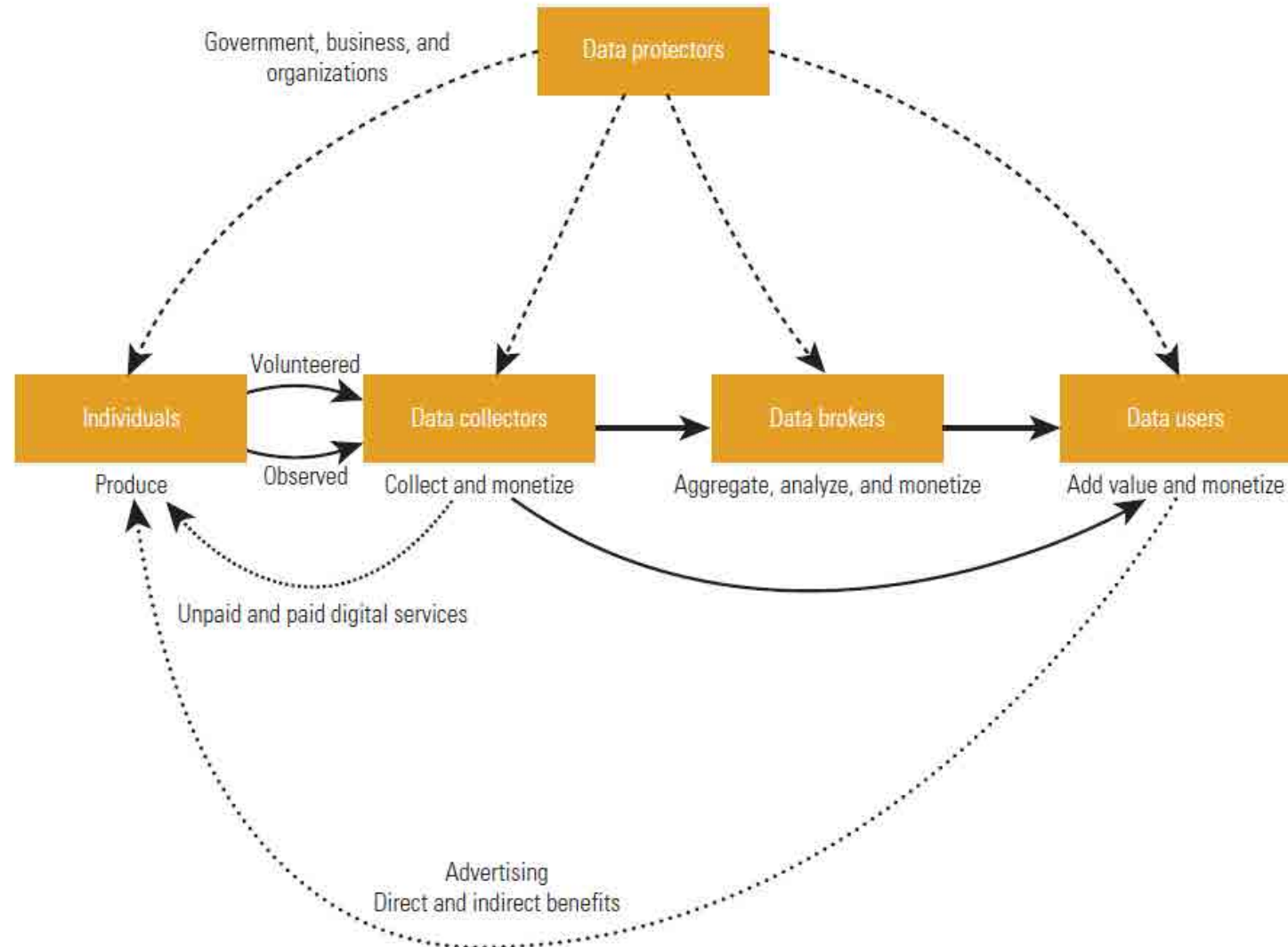
People and Data

While the data revolution brings lots of benefits to citizens, it also brings risks and costs to individuals

- Benefits include better decision making and more convenience. Services can be "free", thanks to advertising, but require data access
- Possible costs include loss of privacy, agency and control
- Risks include the danger of exclusion, as a result of profiling, and erosion of trust

Emerging trends suggest new opportunities for individuals to regain control of their personal data

- Keeping personal data secure may also permit individuals to monetize it
- Average global revenue from advertising is around US$53 per internet user per year
- There is a trade-off between surrendering personal data to gain free services, like email

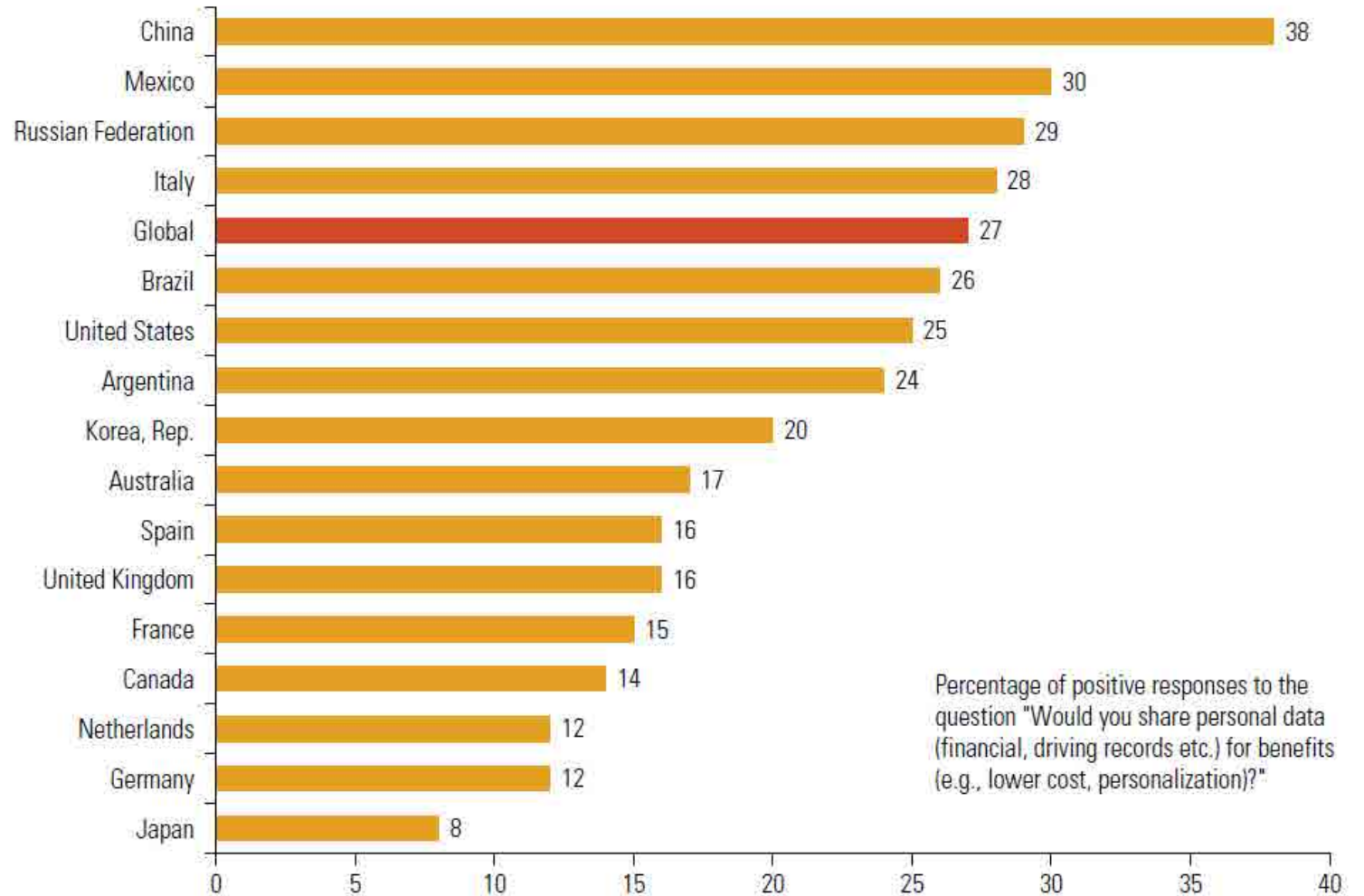**Towards a citizen-controlled market for data**

Willingness to share personal data

Globally, 27 per cent of consumers are happy to share personal data in return for convenience and other benefits

Chinese citizens are most willing. Japanese citizens are the least willing to share

China	38
Mexico	30
Russian Federation	29
Italy	28
Global	27
Brazil	26
United States	25
Argentina	24
Korea, Rep.	20
Australia	17
Spain	16
United Kingdom	16
France	15
Canada	14
Netherlands	12
Germany	12
Japan	8

Percentage of positive responses to the question "Would you share personal data (financial, driving records etc.) for benefits (e.g., lower cost, personalization)?"

Source: GfK 2017.
Note: Based on more than 22,000 consumers online in 17 countries with a response of 7 (on a scale from 1 to 7), where 7 represents full agreement.
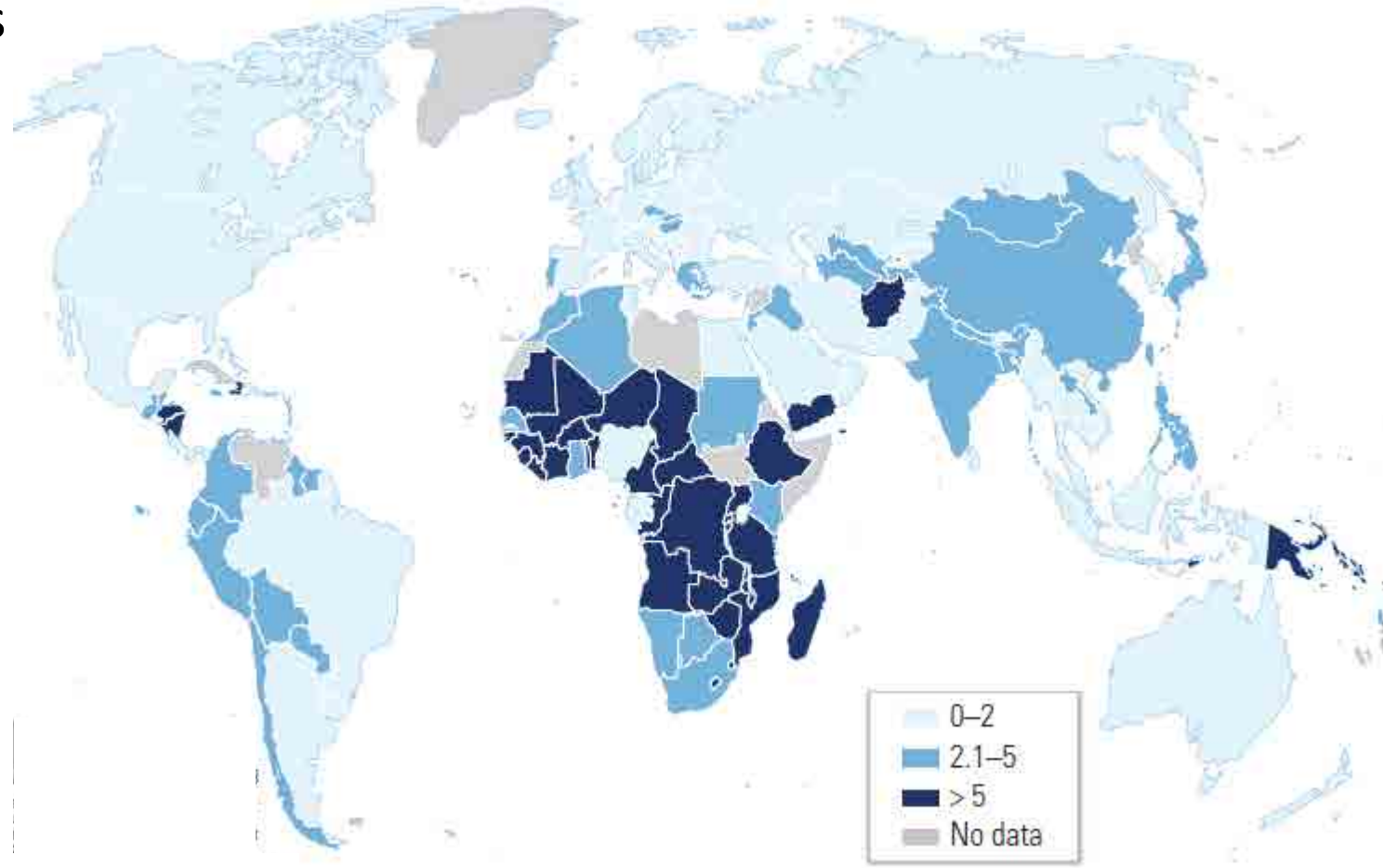
Firms and Data

Data assets are becoming crucial for competitiveness of nations and firms

- The emergence of the data economy is producing new business models
- Digital Platforms are becoming particularly important, but can become gatekeepers
- If data is the new oil, then firms must seek cheaper supplies and must invest in data centers to extract value

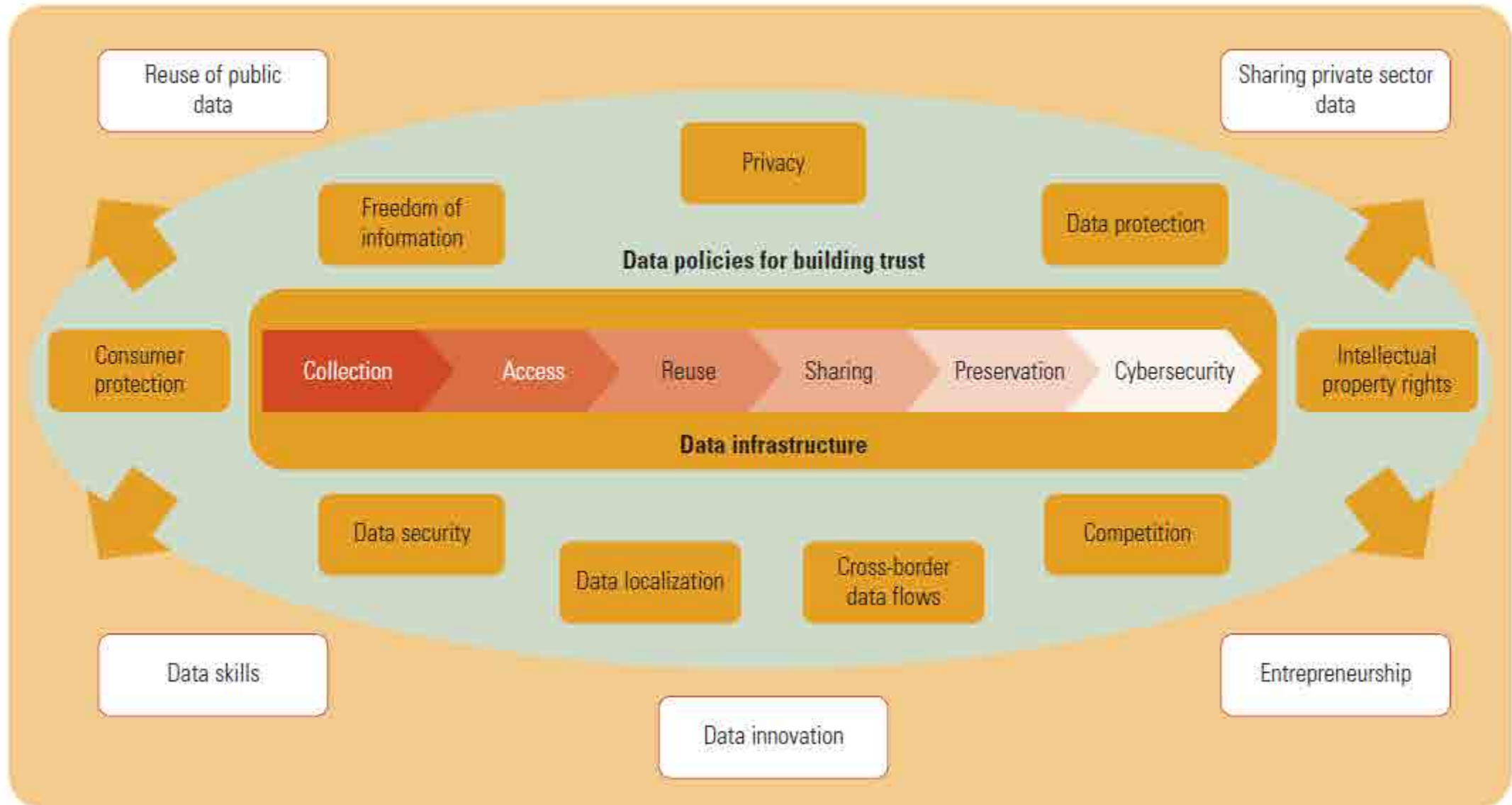## Data inequalities increasingly dominate in global economies, but they need not be permanent

- Countries can take steps to reduce the data talent gap
- Competitive market entry can help to develop data infrastructure and reduce prices
- Clarify the policy environment, especially for data protection, cybersecurity and IPRs
- Promote data entrepreneurship and innovation

Price of 1Gb of mobile data as % of GDP per capita

Legend:
- 0–2
- 2.1–5
- \>5
- No data

Source: ICTdata.org, World Bank.

Policies for the data economy

The full report was published on 30 October 2018 at:
https://openknowledge.worldbank.org/handle/10986/30437

For more information, contact:
Tim Kelly, editor, tkelly@worldbank.org
Christine Howard, choward2@worldbank.org
Roku Fukui, rfukui1@worldbank.org